Colour Constancy: Cues, Priors and Development

WEDGE-ROBERTS, REBECCA,JANE

Colour Constancy: Cues, Priors and Development

A thesis submitted for the degree of Doctor of Philosophy in the

Department of Psychology

Durham University

Rebecca Wedge-Roberts

2021

Abstract

Colour is crucial for detecting, recognising, and interacting with objects. However, the reflected wavelength of light ("colour") varies vastly depending on the illumination. Whilst adults can judge colours as relatively invariant under changing illuminations (colour constancy), much remains unknown, which this thesis aims to resolve. Firstly, previous studies have shown adults can use certain cues to estimate surface colour. However, one proposed cue - specular highlights - has been little researched so this is explored here. Secondly, the existing data on a daylight prior for colour constancy remain inconclusive so we aimed to further investigate this. Finally, no studies have investigated the development of colour constancy during childhood so the third aim is to determine at what age colour constancy becomes adult-like.

In the introduction, existing research is discussed, including cues to the illuminant, daylight priors, and the development of perceptual constancies.

The second chapter contains three experiments conducted to determine whether adults can use a specular highlight cue and/ or daylight prior to aid colour constancy. Results showed adults can use specular highlights when other cues are weakened. Evidence for a daylight prior was weak.

In the third chapter the development of colour constancy during childhood was investigated by developing a novel child-friendly task. Children had higher constancy than adults, and evidence for a daylight prior was mixed.

The final experimental chapter used the task developed in Chapter 3 to ask whether children can use specular highlights as a cue for colour constancy. Testing was halted early due to the coronavirus pandemic, yet the data obtained suggest that children are negatively impacted by specular highlights.

Finally, in the general discussion, the results of the six experiments are brought together to draw conclusions regarding the use of cues and priors, and the development of colour constancy. Implications and future directions for research are discussed.

Contents

List of Tables

List of Figures

Declaration

I declare that no part of the material presented in this thesis has previously been submitted for a degree in this or in any other University. Where material has been generated through collaboration, this has been indicated where appropriate, and the contributions of the authors made explicit. In all other cases, the work of others has been acknowledged and referenced appropriately.

Published and submitted work

The entirety of Chapter 2 has been published in Journal of Vision:

Wedge-Roberts, R., Aston, S., Beierholm, U., Kentridge, R., Hurlbert, A., Nardini, M., & Olkkonen, M. (2020). Specular highlights improve color constancy when other cues are weakened. Journal of vision, 20(12), 4-4.

The entirety of Chapter 3 is under review in Developmental Science with authors Rebecca Wedge-Roberts, Stacey Aston, Ulrik Beierholm, Robert Kentridge, Anya Hurlbert, Marko Nardini, and Maria Olkkonen.

Acknowledgements

Firstly, I want to thank my supervisory team, who have all been amazing throughout: Marko Nardini, Maria Olkkonen, Ulrik Beierholm, and Bob Kentridge. Thank you to Marko, without whom I would never have started this PhD, and who has been the most supportive primary supervisor I could have asked for. He helped me not only with the work itself but with all the emotional stress that comes with a PhD. I am so grateful to Maria who has been enormously kind and helpful throughout. She is always willing to go the extra mile to help and I genuinely couldn't have done this without her. Thanks to Ulrik for all his help with analysis and letting me spend months rendering and re-rendering stimuli on his computer! Thank you to Bob who is the most genuine, kind-hearted, person, and knows about absolutely everything! I am also grateful to Anya Hurlbert for collaborating with us on these projects and sharing her expertise on colour constancy. Thank you to Stacey for teaching me so much and helping programme the early experiments. It has been a privilege to work with all of you.

Thank you to everyone who has participated in my experiments, including the friends who sat in the dark for hours to help me out, and gave me good quality data. Particular thanks to all the children and their families who took time out of their weekends to come and feed a dragon sweets!

Next, I want to thank everyone in my life who led me to do this PhD. Thank you to my A level teacher Geri, without whom I never would have studied psychology at university. My amazing university tutors Nick Yeung, Hannah Smithson, and Kate Nation sparked an interest in studying development and colour vision further.

I also want to thank all the friends I've made who have been on this journey with me – Ali, Liz, Emine, Naz, Reny, Linda, Rachel, Siobhan, and many others. You kept me (relatively) sane, and I'll miss our trips to the Swan.

Thank you to my family for supporting me throughout. Particularly to my parents, Sian and David, who put up with my daily phone calls, and help where they can. Finally, thank you to Jess for being the best companion I could wish for.

Chapter 1

General Introduction

1.1 Colour Constancy: The Problem

A crucial role of the visual system is to recognise objects, so that observers can interact with them appropriately. Whilst this seems effortless for adults, the task is a difficult one, since the environment and viewpoint are constantly changing, yet we require a stable representation of objects to recognise them. Colour is known to improve object recognition (Humphrey et al., 1994; Wurm et al., 1993; Tanaka and Presnell, 1999; Yip and Sinha, 2002); however, the illumination falling on an object or a surface varies across environments and within scenes (e.g. due to shadows). When an observer views an object, the wavelength spectrum of light reaching their eyes is a function of both the surface reflectance function of the object (the proportion of incident light reflected at each wavelength of the visible light spectrum) and the spectral power distribution of the illumination (the amount of light emitted by an illuminant at each wavelength). Specifically, the reflected light, E(λ), is a multiplicative combination of the spectral power distribution of the illumination, I(λ), and the spectral reflectance function of the surface, R(λ):

$$E(\lambda) = R(\lambda)I(\lambda) \tag{1.1}$$

A schematic example of this problem is shown in Fig. 1.1. On the left, a ripe, yellow banana is seen outside under a blueish daylight illumination, and on the right, the same banana is seen indoors under a tungsten light bulb, which appears reddish. The spectrum of light reflected into an observer's eye, E(λ), changes vastly between these two illuminations, with much less short wavelength energy reflected to the observer's eye under the tungsten illumination, despite the surface reflectance of the banana remaining constant. The situation is further complicated by the fact that the human eye only has three cone types: S cones sensitive to short wavelengths of light, with a peak sensitivity at 420nm; M cones sensitive to medium wavelengths with a peak sensitivity at 530nm; and L cones sensitive to longer wavelengths, peaking at 560nm. The spectral sensitivity functions of the three cone types, showing how responsive they are to different wavelengths of light, are also shown in Fig. 1.1. The response of a given cone can only change along one dimension but is affected by both the intensity and wavelength of light (the principle of univariance; Rushton, 1972). Therefore, a single cone would respond equally to a high intensity light at a wavelength it is not optimally tuned to and a low intensity light that it is optimally tuned to. In this way, individual cones are colour blind, and it is only the combination of photon catches in the three cone types (tristimulus values) which gives rise to the percept of colour. Within each cone class, the expected photon catch can be calculated by weighting the spectrum of reflected light, E(λ), by the cone's spectral sensitivity function, K(λ), and integrating this across all wavelengths of the visible light spectrum. Thus, for a given cone type, K, the photon catch is given by:

$$PhotonCatch_k = \int K(\lambda)R(\lambda)I(\lambda)d\lambda \tag{1.2}$$

where K is the S, M, or L cones. The difference in cone photon catches for a yellow banana under tungsten and blueish daylight is shown at the bottom of Fig. 1.1. There is relatively higher S cone activation under the daylight illuminant, in line with the greater proportion of short wavelengths of reflected light compared to under tungsten. However, we would not expect an observer to judge the banana to have a different surface reflectance - and crucially a different level of ripeness - when it is moved from outside to inside. This demonstrates the evolutionary need for colour constancy to determine how to appropriately interact with objects.

One proposed method that people may employ to perceive constant surface reflectances, despite changing light reaching their eyes, is to use knowledge of the illumination's spectral power distribution, I(λ). von Kries proposed a form of adaptation that involves independently adapting the gains for each cone class (see Buldt, 2016). Mathematically, this would be done by using a diagonal matrix, with each element of the diagonal being equal to the ratio of cone activation under the two illuminants (e.g. L_2/L_1). Herbert Ives was the first to suggest that von Kries adaptation could be used for colour constancy in 1912 (see Brill, 1995), by adapting each

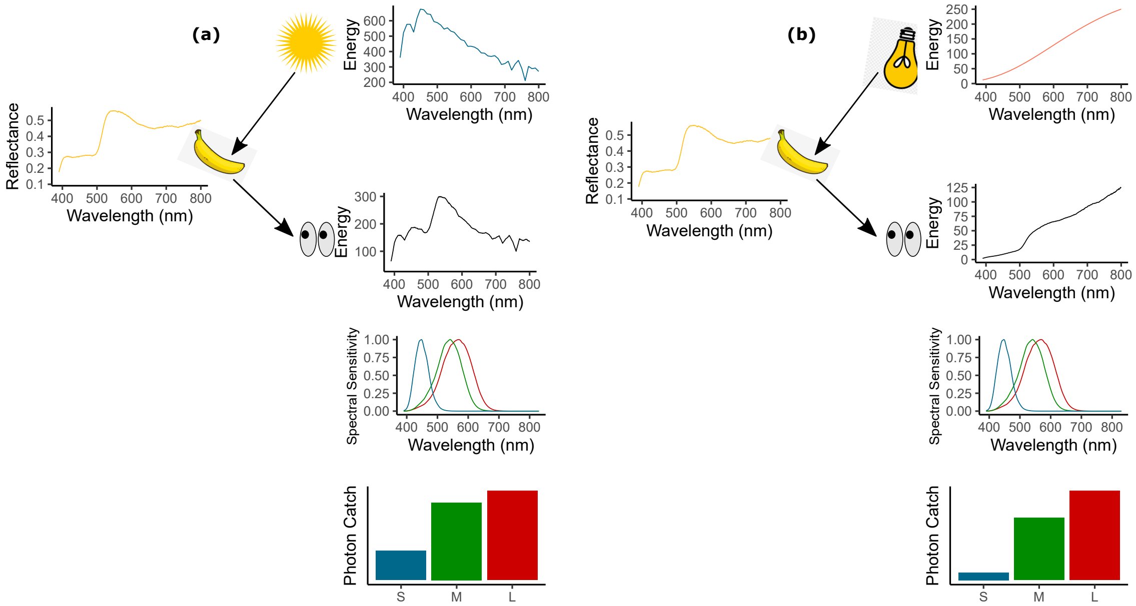


Figure 1.1: Schematic of a ripe yellow banana under (a) a blueish daylight illumination and (b) a tungsten light bulb. Top: the spectrum of light reflected into an observers eye is a combination of the spectral power distribution of the illuminant and the spectral reflectance function of the banana. Middle: Spectral sensitivity function for S (blue), M (green), and L (red) cones. Bottom: the relative photon catch for the three cone types is different for the banana under daylight (a) and tungsten (b) illuminants.

cone class independently under different illuminants to keep relative cone activation for a given surface constant.

However, as the cone activation results from integrating the product of illumination, surface reflectance and cone spectral sensitivity (see Equation 1.2), it is mathematically impossible to recover the reflectance simply by dividing the cones' photon catch by the illumination's spectral power distribution, as suggested by the von Kries-Ives transform. Furthermore, some researchers argue that observers retain information about the illumination and do not simply discount it (Gilchrist, 1979; Zaidi, 2001).

Nevertheless, many proposed colour constancy mechanisms require an estimate of the illumination's tristimulus values - whether explicit or implicit. As the illumination is not always directly visible, it has been proposed that a number of cues and priors may be used to infer the chromaticity of the illumination. These are discussed in the following sections.

1.2 Cues

Many cues and heuristics have been proposed to enable people to estimate the tristimulus values, or chromaticity, of the illumination falling on an object. These range from simple cone adaptation to cognitive cues such as memory. These can be split into so called "low level" cues, which can be used in simple two-dimensional scenes, containing only flat, matte surfaces; and "high level" cues which rely on more complex scenes, containing objects, inter-reflections, etc.

1.2.1 "Low level" cues

One proposed low-level cue is termed the grey world hypothesis. This theory, which is implicit in Land's Retinex theory (Land, 1983, 1986; Land and McCann, 1971), relies on the assumption that the average surface reflectance over a whole scene (the global mean) is neutral, and therefore has a flat spectral reflectance function. In these scenes, when an illumination is reflected from the surfaces, the total reflected light has the same tristimulus values as the illumination (although the luminance will be scaled depending on the lightness of the surfaces). Problems arise from the grey world hypothesis when the average of colours within a scene is not neutral, such as under a green-biased forest canopy or underwater. Simulations assuming a neutral global mean tend to over-correct for chromatic biases of scenes to make the scene look less saturated than it is (Smithson, 2005). Buchsbaum (1980) proposed a modification to the theory to overcome this problem. He suggested that the average spectral reflectance over the scene is not necessarily flat but is known and constant. Using this assumption, he demonstrated that the chromaticity of the illumination can be extracted. However, this model fails when objects are changing along with the illumination, such as when changing environments, as this violates the assumption of a constant surface reflectance.

Another cue that has been proposed is that observers may adapt to the local mean surrounding a surface (Von Kries and MacAdam, 1878). Simulations show that using this cue should lead to over-corrections and reduce differences in appearance between different surfaces that occur on a larger spatial scale than the local mean (Smithson, 2005). However, Smithson and Zaidi (2004) demonstrated that observers are able to adapt to the local mean to demonstrate constancy, even when the global mean cue signals a different illumination. Furthermore, Hurlbert and Wolf (2004) shifted the local surround of a test patch to cause colour contrast induction, and found that local colour contrast can affect colour constancy.

Another assumption that can be made to estimate the illumination's tristimulus values is that the brightest surface within a scene is white. This is often termed the "brightest is white" assumption. As matte white surfaces reflect all wavelengths of light equally, the light reflected from such a surface will have the chromaticity of the illumination. This assumption was also exploited in Land's Retinex theory (Land and McCann, 1971), where the reflectance was normalised to the brightest surface within each channel. Linnell and Foster (2002) investigated whether people use the brightest is white cue or global mean cue when these conflicted, and found that observers used the brightest is white cue when there were a few large patches (one degree of visual angle) in the scene but with a larger number of small patches, observers relied on the global mean instead. This suggests that people can use both cues but weight them differently depending on the context.

Kraft and Brainard (1999) tested whether observers could use the three cues described above (global mean, local mean, brightest surface) for colour constancy by systematically removing each from a scene. They found that observers had close to perfect constancy when all were present and valid, but constancy decreased when each of these cues were individually removed, and was even lower when all were invalid. However, even when observers could not use any of these cues, they still accounted for the difference in illumination to some degree, suggesting residual constancy. This suggests that additional cues may be used.

1.2.2 "Higher level" cues

The three cues discussed previously are applicable to scenes containing only flat, matte surfaces. However, the real world contains three-dimensional objects with varying levels of specularity, orientation, etc. Although this may seem to make the problem more complicated, it provides additional cues to the illumination's chromaticity.

A so-called "higher level" cue which is present in three-dimensional scenes, but not simplified models, is mutual reflections. Within a scene illuminated by a single light source, there will be both direct and indirect illuminations falling on an object. These indirect, or mutual, illuminations arise from reflections from other surfaces. Mutual reflections can provide an additional source of information to the illumination's chromaticity. Gilchrist and Jacobsen (1984) demonstrated this by presenting observers with two identical miniature rooms containing objects. Every surface and object was painted matte white in one room and matte black in the other. In the white room, there were many mutual reflections, as white surfaces reflect a large proportion of light. In contrast, the black room lacked mutual reflections as black absorbs most light. It was found that the intensity profile for the black room had sharper edges compared to the relatively smooth profile of the white room. Changing the intensity of the illumination shifts this profile uniformly whereas changing the colour of the objects changes the shape of the intensity profile. Observers in this study were able to tell whether the rooms were black or white, even when the intensity of the illuminant in the black room was increased such that there was more light reflected from the surfaces than in the white room. As the only difference between the rooms was in the mutual reflections, this suggests that people can make use of the information contained in mutual reflections to determine surface reflectance, at least for achromatic scenes.

Forsyth and Zisserman (1989) have proposed a formal model of how the visual system may exploit mutual reflections to determine surface reflectance. As with Gilchrist and Jacobsen (1984), these researchers demonstrated that mutual reflections change the intensity profile in various scenarios, including a "roof edge" at corners, as opposed to a "step edge" without mutual reflections. The authors proposed that people have a "dictionary", which maps mutual illumination effects to surface reflectance, for known object shapes.

Bloj et al. (1999) demonstrated that people use mutual illumination information in their judgements of surface colour. They presented observers with a concave-folded piece of card with one side coloured magenta and the other side white, such that mutual reflections made the white side appear pink. When viewed normally, observers correctly judged the surfaces as magenta and white. However, when viewed through a pseudoscope so that the card appeared convex, the white side was judged to have pink reflectance, as the mutual reflections were not discounted. This shows that people take three-dimensional shape and mutual reflections into account when determining surface reflectance. Similarly, Doerschner et al. (2004) investigated whether people use mutual reflections to determine surface reflectance, by varying the orientation between a bright orange cube and test patch. It was found that observers did take orientation into account in their judgements of surface reflectance, suggesting that they use mutual reflection information to remain colour constant.

A higher-level cognitive cue used for colour constancy is memory and object knowledge. The world does not contain arbitrary surfaces, but instead objects - some with high colour diagnosticity. Observers can use this knowledge of the canonical colour of objects to infer the illumination's chromaticity in a similar way to using a white surface. Hansen et al. (2006) demonstrated that memory influences perceived colour, by showing observers photographs of natural fruits and asking them to manipulate the colour of the fruit until it appeared grey. It was found that observers overcompensated - for example, they set a banana to have more shorter wavelengths than a neutral surface (making it appear more blue). This may be due to the image of a banana stimulating yellow and this memory adding a yellowish hue to the perceived image, so the actual chromaticity needed for observers to perceive the banana as neutral contained more short wavelengths of light. This shows how memory can affect colour appearance in a top-down manner and may be involved in colour constancy.

Evidence that colour knowledge can improve colour constancy comes from Van De Weijer et al. (2007) who showed that computational models using top-down knowledge, such as that the sky tends to be blue and grass tends to be green, can improve illumination estimation in models of colour constancy. Furthermore, Granzier and Gegenfurtner (2012) found that observers had higher colour constancy in scenes containing diagnostically coloured objects, compared to scenes without. Similarly, Emmerson and Ross (1987) found that observers had better constancy for objects seen underwater when familiarly coloured objects were present. Further uses of top-down prior knowledge to improve colour constancy judgements are given in Section 1.3 below.

Specular Highlights

A cue which may be used to aid colour constancy when glossy objects are present is specular highlights. A perfectly specular object (such as a mirror) will have pure highlights as the surface does not absorb the illuminant, so the light reaching the observer's eye is simply a reflection of the incident illumination. Thus, assuming highlights have been identified as such, there is direct access to the illumination's chromaticity on purely specular objects.

However, most objects are only partially specular, so they have both a diffuse component and a specular component. The diffuse component, or body colour, absorbs a proportion of the illuminant and reflects a combination of the illuminant and object reflectance. This is what is typically referred to as the object's "colour". In contrast, the specular component is as described above for perfectly specular surfaces. The orientation of a surface with respect to the illuminant, as well as its material, will affect the weight of the specular and diffuse components. When light hits a partially specular object, its rays are scattered such that some are absorbed more than others. Therefore, the light reflected to the observer's eye forms a line in colour space between the chromaticity of the diffuse component and the chromaticity

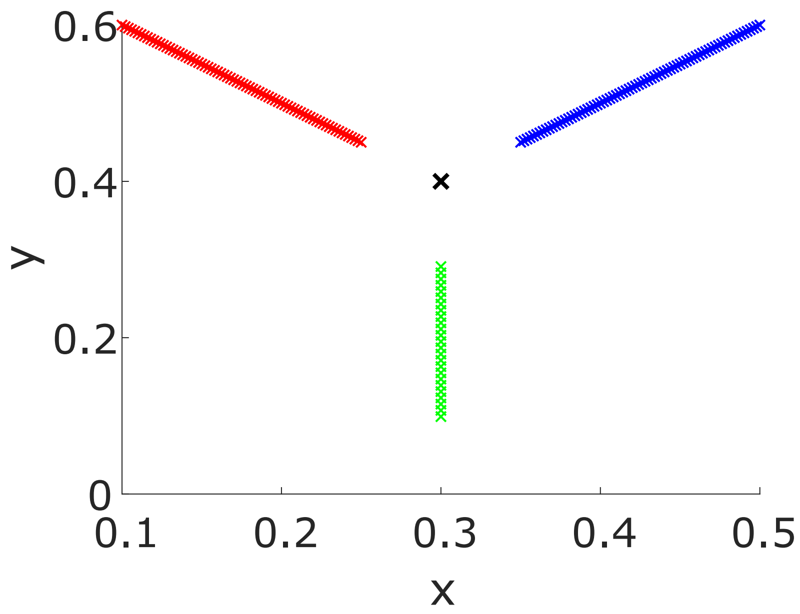


Figure 1.2: Example of chromaticity convergence. The black X is the chromaticity of the illumination; green, blue, and red Xs are chromaticities taken from three objects with different body colours under the same illuminant. For each object, the chromaticites fall on a line pointing towards the illumination chromaticity. If extended, they would intersect at the chromaticity of the illumination.

of the illumination (see Fig. 1.2). Most partially specular objects do not have any perfect specular highlights so the end of this line closest to the chromaticity of the illumination contains some body reflectance, and does not give the chromaticity of the illumination. Therefore, when there is only one partially specular object present, observers will struggle to infer the illumination's chromaticity. However, when there are multiple objects with different diffuse reflectances (i.e. different "colours") in a uniformly lit scene, the specular ends of these lines in colour space will point towards the chromaticity of the illumination. If these lines were extended, they would intersect at the chromaticity of the illumination, as in Fig. 1.2. This was formalised by Hurlbert (1998) as chromaticity convergence.

A number of studies have investigated whether observers can use specular highlights as a cue for colour constancy. For example, Granzier et al. (2014) presented observers with either matte or glossy cylinders under D65 (neutral daylight) or a reddish illumination, and asked them to pick a matching Munsell chip from under a

Tungsten illumination. They found higher colour constancy for the glossy, compared to matte, cylinders, and for smooth, compared to rough, objects, which have more pronounced highlights. Additionally, Nagai et al. (2017) found that observers were better at detecting illumination changes in rendered scenes with abstract shapes containing highlights compared to with matte shapes. Similarly, Lee and Smithson (2016) found that people can use specularity, even without pure specular highlights. In this study, observers saw an isolated sphere, which either experienced a change in illumination or reflectance, and they had to identify the type of the change. Observers could do this well, especially at higher levels of specularity, but failed when the pixels were rearranged to remove the perception of a sphere. As the sphere was isolated, there were no cues present other than specularity, suggesting this is a cue that people can use.

An alternative way to determine the extent to which different cues are used is cue perturbation, which has previously been applied to depth and shape vision (e.g. Maloney and Landy, 1989). This method allows researchers to quantify the extent to which cues are used, while altering the scene less than silencing cues. Yang and Maloney (2001) used cue perturbation to determine the extent to which people use specular highlights and uniform background cues for colour constancy. Scenes containing six partially glossy spheres with specular highlights, on a plain background were rendered under two different illuminants: A (a reddish tungsten light) and D65 (daylight). Matte versions of each scene were also rendered. To create perturbed images, the matte image under illuminant A was merged with the specular image under illuminant D65, and vice versa. This meant that most cues in a scene signalled one illumination (e.g. A) while the specular highlights signalled the other (e.g. D65). Observers set a test patch on one of the spheres to appear grey under the non-perturbed and perturbed conditions (achromatic setting). To determine the extent to which the specular highlight cue was used, observers' settings for the unperturbed scenes were plotted in colour space with a line joining them. The settings made in the perturbed conditions were then plotted along this line. If the specular highlight cue was not used at all, the setting under illuminant A perturbed to D65 would be the same as under A; if the specular highlight cue was the only cue used, the setting would be the same as under D65. It was found that people assigned a large weight to the specular highlight cue when the highlights were perturbed to appear as they would under D65 but only a small weight when perturbed to appear under illuminant A, which may imply use of a daylight prior (see below). In addition, perturbing the uniform background had almost no effect, although this could be because the test patch was on the sphere rather than the background. This study suggests that people can use specular highlights but the weighting applied to it is variable.

Yang and Shevell (2003) also used cue perturbation to investigate colour constancy in a scene lit by two illuminants on either side of a dividing wall. When the wall was so low that the illuminants crossed over it, observers' colour constancy indices were lower than in scenes containing a high wall keeping the illuminants separate. However, when the specular highlights were perturbed so that they were consistent with the illuminant on the same side of the wall, rather than the illuminant on the other side, observers' colour constancy indices improved. This suggests that specular highlights can be used as a cue to improve colour constancy and can remove the ambiguity present when scenes are lit by multiple lights.

Further studies have investigated the extent to which specular highlights can influence colour and material perception. For example, Xiao and Brainard (2008) presented observers with spheres of varying specularity and roughness, and asked them to match the diffuse colour. Whilst observers' matches did vary depending on the material of the test sphere, the effect was only small, suggesting that people are generally good at understanding how gloss works. Boyaci et al. (2006) also determined that observers can use specular highlights to determine the direction of an illuminant, which they can then use to judge the lightness (albedo) of a surface at a given orientation. However, it should be noted that only two out of the six observers tested were able to use this cue.

Taken together, the research looking at the use of specular highlights as a cue for colour constancy does suggest that observers may be able to use them in some circumstances. However, the extent to which they are used can depend on other factors such as the type of illumination. Furthermore, many of the studies discussed above tested only a small number of observers, and these were often not naive. Therefore, the use of specular highlights in colour constancy requires further research.

1.3 Priors

According to Bayesian models, combining current (ambiguous) information with prior knowledge can help to improve precision and reduce variability in judgements. For example, if one was trying to determine whether or not to wear a coat outside on a sunny day in December, they may combine the current knowledge they have of the weather (sunny) with prior knowledge of typical temperatures in December (cold) and determine that a coat is a good idea. Had they simply used the current information available, they may have guessed that it would be warm instead, and chosen not to wear a coat. Bayesian models combine the prior information, with the likelihood of an observation given a hypothesis, to develop a posterior probability:

$$posterior \propto likelihood \times prior \tag{1.3}$$

or:

$$p(a|b) \propto p(b|a) \times p(a) \tag{1.4}$$

In the example above, the *posterior* is the probability that it will be warm, given that it is sunny. This is directly proportional to the probability that it would be sunny if it were warm (the *likelihood*) multiplied by the probability that it would be warm in December (*prior*). The goal would be to estimate the posterior as accurately as possible, so that they wear a coat when it is cold.

There is a wealth of evidence that observers can use prior knowledge to improve perceptual judgements, including in both visual (e.g. Gerardin et al., 2010; Seriès and Seitz, 2013) and auditory (Wang et al., 2019; Sohoglu et al., 2012) domains.

In terms of colour constancy, if observers had a prior for the illuminations that are likely to occur, the posterior would be:

$$p(I|x) \propto p(x|I) \times p(I) \tag{1.5}$$

where I is a given illumination, x is the light reaching the observer's eye, and *p(I)* indicates an illumination prior. Thus, an illumination prior could help to improve estimates of the illumination in a given scene, and constrain the number of likely illuminants.

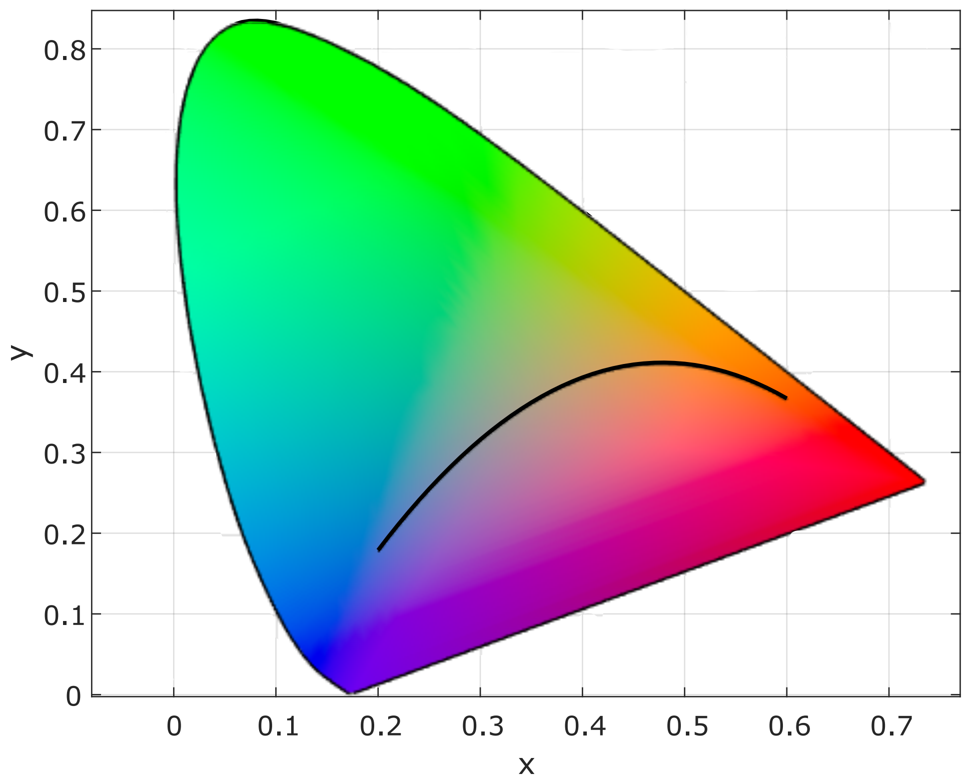


Figure 1.3: Daylight locus (black line) in CIE xy colour space

An illumination prior that has been proposed is a daylight prior. Daylight illuminations fall on a daylight locus which ranges in appearance from blueish (skylight), at high correlated colour temperatures (CCTs), to yellowish (sunlight) at lower CCTs. The chromaticities of the daylight locus are plotted in CIE xy colour space in Fig. 1.3. Judd et al. (1964) first determined the nature of the daylight locus, by combining spectral measurements of daylight (including direct sunlight, and sunlight with skylight) from a number of sources, and calculating the corresponding chromaticities. They found that these chromaticities fell on a line close to the Planckian (Blackbody) locus in colour space, and any chromaticity on this line could be reconstructed using three basis functions. Later, Nascimento et al. (2016), Spitschan et al. (2017), and DiCarlo and Wandell (2000) measured daylight illuminations and found a similar daylight locus.

Brainard and colleagues (Brainard and Freeman, 1997; Brainard et al., 2006) have proposed a Bayesian model in which observers can use a daylight prior to improve their colour constancy. More specifically Brainard et al. (2006) fit a prior to data measured in an earlier paper (Delahunt and Brainard, 2004). They found that a very broad daylight prior along the daylight locus, centred on D65 (typically considered the "average" daylight) explained the psychophysical data well. Many other papers have developed computational models of colour constancy involving a daylight prior, including Liu et al. (2015). In this paper, Liu et al developed a model which combined cues with illumination priors and obtained high levels of colour constancy compared with earlier models using cues alone. This suggests that a daylight prior theoretically can improve colour constancy.

A number of studies have investigated whether people do use a daylight prior, by determining whether they have superior colour constancy for scenes illuminated by daylight, compared to non-daylight, illuminations, with mixed results. For example, Delahunt and Brainard (2004) (mentioned above) recorded observers' achromatic settings in rendered scenes illuminated by either a neutral daylight illumination (D65), or blue, green, yellow or red light equally far away (30 ΔE) from D65 in CIELUV colour space, which is a perceptually uniform space. The blue and yellow lights were on the daylight locus whereas the red and green lights were perpendicular to the daylight locus. When the local surround was held constant under changing illumination, thus providing an invalid cue, observers' colour constancy was greatest under the blue daylight illuminant and worst for red non-daylight illuminant. This suggests that colour constancy is indeed greatest for more natural illuminants. However, colour constancy was higher under the green than the yellow illuminant, despite the yellow illumination being on the daylight locus. A possible explanation is that this is due to experience under foliage which filters sunlight to make it appear green.

Using a similar methodology, Weiss et al. (2017) asked observers to adjust a test patch to appear grey, within a two-dimensional array, simulated under 41 different illuminants of different chromaticities. They found that colour constancy was almost perfect under a blue daylight illuminant, but was worse under illuminants further away from blue. Specifically, they saw a "blue bias", in which achromatic settings under all illuminants were drawn towards the blue illuminants. This suggests that any daylight prior may only apply to blueish daylights and not extend to yellowish daylights.

Bosten et al. (2015) also used achromatic setting to test whether observers have a daylight prior. Unlike the previous studies, they investigated the distribution of achromatic settings made in a void (without any implied illumination). They found that there was the greatest variability in these settings along the daylight locus, suggesting that observers do not detect deviations from neutral along the daylight locus, therefore suggesting better colour constancy. Furthermore, this effect was stronger when the test patch was embedded in a black background than in a white background. This suggests that a daylight prior is more strongly used when there are fewer cues to use (as in the black background condition).

On the other hand, Gupta et al. (2020) has more recently measured achromatic settings and found little evidence for a daylight prior. In this study, observers sat in a "light room", in which the spectrum of the illumination was controlled, and could therefore either be on or off the daylight locus. In this room, observers made achromatic settings at various time points during adaptation. It was found that there was no difference in the time course of adaptation (and correspondingly, levels of colour constancy), for illuminations on or off the daylight locus. Similarly, Brainard (1998) asked observers to make achromatic settings in an immersive environment, under illuminants either on or off the daylight locus, and found no pattern across different illuminants. These studies suggest that observers may not use a daylight prior, or weight it highly enough to have a noticeable effect on colour constancy, in more immersive environments, with many other cues available.

Another way of determining whether observers can use a daylight prior is to measure discrimination thresholds. Higher discrimination thresholds mean that observers are less likely to notice a change in illumination, and therefore imply better colour constancy. Both Aston et al. (2019) and Pearce et al. (2014) employed this method. Aston et al. (2019) found higher discrimination thresholds when the illumination changed from neutral towards a blueish daylight illumination, compared to yellowish (daylight), reddish (non-daylight), or greenish (non-daylight). Pearce et al. (2014) found higher discrimination thresholds along both yellow and blue directions of the daylight locus, although they were higher in the blue than the yellow direction. These both support the notion of a daylight prior, particularly for blueish daylights.

Some studies have found an interaction between a daylight prior and cues. For example, as noted in the ***Specular Highlights*** section above, Yang and Maloney (2001) found an asymmetric effect whereby perturbing specular highlights to a daylight illumination had a strong influence on achromatic settings whereas perturbing to a non-daylight tungsten illumination light did not. Nagai et al. (2017) also found an effect of specular highlights but no effect of illumination, although the illuminants used were not defined as on or off the daylight locus. Furthermore, Delahunt and Brainard (2004) andBosten et al. (2015) found no, or little, difference in colour constancy between daylights and non-daylights when more cues were present. Studies conducted in more immersive environments, with many cues to the illuminant (Gupta et al., 2020; Brainard, 1998) also found no effect of illumination. Taken together, these findings suggest that a daylight prior may not be used as much when other cues can be used.

Overall, some studies have tested whether observers use a daylight prior to improve colour constancy under natural illuminations, with mixed findings. Further studies are necessary to determine the link between illumination and colour constancy, and explore the relationship between priors and cues in colour constancy.

1.4 Development

1.4.1 Development of Colour Constancy

The literature cited above suggests that adults may use a set of cues to the illuminant, and a daylight prior, to assist with colour constancy. However, it is unknown to what extent the ability to perceive surfaces and objects as having a constant colour is learned or innate, or at what age children are able to use the cues and priors discussed above.

Whilst the development of colour constancy has not been studied extensively, there have been a few studies investigating colour and lightness constancy during infancy and early childhood. An early study was conducted by Dannemiller and Hanko (1987), who tested colour constancy in four month old infants, using the preferential looking paradigm. In this, infants were familiarised to a purple surface under incandescent illumination. They were then tested with a novel blue and a familiar purple surface under fluorescent illumination, such that the light reflected from both surfaces was different to the light reflected from the familiarised stimulus. It was found that 12 out of 14 infants preferred the novel (blue) surface, suggesting that they have some rudimentary form of colour constancy which allows recognition of the same surface under different illuminants. A later study conducted by Dannemiller (1989) tested both nine- and 20-week old infants with a habituation paradigm. After familiarisation with a stimulus, the surface reflectance and/or illumination was changed. Whilst nine-week-old infants' looking times increased in response to all changes, 20-week-old infants only looked longer when the surface reflectance changed and ignored changes in illumination alone. This suggests that 20-week-olds, but not nine-week-olds, have some form of colour constancy.

In a more recent study, Yang et al. (2013b) used the preferential looking paradigm, as used by Dannemiller and Hanko (1987), to test colour constancy in four and a half month old infants. In this study, individual infants' luminance contrast was determined and used to make stimuli isoluminant. This ensured luminance could not be used as a clue. After being familiarised to an image of two smiley faces under a neutral illuminant, infants were tested under either a blue or orange illuminant, with one smiley face retaining the same reflectance meaning that the light reflected changed with the illuminant, and the reflectance of the other face changing such that the light reflected from it was the same as during the familiarisation phase. Colour constant infants would expect the reflectance to stay the same, which would result in the reflected light changing. Under both the blue and the orange illumination change, infants looked longer at the face whose reflected light did not change. This either implies that they expected the reflected light to change with the illuminant (due to having colour constancy) and were surprised when it did not; or they perceived the image with a constant reflected light as having a novel reflectance and looked longer due to a novelty preference. Either of these interpretations suggest that they did have colour constancy. However, using the preferential looking paradigm it is impossible to determine whether infants perceived the appropriately changing face as having constant reflectance, or just more constant than the other face.

Chien et al. (2006) investigated the use of two cues for lightness constancy in four-month old infants. Lightness constancy is analogous to colour constancy for achromatic surfaces and illuminants. The initial paradigm was the same as in Yang et al. (2013b) with light and dark grey faces on a white background. In this scenario it was found that the infants did have a novelty preference, thus suggesting lightness constancy. However, when the background was changed from white to black, removing the "brightest is white" cue, infants showed no preference for either stimulus. Preventing use of the local surround as a cue, by changing the local background reflectance between familiarisation and test phases, also eliminated any preference for the novel stimulus. This suggests that four-month-old infants have a form of lightness constancy and can use the brightest is white and local surround cues but cannot compensate when these cues are disrupted.

The study above investigated the use of cues for lightness constancy. Kimura et al. (2010) investigated infants' ability to use a cue to colour constancy: memory for canonical object colour. In this study, five- to eight-month-old infants saw images of items (faces, fruits, and flowers) in their canonical colours alongside hue inverted images of the same objects. Five-month old infants did not differ in their looking times towards the correctly and incorrectly coloured items, whereas all older age groups looked significantly longer at the correctly coloured faces than incorrectly coloured faces. In addition, six and eight-month olds looked longer at correctly coloured fruit than incorrectly coloured fruit. No infants preferred correctly coloured flowers, but these have low colour diagnosticity. These findings suggest that infants as young as six-months old have knowledge of canonical object colour, which they may be able to use as a cue to the illumination for colour constancy.

Whilst no studies have investigated whether infants can use specular highlights as a cue in colour constancy, some have tested whether they can perceive gloss, which may be an important first step to using specular highlights. Yang et al. (2015) used a change-detection paradigm to determine whether three to eight month old infants would notice a change in illumination or material. They used two pairs of stimuli for this task. The first consisted of two glossy snails rendered under different light fields such that the reflected light is different but they appear similar to adult observers due to colour constancy. The second pair contained one of the glossy snails, and one matte snail. Whilst this second pair of stimuli appear different to adults, due to having different materials, they reflect a more similar pattern of light than the first pair. In the change-detection task, infants saw these pairs alternating on one side of the screen with the same image repeatedly shown on the other side. As infants prefer to look at changing stimuli, if they look more to the changing side it suggests that they can detect a difference. They found that three and four month old infants looked longer when the light field was changing than when the material changed whereas seven and eight month olds looked longer when the material changed. As the seven month old infants are less sensitive to light changes, and have looking behaviour consistent with adult perception, this suggests they have some form of colour constancy. Furthermore, they can discriminate between matte and glossy images, which provides an important step towards being able to use specular highlights for colour constancy.

Other studies have also investigated whether infants can perceive gloss. For example, Yang et al. (2011) presented five to six and seven to eight month old infants with two rendered images of gnomes: one matte and the other glossy. They recorded looking times and found no preference for either image in the younger infants but the older infants looked significantly longer at the glossy image. To test whether this was simply due to luminance statistics, they then tested infants' preference for textured matte stimuli with a similar luminance histogram to the glossy stimuli. Again, only seven to eight month olds looked longer at the glossy compared to the textured matte image. Neither age group had a preference for the plain matte or textured matte stimulus. This suggests that by seven months old, infants can perceive glossiness, which is necessary to use specular highlights. Using a similar methodology, Yang et al. (2013a) showed five to six and seven to eight month old infants either a matte green shape next to a matte yellow shape, or a glossy green shape next to a gold shape. The older infants preferred the gold to the glossy green, but had no preference for matte yellow compared to matte green. The younger infants had no preference in either condition. The authors argued that this is evidence that seven month old infants can discriminate yellow from gold. However, the infants were never presented with gold and matte yellow stimuli on the same trial, so it is difficult to conclude anything about their preference for matte vs glossy stimuli from this study.

The above studies have all investigated colour constancy, or gloss perception (which is related to the use of a specular highlights cue), during infancy. One study investigating colour constancy in toddlers was conducted by Witzel et al. (2013). In this, three- and four-year olds' categorical colour constancy – the consistency of category membership for surfaces across changes in illumination – was tested. This age group was selected as they have immature colour categories. To test their categorical colour constancy, coloured Munsell chips were placed on different cut-outs of clothing and children were told that different animals would only wear clothes of one colour. This allowed the experimenters to determine children's colour categories. Observers repeated this sorting task under a neutral, a red, and a green illumination. Children had a greater level of colour constancy in the centre of colour categories than at the edges, similar to adults (Olkkonen et al., 2010), and there was a positive correlation between the consistency maps of children and adults. This suggests that children of this age have similar categorical colour constancy to adults, despite lacking fully developed colour categories.

More recently, Rogers et al. (2020) measured colour constancy in two- to three-year old toddlers. To do this, they presented children with two boxes - one illuminated by a neutral illumination (D65), and the other illuminated by a reddish illumination. A cardboard cutout of a bear was placed in each box. In one of the boxes, the bear was given a pair of trousers. In the other box, observers saw four pairs of trousers and were told to pick the pair which matched the other bear's trousers. The chromaticities of the four pairs of trousers were designed such that one had the same reflectance as the target (perfectly colour constant), one had a different reflectance but reflected the same light as the target (colour inconstant), and the others had chromaticities in between. The task was repeated with the target under the neutral illumination and under the red illumination. From the trousers selected, a colour constancy score could be calculated. They found a range of scores across children, from about 25% constant to perfect constancy, while adults tended to have perfect constancy on this task. This suggests that colour constancy may not be fully developed in most toddlers at this age, although there are large individual differences.

One study which measured colour constancy in older children was anecdotally reported by Katz (1935). It should be noted that, although Katz refers to these as colour constancy experiments, the surfaces tested were achromatic. In this series of experiments, observers viewed two discs under different lighting conditions. One was under daylight while the other was in the shadow of a sheet of paper. The disc in the shadow was the target disc, and observers' task was to adjust the brightness of the other disc to match. In variants of this experimental setup, either one of both of the discs were changed to charts containing 48 shades, from which observers chose which one matched the target. Katz reports on data collected in this set up by Burzlaff and Brunswik on observers from three years old up to adulthood. They found that the nature of the developmental trajectory depended on the experimental setup, such that with two discs, constancy increased with age up to eight to 15 years, after which it decreased with age. However, when both discs were replaced with charts, observers of all ages showed perfect constancy. Whilst these findings are to be taken with caution, as no data is presented, nor statistical analyses reported, they do suggest that children of a certain age could show superior colour constancy to adults with certain stimuli. However, it is clear that the nature of the stimuli is crucial.

Another approach taken to determine the role of developmental experience in colour constancy is to rear animals in unusual lighting conditions, as proposed by Beau Lotto (2004). One such study was conducted by Sugita (2004) who raised Japanese monkeys with only monochromatic lights (which emit only a very narrow band wavelength of light, rather than a full spectrum) until they were one year old. Afterwards, it was found that their brightness matching was intact, but they required extensive training to transfer this ability to colour matching. Furthermore, they completely lacked colour constancy; simply using the colour of the light reaching their eyes to determine surface colour, even nine months after deprivation. This suggests that there is a critical period in which experience with a range of broadband lights is necessary to enable colour constancy.

Similarly, Wagner and Kröger (2005) reared Blue Acara (a breed of fish) under monochromatic lights for a year and found that, although their photoreceptors were relatively intact, much of the neural wiring (e.g. in horizontal cells) was altered. Additionally, their optomotor response, which is a measure of the point of isoluminance between two lights, was affected by the light they were reared in. Although this study did not directly test the fish's colour constancy, it seems likely that it would be disrupted as their sensitivity to different colours was altered. This may help explain why Sugita (2004) found that rearing with monochromatic lights disrupts colour constancy, and further shows the importance of experience for colour vision more generally.

1.4.2 Development of Size Constancy

Whilst research on the development of colour constancy in children, especially on the development of cue-use, is lacking, more studies have been conducted into the development of other perceptual constancies. Considering the development of such constancies is important as they represent the visual system learning to solve a similar problem, and can give some ideas of when colour constancy might be expected to develop. One example is size constancy, which is the ability to recognise objects as having a constant size at different viewing distances, despite changes in the retinal size. As can be seen in Fig. 1.4, we expect objects further away to appear smaller, so when an object has the same projected (retinal) size at different viewing distances, the further object seems bigger. Many studies have shown that this ability is still developing when viewing objects at large distances until around nine years of age, using discrimination learning (Brislin and Leibowitz, 1970), size matching (Granrud and Schmechel, 2006), and size reproduction (Rapoport, 1967). It should be noted that this developmental change does not seem to be present when viewing objects at near distances (Brislin and Leibowitz, 1970; Granrud, 2009; Tronick and Hershenson, 1979), which may be due to the fact that optical cues such as vergence and binocular disparity are reliable enough at near distances to reduce the ambiguity present at further distances.

Káldy and Kovács (2003) used the Ebbinghaus illusion, in which a disk's apparent size is influenced by the size of surrounding disks, to investigate the development of a phenomenon related to size constancy: context integration. It was found that four-year old children were significantly less affected by the illusion than adults. Similarly, Doherty et al. (2010) found that seven year old children are less affected by the illusion than adults. These findings suggest that young children do not have

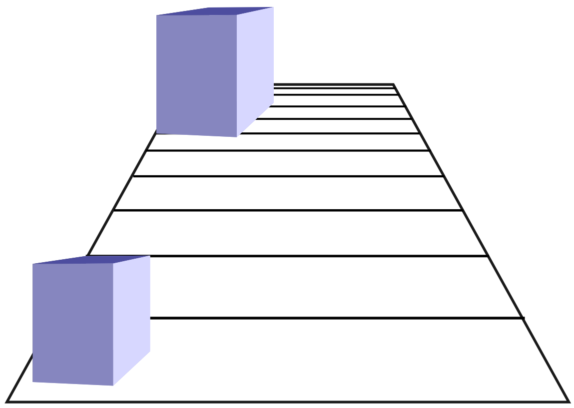


Figure 1.4: The two cubes in this illusion are the same size, yet the further cube appears bigger. This is because size constancy means we expect objects further away to appear smaller.

fully integrated context integration, thus allowing them to perceive the central disk more veridically. This may extend to size constancy, which also requires integration between an object and its surround to perceive its physical size.

Granrud (2009) investigated whether the cognitive cue of knowledge about the size-distance relationship influences size constancy. To test this, five- to ten-year olds were given a size-distance knowledge test to determine whether they were aware of the relationship between distance and perceived size (i.e. that apparent size decreases as distance increases but objects do not change size). The children then viewed a white disc at various distances and were asked to select a disc from a set of comparison discs which matched the test disc in size. There was a positive correlation between error, when matching objects 61 feet away, and size-distance knowledge, even when age was partialled out. This suggests that the reason younger children tend to demonstrate immature size constancy is due to a lack of knowledge about the effect of distance on size and, resultantly, a lack of cognitive compensation for the distance.

On the other hand, Tronick and Hershenson (1979) used the Jastrow illusion to argue that knowledge does not influence size constancy. In this illusion, a small arc is placed above a larger arc such that the smaller arc appears bigger. In this study, children between the ages of three and five years were presented with this illusion and it was pointed out that the smaller arc appeared – but was not – bigger. The children were then split into those who responded based on the actual size of the arcs and those who persisted in responding based on the illusory size. When later asked to match the size of an object viewed nine feet away, there was no difference in accuracy between the groups. Furthermore, using objective (matching the size the object truly is) vs phenomenal (matching the size the object appears to the observer) instructions had no effect on size constancy. This suggests that understanding the difference between apparent and objective size does not influence size constancy. However, it should be noted that the viewing distance in this study was only nine feet, whereas other studies showing an effect of knowledge have employed significantly larger viewing distances (e.g. Granrud, 2009). As noted above, optical cues are sufficient up to a certain viewing distance, above which knowledge may be a more important factor.

Yonas and Hagen (1973) investigated the use of visual, rather than cognitive, cues to size constancy in three- and seven-year old children, and adults. They tested whether observers could use motion parallax and static depth cues. Observers viewed scenes containing two differently sized objects on a textured alley (so they could see which was nearer) and were asked to point to the larger object. The physically larger object was always further away from the observer than the smaller object, so that it subtended an equal or smaller size on the retina. In motion parallax present conditions, observers could move their heads freely, whereas in motion parallax absent conditions, observers viewed the scene through a peephole. To prevent the use of static depth cues, half the observers saw a projection of the scene, rather than the physical scene, meaning the retinal signal indicated the two objects were at an equal distance from the observer. Both three- and seven-year old children chose the retinally larger object, rather than the physically larger object, when viewing a projection of the scene, or when motion parallax was absent, implying poor constancy. In contrast, adults usually picked the physically larger object, regardless of condition. This suggests that children do use motion parallax and static depth cues to size but are less able to inhibit irrelevant cues than adults.

Overall, the size constancy literature suggests that children are able to use cues including knowledge, motion parallax and texture. However, even 10-year olds' constancy is not adult-like. Furthermore, they do not combine cues in an adult-like manner, and different cues develop at different ages, with the more cognitive cues developing later. Whether this finding extends to colour constancy remains an open question.

1.4.3 Development of Shape Constancy

Another perceptual constancy which received some experimental attention in the developmental field in the 1960s and 1970s is shape constancy. Shape constancy is the ability to perceive the shape of objects the same under changes in viewing angle, and was described in a seminal paper by Thouless (1931). As can be seen in Fig. 1.5, when a shape such as a circle is rotated, the shape projected onto the retina changes (into an ellipse in this example). However, observers tend not to judge the physical shape to have changed. Generally, studies into the development of shape constancy have found mixed results, with some finding it increases with age (Kaess et al., 1974); some finding no change (Field and Collins, 1977; Myambo, 1972); and others finding decreasing constancy with age (Meneghini and Leibowitz, 1967).

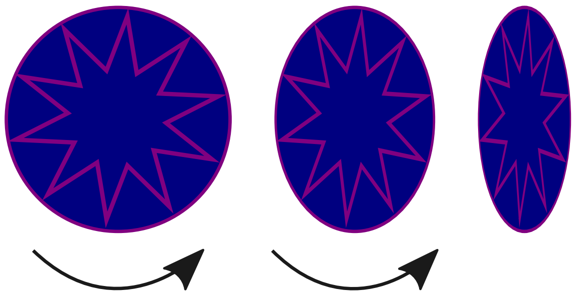


Figure 1.5: As a circular disc is tilted, the projected shape changes. The ability to identify the disc on the right as the same as the disc on the left is called shape constancy.

Kaess et al. (1974) presented observers between four and 75 years old with a target rectangle either near (91cm) or far (457cm) away, which was tilted either 25°or 75°, such that the shape projected onto the retina was different to the shape of the physical object. Observers' task was to pick a rectangle from a set of comparison shapes, which had the same shape as the target. When the targets were viewed from a far distance, it was found that four year old children had significantly poorer shape constancy than 11 year olds, who in turn had worse constancy than 19 year olds. However, when viewing the objects at a near distance, there was no effect of age.

This suggests that shape constancy, for objects seen far away, is still developing up to 19 years.

In contrast, Field and Collins (1977) measured shape constancy in six, eight, 10, 12, and 19 year olds and found no significant effect of age. In this study, observers viewed a target ellipse from 170cm away, which was tilted 39°, 57°, or 72°, and adjusted a comparison ellipse to match in shape (but not size). Similarly, Myambo (1972) measured shape constancy in a tribe from Malawi, and found no significant effect of age for observers aged five to 20 years. In this study, the target and comparison ellipses were viewed from 91cm away - the "near" distance tested by Kaess et al. (1974). Therefore, although these two studies found no effect of age on shape constancy, this could be due to all stimuli being viewed from relatively short distances.

Interestingly, Meneghini and Leibowitz (1967) found *decreasing* shape constancy with age. In this study, observers aged between four and 19 years viewed a test circle at a distance of either 91cm or 457cm, as in Kaess et al. (1974). This test circle was tilted either 10°, 28°, 45°, or 90°. These were the same angles of inclination used by Myambo (1972). Observers were then asked to choose from a set of comparison ellipses which matched the target. At a viewing distance of 91cm (near), younger children chose comparison ellipses closer in shape to the physical target ellipse than adults, who had a tendency to pick comparisons closer to the retinal image. This means that younger children had a higher degree of shape constancy than older children and adults. However, at a viewing distance of 457cm (far), all observers had poor constancy, and there was no change with age.

Measuring shape constancy in infants, Day and McKenzie (1973) presented eight and 14 week old infants with either a cube repeatedly viewed from one angle; the same cube from changing angles; or photographs of the cubes from different angles. They found that infants habituated to both the cube at a single angle, and the cube from changing angles, as shown by a decrease in looking time over successive trials. From this, they concluded that infants perceive the cube as constant despite changes in viewing angle. However, infants did not habituate to the photographs, suggesting that they rely strongly on three-dimensional cues, rather than picture cues, for shape constancy.

Taken together, the research into the development of shape constancy in four to 19 year olds has yielded mixed results, with some finding an intuitively surprising outcome of poorer constancy with age. A possible explanation for the discrepant findings is the distance at which stimuli were viewed. When stimuli are viewed from large distances, shape constancy seems to decrease, or not change, with age, whereas for nearby stimuli, shape constancy may increase with age. This difference could be due to different cues being used at different distances, as for size constancy.

1.4.4 Development of Priors

A recent study conducted by Skelton, Franklin and Bosten (in prep) has measured the use of a daylight prior in four to six month old infants. In this study, observers were eye tracked whilst viewing a computer monitor. A target showed up on the monitor, and if observers successfully detected it (shown by looking at it), a smiley face was shown. The targets varied in both hue and saturation, allowing the experimenters to determine detection thresholds for colours both on and off the daylight locus. Both infants and adults had higher discrimination thresholds for targets aligned with the daylight locus. In previous studies conducted with adult observers, higher discrimination thresholds have been linked to higher colour constancy. Therefore, this finding suggests that by four months old, infants may already have superior constancy for daylight illuminations.

Whilst there is little research into the development of prior-use in colour constancy (such as a daylight prior), the development of some other priors has been studied. For example, when judging whether an ambiguous shape is convex or concave, adults have a light-from-above prior, which may be stronger for above-left (Adams, 2007), and a convexity prior (Langer and Bülthoff, 2001). Stone (2011) tested whether four to 10 year old children have a light from above prior in determining whether naturalistic and symbolic objects are convex or concave. They tested this by presenting images both the right way up and upside down, such that using a light from above prior would imply one was convex and the other concave. They found that observers were significantly more likely to assume a light from above prior for naturalistic, compared to symbolic stimuli, and that the tendency to use this prior increased with age. However, all observers, except four to five year olds judging symbols, used the light from above prior to some extent. This suggests that this prior is present in young children but increases in strength up to at least 10 years.

Similarly, Thomas et al. (2010) tested whether children from four years, and adults, used a light from above and/or a convexity prior. Observers were presented with a polo mint stimulus, in which the segments appears either convex or concave, depending on the assumed lighting direction. In some trials the convexity and light from above prior pointed to the same interpretation whereas on the other trials, these priors were in conflict. They found that all observers used both priors when in agreement. However, when these priors were conflicting, younger children put more weight on the convexity prior while older children and adults put more weight on the light from above prior. This study suggests that young children can use lighting priors to influence their perception, but their weighting of competing priors is not adult-like until they are much older.

Pickard-Jones et al. (2020) also investigated the development of priors for convexity/concavity. In this study, five to 11 year old children saw a honeycomb stimulus in which the light/dark lines signalled concave or convex segments depending on assumed lighting. This stimulus was shown at 12 different orientations so that observers' judgement of convexity would vary by orientation if they were using a light from above prior. They found that younger children were less sensitive to the orientation than older children. Of the individuals who were sensitive to the orientation, there was no significant effect of age on strength of the light from above prior. This builds on the findings from Thomas et al. (2010) that younger children are more likely to assume convexity over light-from-above, shown by being less sensitive to the orientation. Taken together, these three studies suggest that young children can use illumination priors, although they are still developing with age.

Other studies investigating the development of prior use have taught children priors during the course of the experiment. For example, Bejjanki et al. (2020) presented six and seven year old children with a cluster of dots on a computer screen, which were supposed to represent previous guesses of the location of a bucket. The children's task was to guess where the hidden bucket was located by pressing on the screen. The true location of the bucket on each trial was drawn from an underlying prior distribution, which meant it was more likely to be in some locations than others. They found that children could learn this prior distribution, and use it to improve their guesses of the bucket's location, in a similar manner to adults. However, when there were two different prior distributions, children could not integrate the prior knowledge with the current sensory information. This suggests that children can learn and use simple priors, but struggle when there is too much information.

Chambers et al. (2018) used a similar task to test whether six to 11 year olds could use a prior distribution, and combine it with sensory information in a Bayes optimal fashion. In this task, unlike in the previous study, the prior distribution was shown on the screen during each trial so observers did not need to learn it. There were two priors with different variances: a narrow one and a wide one. If observers were using these priors optimally, they would assign more weight to the narrow prior than the wide one. It was found that the youngest age group (six to eight year olds) did not use the prior differently depending on its variance. However, all observers' data were better fit by a Bayesian model (using both prior and sensory information) than any alternative models tested, suggesting that even the youngest children were using the prior to some extent. The parameters of this model were less optimal in younger children and improved with age. Both this study and Bejjanki et al. (2020) suggest that young children can learn and use a prior, although their use is not as optimal as in adults.

Taken together, the studies into the development of prior use in children suggest that children can use priors, but not in an adult-like manner until at least 10 years old. Whether children can use a daylight prior for colour constancy remains unknown.

1.5 Methods

As alluded to in the previous sections, many methods have been used to measure colour constancy in adult observers. One reason that colour constancy has not been investigated in children is that many of these methods are inappropriate for testing children. These methods include achromatic matching, asymmetric matching, measures of operational colour constancy, and object selection.

Achromatic matching (Brainard, 1998) involves setting a patch to appear neutral, or grey. Assuming that there are no internal biases, this setting of grey is a proxy for observers' perception of the illumination, as a neutral surface will reflect the chromaticity of the illuminant. If the setting is close to the chromaticity of the illumination, it suggests the observer has good colour constancy whereas if it is far away, it suggests they have poor constancy.

Asymmetric matching (Brainard et al., 1997) is similar to achromatic matching, in that observers set a patch to appear a certain colour. However, instead of matching to an internal representation (grey), they are presented with a target patch under one illumination and are asked to adjust a test patch seen under a different illumination to the same colour. Good colour constancy would be achieved by setting the test patch to have the same reflectance as the target, despite it being seen under a different illumination. Asymmetric matching can involve the two illumination conditions being presented simultaneously, side-by side; successively; or haploscopically so that each eye sees one illumination and can be fully adapted.

Operational colour constancy is the ability to distinguish between a surface reflectance change and and illumination change (Craven and Foster, 1992) . This is relatively straightforward to measure as observers simply have to say whether the illumination or the reflectance has changed between trials. This is also a more ecologically valid measure than matching, as colour constancy in the real world is used to identify objects as unchanging under different lighting conditions.

Another more realistic way of measuring colour constancy is object selection (Radonjić et al., 2015b). In these tasks, observers see a target stimulus under one illumination and pick a matching object from an array seen under a different illumination. This is similar to asymmetric matching but considerably more simple for observers to complete, and does not require fine motor skills. This is therefore a more appropriate way of measuring colour constancy in children.

1.6 Outstanding Questions

Although some studies have investigated whether adults can use specular highlights as a cue for colour constancy, they have typically not considered how the ability to use this cue may interact with other factors such as validity of other cues, or type of illumination. Furthermore, many studies investigating the use of specular highlights have lacked the power to conduct formal statistical analyses. Therefore, this thesis contains three experiments to answer the outstanding questions: does the weight applied to the specular highlights cue depend on the presence of other cues? Does the weight applied to specular highlights depend on the type of illumination (natural or unnatural)?

Whilst many computational models have been developed to incorporate a daylight prior to improve colour constancy, the research into the use of such priors in behavioural studies has been inconclusive. Whilst some studies have found superior constancy under daylights in certain conditions, others find no benefit. A possibility for this discrepancy is that observers rely more on a daylight prior when other sources of information (such as from cues) is lacking. In all six experiments presented in this thesis, both daylight and non-daylight illuminations are used. This allows us to ask (a) whether adults can use a daylight prior with many cues present and (b) whether adults' use of this prior varies depending on the presence of other cues.

Research into the development of colour constancy in children is severely lacking. Although there have been some studies investigating the development of colour constancy, most of these have studied either infants or non-human animals. The only studies investigating colour constancy in toddlers have tested categorical colour constancy (Witzel et al., 2013) or the link between colour constancy and colour knowledge (Rogers et al., 2020), rather than explicitly testing the degree of colour constancy. No studies have measured colour constancy in older children. Therefore, in this thesis we aim to determine at what age children develop adult-like colour constancy.

Furthermore, no studies have tested at what age children can use the cues to colour constancy discussed above and, if so, whether they use them in an adult-like manner. It therefore remains unknown whether these cues are learned or innate. Whilst it has been proposed that the nature of the illumination can have an influence on the degree of colour constancy in adults, the development of a daylight prior remains to be investigated. As children have less experience with daylights than adults, it is plausible that they will be less influenced by the type of illumination in a scene. In the last three experiments of this thesis, we explore how use of both a daylight prior, and specular highlights cue develop during childhood. These studies can inform of the relative importance of nature vs nurture in visual perception, as well as help to understand how children's visual world may differ from that of adults.

Preface to chapter 2

As there was little research into the use of specular highlights in colour constancy in adults previously, the first goal was to determine whether we could find evidence of adults using specular highlights. This was necessary to do before testing children, as it would be unethical to test children if we did not expect to find use of specular highlights in adults, especially given that children are harder to recruit. In the following series of experiments, we created three-dimensional rendered scenes containing either matte (no highlights) or glossy (containing specular highlights) objects, to test whether adults would have better colour constancy with specular highlights.

Chapter 2, which follows, has been published in Journal of Vision: Wedge-Roberts, R., Aston, S., Beierholm, U., Kentridge, R., Hurlbert, A., Nardini, M., & Olkkonen, M. (2020). Specular highlights improve color constancy when other cues are weakened. Journal of vision, 20(12), 4-4.

Chapter 2

Specular highlights improve colour constancy when other cues are weakened

Previous studies suggest that to achieve colour constancy, the human visual system makes use of multiple cues, including a priori assumptions about the illumination ("daylight priors"). Specular highlights have been proposed to aid constancy, but the evidence for their usefulness is mixed. Here, we used a novel cue-combination approach to test whether the presence of specular highlights or the validity of a daylight prior improves illumination chromaticity estimates, inferred from achromatic settings, to determine whether and under which conditions either cue contributes to colour constancy. Observers made achromatic settings within three-dimensional rendered scenes containing matte or glossy shapes, illuminated by either daylight or non-daylight illuminations. We assessed both the variability of these settings and their accuracy, in terms of the standard colour constancy index (CCI). When a spectrally uniform background was present, neither CCIs nor variability improved with specular highlights or daylight illuminants (Experiment 1). When a Mondrian background was introduced, CCIs decreased overall, but were higher for scenes containing glossy, as opposed to matte, shapes (Experiments 2 and 3). There was no overall reduction in variability of settings, and no benefit for scenes illuminated by daylights. Taken together, these results suggest that the human visual system indeed uses specular highlights to improve colour constancy, but only when other cues, such as from the local surround, are weakened.

2.1 Introduction

2.1.1 General Overview

Observers generally have little difficulty judging objects as having a relatively stable colour under changes in illumination - an ability termed colour constancy. However, it is still unknown to what extent they are able to make use of specular highlights to assist in this. In this series of experiments, we tested whether observers' colour constancy improves when specular highlights are present. We also investigated whether judgements are less variable and more accurate along the daylight locus, consistent with the use of daylight priors, and whether these would interact with specular highlights.

2.1.2 Colour Constancy and Cue Use

Perceiving objects to have a stable surface colour under changing illuminations is a remarkable computational challenge that our brains solve every day. This problem is demonstrated in Fig. 2.1. In Fig. 2.1a, a purple flower is seen outside under daylight (D65). The light reaching an observer's eye is the product of the spectral power distribution of the illumination and the spectral reflectance of the flower. Moving the flower inside under a tungsten light bulb (illuminant A; see Fig. 2.1b) changes the spectrum of light reaching the observer's eye. Whilst this changes the relative photoreceptor activation in the retina, our observer would probably not judge the flower to have changed colour appreciably. The extent to which the colour appearance of the flower changes depends on multiple factors, including the extremeness of the illumination change, the duration of exposure and adaptation to each illumination, the surface properties of the flower, and its context. Yet, in everyday life, people would tend to attribute a constant colour to the flower despite changes in the light it reflects caused by changes in illumination.

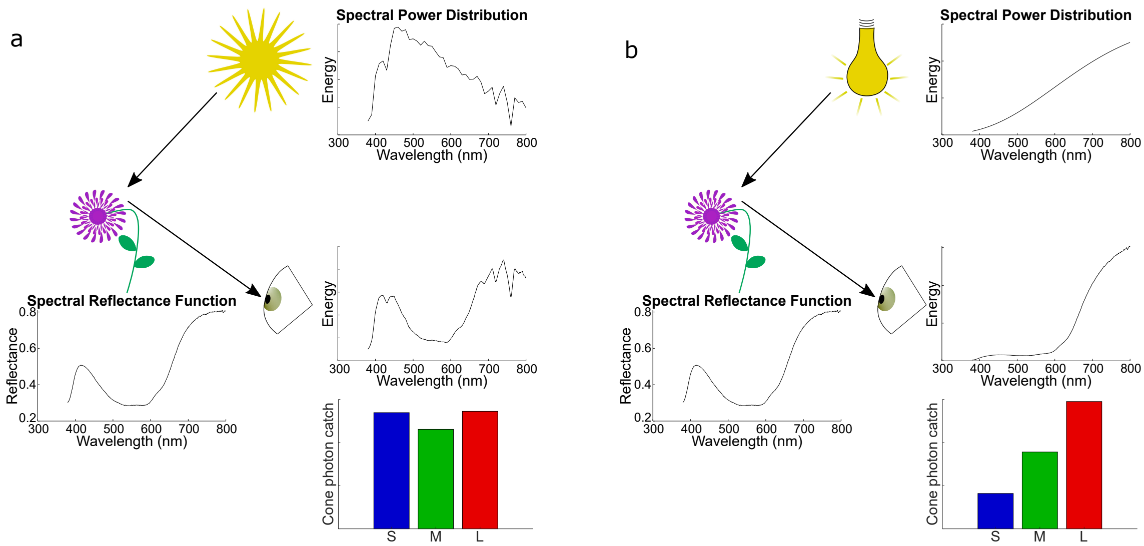


Figure 2.1: a. Spectra of light reaching observer from a flower under D65; b. Spectra of light reaching observer from the same flower under illuminant A. Relative cone activations for each scenario are shown at the bottom.

As the signal reaching the eye confounds illumination with reflectance information, our brains must compensate for any change in illumination in order to create stable perceptions corresponding to the object's invariant surface reflectance properties. This is computationally difficult as an infinite number of combinations of reflectances and illuminations can yield the same photoreceptor activation, and surfaces which are metamers (appear identical) under one illumination may not be metamers under a different illumination (Logvinenko et al., 2015). However, given the constraints of natural illuminations and surfaces, the problem becomes more tractable (see Hurlbert, 1998, for a review of computational models).

Many computational color constancy algorithms have been proposed which all incorporate some form of illumination estimation from regularities in the input image (for reviews, see Smithson, 2005; Lee and Plataniotis, 2014). Although most of the algorithms have not been explicitly proposed or tested experimentally as models of human visual perception, a small number of studies suggest that certain proposed cues do influence perceived illumination chromaticity, such as local surround (Valberg and Lange-Malecki, 1990), the global mean chromaticity ("grey world") (Land, 1983, 1986) or the chromaticity of the brightest surface (e.g. "brightest is white", Hurlbert, 1998). Kraft and Brainard (1999) demonstrated that these three cues can all be used in combination and that none of them is alone sufficient to account for colour constancy performance.

Specular highlights may also be used to gain information about the illumination chromaticity. Perfectly specular highlights are perfect reflections of the incident illumination and therefore have the same chromaticity as the illumination. However, most objects in the real world are not purely specular. A model proposed by Lee (1986) and D'Zmura and Lennie (1986), coined chromaticity convergence (Hurlbert, 1998), explains how partially specular objects (more precisely, optically inhomogeneous materials) can assist in estimating illumination chromaticity. The light reflected by partially specular objects has two components: a diffuse component, in which light has been scattered throughout the material before being reflected back - typically referred to as diffuse reflectance, or "colour" - and a specular component. From a single object, the reflected light is a mixture of these two components, and therefore its chromaticity falls on a line in colour space between the diffuse and specular components. When there are two or more surfaces present, these lines would point towards the illumination's chromaticity and, if extended, they would intersect at the chromaticity of the illumination (see Fig. 2.2). It should be noted that the specular highlights do not need to be visible for chromaticity convergence; as long as there are multiple partially specular surfaces with different diffuse components present, the chromaticities will form lines towards the illumination chromaticity.

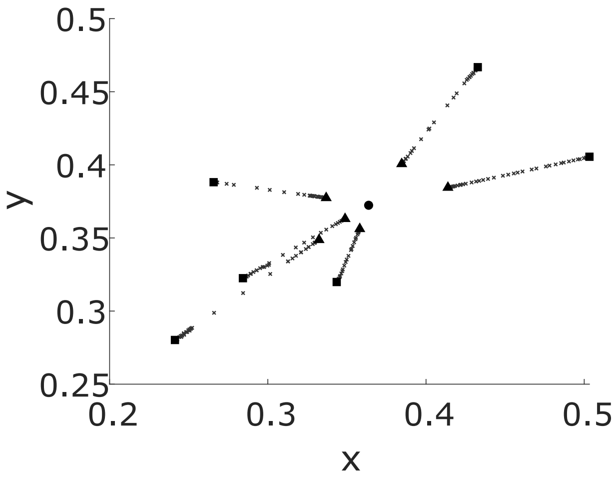


Figure 2.2: Chromaticity in CIE xy colour space, taken from six objects with different reflectances under the same illumination. Grey +s show individual pixel chromaticities and fall on lines between the diffuse reflectance (squares) and specular reflectance (triangles). The black circle shows the illumination chromaticity. As can be seen, the chromaticity of the light reflected from the six objects converges towards the illumination chromaticity.

2.1.3 Research into the Use of Specular Highlights

In one of the few studies into the use of specular highlights, Yang and Maloney (2001) rendered a set of specular spheres under two illuminations (D65 and A). They then swapped the highlights from scene A with those in scene B and vice-versa. The observer's task was to adjust the colour of a patch in the scene so that it appeared to be grey. These achromatic settings shifted towards the illumination chromaticity when perturbed highlights signalled a daylight illumination (D65), but not when they signalled a non-daylight illumination (A). Furthermore, specular highlights were only used when the test patch was on one of the specular objects in the scene; not when it was on the back wall. This suggests that observers segment scenes into different illumination frameworks, which could be achieved based on the depth plane in which an estimate is being made (see Werner (2006) and Snyder et al. (2005)).

Yang and Shevell (2003) also perturbed specular highlights in an asymmetric matching task under D65 and A, and found observers' colour constancy to improve with consistent highlights. Additionally, Yang and Shevell (2002) found better colour constancy for scenes containing specular highlights compared to scenes without. Lee and Smithson (2016) found that observers can use specular highlights, even without any other information, for operational colour constancy (distinguishing a reflectance change from an illumination change). Xiao et al. (2012) found better color constancy for glossy, compared to matte, shapes but only when other cues were reduced. Measuring lightness constancy, Boyaci et al. (2006) found that some (two out of six) observers could use highlights on spheres in a void to determine an illuminant's direction and brightness. Taken together, these few studies suggest that observers can use specular highlights to improve their colour and lightness constancy under certain conditions.

2.1.4 Cue Combination

In non-colour-related domains, observers are able to optimally combine multiple cues to improve estimates of ambiguous stimulus properties (Ernst and Banks, 2002; Ernst, 2006) as shown by decreased variable error compared to estimates made using single cues alone. Optimal integration, according to Maximum Likelihood

Estimation and Bayesian models, means that the reliability of an estimate based on multiple cues is the sum of the reliabilities of each individual cue (Van Beers et al., 1999; Landy et al., 1995). Thus, having more cues available should increase the reliability, thereby decreasing the variability. So far, none of the research into cue use in colour constancy has considered this model of optimal cue integration, instead focusing only on colour constancy indices. Here, we ask whether adding a specular highlight cue decreases the variability of achromatic settings in addition to improving colour constancy.

2.1.5 Daylight Priors

Bayesian models also predict that combining priors (previous knowledge) with current information should aid in the perception of ambiguous stimuli by increasing accuracy and decreasing variability of estimates (Knill and Richards, 1996). Colour constancy may benefit from a prior for daylight illuminations (Brainard et al., 2006). Daylight illuminations are broadband mixtures of sunlight and skylight, and vary in a regular, predictable way. Their chromaticities fall along a curve - the "daylight locus" (Judd et al., 1964; Spitschan et al., 2017) - in the chromaticity plane, ranging in appearance from orangeish to blueish.

Research tentatively supports the notion that observers have a prior for daylights, although the prior seems to be asymmetric such that the improvements are usually found for bluish illuminations. Pearce et al. (2014) found lower sensitivity to changes in illumination along the daylight locus than along the opposite reddish-greenish direction, implying better colour constancy for daylight illuminations, especially in the bluish direction (see also Radonjić and Brainard, 2016; Radonjić et al., 2018; Aston et al., 2019). Additionally, Weiss et al. (2017), using a different technique (achromatic adjustment) found the highest degree of colour constancy under a blueish daylight illumination and achromatic settings made under all other illuminations were skewed towards blue chromaticities. Finally, Delahunt and Brainard (2004) found the highest color constancy for bluish daylight illuminations, but the effect was robust only when the local contrast cue was silenced (also see Brainard et al., 2006). As noted above, Yang and Maloney (2001) found that specular highlights were only used as a cue to the illumination when they signalled a daylight, and not when they signalled a tungsten illumination. This, together with the finding by Delahunt and Brainard (2004) suggests that the effect of a daylight prior interacts with the effect of cues, such that the prior is used when cues are weakened, or, in the case of Yang and Maloney (2001), conflicting.

2.1.6 Current Study

Taken together, previous research suggests that the visual system may be able to use specular highlights and daylight priors to support colour constancy in certain circumstances. However, there is little data on the interaction of a specular highlights cue with a daylight prior. Furthermore, the research into both areas is limited and much of it lacks the power to conduct formal statistical analyses or to quantify the benefit of either. Therefore the contexts in which specular highlights may be useful remain unclear. Here, we take a novel cue combination approach to ask whether, in addition to supporting color constancy, specular highlights help decrease variability of illumination estimates. We characterized color constancy with achromatic adjustment and used the equivalent illuminant modeling approach to obtain observers' illumination estimates for scenes containing matte or glossy shapes, under illuminants either on or off the daylight locus.

2.1.7 Hypotheses

1. Specular highlights will be used as a cue to the illumination for colour constancy, shown by improved colour constancy and decreased variable error for scenes containing specular highlights compared to scenes without.

2. The effect of specular highlights will be mediated by (a) the type of illumination, such that observers will use highlights more when they signal a daylight (blueish/ yellowish, but potentially more so for blues), and (b) position of achromatic matching patch, such that observers will only use the highlights when making settings on a specular object that lies in the same illumination framework.

3. Observers will show a higher degree of colour constancy for scenes illuminated by daylights (particularly blues). This effect will be mediated by the presence

of other cues so the daylight prior will be relied upon more for scenes with no specular highlights, and an invalid or biased local surround cue.

2.2 Experiment 1

To test our hypotheses, we asked observers to set a patch in a 3D rendered scene to look grey (achromatic adjustment) when the objects were either matte (and therefore had no specular highlights), or when they were glossy (and had a valid specular highlights cue). The scenes were illuminated by either blueish or yellowish illuminants on the daylight locus, or reddish or greenish illuminants off the daylight locus.

2.2.1 Methods

Observers

Fourteen paid volunteers aged between 20 and 36 years (mean age = 24.6; 9 females) took part in this study. All observers were naïve to the purpose of the study. All observers had normal colour vision, as screened by Ishihara plates (Ishihara, 2006), and all had normal or corrected to normal visual acuity.

Materials and Apparatus

Stimuli were presented on a 10-bit ASUS PA382Q 23" monitor controlled by a Nvidia quadro k600 graphics card. The monitor was calibrated by generating a gamma-corrected lookup table (LUT) converting XYZ to RGB, accounting for output non-linearities, based on the monitor primaries recorded with a Konica Minolta CS-2000 spectroradiometer (Minolta, 2008). Observers sat approximately 60cm from the monitor and viewed the screen binocularly with free head movement. At this distance, the screen subtended 41 $\times$ 23 degrees of visual angle. The computer presenting stimuli was changed after the first ten observers as the original PC malfunctioned. The monitor was recalibrated following this change.

Stimulus Generation

Three unique scenes were pre-generated, each having a different arrangement of six different three-dimensional shapes. Each shape in each scene had a different surface reflectance, except the central shape which kept the same surface reflectance in each scene. For each scene, two versions were generated - one glossy and one matte. Each of these scenes was shown under five different illuminations (neutral, red, green, blue, and yellow). This resulted in a total of thirty unique stimuli.

The three-dimensional scenes were created using Blender (`https://www.blender.org/`) and rendered using Mitsuba (`http://www.mitsuba-renderer.org/`), compiled for spectral rendering with 30 spectral bands. They were then scaled, such that each scene had a mean luminance of 60cd/m^2, which ensured most pixels were in gamut. The scenes were tone mapped by clipping any out of gamut values (negative RGB values were clipped to 0 and RGB values greater than 1 were clipped to 1), to prevent saturation at high luminance levels. The pixels which were clipped in each glossy scene are shown in the Supplementary Material (Fig. 2.15). Most were on green/blue shapes, and none of the pixels making up the specular highlights were truncated. Clipping did not significantly shift the chromaticities, as can be seen in the Supplementary Material which shows the mean scene chromaticities against the illumination chromaticities. Finally, the scenes were converted to RGB images using the LUT constructed from the calibration described above.

The scenes consisted of six shapes sitting in a box with grey walls (see Fig. 2.3). The background was the brightest surface in the scene with a flat spectral reflectance of 0.8. This ensured the background could be used for the "brightest is white" cue, even when specular highlights were present. The specularity of the background was 0.5 and the alpha (a measure of roughness) was 0.05 - slightly rough. A square area light, the same size as each wall of the room, was positioned just behind and above the camera, angled towards the junction of the two vertical walls. This produced a diffuse light across the scene.

Fourteen unique naturalistic looking "blobby shapes" were generated for the scenes using ShapeToolbox (Saarela, 2018). Each shape was created using a sphere base shape with five sets of twenty randomly positioned bumps. For each set of bumps, the amplitude was randomly selected with an upper limit of 0.5 radians,

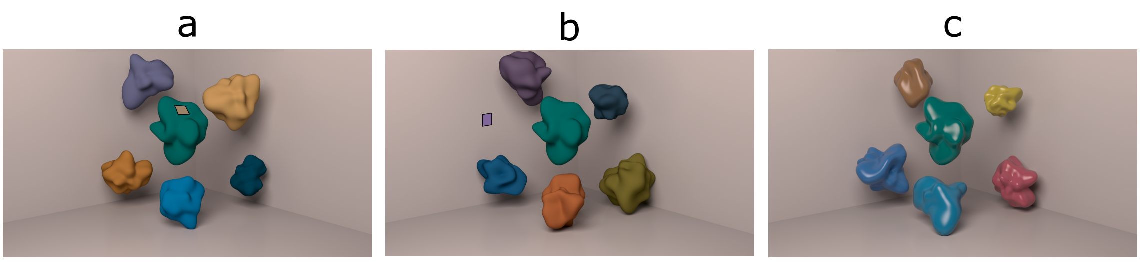


Figure 2.3: Example of matte scenes containing matching patch on shape (a) and wall (b). Example of a glossy scene (c).

and a SD between 0.2 and 0.3. In addition, each of the fourteen shapes were modulated using four sinusoidal modulations with a random frequency below 10 cycles/2π radians and an amplitude between 0.05 and 0.2 radians. One of these shapes was selected to be the standard shape, which was always in the centre of the room, and had a surface reflectance of Munsell chip 5BG4/6. This shape subtended approximately 8 degrees of visual angle. The reflectance functions of each Munsell chip used were retrieved from `https://www.uef.fi/web/spectral/munsell-colors-matt-spectrofotometer-measured`. The standard shape's reflectance was consistent over all stimuli to reduce any noise between trials, and the shape was selected to have a flat surface for the matching patch to sit on.

Three unique scenes were generated from combinations of the 14 unique shapes. Scene acted as a random variable, as we were interested in effects common to different shape, surface reflectance, and location combinations. In each scene, the five shapes other than the standard shape were taken randomly from the set of 13 remaining shapes, and randomly rotated so that a different side was viewed in each scene. The shapes were positioned randomly within a region which ensured no overlap so all the shapes were visible.

For each of the three scenes, two versions were created using the same arrangement of shapes - one glossy and one matte. Within each pair, the corresponding glossy and matte shapes had the same diffuse surface reflectance but different levels of specular reflectance. Despite the presence of specular highlights on the glossy shapes, the mean chromaticity across scenes did not vary much between matte and glossy scenes (see Supplementary Material Fig. 2.14 for details). Both matte and glossy materials were created using the plastic material in Mitsuba. The roughness (alpha) of both materials was set to 0.1, which is relatively rough. This roughness meant that the specular highlights on the glossy shapes were not too bright to be presented on the monitor. For the matte material the specular reflectance was 0, meaning they were perfectly matte and contained no specular highlights. The glossy material had a specular reflectance of 1, which ensured the shapes had specular highlights. The surface reflectance of the six shapes within each scene was chosen such that the average surface reflectance was neutral (not differing by more than 1ΔE_{u*v*} from a flat surface reflectance). Each of the three scenes contained a different set of six surface reflectances to ensure any findings were not specific to the stimuli used. The Munsell chips used in each of the three arrangements are given in Table 2.8 in the Appendix.

The illuminations selected comprised a neutral D57 (CCT 5698), which has been proposed to be the mean of all daylights (Nascimento et al., 2016), and four other illuminations, all 30ΔE_{u*v*} away from neutral (using the neutral illumination as the white point). These illuminations included two along the daylight axis in the blue and yellow directions and two illuminations with chromaticities perpendicular to the daylight axis at CCT 5698K in CIE L'u'v' space. These illuminations will hereafter be referred to as 'neutral', 'blue', 'yellow', 'red', and 'green' respectively. See Table 2.1 and Fig. 2.4 for chromaticity co-ordinates of the illuminations used in CIE Yxy and CIE L'u'v' colour space.

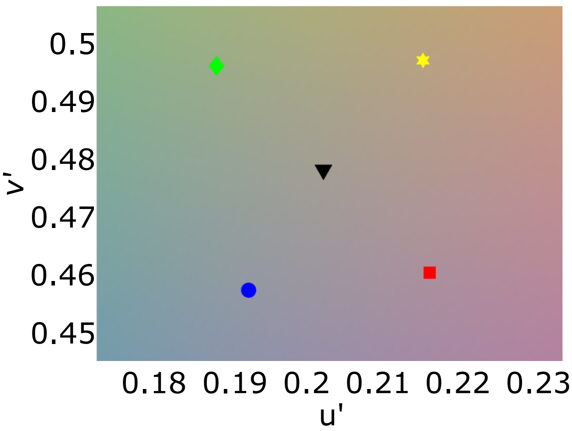


Figure 2.4: Illumination chromaticities in L'u'v' space

Both matte and glossy versions of the scenes were rendered under all five illuminants for a total of 10 scenes, each with three different arrangements of shapes and reflectances.

Task

We used an achromatic adjustment task in these experiments, as this has been used extensively to measure colour constancy (Werner, 2007; Boyaci et al., 2004; Yang

Illumination	x	y	u’	v’
Neutral (D57)	0.328	0.344	0.203	0.478
Blue	0.297	0.314	0.193	0.458
Yellow	0.364	0.372	0.216	0.497
Red	0.329	0.310	0.217	0.460
Green	0.327	0.382	0.189	0.496

Table 2.1: Co-ordinates of the illuminations used in CIEYxy and CIEL’u’v’ colour spaces

and Maloney, 2001; Delahunt and Brainard, 2004; Brainard, 1998), and Speigle and Brainard (1999) have shown that results from achromatic adjustment agree with those from asymmetric matching. Observers were shown a scene with a square patch (approximately 1.4 degrees of visual angle) either on the central shape in the scene (see Fig. 2.3a), or on one of the background walls (Fig. 2.3b). The starting chromaticity of the patch was chosen randomly from an area within $38\Delta E_{u*v*}$ from neutral (D57). The luminance of the matching patch was fixed at 60cd/m^2, to match the mean luminance of the scenes. Observers adjusted the colour of this patch until it appeared grey, using buttons on an Xbox controller which altered the CIE u*v* colour co-ordinates. There were three different step sizes such that observers could shift the chromaticity by 10 (big), 5 (medium), or 1 (small) ΔE in the u* or v* directions. Observers were free to choose which step sizes to use. The instructions given to observers were adapted from Radonjić and Brainard (2016), which emphasised considerations of reflectance properties, and are thus aimed to optimise colour constancy. These are given in the appendix.

Procedure

Each of the illumination-specularity combinations was presented nine times within a session, for a total of 90 trials per session. This consisted of three repeats of each of the three unique scenes. Half the observers completed the first session with the patch on the wall followed by the second session with the patch on the shape; the other half did the sessions in reverse order. Each session lasted approximately one hour. The two sessions were separated by a gap ranging from one day to 22 days between observers.

Observers were first dark adapted for two minutes. They were then given five practise trials in which they adjusted the colour of a circular patch on a black background to appear grey, green, blue, yellow, and red. This was to ensure observers understood how the controllers worked. They were then presented with an empty scene, consisting of the box with grey walls and floor but no shapes, under one of the five illuminants. The order of the illumination colours was randomised between observers. Following two minutes of adaptation to an empty room, observers made achromatic settings for the eighteen trials under this illumination colour (9 glossy; 9 matte). For each illumination, the nine matte scenes were presented followed by the nine glossy scenes or vice versa. This order was randomised for each illumination and observer. After completing all trials under one illuminant, observers were adapted to the next illumination for two minutes before completing all trials under that illuminant, until all conditions had been completed. Although two minutes is not sufficient for complete global adaptation, each illumination condition lasted approximately ten minutes following adaptation, meaning most achromatic settings were made after at least five minutes of adaptation. By this time global adaptation would have stabilised.

Data Analysis

For all data analysis, settings were converted from L*u*v* to L'u'v'. Both colour spaces are designed to be perceptually uniform (that is, any two points at a fixed distance from each other will theoretically be approximately equally discriminable anywhere in the colour space), but u* and v* vary with changes in L* whereas u' and v' remain constant for any value of L' (Wyszecki and Stiles, 1982). Therefore, conducting analysis in L'u'v' ensures that the chromaticity does not depend on luminance.

Exclusion Criterion

In 95% of trials pooled across observers, at least four adjustments (button presses) were made to the patch colour before a setting was submitted as grey. We called any trial with fewer than four adjustments an outlier and removed it from the analysis. After exclusion, there remained at least seven valid trials for each observer and condition.

Variable Error

A 2D Gaussian was fit to the settings made for each condition and for each observer separately. This was done by calculating a covariance matrix of the data, and calculating the eigenvalues and eigenvectors of this matrix. To calculate a single measure of variable error, the area of an ellipse encompassing the eigenvectors of this Gaussian was calculated:

$$Area = \pi \times \sqrt{Eigenvalue_1} \times \sqrt{Eigenvalue_2} \tag{2.1}$$

As these Gaussians were fit to only nine settings, they could have resulted in poor fits. Therefore, in addition, the standard deviation of the settings was calculated separately in the u' and v' dimensions. This additional measure also allowed us to further explore any significant effects found for the overall variable error.

Colour Constancy Index (CCI)

Colour constancy indices were calculated using the equivalent illumination method of Brainard (1998), which has been used in many studies of colour constancy (Yang and Maloney, 2001; Delahunt and Brainard, 2004; Kraft et al., 2002). The equivalent illumination method effectively re-centers individual observer's settings on their individual settings under the neutral reference illumination. This takes into account the fact that the observers' internal representations of grey may not be spectrally uniform and predicts what chromaticity observers' setting would have been if they did have a spectrally uniform representation of grey. These were calculated relative to the settings made under the neutral illumination as follows. All settings and illumination chromaticities were first transformed to LMS space. The reference illumination ($\mathbf{r_i}$) was defined as the cone co-ordinates of the neutral illumination. The reference match ($\mathbf{r_m}$) was the cone co-ordinates of the mean setting

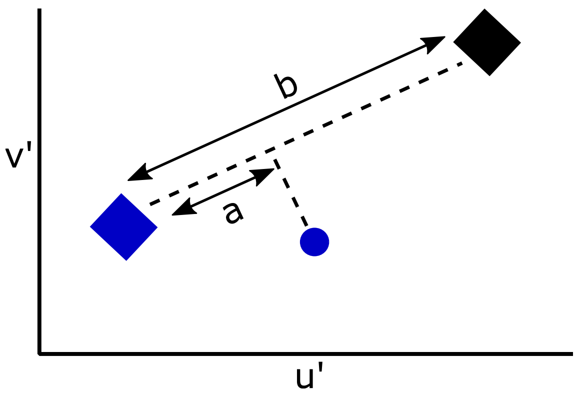


Figure 2.5: Example settings for calculating a CCI. The blue circle is $\mathbf{e_i}$, the blue diamond is $\mathbf{t_i}$, and the black diamond is $\mathbf{r_i}$. See text for explanation of abbreviations.

made under the neutral illumination. The test illumination ($\mathbf{t_i}$) was defined as the cone co-ordinates of the chromatic illumination for this condition (blue, green, red, or yellow). The test match ($\mathbf{t_m}$) was the mean cone co-ordinates of the settings made under this condition. The equivalent illumination ($\mathbf{e_i}$) was then calculated as:

$$\mathbf{e_i} = \mathbf{r_i}\frac{\mathbf{t_m}}{\mathbf{r_m}}, \tag{2.2}$$

for each cone class (L, M, and S) separately.

The equivalent illumination was then converted to CIE L'u'v' in order to calculate a colour constancy index, CCI, as:

$$CCI = 1 - \frac{a}{b}, \tag{2.3}$$

where a is the length of the vector projection of $\mathbf{e_i} - \mathbf{t_i}$ onto $\mathbf{r_i} - \mathbf{t_i}$ and b is the length of $\mathbf{r_i} - \mathbf{t_i}$ (Fig. 2.5).

2.2.2 Results

The average achromatic settings over all observers and repeats is shown in Fig. 2.6 (top row), with error bars representing ± 1 standard error of the mean. As can be seen, settings are all drawn strongly towards the blue direction of u'v' colour space, with the mean setting under neutral, when the patch was on the shape, being almost on the blue illumination's chromaticity. Therefore, for clarity, the equivalent illuminants (described above) are shown on the bottom row of Fig. 2.6. The mean colour constancy index across all observers and conditions was 0.519 $\pm$0.0240.

To test whether people were using specular highlights as a cue to improve their colour constancy and decrease variability, we ran a 4 (illumination) $\times$ 2 (material) $\times$ 2 (position) ANOVA on the CCIs and a 5 (illumination) $\times$ 2 (material) $\times$ 2 (position) ANOVA on the variable error. The results are shown in Tables 2.2 and 2.3 respectively.

There was no evidence for a main effect of specular highlights on either CCI or variable error, against our first hypothesis. In addition, there were no interactions of specular highlights with illumination or position, thereby not supporting hypothesis 2. Against our third hypothesis, there was no main effect of illumination on CCI, or any significant interactions.

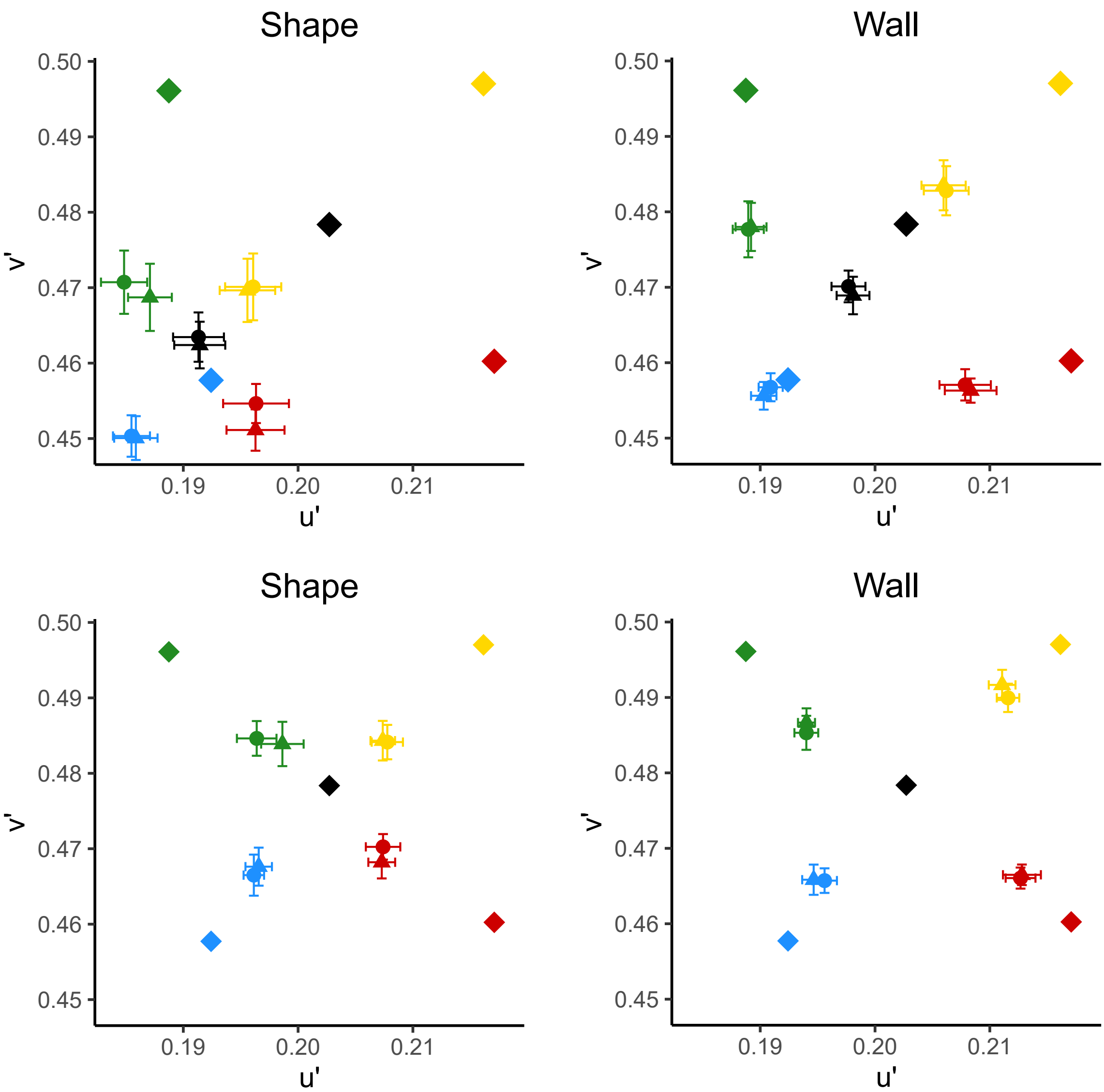


Figure 2.6: Top: raw achromatic settings in u'v'; Bottom: equivalent illuminants in u'v'. Diamonds represent illuminant chromaticities; circles represent mean settings made in glossy scenes and triangles represent mean settings made in matte scenes. Colours represent illuminant colours. Error bars show ±1 standard error of the mean.

We found a significant main effect of patch position on both CCI and variable error. When the patch was on the wall, there was a significantly higher CCI (mean = .619, SD = .305) than when the patch was on the shape (mean = .418, SD = .383). Similarly, there was a smaller variable error when making adjustments to the patch on the wall (mean = 6.24×10^{-5}, SD = 4.31×10^{-5}) than on the shape (mean = 1.24×10^{-4}, SD = 7.26×10^{-5}), suggesting the estimates were more reliable when made on the wall than on the shape. The effect on variable error was further

Effect	F (df)	p	η_p^2
Illumination (main)	2.156 (1.575,20.472)	.149	.142
Position (main)	**8.361 (1,13)**	**.013***	**.391**
Material (main)	.067 (1,13)	.799	.005
Illumination × material	.375 (3,39)	.771	.028
Illumination × position	1.265 (3,39)	.300	.089
Material × position	2.053 (1,13)	.175	.136
Illumination × material × position	1.553 (3,39)	.216	.107

Table 2.2: Results from ANOVA on CCIs

Effect	F (df)	p	η_p^2
Illumination (main)	2.409 (4,52)	.061	.156
Position (main)	**55.965 (1,13)**	**<.001*****	**.811**
Material (main)	1.444 (1,13)	.251	.100
Illumination × material	1.050 (4,52)	.390	.075
Illumination × position	.267 (4,52)	.898	.020
Material × position	.295 (1,13)	.596	.022
Illumination × material × position	1.522 (4,52)	.210	.105

Table 2.3: Results from ANOVA on Variable error (area of ellipse fit to 2D Gaussian)

investigated by looking at the error made in u' and v' separately. For u', there was a significant main effect of patch position ($F(1,13) = 36.501$, $p < .001$, $\eta_p^2 = .737$), with a higher variable error on the shape (mean = $.564\times10^{-3}$, SD =$.200\times10^{-3}$) than on the wall (mean = $.381\times10^{-3}$, SD= 1.69×10^{-3}). Similarly, for v' there was a significant main effect of position ($F(1,13) = 15.513$, $p = .002$, $\eta_p^2 = .544$), with a higher variable error on the shape (mean = 7.62×10^{-3}, SD = 2.87×10^{-3}) than on the wall (mean = 5.84×10^{-3}, SD = 2.74×10^{-3}). The difference between settings on wall and shape suggest that the effects of simultaneous contrast between the patch and its local surround influence observers' setting.

2.2.3 Interim Discussion

In this experiment we tested fourteen observers, which is more than have been used in much of the previous research into colour constancy (five naive in Yang and Maloney (2001); eight in Experiment 1 of Lee and Smithson (2016) and four in Experiment 2). This allowed us to conduct inferential statistical tests. Overall, the mean CCI was slightly lower than other studies measuring colour constancy in 3D rendered scenes (e.g. Yang and Maloney (2001) found average CCIs of 0.65; Delahunt and Brainard (2004) found CCIs ranging from 0.67 to 0.81 when cues were consistent and valid).

Against our hypotheses, there were no main effects of specular highlights on either CCIs or variable error, or any significant interactions. It should be noted, however, that in these scenes there was a strong cue from the uniform background which reflected the chromaticity of the illumination. This strong cue may have hidden any smaller effects of specular highlights, which were predicted to improve colour constancy. This is in agreement with findings from Xiao et al. (2012) who found specular highlights only improved colour constancy when a local surround cue was silenced.

It was also predicted that observers would show a higher degree of colour constancy for scenes illuminated by daylights (blue and yellow) than non-daylights. Although there was no effect of illumination on CCI, there was a noticeable bias towards blue for the raw settings, consistent with previous studies (Winkler et al., 2015; Weiss et al., 2017). This could reflect observers' internal representations of grey being blueish (i.e. containing relatively more short wavelengths) rather than spectrally neutral. Alternatively, it may result from having a prior for blueish illuminations (Pearce et al., 2014; Delahunt and Brainard, 2004), which causes observers to attribute blueish components of reflected light to the illumination rather than surface reflectance, and thus perceive blueish surfaces as neutral (see Aston and Hurlbert, 2017).

A noticeable finding from this experiment was that observers performed significantly better (in terms of both CCI and variable error) when the patch was on the back wall than on the turquoise shape. In fact, when the patch was on the back wall, CCIs were close to those of previous studies, many of which had the matching patch on a back wall. One possible explanation for this difference is the difference in lightness and chromatic contrast. When the matching patch was on the back wall, it had a luminance slightly lower than that of the local surround, which ranged from 70.31 to 70.32cd/m^2 across different illuminations. However, on the turquoise shape, the patch had a much higher luminance than the surround, which had a mean luminance ranging from 20.73 to 22.34cd/m^2 across the different illumination conditions. Using a matching patch with a higher luminance than the local surround has been shown to decrease colour constancy (e.g. Bäuml, 2001; Helson, 1938), and having greater luminance contrast between matching patch and surround (in either direction) results in poorer colour constancy (Werner and Walraven, 1982; Kuriki, 2006). Similarly, chromatic contrast has been found to affect appearance (Shevell and Wei, 1998; Werner, 2014; Faul et al., 2008) which could explain the difference in performance across patch positions. Whilst the black border between matching patch and surround will have decreased the effect of simultaneous contrast (Faul et al., 2008; Blackwell and Buchsbaum, 1988), it was not sufficient to eliminate the effect. An alternative explanation for this difference is that the uniform grey background is giving strong contributions of global and local adaptation to constancy, and thus masking any subtler effects of the specular highlights.

2.3 Experiment 2

In Experiment 2 we tested whether we would see an effect of specular highlights when the uniform background was replaced with a checkerboard, Mondrian-like background. This was designed to prevent use of a uniform background colour as a direct reference cue, to limit the contribution from local surround adaptation on the wall, and to more closely equate the local surround across patch positions. The average chromaticity of the wall's reflectance remained neutral so observers could theoretically still use the global mean cue.

2.3.1 Methods

Observers

A different group of fourteen naïve psychology undergraduates participated in this experiment to earn course credits. They had a mean age of 20.93, ranging from 18 to 41, and included 11 females. All observers were screened for colour blindness using Ishihara plates, and had normal or corrected to normal visual acuity.

Materials and Apparatus

The materials and apparatus were the same as in Experiment 1.

Stimulus Generation

The stimuli were the same as in Experiment 1 except the background wall consisted of a checkerboard pattern with multiple instances of 20 distinct surface reflectances (henceforth, Mondrian background). To generate this, 10 Munsell chips were selected (5B7/8, 5BG7/8, 5G7/8, 5GY7/8, 5Y7/8, 5YR7/8, 5R7/8, 5RP7/8, 5P7/8, and 5PB7/8). These were reconstructed using the basis functions of Parkkinen et al. (1989). To create the remaining 10 surfaces, the weights on the basis functions were inverted, and the spectra shifted to have the same mean reflectance as the original surface. As a result, the average of the 20 surfaces has a flat spectral reflectance. In addition to these twenty reflectances, a spectrally non-selective surface was also used which reflected 80% of light at all wavelengths, which is more light than any of the other surfaces. This ensured the brightest is white cue was still valid. Each wall

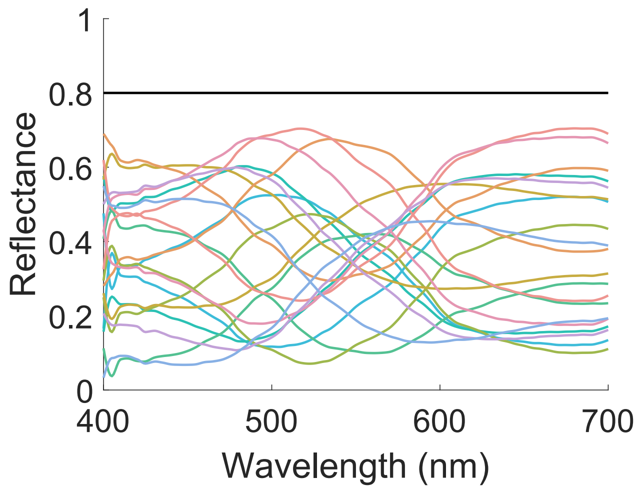


Figure 2.7: Spectral reflectance functions of 21 patches used in Mondrian. The 20 coloured lines are the original and inverted Munsell chip reflectances. The black line is the reflectance of the white patches.

and floor consisted of a 40×40 array of small planes (varying from 0.4 degrees at the furthest point to 0.6 degrees of visual angle at the nearest point), each randomly assigned one of the reflectances described above. An example scene is shown in Fig. 2.8.

Figure 2.8: Example scene

The rendered scenes were scaled to have a mean luminance of 40cd/m^2 – lower than in Experiment 1. This was to keep the luminance of the shapes similar to in Experiment 1, despite the darker background.

The rendered adaptation rooms were the same as in Experiment 1, with blank neutral walls. They were scaled to have a mean luminance of 40cd/m^2 to match the experimental stimuli used.

Task

The task was to set a patch to appear grey, on either the central shape or the back wall, as in Experiment 1. In order to overcome the difference in local lightness contrast in Experiment 1, in which the matching patch was lighter than the surround on the shape but darker than the surround on the wall, in the present experiment the luminance of the matching patch was set to 20cd/m^2. The luminance of the local surround on the shape ranged from 20.06 to 21.54cd/m^2 and on the wall ranged from 41.80 to 42.15 across the different illumination conditions. Therefore, decreasing the luminance of the matching patch ensures it is close to, or lower than, the local surround in all conditions.

Procedure

The procedure for each session was identical to Experiment 1, and the order of the sessions was counterbalanced between observers. However, this time, observers were given the option to do both sessions in immediate succession, and 11 observers chose to do this. For the remaining 3 observers, the sessions were completed between 7 and 19 days apart. Each session lasted approximately 1 hour.

Data Analysis

The data were analysed in the same way as in Experiment 1. Using the same exclusion criterion, only three trials were included for one observer under one condition. All other conditions had at least five valid trials.

Modelling Local Surround

In further analyses, we wished to compare observers' matches with those predicted from the local adaptation cue alone. We modelled the predicted CCIs which would be obtained using only local surround information, as follows. The local surround for the patch on the wall was defined as the area encompassing at least one square of Mondrian in every direction. On the shape, it was defined as an area the same size as the wall's local surround, minus any pixels making up the Mondrian. This local surround subtended approximately 0.5 degrees of visual angle on each side of the matching patch. The mean chromaticity of the local surround was calculated from the hyperspectral rendering for one matte scene under each of the five illuminants. It should be noted that the local surround did not extend onto any highlights, meaning there was no difference in the local surround on matte compared to glossy shapes. A CCI was calculated for each patch position under each illuminant according to the equivalent illumination calculation described in Equations 2.2 and 2.3 above. $\mathbf{t_m}$ was defined as the chromaticity of the local surround under the chromatic illuminants, and $\mathbf{r_m}$ was defined as the chromaticity of the local surround under the neutral illuminant. $\mathbf{t_i}$ and $\mathbf{r_i}$ were defined as in Equation 2.2. Note that the equivalent illumination calculation ensures the overall shift away from neutral in local surround on the shape does not impair predicted CCIs.

2.3.2 Results

The raw achromatic settings averaged across all fourteen observers are shown in Fig. 2.9 (top row). As in Experiment 1, and in line with some previous research (Weiss et al., 2017), there is a clear bias in the settings towards a blue daylight illumination. However, there is now a stronger bias towards blue when the patch was on the back wall than in Experiment 1. The equivalent illuminants are plotted in Fig. 2.9 (bottom row) for clarity. The mean CCI over all observers and conditions was 0.255 ± 0.0160.

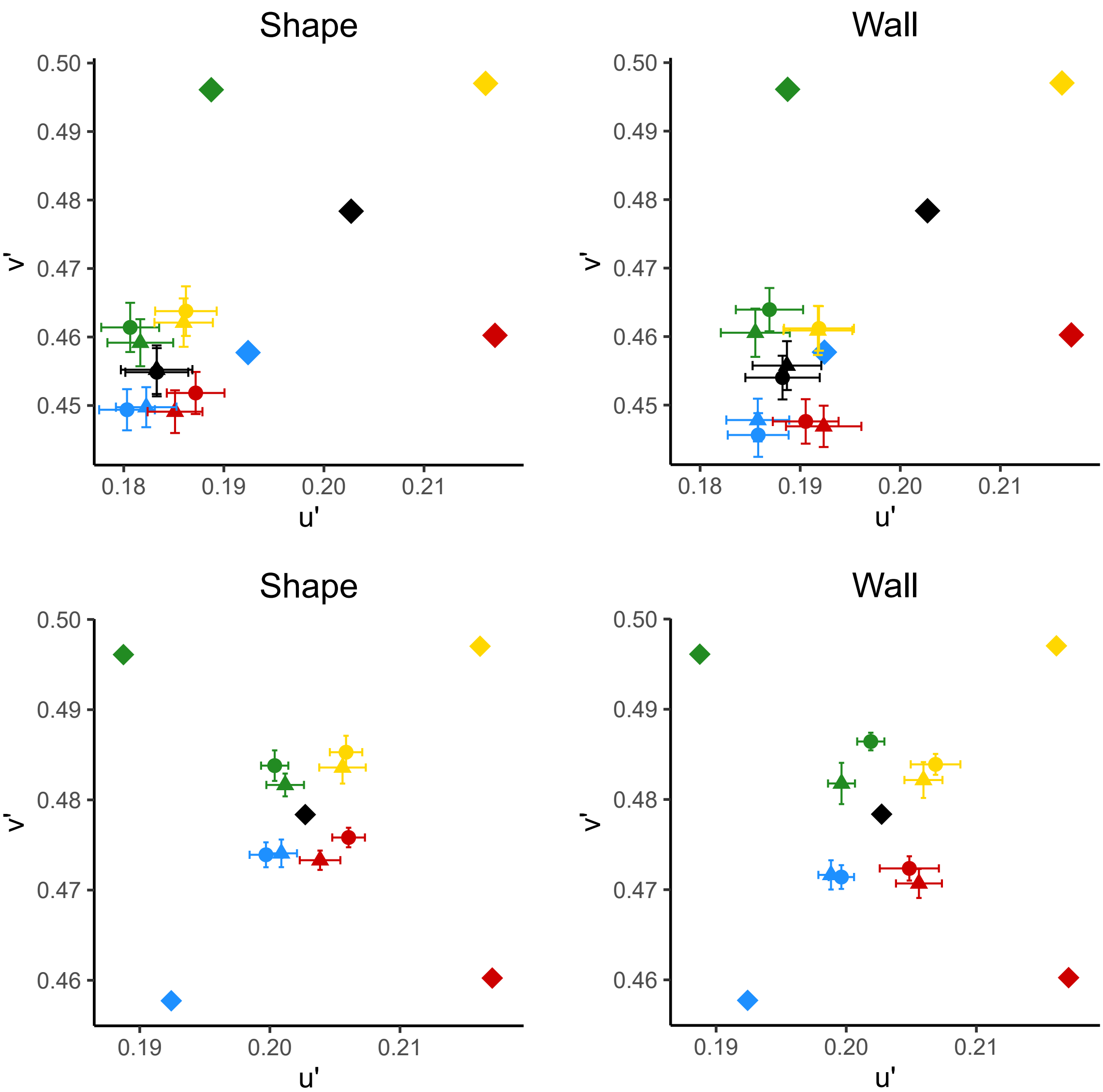


Figure 2.9: Top: raw settings in u'v'; Bottom: equivalent illuminants in u'v'. Diamonds represent illuminant chromaticities; circles represent mean settings made in glossy scenes and triangles represent mean settings made in matte scenes. Colours represent illuminant colours. Error bars show ±1 standard error of the mean.

The results of the 4 (illumination) × 2 (position) × 2 (material) ANOVA on CCIs and 5 (illumination) × 2 (position) × 2 (material) ANOVA on variable error are shown in Tables 2.4 and 2.5 respectively. In line with our first hypothesis, we found a significant main effect of material on CCI, with a higher CCI for scenes containing glossy shapes (mean = .272, SD = .220) compared to matte shapes (mean = .239, SD = .237). This effect was not found on variable error.

Effect	F (df)	p	η_p^2
Illumination (main)	.302 (3,39)	.823	.023
Position (main)	**5.469 (1,13)**	**.036***	**.296**
Material (main)	**5.958 (1,13)**	**.030***	**.314**
Illumination × material	1.863 (2.097,27.258)	.173	.125
Illumination × position	.773 (3,39)	.516	.056
Material × position	.526 (1,13)	.481	.039
Illumination × material × position	.091 (2.002,26.023)	.914	.007

Table 2.4: Results of ANOVA on CCIs

Effect	F (df)	p	η_p^2
Illumination (main)	1.188 (4,52)	.327	.084
Position (main)	**7.219 (1,13)**	**.019***	**.357**
Material (main)	2.374 (1,13)	.147	.154
Illumination × material	1.364 (4,52)	.259	.095
Illumination × position	.637 (4,52)	.638	.047
Material × position	**7.531 (1,13)**	**.017***	**.367**
Illumination × material × position	.270 (4,52)	.896	.020

Table 2.5: Results of ANOVA on Variable Error (Area of ellipse fit to 2D Gaussian)

In support of our second hypothesis, we also found a significant interaction between material and position on variable error. This interaction is shown in Fig. 2.10a. To further explore the interaction, paired samples t-tests were conducted separately for patch on shape and patch on wall conditions. When the patch was on the shape, but not on the wall, there was a significant effect of material (t(13) = 2.376, p = .034), with less variability of estimates for scenes containing glossy shapes (mean = 1.30×10^{-4}, SD = 4.5×10^{-5}) than for scenes containing matte shapes (mean = 1.56×10^{-4}, SD = 6.0×10^{-5}). To determine whether this effect was driven by variability in u', v', or both, we ran further ANOVAs on the standard deviation in both directions separately. There was a significant material × position interaction in u' (F(1,13) = 4.651, p = .050, η_p^2= .264) but not in v'. The interaction in u' is shown in Fig. 2.10b, and is driven by significantly less variable error in scenes containing glossy shapes (mean = .0060, SD = .00143) than in scenes containing matte shapes (mean = .0068, SD = .00180), when the patch is on the shape: t(13) = 3.106, p = .008. There was no significant difference in variable error in u' between scenes containing matte and glossy shapes when the patch was on the back wall.

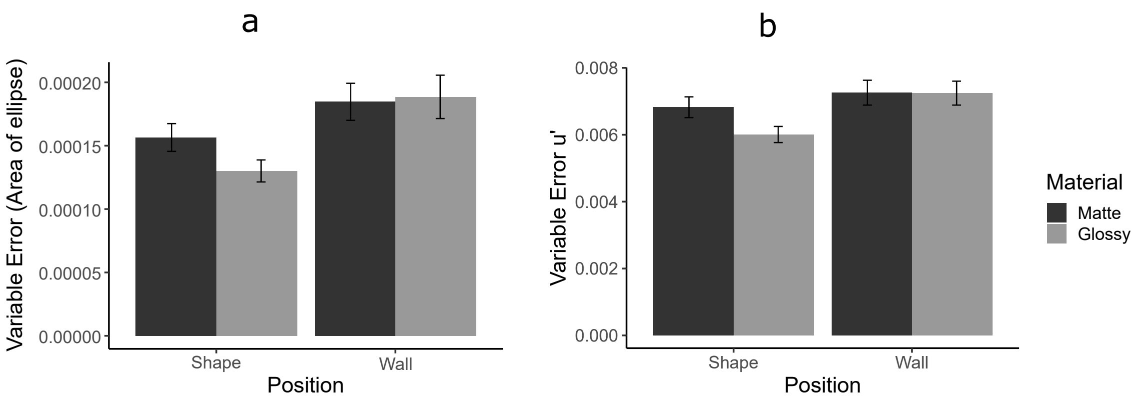


Figure 2.10: (a)Variable error (area of an ellipse fit to Eigenvectors of Gaussian) when the matching patch was on the shape (left) and wall (right). Dark grey bars are for scenes containing matte shapes; light grey bars are for glossy shapes. Error bars are ±1 standard error of the mean. (b) The same interaction for variable error in u', measured as the standard deviation across all settings.

Against our third hypothesis, there was no significant main effect of illumination on CCI, or any significant interactions involving illumination.

As in Experiment 1, a significant main effect of position was found on CCIs, with significantly higher CCIs when the patch was on the back wall (mean = .286, SD = .253) than on the shape (mean = .225, SD = .222). To investigate this further, we modelled the predicted CCIs for each patch position, using local surround information only, as described above. The predicted equivalent illuminants from this analysis are plotted in Fig. 2.11. As expected, there was a greater bias in the local surround when the patch was on the turquoise shape than on the back wall. It should be noted, however, that due to the random nature of the Mondrian background, the local surround on the wall was slightly biased towards blue. Modelling optimal performance using only local surround information, higher CCIs are predicted when the patch is on the back wall (mean = 0.980 across all illuminations) than on the shape (mean = 0.744 over all illuminations).

In addition, there was a significant main effect of position on variable error, with significantly more variable error when the patch was on the back wall (mean = 1.866 $\times 10^{-4}$, SD = 1.326 $\times 10^{-4}$) than on the shape (mean = 1.433 $\times 10^{-4}$, SD = 8.347 $\times 10^{-5}$). To further explore this, we looked at the ANOVA conducted on variable error in u' and v' separately. There was a significant main effect of position on variable error in u' (F(1, 13) = 10.009, p = .007, η_p^2 = .435), but not v' (p = .093). In u' there was significantly more variable error when the patch was on the wall (mean = 7.251×10^{-3}, SD = 3.040×10^{-3}) than on the shape (mean = 6.416×10^{-3}, SD = 2.340×10^{-3}).

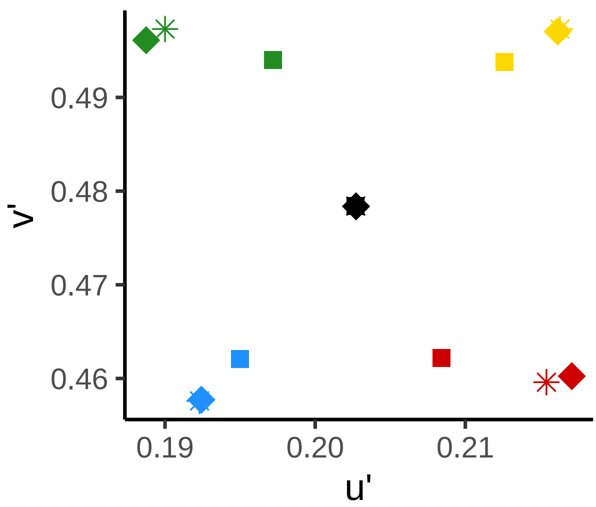


Figure 2.11: Predicted optimal performance using local surround information only. As in previous figures, diamonds indicate actual illumination chromaticities. Squares reflect optimal performance when making estimates on the central shape; asterisks are optimal performance when making estimates on the back wall (using local surround information only).

2.3.3 Interim Discussion

In the present experiment, unlike Experiment 1, we did find a significant main effect of material on CCIs, supporting the first hypothesis that the presence of specular highlights can improve colour constancy. In addition, there was a decrease in variability of estimates for scenes containing specular highlights when observers were making illumination estimates on the shapes but not on the back wall, in terms of both overall variability and standard deviation in u'. This is in agreement with Yang and Maloney (2001) who found no effect of perturbing specular highlights when judgements were made on a back wall. Taken together, these findings suggest that observers are able to make use of specular highlights when estimating illumination chromaticities, but segment the scene into different illumination frameworks before making estimates. An alternative explanation is that observers may be relying on specular highlights more when making estimates on the shape because the local surround cue is harder to use so more weight is applied to the highlights. With the Mondrian background, observers appear to rely more on the specular highlights than when scenes contained a uniform neutral background (Experiment 1). It is important to note that CCIs were, on average, much lower here than in Experiment 1. This could be due to the Mondrian background making the local and global mean cue harder to use. Nevertheless, the fact that an effect of specular highlights was found here suggests that the high overall performance in Experiment 1 may have masked any potential effect of highlights.

As in Experiment 1, we did not find the predicted benefit for scenes illuminated by daylight illuminants. There was still a strong bias towards blue in the raw settings, but when converted to equivalent illuminants, there was no difference in constancy between the illuminations. This suggests observers' internal representations of grey are not spectrally uniform but biased towards spectra containing more shorter wavelengths, which appear blueish-grey. However, an alternative explanation - at least when the matching patch was on the shape - is that the turquoise local surround from the shape could have biased settings towards blue.

2.4 Experiment 3

In Experiment 3 we aimed to test to what extent the bias in local surround could have driven a difference between performance on the wall vs on the shape in Experiments 1 and 2. To this end, we varied the reflectance of all shapes, including the one containing the matching patch, on every trial. This allowed us to determine whether the effect of position depended on having the same local surround on the shape on every trial. This also ensures any findings are not specific to the limited range of stimuli used in Experiments 1 and 2. However, changing the colour of the shape on a trial-by-trial basis is likely to add variability and noise to estimates. In order to counteract this, we ensured observers used the smallest adjustment step size, thus making more finely tuned settings overall than in the previous experiments.

2.4.1 Methods

Observers

Fourteen naïve paid volunteers participated in this study. They had a mean age of 24.79, ranging from 20 to 35, and included eight females. All observers were screened for colour blindness using Ishihara plates, and had normal or corrected to normal vision.

Materials and Apparatus

The materials and apparatus were the same as in Experiment 2.

Stimulus Generation

The same Mondrian room from Experiment 2 was used in the present experiment. Rather than having three scenes, we selected one scene from Experiment 2 (using the reflectances in Set 1 in Table 2.8 in the Appendix). This meant all stimuli had identical shapes. For each trial and each observer, the six reflectances were randomly assigned to each of the six shapes. This meant that the scenes were more variable, and the shape containing the matching patch could have one of six different reflectances. It is important to note that the average chromaticity of all six reflectances under all illuminants used did not differ by more than $1\Delta E_{L*u*v*}$ from a neutral surface with the same mean luminance. This means that, over all trials, on average, there should be no biased local surround when the patch is on the central shape.

Task

The task was the same as in Experiment 2. However, in order to decrease the noise in observers' settings, they were now required to use the smallest adjustment step size ($1\Delta E$) at least once before a match could be recorded.

Procedure

There were nine repeats under each condition, with a different arrangement of reflectances on each repeat, individually defined for each observer.

Initially, observers completed a session with the patch on a wall and a session with the patch on the shape, as in Experiments 1 and 2, spread out between one and 22 days. However, when running the Wall condition some data were lost for 11 observers due to computer error. To rectify this, a third session was conducted in which observers made settings on both the wall and the shape, with the position changing halfway through. This was conducted at a later date, between 49 and 82 days after the latter of the first two sessions. For this, half the trials from the first session, and half the trials from the second session were used. The order was matched to the original sessions such that if an observer participated in the shape session before the wall session, in the final session they would see the shape trials first. Although this was conducted to replace missing data, it also allowed us to test for consistency in achromatic settings over time.

No data were lost for the remaining three observers, so they just took part in two sessions each.

Data Analysis

The same exclusion criterion was applied as in Experiments 1 and 2, and CCI and variable error were calculated in the same way. After exclusion, all conditions had at least eight valid trials.

The data collected on the shape session and shape half of the third session were compared to test for a difference across time points. To determine whether there was a difference between the sessions, a 2 (session) $\times$ 4 (illuminant) ANOVA was conducted on the CCIs, collapsed across both material types as only one material was used under each illuminant in the third session. There was no significant main effect of session, illuminant, or interaction. From this, it was concluded that observers' illumination estimates are consistent over time.

For the full analysis below, data collected on the shape session were analysed as in Experiments 1 and 2. For the 11 observers who completed three sessions, the data from the wall session and the wall half of the final session were combined for analysis. The data from the remaining three observers were analysed as normal.

2.4.2 Results

The mean raw settings over all observers and repeats are shown in Fig. 2.12 (top row). As can be seen, there is still a strong pull towards blue in these settings in both the wall and shape conditions. The equivalent illuminants are shown for clarity in Fig. 2.12 (bottom row). The mean CCI over all conditions and observers was 0.361 ± 0.0193 – higher than in Experiment 2 but lower than in Experiment 1.

As before, a 4 (illumination) $\times$ 2 (position) $\times$ 2 (material) ANOVA was run on CCIs, and a 5 (illumination) $\times$ 2 (position) $\times$ 2 (material) ANOVA was run on the variable error. The results of these are shown in Tables 2.6 and 2.7 respectively.

In support of our primary hypothesis, a main effect of material on CCIs was once again found, with a higher CCI for scenes containing glossy (mean = .391, SD = .282) than matte shapes (mean = .331, SD = .295). There were no significant interactions between material and position or illumination on the CCIs, against our second hypothesis. In addition, this time there was no significant main or interaction effects of material on the variable error.

There were no significant main or interaction effects of illumination on either CCIs or variable error, against our final hypothesis that scenes illuminated by daylight illuminants would result in higher CCIs.

A significant main effect of position was found on both CCIs and variable error. There was a higher CCI when the patch was on the wall (mean = .385, SD = .263) than on the shape (mean = .337, SD = .313), as in Experiment 2. To test whether the difference in local surround could explain this finding, we modelled the CCIs that an ideal observer would achieve using only local surround information on either the back wall or the shape. Unlike in Experiment 2, the shape containing the matching patch now changes reflectance on every trial. Therefore, the mean predicted settings over all six reflectances were averaged before calculating a predicted CCI on the shape. As before, CCIs on the wall were predicted to be higher (0.980) than on the shape (0.967). However, these are a lot more similar than in Experiment 2, which may explain why the effect size is smaller ($\eta_p^2 = .264$ in Experiment 3 vs $\eta_p^2 = .296$ in Experiment 2).

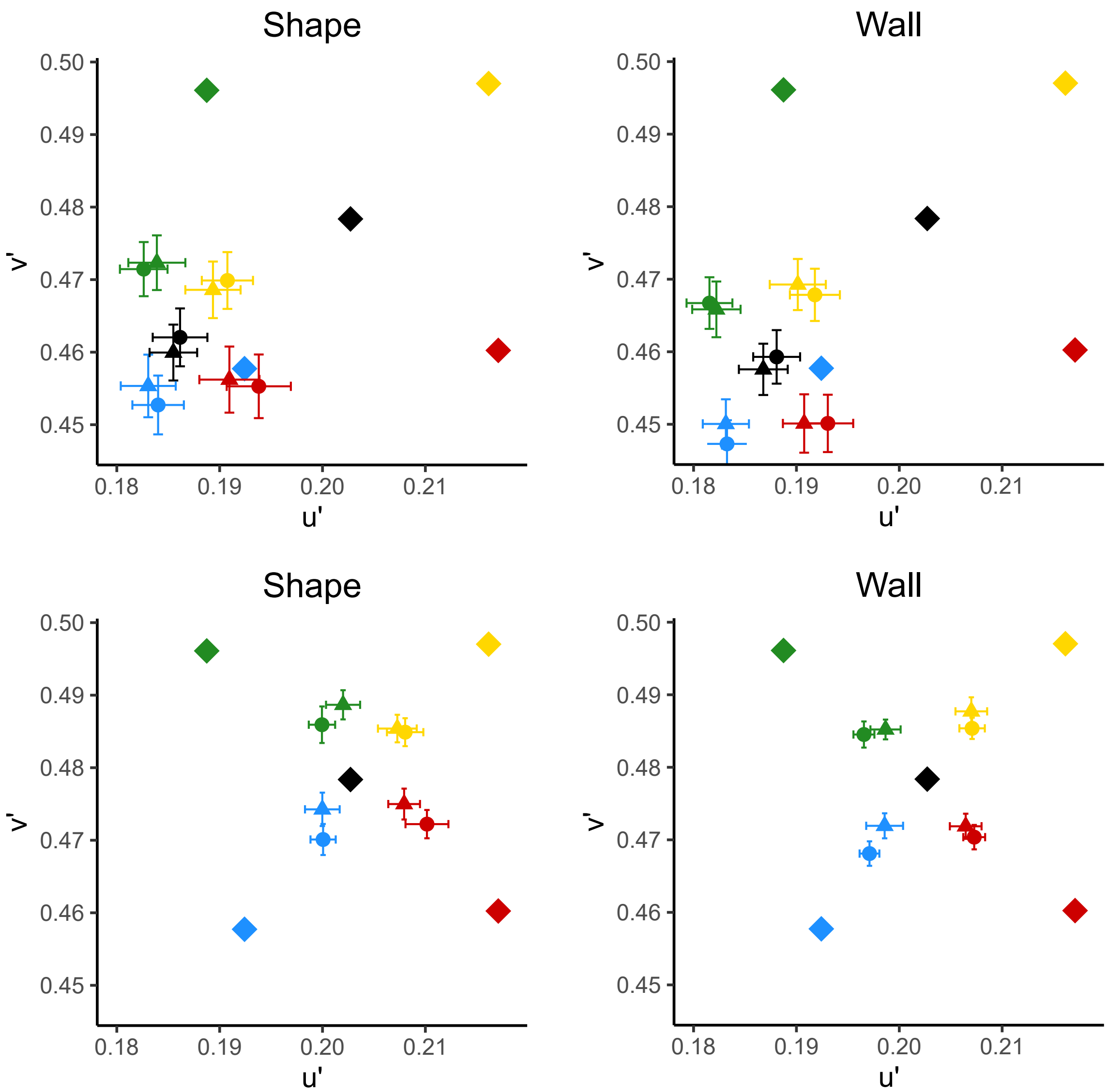


Figure 2.12: Top: raw settings in u'v'; Bottom: equivalent illuminants in u'v'. Diamonds represent illuminant chromaticities; circles represent mean settings made in glossy scenes and triangles represent mean settings made in matte scenes. Colours represent illuminant colours. Error bars show ±1 standard error of the mean.

In addition, there was more variable error when the patch was on the shape (mean = 1.979×10^{-4}, SD = 1.455×10^{-4}) than on the wall (mean = 1.238×10^{-4}, SD = 8.87×10^{-5}). To further explore this, we analysed the variable error separately in u' and v'. A significant main effect of position was found both on u' ($F(1,13) = 20.850$, $p = .001$, $\eta_p^2 = .616$) and v' ($F(1,13) = 23.754$, $p < .001$, $eta_p^2 = .646$). Variable error in u' was higher when the patch was on the shape (mean = 7.08×10^{-3}, SD = 3.07×10^{-3}) than on the wall (mean = 5.21×10^{-3}, SD

Effect	F (df)	p	η_p^2
Illumination (main)	.119 (3,39)	.948	.009
Position (main)	**4.654 (1,13)**	**.050***	**.264**
Material (main)	**35.477 (1,13)**	**<.001*****	**.732**
Illumination × material	1.958 (3,39)	.136	.131
Illumination × position	.241 (1.514, 19.683)	.726	.018
Material × position	.342 (1,13)	.569	.026
Illumination × material × position	.302 (3,39)	.824	.023

Table 2.6: Results of ANOVA on CCIs.

Effect	F (df)	p	η_p^2
Illumination (main)	1.877 (2.356, 30.631)	.165	.126
Position (main)	**19.662 (1,13)**	**.001****	**.602**
Material (main)	.017 (1,13)	.899	.001
Illumination × material	.866 (4,52)	.491	.062
Illumination × position	1.565 (4,52)	.197	.107
Material × position	.539 (1,13)	.476	.040
Illumination × material × position	.683 (1.963, 25.520)	.512	.050

Table 2.7: Results of ANOVA on Variable Error (Area of ellipse fit to 2D Gaussian)

= 2.18×10^{-3}). This pattern was also replicated in v' with a higher variable error on the shape (mean = 9.83×10^{-3}, SD = 3.52×10^{-3}) than on the wall (mean = 8.02×10^{-3}, SD = 2.99×10^{-3}).

2.4.3 Interim Discussion

The findings of this experiment are generally in agreement with those of Experiment 2. Again, we found evidence that observers are able to benefit from the presence of specular highlights, as shown by an increase in CCI. In fact, the effect was even stronger here, suggesting the effect of specular highlights found in Experiment 2 may have been underestimated. In the present experiment observers were required to make more fine-tuned settings, resulting in more precise estimates, giving us more power to detect the true effect of highlights. However, there was still no decrease in the variability of settings for scenes containing glossy, compared to matte, shapes. In addition, the interaction between material and position on variable error which was found in Experiment 2 was no longer significant ($p = .071$). The lack of an interaction between material and position or illumination suggests that the use of specular highlights is not mediated by either of these factors.

The CCIs were generally higher than in Experiment 2, which could be due to observers using the smallest step size for each achromatic setting. This encouraged observers to make settings closer to their true perceived grey, compared to in the previous experiments. Furthermore, this resulted in less variability when observers were making settings on the back wall. It should be noted that the variability when the matching patch was on the central shape remained high as the reflectance of the shape, and thus the local surround, changed on every trial.

As in both Experiments 1 and 2, there appears to be no benefit for scenes illuminated by daylights as opposed to non-daylights, against our final hypothesis, and previous research (e.g Delahunt and Brainard, 2004). As there were no effects of illumination found on CCIs in any of the experiments, we ran one more analysis to determine whether this was due to lack of power. For this, the data from Experiments 2 and 3 were combined for a total of 28 observers. It should be noted that the data from Experiment 1 were not included in this analysis as the methods used were so different to the other experiments that it would not be appropriate to combine the data. A 4 (illumination) $\times$ 2 (position) $\times$ 2 (material) $\times$ 2 (experiment) way ANOVA on the combined data still found no significant main effect of illumination ($p = .688$), but the effects of material ($p < .001$) and position ($p = .004$) remained significant. This suggests the lack of effect of illumination found in any individual experiment was not due to a lack of power.

The similarity of findings in this experiment and Experiment 2 suggests that the findings are not specific to one particular scene with certain reflectances. In addition, the fact that raw settings in L*u*v* are still drawn toward the blue illumination chromaticity suggests that this is not caused by the local surround of the central shape, which previously always had the same surface reflectance but here averages out to a neutral chromaticity.

2.5 Discussion

This series of experiments was designed to test three primary hypotheses: that specular highlights would improve colour constancy and decrease variable error; that the effect of highlights would be mediated by the type of illumination (daylight or non-daylight) and position of a matching patch; and that scenes illuminated by daylights would result in a higher degree of colour constancy than scenes illuminated by non-daylights. A summary of findings across all experiments, in terms of both variable error and CCIs, is shown in Fig. 2.13. Observers had higher CCIs when making achromatic settings on the back wall than on the shape in all three experiments. In addition, in Experiments 2 and 3 (when a Mondrian background was introduced), CCIs were higher for scenes containing glossy, compared to matte, shapes. In both Experiments 1 and 3, estimates were less variable when the matching patch was on the back wall compared to on the shape. In Experiment 2, this effect was reversed, with higher variable error on the wall than on the shape. In addition, there was an interaction between material and patch position on variable error in Experiment 2.

In support of our primary hypothesis, we found highlights to improve colour constancy in Experiment 2, when scenes contained a Mondrian background, in agreement with previous research (Yang and Shevell, 2002; Lee and Smithson, 2016). This finding was replicated with another group of observers in Experiment 3, when the reflectance of the shapes was changed on each trial. However, there was no effect of specular highlights when there was a uniform background (Experiment 1). The uniform light grey background in Experiment 1 could have provided a strong cue to the illumination, in terms of both local and global adaptation, thereby weakening reliance on specular highlights. Removing this cue resulted in an overall reduction in CCIs, allowing an otherwise masked effect of specular highlights to be revealed, in line with Xiao et al. (2012). Whilst observers could theoretically have used the global mean or local surround cue with the Mondrian background, it seems that they found it more difficult to use such cues for colour constancy. This is in contrast with previous research (Linnell and Foster, 2002) suggesting that having multiple surfaces of different reflectances in the scene (as in our Mondrian backgrounds) should improve colour constancy. However, it should be noted that observers were probably

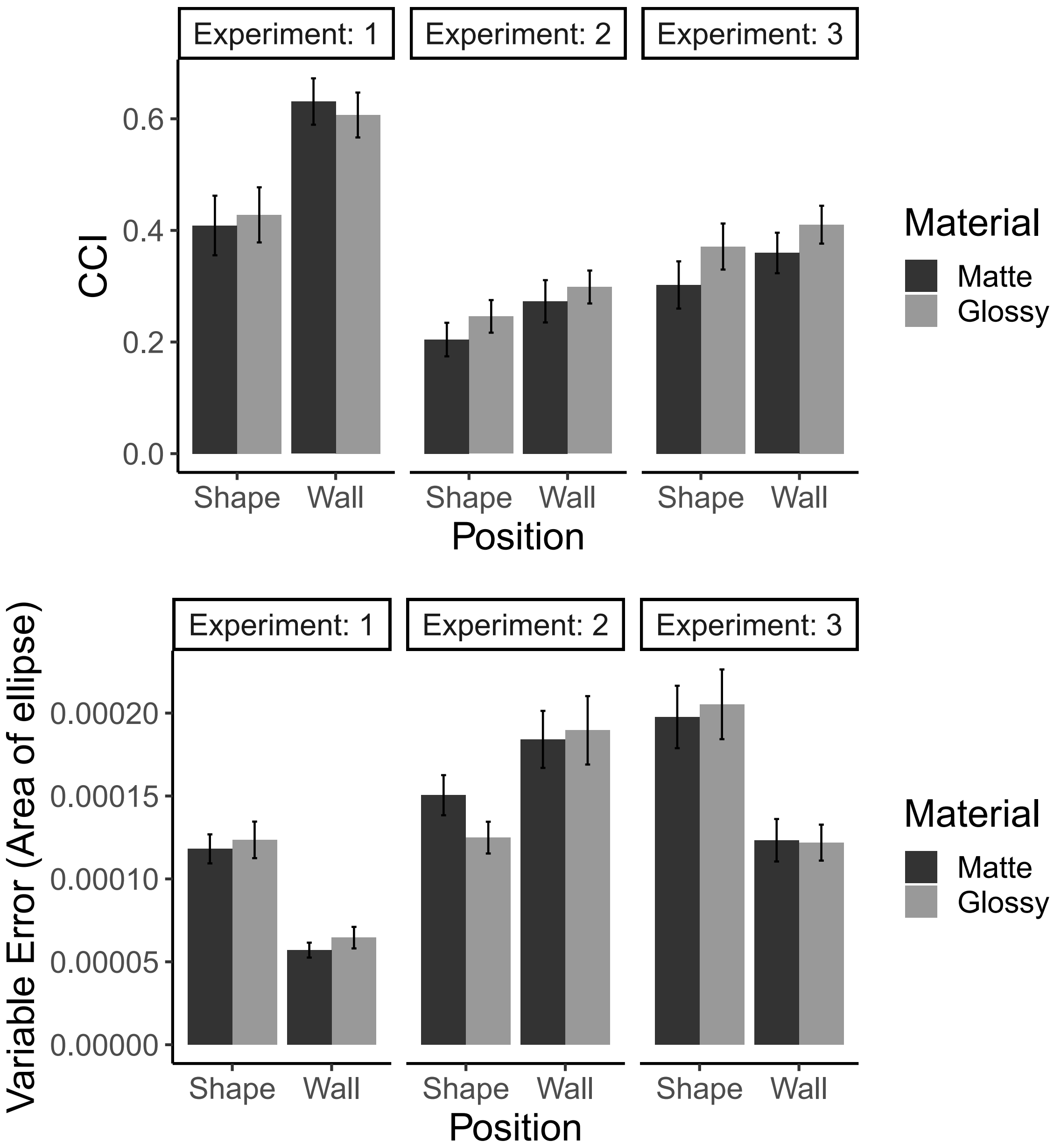


Figure 2.13: Summary results across all three experiments. Top: mean CCIs for matching patch on shape and wall, with matte or glossy shapes. Bottom: mean variable error, as measured by the area of an ellipse fit to a Gaussian. For both graphs, error bars show ±1 standard error of the mean

utilising the global and local mean, and brightest is white cues to some extent in Experiments 2 and 3 as shown by the non-zero CCIs for scenes containing matte shapes.

Whilst it could be argued that the benefit seen for scenes containing specular highlights could be explained simply by the difference in local surround, there are at least three reasons to believe this not to be the case. When modelling the effect of local surround (subtending roughly 0.5 degrees from the matching patch) we found no difference between matte and glossy shapes. If the region of local surround were increased to include the specular highlights when the patch was on the shape, it could still not explain the increase in CCIs we see for scenes containing glossy shapes when the matching patch is on the back wall. Furthermore, if observers were simply using the local surround, we would expect to find the same results when the matching patch is on the shape in Experiments 1 and 2. The fact that there is no benefit for specular highlights with a uniform background suggests that observers are using more information than simply the local surround.

There was no overall reduction in variable error for scenes containing specular highlights on any of the three experiments. Whilst we did find a position × material interaction in Experiment 2, this was not robust enough to show up in Experiment 3. This is against our prediction, based on the cue combination literature (Ernst and Banks, 2002; Ernst, 2006), that adding valid cues should decrease variability of estimates. A possible explanation for this null finding is that the variable error is sensitive to the number of trials, and nine per condition may be insufficient. Due to the many conditions in these experiments, it would be unfeasible to add more repeats per condition as each experiment already lasts an hour. Future experiments using more trials per condition, but fewer conditions, are needed to determine whether an effect of specular highlights on variability of estimates can be found.

In support of our second hypothesis, there was an interaction between material and position on variable error in Experiment 2, such that specular highlights only reduced variance when estimates were made on the shape and not the back wall. Whilst this lends support to the notion that scenes are separated into different illumination frameworks, and agrees with Yang and Maloney (2001), the finding was not replicated in Experiments 1 or 2, so should be taken with caution. Furthermore, there were no such interactions on CCIs in any of the three experiments. Therefore, our results do not strongly support the prediction that the effect of specular highlights on constancy will be mediated by patch position.

In addition, there were no significant interactions between specular highlights and illumination, suggesting the effect is also not mediated by type of illumination. This is in contrast to Yang and Maloney (2001), who found observers only used specular highlights when they signalled a daylight illumination. However, it should be noted that the illuminations used in the present study were more controlled, such that they were all equally discriminable from neutral. Furthermore, as highlights were not perturbed in the present study, all scenes were physically possible - even those illuminated by non-daylights - which may explain the lack of interaction found here.

Against our final hypothesis, there were no significant effects of illumination, even when data from Experiments 2 and 3 were pooled. Unlike Delahunt and Brainard (2004), we did not find an improvement in colour constancy for scenes illuminated by daylights compared to non-daylight illuminants. However, it should be noted that in Delahunt and Brainard (2004), the effect of illumination was only statistically significant when a local surround cue was silenced. Whilst the specular highlight cue was absent in half the trials of the present experiments, there were no inconsistent cues, and the global mean and brightest is white cues were always present and consistent.

Whilst there were no effects of illumination on the equivalent illuminant calculations, a robust finding was a bias in raw achromatic settings towards blues. This was not an artefact of the local surround, as in Experiment 3 the local surround on the shape was varied on a trial-by-trial basis. A plausible explanation for this bias is that observers assume illuminations are blue-ish. Therefore, when they see a surface containing more shorter wavelengths, they attribute the extra blue to the illumination and discount it in their achromatic settings.

A further finding was the overall effect of patch position on CCIs. In all three experiments, observers had a higher degree of colour constancy when the matching patch was on the back wall than on the shape. This was most likely due to the difference in local surround, rather than differences in local lightness contrast between the patch and surround across positions. In Experiments 2 and 3, the patch luminance was lower than that of the surround in both positions, and the absolute luminance contrast was greater on the wall than the shape, which should have resulted in lower CCIs on the wall, counter to what we found. In fact, the effect we found was qualitatively predicted by modelling the effect of local surround.

Over all three experiments - but particularly in Experiments 2 and 3 - we found CCIs lower than reported in many previous studies. There are a number of possible explanations for the low CCIs found here. In many of the previous studies, the environment was more immersive than here - either using an immersive room (Gupta et al., 2020), or stereoscopic viewing (Yang et al., 2013a; Delahunt and Brainard, 2004; Xiao et al., 2012). Yang and Shevell (2002) found that binocular disparity, achieved through stereoscopic viewing, improves colour constancy, with CCIs under monocular viewing similar to those found here. Furthermore, there are large individual differences in CCIs, with non-naive observers often outperforming naive observers. Here we used only naive observers, which may have contributed to the lower CCIs. Finally, the specific setup of the stimuli used, as well as the instructions (Arend and Reeves, 1986) can have a large impact on the resulting degree of colour constancy. Indeed, this is what we found here, with much lower CCIs when a Mondrian background wall was introduced, as opposed to a uniform, spectrally neutral wall.

Many of the previous studies into the use of specular highlights as a cue are not suitable for formal statistical analyses, and did not consider how specular highlights may interact with other factors. Here, we increased the number of observers and extended previous findings that observers are able to use specular highlights. However, this was only the case when a uniform background cue was weakened. Furthermore, controlling the illuminations, such that those on and off the daylight locus were equally discriminable, removed the interaction between illumination and specular highlights found previously (Yang and Maloney, 2001). We did not find that observers used a daylight prior to improve colour constancy estimates. The novel cue combination approach taken here revealed some interesting results that would have otherwise been overlooked. However, further studies investigating the effect of cues on variable error in colour constancy using different methods, such as asymmetric matching, or using more immersive environments, are needed to draw firmer conclusions. Additionally, using the cue combination approach to study how variable error is affected by other cues (such as global mean), which we did not manipulate here, would be of great interest.

2.6 Appendix

2.6.1 Munsell reflectances

Set 1	Set 2	Set 3
5BG4/6	5BG4/6	5BG4/6
10B3/4	2.5Y7/6	10Y7/8
10Y5/6	7.5B3/6	10RP5/8
2.5PB4/8	10YR6/8	5PB5/10
5YR5/8	2.5RP5/10	10B5/8
2.5P4/4	7.5PB5/6	10YR5/4

Table 2.8: The sets of Munsell chips used for the six blobby shapes in Experiments 1 and 2. In Experiment 3, only Set 1 was used.

2.6.2 Instructions

"In the following trials, you will see an image of a scene. In each scene, there are six randomly coloured shapes sitting in a box. On top of the central shape (on one wall) you will see a small patch, indicated by a black outline. Your task is to adjust the colour of this patch so that it looks like a grey piece of paper sitting on top of the shape in the scene. That is, adjust the patch so that it looks like it has the same reflectance properties as a grey surface would in this scene. You may notice an overall colour change of the scenes on some trials. Think of this as a change to a different colour of illumination and focus on adjusting the patch so that it looks like it has the same surface reflectance as a grey surface. That is, adjust the patch so that it looks like a grey surface under the changed illumination."

2.7 Supplementary Material

To ensure there were no differences in cue validity between the matte and glossy scenes used, we modelled the chromaticity which would be predicted using the global mean and brightest is white cues. We did not model the local surround for these purposes as this was not affected by the specularity of the shapes, and was modelled separately. We modelled the chromaticity predicted by these cues for one scene illuminated by each of the five illuminants used Experiments 1 and 2. It was not necessary to perform this modelling for Experiment 3 as the stimuli were so similar to those used in Experiment 2.

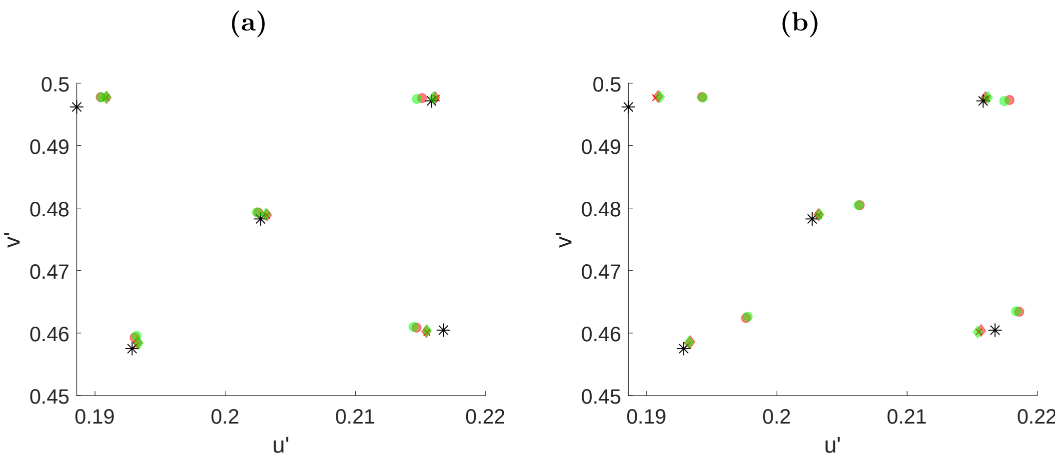


Figure 2.14: chromaticity implied by various cues. Black *s are illumination chromaticity; green symbols are cues in matte scenes, red symbols are cues in glossy scenes. Circles are global mean; diamonds are the brightest point; red Xs are chromaticity convergence (in glossy scenes only). (a) is Experiment 1; (b) is Experiment 2.

To model global mean chromaticity, we took the average chromaticity of all pixels in the scene. For the glossy shapes this included the specular highlights. This is shown by the circles in Fig. 2.14. To model brightest is white, we identified the pixel with the highest luminance (Y), and calculated the chromaticity of this pixel. For both Experiments 1 and 2, with matte and glossy scenes this was a pixel on the back wall. This is shown by the diamonds in Fig. 2.14. In addition, we modelled the chromaticity convergence cue for the glossy scenes. To do this, we identified 40 pixels lying on a line on each shape which included a specular highlight and the body colour. For each of the six shapes, a line was fit to the chromaticities of these 40 pixels, and the intersection of the six lines was calculated. This chromaticity is shown by the Xs in Fig. 2.14

To ensure truncating the out of gamut pixels did not affect the chromaticity of the specular highlights, we plotted the rendered scenes with the truncated pixels shown in black. This can be seen in Fig. 2.15, for Experiment 1 on the top row and Experiment 2 on the bottom row. None of the truncated pixels coincide with the specular highlights.

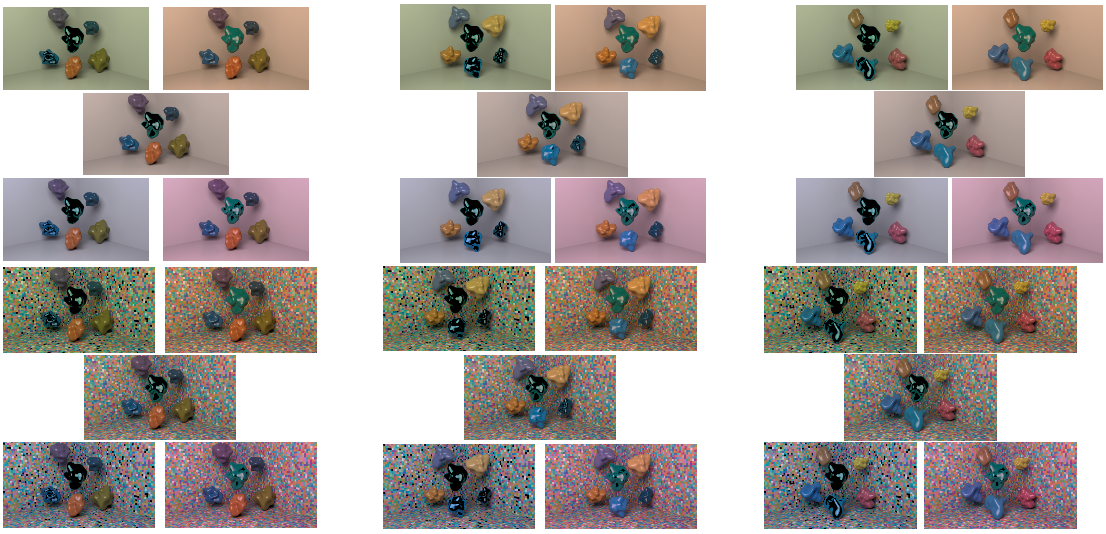


Figure 2.15: The scenes used in Experiment 1 (top) and 2 (bottom), with the location of the truncated pixels shown in black. The scenes used in Experiment 3 are not shown, as the reflectances are the same as those used in the leftmost image of experiment 2.

Preface to chapter 3

Following the previous set of experiments, we determined that adults could use specular highlights as a cue for colour constancy. However, it was evident that achromatic setting would be inappropriate for testing children for a number of reasons. Firstly, the data were noisy even in adults, and it would be expected that children's data would be even more noisy. Secondly, the task was not easy to do as it took adults a while to learn how the controller worked to change the colours, with many complaining that they were "lost in colour space". Finally, while achromatic settings are useful for measuring colour constancy, the task is not fun and children would quickly become bored. Therefore, it was necessary to develop a different task that children wold find fun and engaging, and easy to do. In the following chapter, we developed such a task using a dragon who liked to eat "sweets". The first experiment of this chapter used simple two-dimensional stimuli which were relatively easy to create, in order to determine whether the task would be appropriate. The simplicity of this task also allowed for a baseline measure of colour constancy which was comparable to many previous studies. Following on from this, we developed a three-dimensional rendered version. It required more time to generate the stimuli for this task, but meant that it would allow us to manipulate more cues in future versions.

Chapter 3, which follows, has been submitted for publication in Developmental Science with authors Rebecca Wedge-Roberts, Stacey Aston, Ulrik Beierholm, Robert Kentridge, Anya Hurlbert, Marko Nardini, and Maria Olkkonen.

Chapter 3

Developmental changes in colour constancy in a naturalistic object selection task

3.1 Research Highlights

- Six- to 11-year-old children demonstrated better colour constancy than adults in an object selection task
- Colour constancy decreased with age over childhood
- These findings may indicate development of cognitive strategies used to overcome automatic colour constancy mechanisms

3.2 Abstract

When the illumination falling on a surface changes, so does the reflected light. Despite this, adult observers are quite good at perceiving surfaces as relatively unchanging – an ability termed colour constancy. Very few studies have investigated colour constancy in infants, using preferential looking, and even fewer in children. Here we asked whether there is a difference in colour constancy between children and adults; what the nature of the developmental trajectory is between six and 11 years; and whether the pattern of constancy across illuminations and surface reflectances differs between adults and children. To this end, we developed a novel, child-friendly computer-based object selection task. In this, observers saw a dragon's favourite sweet under a neutral illumination and were required to pick the matching one from an array of eight sweets seen under a different illumination (blue, yellow, red, or green). This set contained a reflectance match (colour constant) and a tristimulus

match (colour inconstant). We ran two experiments, with two-dimensional scenes in the first experiment and three-dimensional renderings in the second. Twenty-six adults and 33 children took part in the first experiment; 26 adults and 40 children took part in the second. Children were more colour constant than adults and their constancy decreased with age in both experiments. We found differences across illuminations and sweets, but a similar pattern across both age groups. This surprising finding suggests that colour constancy may be fully functional from a young age, and that overriding it requires the development of other, possibly cognitive, mechanisms.

3.3 Introduction

The light reflected from a surface depends on both the surface reflectance and incident illumination. Colour constancy is the ability to judge surfaces as relatively invariant under different illuminations. This is crucial for object recognition - without it, objects would appear to change colour radically in different environments. Empirical measurements of colour constancy find varying levels depending on the method of measurement (including instructions; Arend and Reeves, 1986), surfaces and illuminations used, and the individuals tested (see Foster, 2011; Smithson, 2005, for a review). Typical methods of measuring colour constancy include achromatic setting (Brainard, 1998) in which observers adjust a patch to appear achromatic; asymmetric matching (Brainard et al., 1997) in which observers adjust a test patch to match a target patch under a different illumination; and object selection (Radonjić et al., 2015b) in which observers select one object from several to match a target seen under a different illuminant. Object selection tasks are more naturalistic, and generally elicit the highest levels of colour constancy (Radonjić et al., 2016).

An almost infinite variety of illumination and surface reflectance combinations may give rise to the same light signal at the eye. Bayesian models (Brainard et al., 2006; Brainard and Freeman, 1997; Olkkonen et al., 2016) propose that the visual system narrows down the possibilities and estimates the actual physical stimulus by learning the most likely combinations. Prior experience, acquired during development, is arguably necessary to learn these mappings (Beau Lotto, 2004). Developmental studies into colour constancy are therefore needed to empirically determine the relationship between experience and perception. Whilst animal studies have shown that experience with broadband illuminations is essential for colour constancy (Sugita, 2004), little is yet known about the development of colour constancy in children.

Many low-level systems necessary for colour vision develop during the first months of life (see Brown, 1990, for a review), including macular pigment density (Bone et al., 1988), cone contrast sensitivity and acuity (Morrone et al., 1990). Two-month-old infants can discriminate chromatic from achromatic surfaces (Peeples and Teller, 1975), and four-month-old infants have colour categories similar to those of adults (Franklin et al., 2005). Additionally, young infants exhibit other complex aspects of perception including transparency perception (Johnson and Aslin, 2000) and certain visual illusions (Yang et al., 2010). However, other aspects continue to develop, including detecting colour-defined form, which is not adult-like until teenage years (Hollants-Gilhuijs et al., 1998). This ability is likely to depend on processes independent of those giving rise to the experience of colour.

Few studies have investigated colour constancy in children. Dannemiller and Hanko (1987), Dannemiller (1989) and Yang et al. (2013a, 2015) used a preferential looking paradigm to study colour constancy in three- to seven-month-old infants and found rudimentary colour constancy by 4.5 months. Similarly, Chien et al. (2006) and Kavšek (2011) found looking behaviour consistent with lightness constancy in four- and six-month-old infants, respectively. Rogers et al. (2020) collected asymmetric colour matches with two- to four-year-old toddlers and concluded that they have poorer colour constancy than adults in these conditions, although there were large individual differences.

On the other hand, Katz (1935) anecdotally reports that children aged eight to 15 years may have equal or superior colour constancy to that of adults. However, this report is not supported by explicit data, so should be interpreted with caution. Similarly, Beck (1966) found that five-year-old children had poorer lightness constancy than adults when comparing arrays of chips under light vs in shadow, but no difference between age groups when judging a single chip's lightness.

The development of size and shape constancy has received more experimental attention. Some studies report that size constancy improves with age up to seven to nine years (Brislin and Leibowitz, 1970; Granrud and Schmechel, 2006; Kavšek and Granrud, 2012), although others find adult-like levels in three-year-olds (Tronick and Hershenson, 1979). This discrepancy may be explained by differences in instructions and strategies (Granrud, 2009; Rapoport, 1967). Kaess et al. (1974) found increasing shape constancy with age up to 19 years, whereas others found no effect of age (Field and Collins, 1977), or decreasing constancy with age, but only for small viewing distances (3 vs. 15 feet; Meneghini and Leibowitz, 1967). In shape and size constancy, the viewing distance matters as different mechanisms are involved in judging the depth and distance of near (e.g. stereopsis) vs. far objects.

According to a popular hypothesis, adults' colour constancy may be optimised for daylight illuminations, which vary in appearance from yellowish to blueish (Judd et al., 1964; Hernández-Andrés et al., 1999; Spitschan et al., 2017). Through experience, observers may have developed a "daylight prior", in line with Bayesian models (Brainard et al., 2006). The finding of higher constancy for scenes illuminated by blueish daylights than non-daylights (Delahunt and Brainard, 2004; Weiss et al., 2017; Pearce et al., 2014) partially supports this hypothesis. Whilst the development of a daylight prior has not been investigated, studies have found other perceptual priors developing between four and 12 years (Thomas et al., 2010; Yonas and Hagen, 1973; Stone, 2011; Chambers et al., 2018).

Taken together, the limited research on the development of colour constancy is inconclusive, with some studies suggesting children are better than adults and others finding the opposite. Furthermore, few of these studies have investigated the developmental trajectory of colour constancy with age, or used a comparable task for different age groups. No experiments have measured colour constancy in children older than four years, possibly because many methods of measuring colour constancy in adults (achromatic matching, asymmetric matching) are inappropriate for young children, either requiring observers to remember a colour, and/or perform fine-tuned matches which require a large attention span and fine motor skills. Therefore, we have adapted an object selection task from Radonjić et al. (2015a) to develop a novel, child-friendly measure of colour constancy with no memory demands, and no need to make explicit judgements about colours. We used this task to better understand the role of development in perceptual constancies. Specifically, we aimed to answer three research questions: (1) is there an overall difference between six- to 11-year-old children and adult observers' colour constancy? (2) Is there a developmental trajectory in colour constancy from six to 11 years? and (3) Does the pattern of colour constancy across surfaces and illuminations, such as effects driven by a daylight prior, differ between adults and children? We focussed on the age range (six to 11 years) in which developmental changes in size and shape constancy occur. In the first experiment, we aimed to measure colour constancy in children and adults using simple two-dimensional stimuli, comparable to those used in much previous research. In the second experiment, we measured colour constancy with more realistic three-dimensional rendered stimuli, to ask whether the findings apply to scenes which more closely resemble the real world.

3.4 Experiment 4

We developed a novel computer-based object selection task, which children would find engaging, based on the materials and methods of Radonjić et al. (2015a), using simple two-dimensional scenes. This involved finding a dragon's favourite "sweet" from a set seen under one illumination to match a target sweet seen under a neutral illumination. This experiment allowed us to measure baseline performance with minimal cues, and to confirm whether the task was appropriate for children. Six- to 11-year-old children, and adults participated, with scenes simulated under either daylight or non-daylight illuminants to determine whether overall colour constancy, or the pattern across illuminations, differs between children and adults.

3.4.1 Method

Observers

Twenty-six adults (mean age = 22.57 years, SD = 6.51; 5 male, 21 female) and 33 children aged between six and 11 years (11 male, 22 female) participated. Informed consent was given by adults and parents of children. The children assented to take part, and were repeatedly asked during the experimental session if they were happy to continue. All observers were screened for colour vision deficiencies using Ishihara plates (38 plates edition Ishihara, 2006). Two children had scores outside the normal range (more than two errors) so their data were not included in analyses. Adults were either psychology undergraduate students who took part for course credits, or paid volunteers. Children were rewarded with a small prize at the end of the experiment. All observers had normal or corrected to normal visual acuity.

Materials and Apparatus

Scenes were presented on an ASUS PA382Q 23" monitor with 10 bits per channel, controlled by an Nvidia quadro k600 graphics card. The monitor was characterized with a Konica Minolta CS-2000 spectroradiometer and monitor linearization and colour conversions were achieved with standard methods (Brainard et al., 2002). The computer presenting the stimuli and the testing room were changed after the first 20 participants (five adults and 15 children) due to availability of laboratory space and occasional computer malfunction. The monitor was recalibrated after moving, to maintain stimulus properties. To check this did not affect results, we ran a linear mixed effects (lme) model including computer setup as a predictor of colour constancy indices and found no significant main effect of computer setup, and no significant interaction with age group. For all conversions to CIE L*u*v*, the white point was defined as the chromaticity of the neutral illumination (D57), at a luminance of 60 cd/m^2 (Yxy = 60, 0.328, 0.344). In each setup, observers sat in the dark, approximately 60cm from the monitor, with free head movement. Experiments were run on MATLAB (The Math Works, Inc. , 2018), using functions from the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007).

Stimuli

Each experimental scene filled the monitor, subtending 46 $\times$ 27 degrees of visual angle. The scenes consisted of two halves (each 23 $\times$ 27 degrees) - a square target "sweet" was shown against a neutral background on the left-hand side, and eight competitors were shown against a "chromatic" background on the right-hand side (Fig. 3.1). All nine sweets were identical in shape and size (129 pixels^2; 3.3 degrees of visual angle). The target sweet was centred on the left side of the screen. The eight competitor sweets were aligned in two equally spaced rows of four, symmetric above and below a horizontal line at the vertical middle of the screen. Numerals were added above the top row and below the bottom row of competitors so that children unable to use the mouse could say which number to pick.

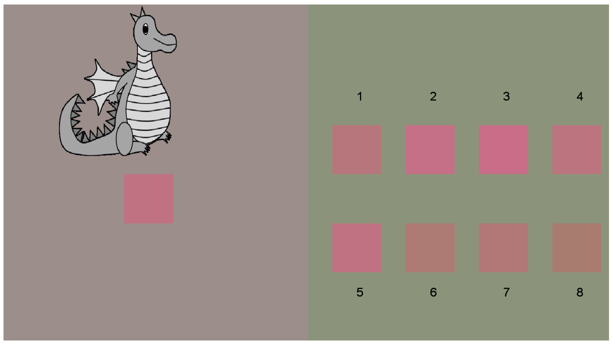


Figure 3.1: Example of a scene with a rose sweet under green illumination. Note that Derek the dragon was not present in scenes in the main experiment. Sweet 5 is the tristimulus match and sweet 6 is the reflectance match.

The backgrounds were designed to simulate different illuminations, so will hereafter be referred to as illuminations. The neutral illumination had a CCT of 5698K (D57), close to the peak chromaticity of the measured daylight distribution from 30 natural scenes (Nascimento et al., 2016). Four chromatic illuminations were used: two along the daylight locus in the blueish and yellowish directions, and two perpendicular to the daylight locus at CCT 5698K (in u*v*), which appear reddish and greenish. In all conditions the background was scaled to have a luminance of 60cd/m^2. At this luminance, all chromatic illuminations were $30\Delta E_{u*v*}$ from D57. Illumination chromaticities are given in Table 3.1.

Illumination	x	y	u'	v'
Neutral (D57)	0.328	0.344	0.203	0.478
Blue	0.297	0.314	0.193	0.458
Yellow	0.364	0.372	0.216	0.497
Red	0.329	0.310	0.217	0.460
Green	0.327	0.382	0.189	0.496

Table 3.1: Co-ordinates of the illuminations used in CIEYxy and CIEL'u'v' colour spaces

We used four target reflectances which appeared grey, green, rose, and teal under neutral illumination (Radonjić et al., 2015a). For each target and illumination, we generated eight competitors, equally spaced on a line in u'v', which contained a tristimulus match (T; indicating no constancy) and a reflectance match (R; indicating perfect constancy). An example of the competitors' chromaticities is shown in Fig. 3.2. Full details of how these were generated is given in the Supplementary Material. Although the competitors should have slightly different luminances due to the interaction between the illumination and reflectance, for simplicity we fixed the luminance of each competitor at 50cd/m^2. The difference in luminance between competitors before fixing was small, with a maximum discrepancy between T and R of 1.8cd/m^2.

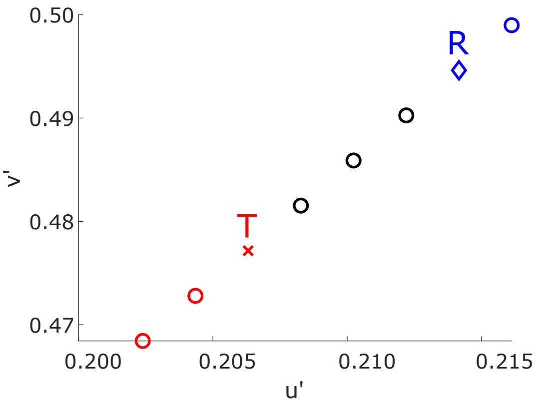


Figure 3.2: Example chromaticity of eight competitors in u'v'. Red X represents the tristimulus match, T; blue diamond represents the reflectance match, R; the black Os are the competitors in between T and R; red Os are underconstant and blue O is overconstant.

In addition to these sixteen experimental scenes (four illuminations × four sweet colours), there were two control conditions for each target sweet, in which both sides of the screen were illuminated by D57, to test for any internal biases - such as a preference to pick more saturated sweets. In these scenes, the background was uniform across the screen. Full details of how the eight competitors in these scenes were generated are given in the supplementary Material.

Overall, there were 24 conditions (6 illuminations × 4 sweet colours). Each observer was presented with each condition ten times, for a total of 240 trials, with the eight competitors positioned in a random order on each trial.

Task

The observers' task was to feed "Derek the dragon" his favourite sweets. They saw his "favourite sweet" – the target sweet – on the left-hand side of the screen under neutral illumination (D57). From the eight competitors on the right-hand side of the screen, observers used a mouse to click on the sweet they thought was his favourite. The instructions (see Supplementary Material) did not mention colour, or explain that the backgrounds simulated illumination changes. Therefore, no specific strategy was encouraged. Children who could not use a mouse spoke the number of the sweet to select, and the experimenter clicked on it. The selected sweet was then indicated by a small, uniform increase in size. Observers were allowed to change their selection as often as required, with unlimited time. To feed Derek the sweet, observers pressed the space bar.

Procedure

All adults completed all 24 conditions in a single session, except the first five who did not complete the red-green control condition, as this was added later. Fifteen children were also tested before the red-green control condition was added. Children completed as many trials as they could in a single session before becoming fatigued. Generally, the older children were more likely to complete all conditions than the younger children. Over all 33 children, 28 blocks (out of a possible 183) were not completed due to fatigue. The testing sessions lasted approximately an hour, with large variability between observers.

The full sessions were split into six blocks - one for each illumination. Within a block, each sweet was shown 10 times, resulting in 40 trials per block. Trial and block orders were randomised for each observer.

Each block began with two minutes of dark adaptation. In the first block immediately following the adaptation period, the instructions were presented on the screen by an animation of Derek the Dragon. The illumination of the first block was used in the example scene in the instructions. For children who struggled to read, the experimenter read out the instructions, and pointed to the relevant parts of the scene. Following the instructions, the first trial was presented.

Children were given a star chart at the beginning of the session and were rewarded with a star sticker after a set of 6, 7, or 8 trials, with the set length randomised each time. Between every trial, observers saw a black screen containing an energy bar, alongside text telling them to continue. The energy bar was filled incrementally, either 1/6, 1/7 or 1/8 after each trial depending on the set length. When the bar was full a black screen containing a large silver star was shown, alongside an animation of Derek talking with the text "Thank you for feeding me, you have earned a star!". This reward was designed to keep children motivated, without giving any meaningful feedback.

At the end of each block, an animation of Derek appeared alongside text saying "Thank you for feeding me so many sweets!". Between each block, observers were asked if they wanted to play the next round.

Data Analysis

We developed two criteria to exclude random responses. This was done for each illumination condition, but not separated by sweet colour, as sufficient responses were needed to test for potential randomness. There were, therefore, 40 responses (each a number between 1 and 8 indicating the competitor selected on that trial) to assess. A uniform and a Gaussian distribution were fitted to these responses. Sets of responses where the Bayesian information criterion (BIC) of the Gaussian fit was larger than, or within six units from, the BIC for the uniform fit were deemed as equally or better fit by the uniform than the Gaussian distribution, and were excluded from analysis. This excluded three conditions, all from children. Additionally, any set of responses with SD greater than 2.32 (the SD of 40 responses uniformly distributed amongst eight competitors) were excluded. This excluded five conditions, all from children. Due to overlap between the exclusion criteria, five conditions in total were excluded, from four children, representing 3.23% of conditions completed by children. No data were excluded from adults.

To analyse the remaining data, all responses were first converted to the associated chromaticities of the competitors in CIE u'v'. T had the same chromaticity as the target sweet, despite the change in illumination, and therefore indicated no colour constancy, while R simulated the target sweet under the chromatic illumination, and therefore indicated perfect constancy. The observer's colour constancy index (CCI) was calculated for each trial as:

$$CCI = 1 - \frac{a}{b}, \tag{3.1}$$

where a is the signed Euclidean distance, in u'v', from R to the competitor selected, which was negative for over-constancy matches, and b is the Euclidean distance from R to T. Possible CCIs ranged from -0.25 for under-constancy matches to 1.5 for over-constancy matches, with CCIs of 1 indicating perfect constancy.

Fig. 3.3 shows an example of how the chromaticity of the sweets selected corresponds to the CCIs.

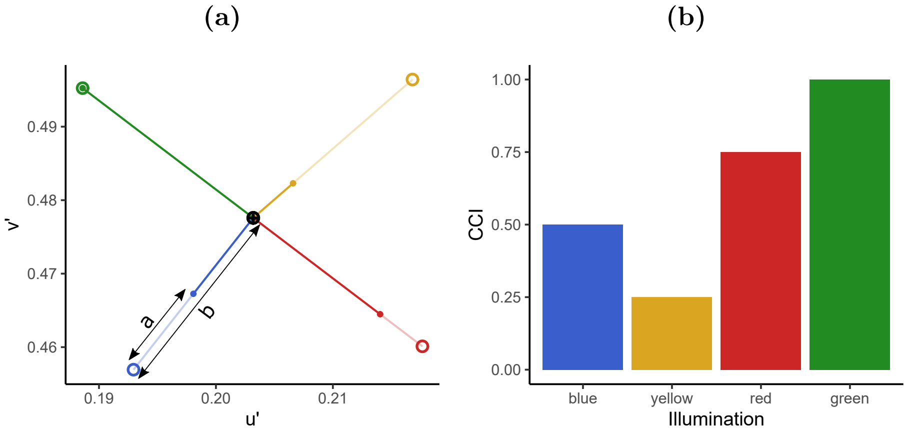


Figure 3.3: (a): Example of an observer's matches for grey sweets in u'v'. The black filled O is the tristimulus match (T). The open coloured Os are the reflectance matches (R) under each of the four illuminants, with the colour of the O corresponding to the colour of the illumination. The filled coloured Os are the observer's matches under each illuminant, with the colour corresponding to the illumination. Pale lines show the whole distance from T to R (b in Eq 3.1); darker lines show the distance from T to the match. (b): the colour constancy indices associated with the chromaticities in (a). As this observer selected R under green, they had perfect colour constancy under green, but they chose a competitor close to T under yellow, resulting in poorer constancy.

In order to test our three research questions, we ran lme models in R (R Core Team, 2020) using the lme4 package (Bates et al., 2015). As many children did not complete all 16 conditions and some conditions were excluded, it was not possible to conduct an ANOVA. We created two models: one to investigate main effects of age group, illumination, and sweet colour:

$$CCI \sim ageGroup + block + illumination + (1|observer) \tag{3.2}$$

And a second to test for interactions between these variables:

$$CCI \sim ageGroup * block * illumination + (1|observer) \tag{3.3}$$

3.4.2 Results

The mean chromaticities of the sweets selected by adults and children under each of the 16 conditions are shown in Fig. 3.4. These graphs are in the same format as Fig. 3.3(a) but show all four sweet colours, with the corresponding target sweet shown above. The pattern across adults and children is similar, with large differences between sweet colours and illuminations.

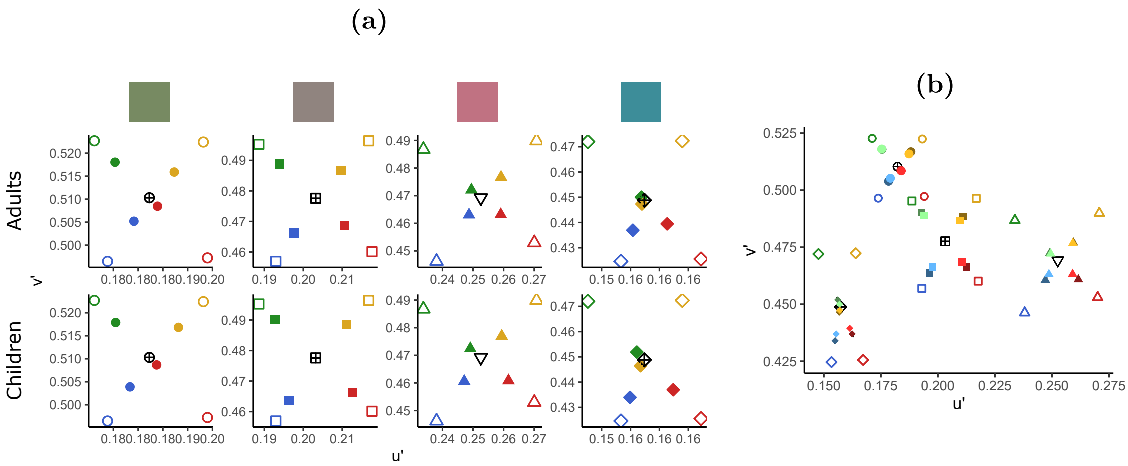


Figure 3.4: Mean chromaticities of sweets selected in u'v', formatted as in Fig. 3.3(a) for each sweet colour separately (a), with the shape families representing different sweet colours: Os are green; squares are grey; triangles are rose; diamonds are teal. Adults' results are in the top row and children's are in the bottom row. (b) shows the mean chromaticities selected by both adults (pale symbols) and children (dark symbols) for all sweet colours. In all plots, black symbols are the targets/tristimulus matches, open coloured symbols are reflectance matches, and filled coloured symbols are mean matches across observers. The colour of the symbols indicates the colour of the illumination. Error bars (too small to be visible in many cases) show ±1 SEM.

In order to meaningfully analyse colour constancy performance, we first needed to ensure that observers could discriminate between the competitors, and accurately match without colour constancy demands. To this end, we analysed the neutral control conditions. Observers performed generally well in these, with a mean $\Delta E_{u'v'}$ (Euclidean distance) from the target of 0.00242, corresponding to a just-noticeable-difference (L*u*v*) of 3.74. In a 2 (condition) × 4 (sweet) × 2 (age group) ANOVA, deviations were significantly larger in the blue-yellow control than the red-green control condition (p = 0.0152), with a bias towards yellower competitors for Green, Grey, and Rose sweets, and a bias towards bluer competitors for Teal sweets. Observers had a small bias towards green for Green and Grey sweets, and towards red for Rose and Teal sweets. There was no significant difference between adults and children (p = 0.603), and both followed the same pattern.

Our first research question was whether CCIs differed between adults and children. Fig. 3.5(a) shows that children (dark violins) had, on average, higher CCIs (closer to 1) than adults (light violins) under all four illuminants (collapsed across sweet colours). Fig. 3.5 shows the same pattern, comparing individual children (b) with adults (c). In a lme model with age group (children or adults), illumination, and sweet colour as predictors of CCI (Equation 3.2), the main effect of age group was significant ($p<.001$, $t=9.605$), with higher CCIs in children (mean = 0.436, SD = 0.373) than adults (mean = 0.364, SD = 0.303) (Fig. 3.6).

To determine whether CCIs changed over childhood, addressing our second question, we ran robust regressions on the children's data using the rlm function of the MASS package in R (Venables and Ripley, 2002). We used age as a predictor of CCI for each illumination separately, collapsed across sweet colour and repetitions. All four regressions (Fig. 3.5(b)) were significant and negative (Blue: B = -0.0390, p = 0.0147; Yellow: B = -0.0252, p = 0.0054; Green: B = -0.0241, p = 0.0061; Red: B = -0.0436, p <0.001), suggesting that CCIs decrease with age between six and 11 years under all four illuminations. B denotes the change in CCI for each year of age; for the Blue illumination, a 7-year-old's CCI is 0.0390 lower than a 6-year-old's, on average.

To determine more specifically the shape of the developmental trajectory, we asked which of four distinct models best fit the CCIs as a function of log(age) for all data (children and adults): linear; jump model, which predicts a step-like change in CCIs; hockey stick model, which predicts a linear change up to a certain age, after which CCIs plateau; and no change with age. We used log(age) to reduce the age gap between adult and child observers. The models were fitted to the data for each illumination condition separately, to determine the interaction between age and illumination. The BIC was determined for each model to determine which best fitted the data, and are all given in the Supplementary Material. The models with

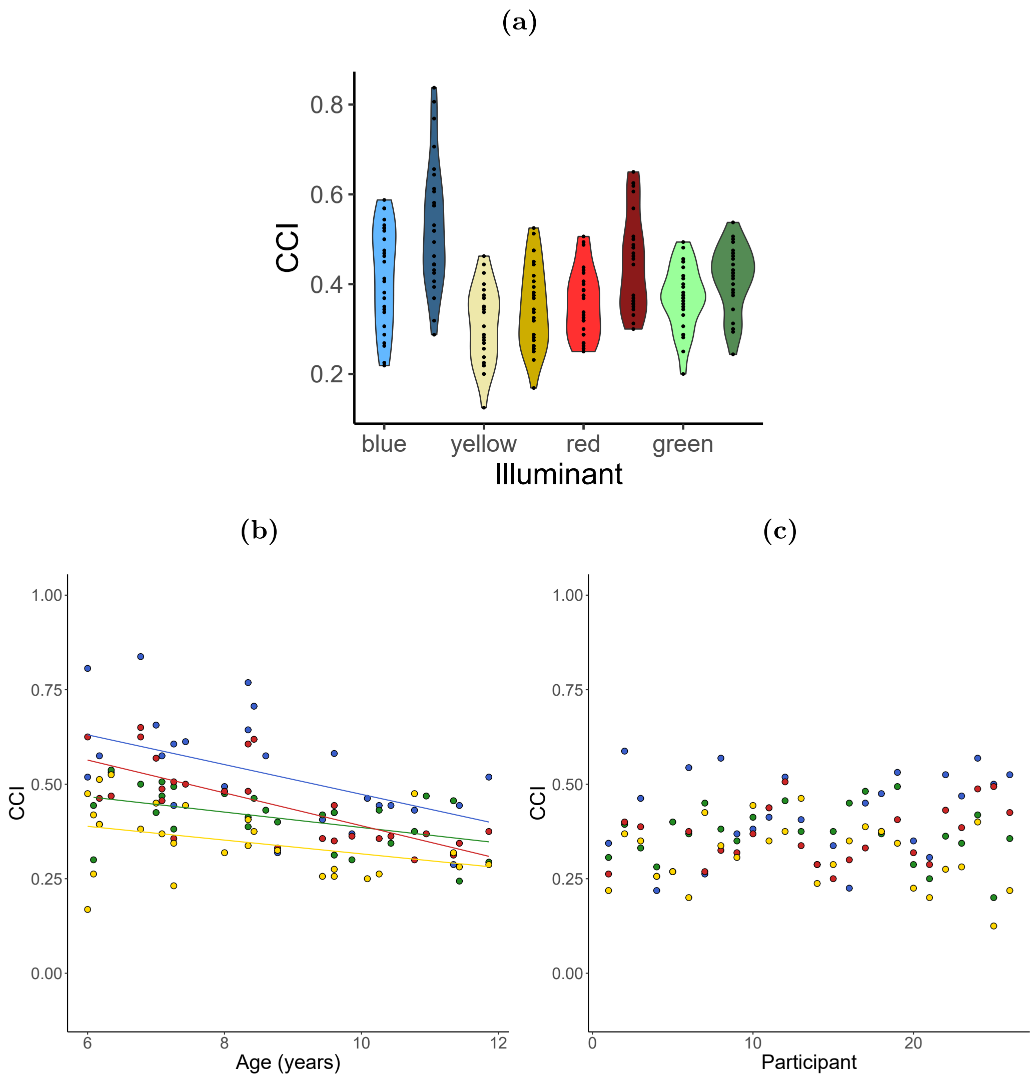


Figure 3.5: (a): Mean colour constancy indices (CCIs) for children (dark violins) and adults (light violins) collapsed across sweet colours, for each illumination condition separately. (b): Age against mean CCI across all sweet colours for each child and illumination condition separately. Colours of points represent illumination colour. Regression lines are shown for each illumination condition. (c): Mean CCI across all sweet colours for each adult and illumination condition. Participants are sorted in ascending age order.

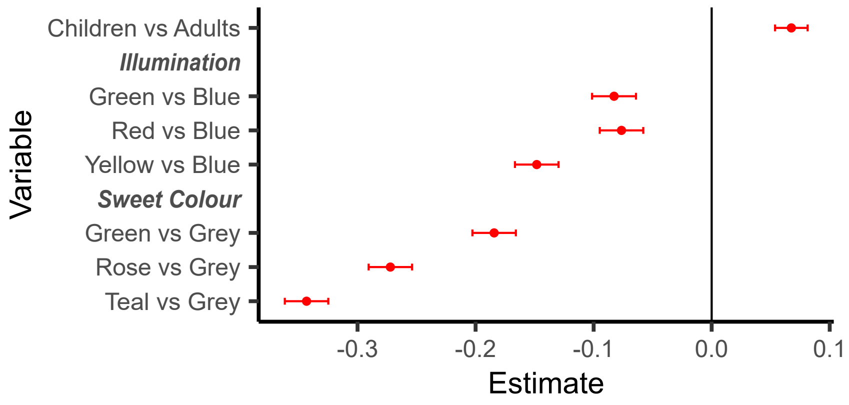


Figure 3.6: Main effects model results. Red circles show estimates; error bars show 95% confidence intervals. Black vertical line at Estimate = 0 reflects non-significance. Estimates greater than 0 indicate the first variable (e.g. "green" in "green vs blue" predicts higher CCIs; estimates less than 0 suggest the variable on the right hand side predicts higher CCIs.

the lowest BIC, and therefore the best fits, were the jump model for Blue and Red illuminations, and the linear model for Green and Yellow. For Blue, the jump was at 8.6 years, with the mean CCI 0.609 below this age and 0.422 above. For Red the jump was at 8.4 years, with CCIs higher below this age (mean = 0.526) than above (mean = 0.357). For green, β was -0.0599, and for yellow β was -0.576. These models are plotted alongside the data in Fig. 3.7.

To test whether the pattern of CCIs across illuminations and reflectances differed between adults and children - our third research question - we ran a second lme model to look for interactions (Fig. 3.8), described in detail in the Supplementary Material.

The two lme models show that CCIs differ significantly across illuminations (highest under Blue and lowest under Yellow) and sweet colours (highest for Grey and lowest for Teal), and that there are significant interactions between the two. None of the two-way interactions involving age group are significant, indicating that the patterns across illuminations and reflectances are similar across adults and children. The two significant three-way interactions (children|Green$_{sweet}$|Red$_{illumination}$ (p = 0.007); children|Teal$_{sweet}$| Yellow$_{illumination}$ (p = 0.003) indicate that the two-way interactions for Green$_{sweet}$ |Red$_{illumination}$ and Teal$_{sweet}$|Yellow$_{illumination}$ are more pronounced for children than adults.

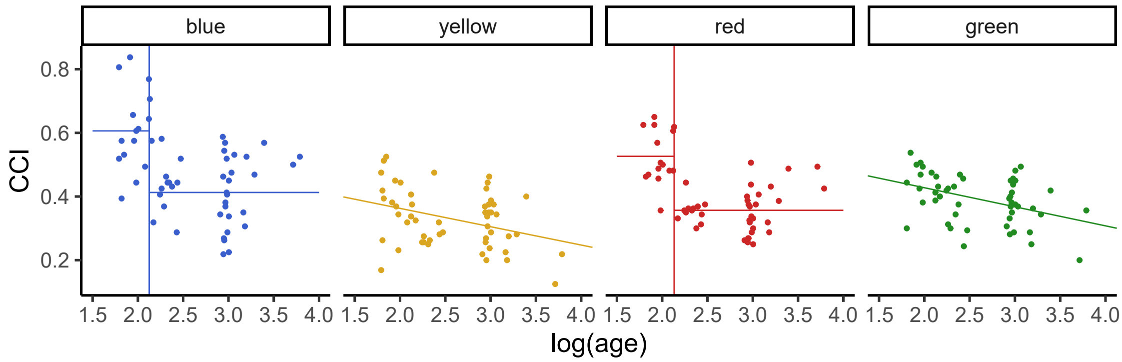


Figure 3.7: CCIs against log(age) with each of the best fitting models for each illumination. For Blue and Red illuminations, the best fitting model is the jump model, with the vertical line indicating the jump and the horizontal lines indicating the mean CCI below and above this age. For Green and Yellow, the best fitting model is linear, with the regression line shown.

3.4.3 Interim Discussion

In this object selection task, we found that children's colour constancy indices were higher than adults' and decreased with age between six and 11 years. Although surprising, the result agrees with some previous reports of superior constancy in children (Meneghini and Leibowitz, 1967; Katz, 1935).

As well as the overall change in CCIs with age, we were interested in the pattern across illuminations and reflectances for adults and children. Both adults and children had the highest CCIs under the Blue illumination, possibly due to a daylight prior, in agreement with Delahunt and Brainard (2004). Furthermore, children and adults had the highest CCIs for grey sweets, in line with previous research (Olkkonen et al., 2009). Interestingly, there was a strong interaction between sweet colour and illumination, such that CCIs were generally higher when the sweet's reflectance and illumination had similar chromaticities (e.g. Green sweets under Green illumination). As with the main effects of sweet colour and illumination, this interaction did not vary by age group. Whilst there were few significant interactions involving age group in the lme model, different models best explained the developmental trajectory under Blue and Red illuminants (jump model) compared to Yellow and Green (linear model). The developmental trajectory may therefore vary across illuminations.

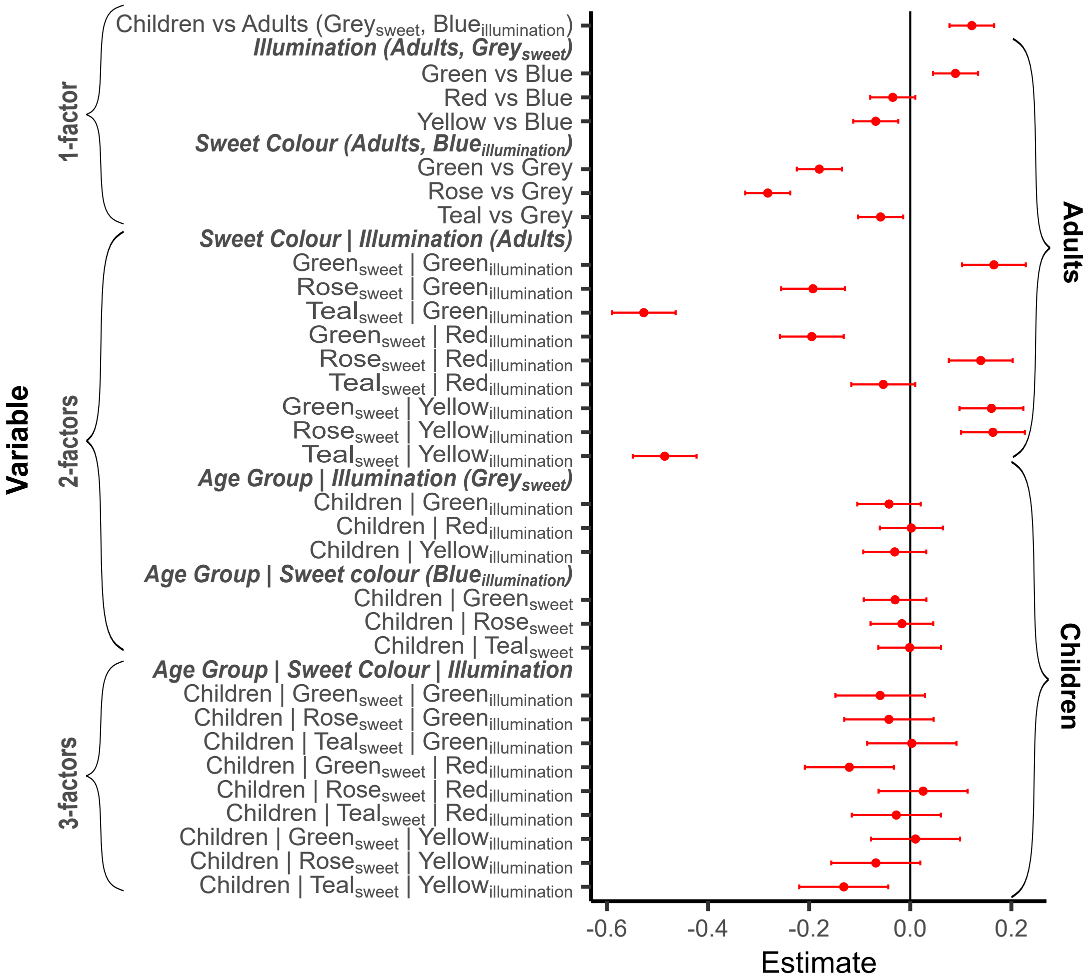


Figure 3.8: Interaction effects model results. Red circles show estimates; error bars show 95% confidence intervals. Bold, italic text is added to divide the different factors. Parentheses show the level of the fixed factors, such that the first effect is a comparison of children compared to adults with grey sweets under the blue illumination. See text for full explanation of results.

This experiment provided a useful way to measure colour constancy in children and adults in a simple, controlled scene, using a novel task. However, the simplicity of the stimuli limits their relevance to colour constancy in real life, as many cues are missing, and observers might not interpret the backgrounds as different illuminations. Therefore, in Experiment 5 we used three-dimensional rendered scenes, to determine whether their additional cues to the illumination geometry would be exploited by children and/or adults.

3.5 Experiment 5

To measure colour constancy in a more realistic environment, in Experiment 5 we used three-dimensional computer rendered stimuli. The task, illuminations, and sweet colours were the same as in Experiment 4.

3.5.1 Method

Observers

A further 26 adults (mean age = 25.2 years, SD = 10.1; seven male, 19 female) and 40 children aged six to 11 years (15 male, 25 female) participated. All observers were naive to the purposes of the study, apart from one adult who was involved in collecting data.

Materials and Apparatus

The same materials were used as in Experiment 4. The computer and testing room were those employed in the second half of Experiment 4.

Stimuli

An example of the stimuli used is shown in Fig. 3.9. The geometry of the three-dimensional scenes was created using Blender (`https://www.blender.org/`), and they were rendered using Mitsuba (`http://www.mitsuba-renderer.org/`) compiled for spectral rendering. Mutual illumination was excluded from the rendering

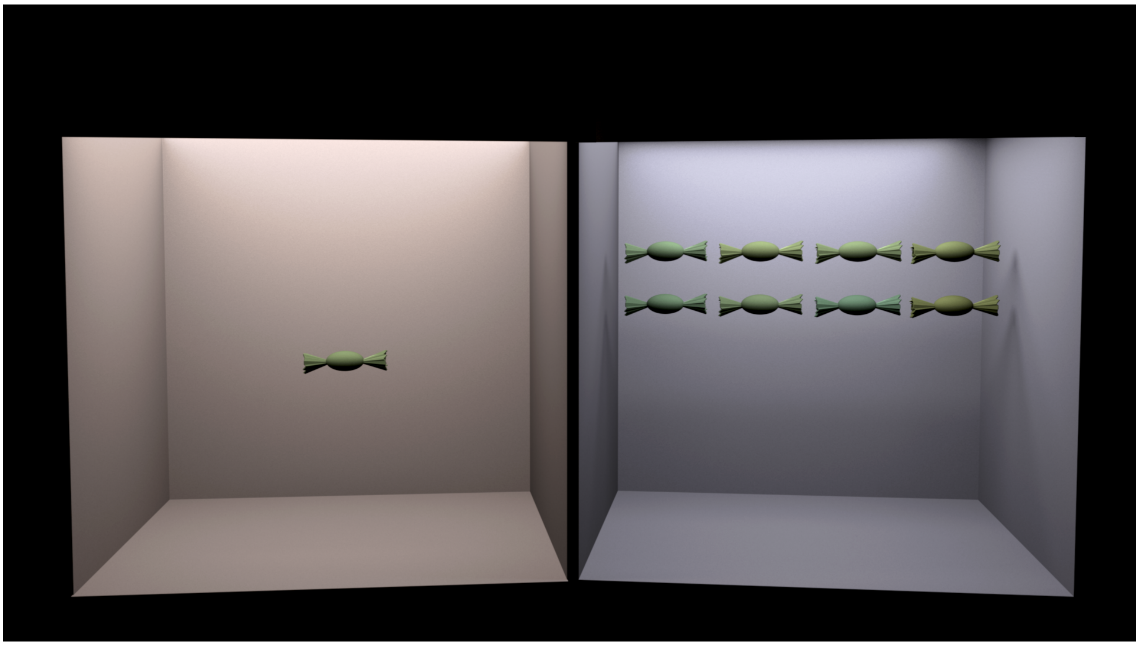


Figure 3.9: Example of a 3D scene containing green sweets under the blue illumination on the right.

process by limiting the number of bounces to one, to ensure the chromaticities of the sweets were not affected by surrounding sweets or walls.

The scenes consisted of two boxes with open fronts. A square area light encompassed the ceiling of each box. The lights were hidden by wall fragments on the top fronts. As in Experiment 4, the left box was always illuminated by D57, and the right box was illuminated by either a Blue, Yellow, Red, or Green illumination - each $30\Delta E_{u*v*}$ from D57. The boxes subtended approximately 11 $\times$ 10 degrees of visual angle.

All surfaces within the boxes had a spectrally neutral reflectance of 0.5. The sweet shape was identical for all competitors and was generated by shrinking a UV sphere in two dimensions to create an ellipse. Cones were added to either side with the vertices randomly deformed to look like the edge of sweet wrappers. A single target sweet was placed in the left-hand box, roughly in the middle, with two rows of four competitor sweets in the right-hand box. Each sweet subtended approximately 1.8 $\times$ 0.4 degrees of visual angle at a viewing distance of 60cm.

The sweets' reflectances were generated using the rough plastic material in Mitsuba, with an alpha of 1, which is perfectly rough, and a specular reflectance of 0, which is perfectly matte. The chromaticities of the target sweets under the neutral illumination were the same as in Experiment 4: Grey, Green, Rose, and Teal. For each chromatic illumination and each target sweet, the chromaticities of the eight competitor sweets were evenly spaced in L*u*v*, and corresponded to a reflectance match (R), a tristimulus match (T), and 6 alternative surfaces, including one over-constant and two under-constant matches, as in Experiment 4. To generate the 3D rendered sweets, spectral reflectance functions for each alternative sweet were computed, which would deliver the desired evenly spaced chromaticities; for details see Supplementary Material.

As in Experiment 4, in two control conditions both boxes were illuminated by the Neutral illuminant (D57). See Supplementary Material for details.

For each of the 24 conditions (4 sweets $\times$ 6 illumination conditions), 10 scenes were rendered with the eight competitors in a different, random, order. The hyperspectral rendered scenes were converted to RGB images using the calibration file described in Experiment 4. As some pixels in the rendered images were out of gamut, these were truncated by converting any pixels with R, G, or B greater than 1 to 1, and converting any values less than 0 to 0.

During rendering and truncation, the sweets' chromaticities shifted slightly from the predicted chromaticities. Therefore, during analysis, the chromaticity of the rendered sweet selected was used, as shown in the Supplementary Material.

Task

The task was the same as in Experiment 4. The observer's selected sweet was indicated by a black arrow instead of an enlargement.

Procedure

All adults completed all conditions within a single session lasting roughly an hour. Most children completed all conditions. Four children, all aged six or seven years, did not complete all conditions due to fatigue, for a total of 12 conditions (out of a possible 240).

The procedure was identical to that in Experiment 4, except additional instructions were given before the main instructions. These consisted of a demonstration scene containing the boxes from the main experiment with perfectly reflective

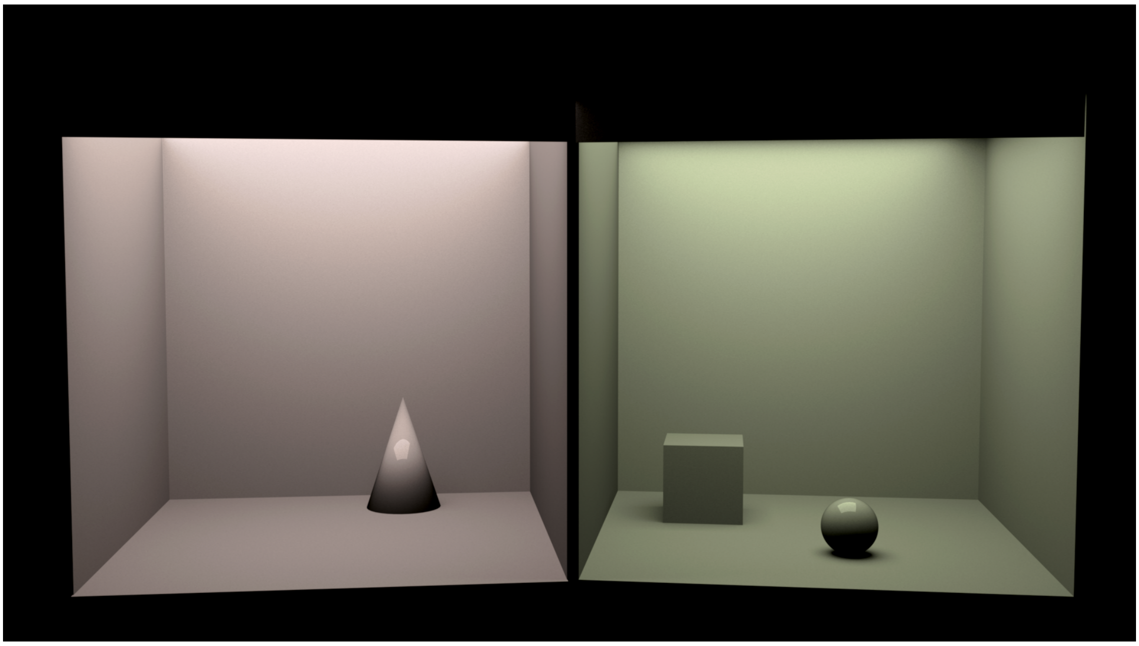


Figure 3.10: Demonstration scene used to show the light changing colour. In this example, the light in the right-hand room is green, and the light in the left-hand room is neutral. In this and all other representations of stimuli, images have been tonemapped for illustrative purposes.

(white), specular, objects inside: a cone, a sphere, and a cube (Fig. 3.10). The experimenter explained to observers that the light would change colour on every round, while manipulating the illumination in the right-hand box to illustrate the effects. This ensured that observers understood that it was the illumination, rather than the wall, changing colour. A three-dimensional rendered image of Derek the dragon gave the remainder of the instructions, with the same text as in Experiment 4.

Data Analysis

The same exclusion criteria as in Experiment 4 were applied. Ten illumination conditions from children were better fit by a Uniform than a Gaussian distribution, and the SD of responses was greater than 2.32 in 22 conditions from children. Due to overlap between the two criteria, 24 conditions, representing 10.5% of those completed by children, were excluded. No adults' data were excluded by either of these criteria. However, two adults were removed from analysis as their mean CCIs were more than 3 SD from the mean of all adults; one of these was a non-naive observer who collected part of the data.

When calculating CCIs, the responses were converted to the chromaticity (in u'v') of the sweet in that particular rendering. "T" was defined as the chromaticity of the rendered target sweet. "R" was the chromaticity the rendered target sweet would have under the chromatic illuminant (see Supplementary Material for further details).

As the chromaticities of the rendered sweets did not always fall on a straight line between T and R (as defined above), CCIs were calculated by defining a as the (signed) vector projection of the line joining R to the response, onto the line joining R to T, using Equation 3.1. In separate analyses, we calculated CCIs by defining a as the (signed) Euclidean distance from R to the response, but the results were no different to those reported here.

3.5.2 Results

The mean chromaticities of the rendered sweets selected by adults and children are shown separately for each sweet colour in Fig. 3.11(a), and together in Fig. 3.11(b). These show a similar pattern across adults and children which varies dramatically across sweet colours.

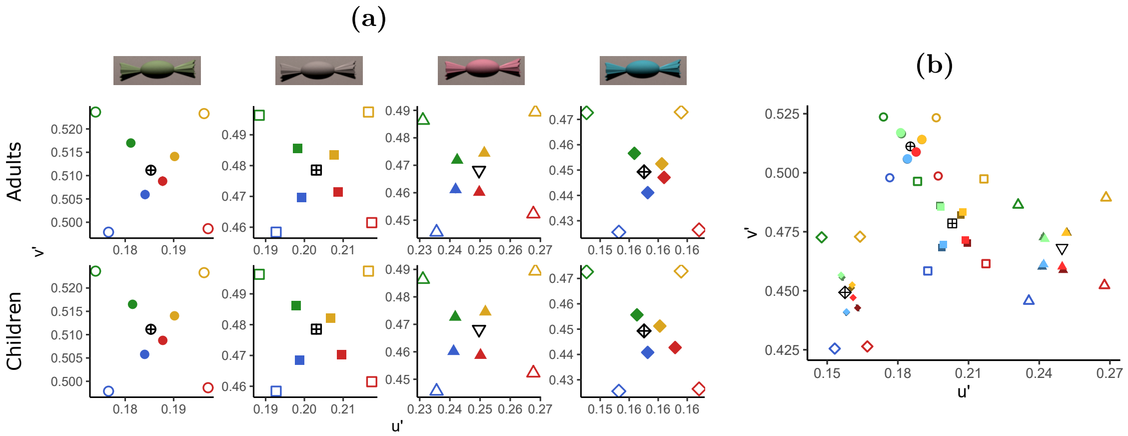


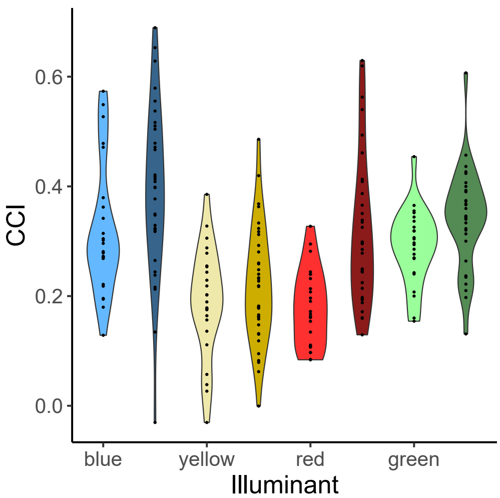
Figure 3.11: Mean chromaticities of sweets selected. (a) shows the chromaticities for each sweet colour separately: Os are green; squares are grey; triangles are rose; diamonds are teal. The top row shows the adults' results and the bottom row shows the children's results. (b) shows the mean chromaticities selected by both adults (pale symbols) and children (dark symbols) for all sweet colours. In all plots, black symbols are tristimulus matches, open coloured symbols are reflectance matches, and filled coloured symbols are mean matches.

To check that observers could discriminate the sweets and perform accurate matches without colour constancy demands, we ran an ANOVA on the data from the neutral control conditions. Across all observers and both control conditions, the mean $\Delta E_{u'v'}$ was 0.00494. Deviations did not differ significantly along the blue-yellow vs red-green axes. There was a significant difference between the sweet colours, with the smallest error for Grey sweets, and the largest for Teal sweets. Children had overall higher $\Delta E_{u'v'}$ than adults ($p<.001$), but the pattern across sweet colours was consistent across age groups.

To answer our first research question, and determine whether CCIs differed between children and adults, we ran a lme model with age group, illumination, and sweet colour as predictors of CCI. As can be seen in Fig. 3.12(a), children (dark violins) have higher CCIs, on average, than adults (light violins) in all four illumination conditions. Individual children's CCIs for each illumination condition are shown in Fig. 3.12(b), and adults' CCIs in 3.12(c). The results of the lme model in Fig. 3.13 show a significant main effect of age group ($p<.001$), with higher CCIs in children (mean = 0.315, SD = 0.362) than in adults (mean=0.245, SD = 0.294).

To determine whether colour constancy changes over childhood (our second re-

(a)



(b) **(c)**

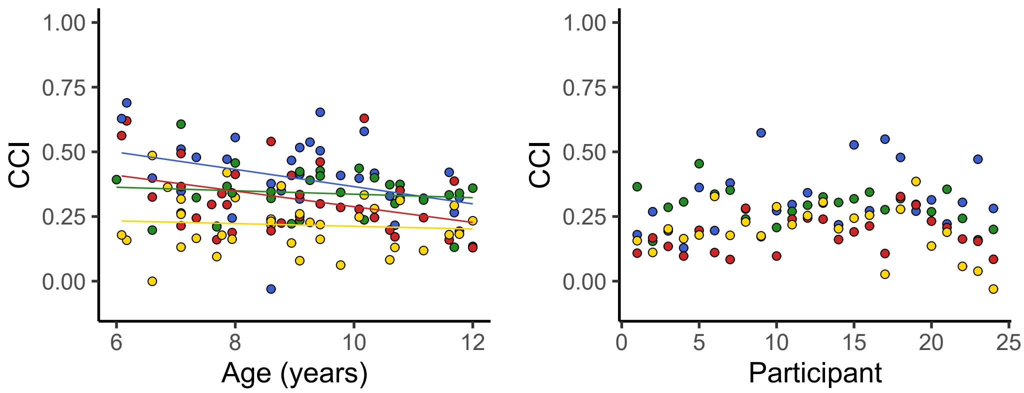


Figure 3.12: (a): Mean CCIs for adults (pale) and children (dark) for each illumination (b): Individual children's CCIs against age, with regression lines for each illumination. Colours reflect illuminations. (c): Individual adult's CCIs for each illumination condition. Horizontal axis is age rank of observer

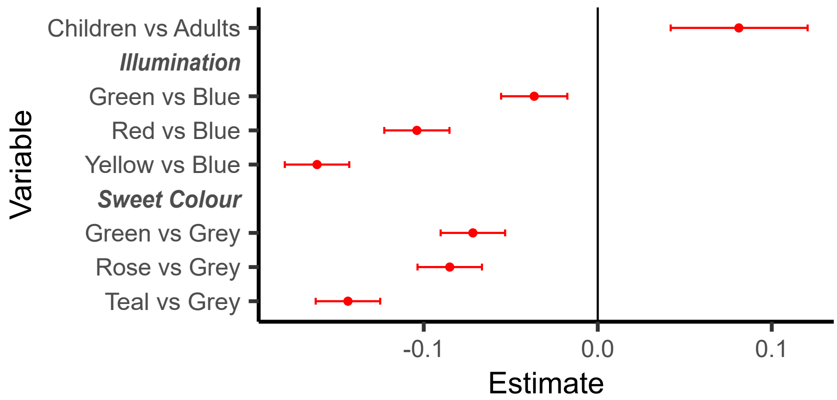


Figure 3.13: Main effects model results. Red circles are estimates, with error bars showing 95% CIs. For estimates greater than 0, the comparison condition predicts higher CCIs; for estimates lower than 0 the reference condition predicts higher CCIs.

search question), robust regressions were fitted to the children's data for each illumination condition separately. Under both Blue (β = -0.036, p = 0.019) and Red (β = -0.031, p = 0.020), but not Yellow or Green, age negatively predicted CCI. To further explore the developmental trajectory, we fit four models (linear, jump, hockey stick, and no change) to the CCIs as a function of log(age). The linear model was the best fit for the Blue (β = -0.0644), Green (β = -0.0515), and Red (β = -0.116) illuminations. Under the Yellow illumination, the no change model was the best fit. The best fitting models are plotted in Fig. 3.14. The BIC for each model are shown in the Supplementary Material.

The main effects model found significant differences in CCIs across illuminations (highest for Blue) and sweet colours (highest for Grey). To determine whether the pattern across illuminations and sweets differed across age groups, addressing our third research question, we ran a lme model to test for any interaction effects. Illumination, sweet colour, and age group were added to the model as interacting predictors of CCIs (Fig. 3.15). There was a significant interaction between illumination and sweet colour, visible in Fig. 3.11. For Grey sweets, children had higher CCIs than adults under the Blue illumination, but the Age Group | Illumination interaction suggests that the difference between children and adults was significantly smaller under the Yellow illumination. Furthermore, the difference between adults and children under the Blue illumination did not depend on sweet colour, but there

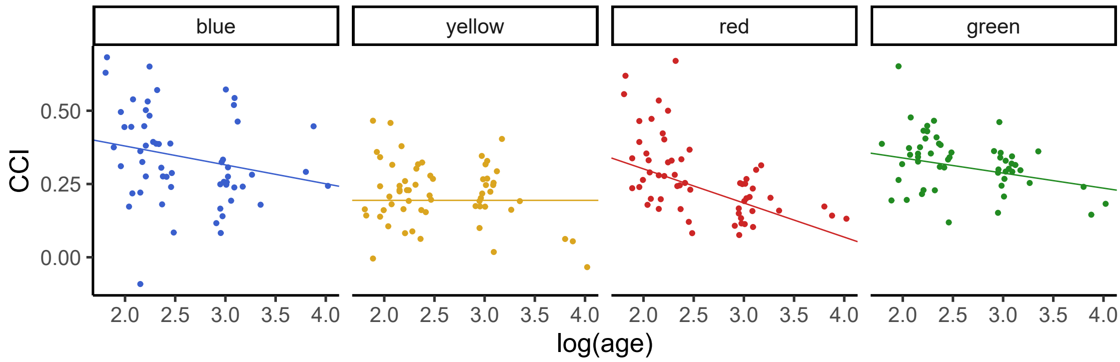


Figure 3.14: CCIs against log(age) with the best fitting models for each illumination. For Blue, Red, and Green the best fitting model is linear, with the regression line shown. For Yellow, the best fit is no change, with the mean across all ages plotted as a horizontal line.

was a significant positive estimate for children|Teal$_{sweets}$|Red$_{illumination}$. This means that, while adults have a negative interaction for Teal$_{sweet}$|Red$_{illumination}$, this effect is either smaller, non-significant, or reversed in children.

3.5.3 Interim Discussion

In this experiment we measured colour constancy in children and adults using realistic three-dimensional computer rendered stimuli, to answer three research questions. First, we found that children had higher colour constancy than adults, as in Experiment 4. Second, we found that under the Red and Blue, but not Yellow or Green, illuminations constancy decreased with age over childhood. This is in contrast to Experiment 4 where constancy decreased with age under all four illuminations.

Third, considering the pattern across illuminations and sweets, the highest CCIs were under the Blue illumination and for Grey sweets, in agreement with Experiment 4. The superior performance under the Blue illumination may indicate use of a daylight prior, although it does not extend to the Yellow illumination used here, which elicited the lowest CCIs. As in Experiment 4, there was a significant interaction between sweet colour and illumination. In Fig. 3.11, CCIs appear to be highest when the chromaticities of the illumination and sweet are similar. There were few interactions involving age group, suggesting a similar pattern across sweets and illuminations. However, there was a significantly smaller difference between the

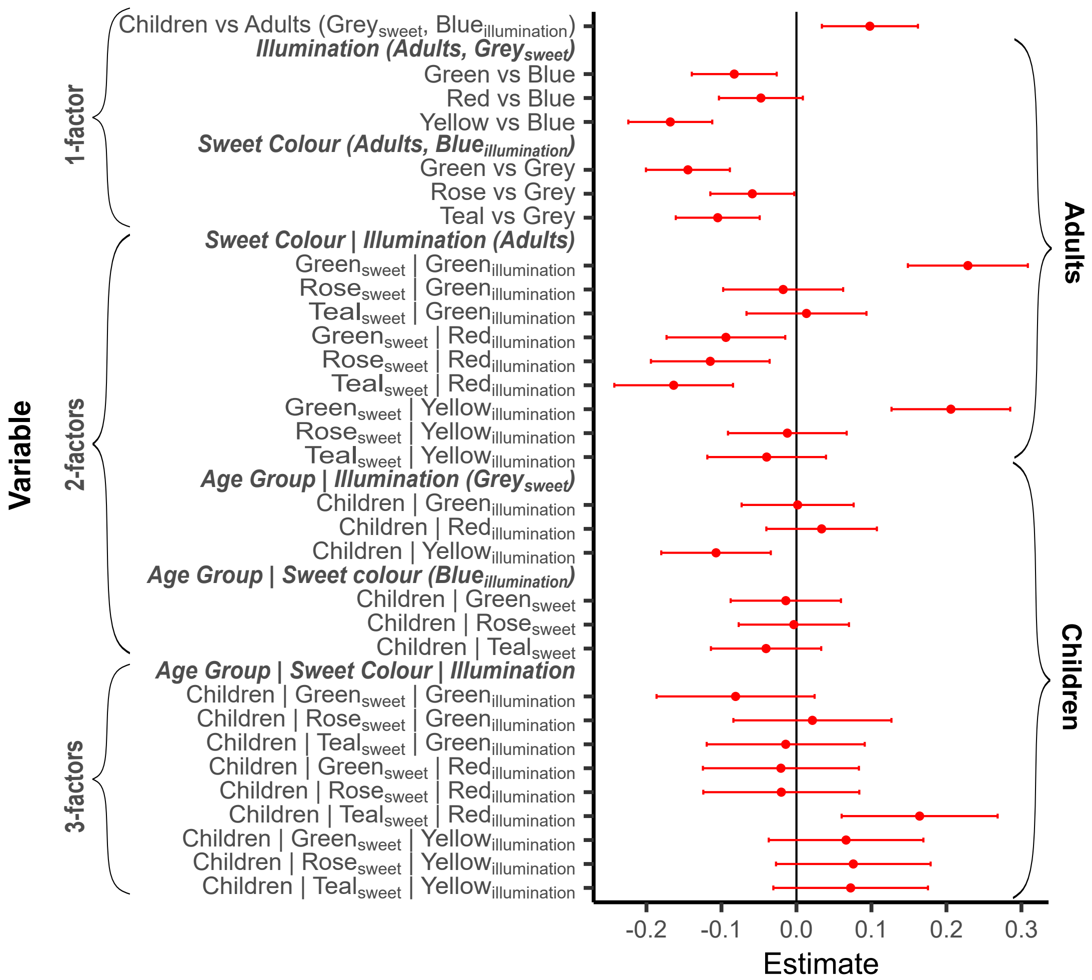


Figure 3.15: Interaction model effects. Red circles show estimates, with 95% CI error bars. See text in Experiment 4 for full explanation.

Yellow and Blue illuminations for children compared to adults. Additionally, model fitting found the developmental trajectory was best fit by a linear model for the Blue, Green and Red illuminations whereas the data were best fit by a no change model under the Yellow illumination, suggesting CCIs do not change with age. Taken together, these findings suggest that the developmental trajectory is different under the Yellow illumination compared to the others.

Overall, the results generally agreed with those from Experiment 4, although the CCIs were somewhat lower.

3.6 Discussion

In this study, we ran two experiments, using a novel measure of colour constancy, to determine the nature of the development of colour constancy. This task was designed to be appropriate for measuring colour constancy in six- to 11-year-old children. Overall, the pattern of results was similar across the two experiments.

We first asked whether constancy differed between adults and children, and examined the developmental trajectory across childhood. In both experiments we found a significant difference between age groups, with children demonstrating better colour constancy than adults, as shown by picking sweets closer to the reflectance match, resulting in higher CCIs. Robust regressions found that CCIs decreased with age from six to 11 years under all illumination conditions with two-dimensional stimuli (Experiment 4), and under Blue and Red illuminations with three-dimensional rendered stimuli (Experiment 5). In Experiment 4, model fitting showed a step-like decrease in colour constancy at around 8.5 years under Blue and Red illuminations and a linear decline with age under Green and Yellow. In Experiment 5, the data were best fit by a linear model under all illuminants except Yellow, in which CCIs did not change with age. Taken together, these findings suggest that colour constancy decreases with age. Whilst this finding may appear counter-intuitive, and differs from the findings in toddlers (Rogers et al., 2020), it agrees with findings described by Katz (1935). Katz reports that Brunswik found colour constancy to peak between eight and 15 years, after which it decreases up to adulthood. Interestingly, in another set of experiments, Katz notes that Burzlaff found that the developmental trajectory depends on the experimental setup, such that children are as good as adults on an object selection task, with almost perfect constancy, but performance on an adjustment task improves with age. The task used here is more like the former, in which children demonstrate high degrees of colour constancy. It would therefore be interesting to determine whether the results found here apply to other experimental setups. However, matching tasks have additional cognitive and motor demands, which could explain the developmental trajectory found by Katz. Furthermore, the findings reported by Katz are to be taken with caution, as he reports the constancy indices for only one child and no statistical analyses.

It is important to consider other possible explanations for the decreasing CCIs with age, apart from declining colour constancy. Firstly, children's data are generally more noisy than adults' data, which could be due to them responding more randomly. If an observer were responding entirely at random, we would predict CCIs of 0.375. In Experiment 4, adults' CCIs were closer to this (0.361) than children's (0.425), whereas in Experiment 5, the lower CCIs meant children were closer to random (0.315) than adults (0.245). It is, however, unlikely that children would respond randomly in one experiment only. Furthermore, we excluded conditions with random responses before analysis.

A more likely alternative explanation is a difference in the interpretation of the deliberately ambiguous instructions. Observers were told to find the sweet that Derek the Dragon would "like best". They were not told whether to match based on the hue and saturation, or to pick the sweet which has the same reflectance as the target (paper match), and there was no explicit mention of colour. This was to determine what observers would do in the real world, without an explicit strategy. However, if observers were attempting to make a hue/saturation match, the "correct" response would be the tristimulus match, whereas for a paper match the correct response would be the reflectance match. The greater cognitive effort required in making a hue/saturation match might make it harder for children to adopt that strategy, since they must override in-built colour constancy mechanisms. Future studies manipulating instructions may help to determine whether different strategies drive differences in performance, at least in adults.

A difference in adaptation to the test illumination may also explain the difference in CCIs between children and adults. Adaptation is known to be one of the major mechanisms of colour constancy (Werner, 2014; Fairchild and Lennie, 1992; Smithson and Zaidi, 2004). Children may have adapted more to the test illumination, either due to having poorer divided attention and executive control (Rueda et al., 2004; Shepp and Barrett, 1991; Klenberg et al., 2001), causing more looking to the right-hand side of the screen, having smaller receptive fields, or using more local, as opposed to global processing (Poirel et al., 2008; Balas et al., 2020). However, there is evidence that by five years children's receptive fields from V1 to VO1 are adult-like (Gomez et al., 2018). To test whether looking behaviour can explain the results, future studies might use eyetracking to determine whether there is a difference in looking behaviour between age groups, or haploscopic viewing to ensure all observers are fully adapted in one eye.

In summary, the developmental findings suggest that, by the age of six years, children are better able to take the illumination into account when selecting objects than adults. This extends findings from infant studies which suggest that infants have a rudimentary form of colour constancy at a couple of months old.

Our use of four different sweet colours and four illuminations allowed us to investigate the pattern of colour constancy across different conditions, and determine whether this pattern changed with age. In both experiments, colour constancy was significantly higher under the blue daylight illuminant than any other illuminants. Such a "blue bias" has been found in previous research (Delahunt and Brainard, 2004; Weiss et al., 2017), and is in line with a broad daylight prior, as modelled by Brainard et al. (2006). The fact that this effect did not extend to the Yellow illumination, under which observers had the lowest CCIs, could be due to a skew in the distribution of daylights such that highly saturated blues are more common than highly saturated yellows (Nascimento et al., 2016; Hernández-Andrés et al., 1999). Thus, observers may be more likely to attribute the blue illuminated box to a difference in illumination while attributing the other illuminations to a difference in wall colour. In both experiments the pattern across illuminations was similar for adults and children. However, in Experiment 5 (three-dimensional stimuli) there was a significant interaction between illumination and age group such that children had a smaller advantage over adults under the Yellow illumination than all other illuminations. This is in agreement with the model fitting which suggested no change in colour constancy with age under the Yellow illumination. Considering the developmental trajectory, we found a divide between Red and Blue vs Yellow and Green illuminations, such that there was a step-like change in colour constancy with age under Blue and Red illuminations but a linear change under Green and Yellow in Experiment 4. In Experiment 5 there was no change with age across childhood under Yellow or Green but a linear decline with age under Blue and Red. Overall, these results suggest that any daylight prior is already present in children, but the divide between daylight and non-daylight illuminations requires further investigation.

As with illuminations, the pattern across sweet colours remained fairly consistent across age groups and experiments, with both adults and children demonstrating the highest constancy for the Grey sweets, followed by Green, Rose, and poorest constancy for Teal. There were no significant interactions between age group and sweet colour in either experiment. Interestingly, we found a consistent significant interaction between surface reflectance and illumination, which was similar across age groups. When the sweet and illumination had similar chromaticities, such as the Green sweet under Green illumination, CCIs were higher than when the chromaticities were opposed. This effect is clearly visible for adults and children in Figs 3.4 and 3.11(b). A possible explanation is that observers have a bias towards more saturated competitors. Under the Green illuminant, the most saturated Green competitor would be the over-constant one whereas the most saturated Rose competitor would be under-constant. An alternative explanation with similar predictions is that observers are biased towards competitors which have a greater cone contrast against the background. Clearly observers are not simply picking the sweet with the highest saturation or greatest cone contrast, as these would result in CCIs of 1.25 or -0.5 depending on the condition, but they may have a bias in that direction. (Radonjić et al., 2015b) used a similar methodology to that used here, with blue and yellow illuminations, and found interactions between the set of reflectances and illuminations. Under their blue illumination, colour constancy was higher for the set of reflectances used in the present experiments than for more natural reflectances, whereas for the natural reflectances, colour constancy was higher under the yellow illuminant. The natural reflectances they used appear more yellow under a neutral illumination. Therefore, although not discussed or explicitly tested, their findings are broadly in agreement with those found in the present study. However, future studies are needed which manipulate the surface reflectances in a more controlled manner, such as using chromaticities aligned with, or orthogonal to, the illuminations, to determine whether saturation or cone contrast can explain these findings. Furthermore, the interaction seen here suggests future experiments investigating a daylight prior would benefit from manipulating surface reflectances in addition to illuminations.

The CCIs found in these experiments were generally low, with mean CCIs of 0.425 and 0.361 in Experiment 4, and 0.315 and 0.245 in Experiment 5, for children and adults respectively. Several factors might explain this level of performance. Firstly, as noted above, the instructions to observers were deliberately ambiguous and did not mention colour. Instructions have been shown to influence measured levels of performance. Hue/saturation matches generally show poorer constancy than paper matches (Arend et al., 1991). It is possible that both children and adults opted for hue/saturation matches, with adults coming closer to this tristimulus match, as discussed above. This cannot be the whole explanation, because CCIs would then be even closer to zero for both (Arend and Reeves, 1986). It is more likely that the constancy task was difficult for both because of the nature of the stimuli. In Experiment 4, the stimuli were two-dimensional and the backgrounds might not have been perceived as differing in illumination. Radonjić et al. (2015b) found very low CCIs for two-dimensional stimuli. It is surprising that the CCIs in Experiment 5, with three-dimensional rendered stimuli and an explanation of the changing illumination, were lower than in Experiment 4. Hedrich et al. (2009) found higher constancy for three-dimensional compared to two-dimensional scenes, and de Almeida et al. (2010) found no effect of dimensionality, both for real, rather than rendered scenes. In some conditions of Experiment 5, it was impossible to achieve perfect constancy as the chromaticities of the competitors did not contain a reflectance match, although CCIs above 0.9 in all but 1 condition were still possible (see Supplementary Materials). Additionally, the 3D realism of these rendered scenes might have been reduced by the lack of mutual reflections. Because the sweets were "floating", they also might have seemed to belong to a different illumination framework from the boxes. Furthermore, the stimuli were not immersive in either experiment; they were viewed on a monitor without the stereoscopic viewing deployed in Radonjić et al. (2015b). The task used here was also harder than in Radonjić et al. (2015b) where observers had only two rather than eight competitors to choose from on a given trial. Many observers commented on the difficulty of the task.

In summary, we found a consistent pattern in colour constancy across illuminations and sweets across ages, with adults and children showing a "blue bias", and higher constancy for sweets with a more neutral reflectance. Furthermore, for both age groups the effect of illumination was mediated by the surface reflectance. Importantly, we found a surprising decline in colour constancy with age from children to adults. This suggests that children may be using a different - less cognitive - strategy to adults.

3.7 Supplementary Material

3.7.1 Competitor Generation

In both experiments, the eight competitor reflectances were generated as follows. A tristimulus match (T; red X in Fig. 3.16) was generated by calculating the chromaticity (u'v' in Experiment 4; u*v* in Experiment 5) that the target sweet would have under the neutral illumination. This is the competitor with the same tristimulus value as the target. Under a different illumination, these tristimulus values would correspond to a real surface with a different surface reflectance from the target. Therefore, this is what observers would select if they did not correct for the difference in illumination, i.e. indicating no constancy. In addition, a reflectance match (R; blue diamond in Fig. 3.16) was calculated by determining the chromaticity that the target sweet would have under the chromatic illuminant. If observers selected this sweet, it would suggest they have perfect constancy, as this is what an object with the same surface reflectance as the target would look like. Three additional competitors were generated to have chromaticities equally spaced on a line between T and R in u'v' (Experiment 4) or L*u*v* (Experiment 5) (black Os in Fig. 3.16). In addition, an "overconstant" competitor (blue O in Fig. 3.16) was generated whose chromaticity was beyond R, with a distance from R equal to the distance between all other competitors. Finally, two "underconstant" competitors were made in the same way, whose chromaticities were beyond T (red Os in Fig. 3.16).

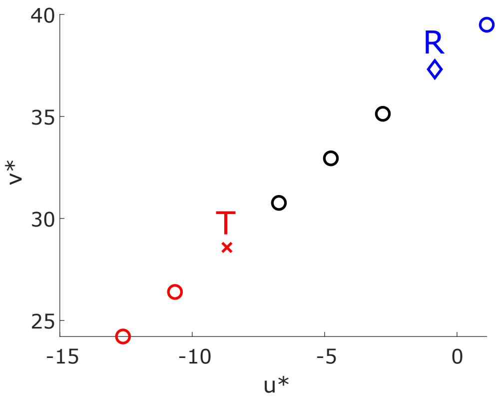


Figure 3.16: Example chromaticity of eight competitors in u*v*. Red X represents the tristimulus match, T; blue diamond represents the reflectance match, R; the black Os are the competitors in between T and R; red Os are underconstant and blue O is overconstant.

In the Neutral control conditions, the T and R matches were the same, so the sets of competitors had to be constructed differently. We chose to construct competitor sets that allowed us to assess bias along chromaticity directions aligned with the experimental stimuli: blue-yellow and red-green. In Experiment 4, to create the competitors along the blue-yellow axis, for each target sweet, we used T (which is identical under each illumination); the three competitors between T and R under the yellow illumination; and the two competitors between T and R, closest to T, under the blue illumination. The remaining two competitors were generated such that their chromaticities were directly between T and its adjacent competitors. The competitors along the red-green axis were generated in a similar manner, with three competitors the same as those under the red illumination and two closest to T under green. As the chromaticity of the three middle competitors in these sets were so similar, in Experiment 5 we used T, three competitors between T and R under both blue and yellow, and R under yellow. For the red-green control, we used T, three competitors between T and R under green and red, and R under red. In these control conditions, T is the "correct" answer, which observers would select if they were correctly matching without any bias.

Experiment 5

As the stimuli in Experiment 5 were three dimensional computer renderings, for which the spectral reflectance functions of each alternative sweet were needed, additional steps were required to generate these. Once the desired reflected chromaticity for each competitor sweet under each illuminant had been determined (see above), the reflectance spectrum for each sweet necessary to elicit this was generated. To do this, we first used least-squares fitting to generate spectra with the desired reflected chromaticities. The basis functions for this procedure were Gaussian curves with peaks between 400nm and 700nm. To derive reflectance spectra for each competitor which would reflect this light spectrum under the chromatic illuminants, the desired reflected spectra were divided by the spectrum of each corresponding illumination. It should be noted that this procedure meant the reflectance match sweet, R, did not have exactly the same spectral reflectance function as the target sweet. However, its appearance was metameric to the target sweet under the chromatic illuminant. That is, its reflected light under the chromatic illuminant was the same chromaticity as the target sweet would be. It therefore would be indistinguishable from a colour constant match.

The eight competitors in each scene were designed to have systematically varying luminance as well as chromaticity, which is why the calculations to determine reflected chromaticity were conducted in L*u*v*. However, due to the scene geometry, the bottom row of sweets were in the shadow of the top row. This meant that the luminance varied more randomly than intended, depending both on the competitor and the location within the scene. Therefore, analyses were conducted in u'v', which does not depend on luminance.

3.7.2 Instructions

"Hi, I'm Derek! I like to eat sweets! On the left of the screen you will see the sweet I want to eat." – *accompanied by a target sweet under neutral illumination on the left-hand side of the screen.*

"My friend has picked some sweets that he thinks I will like. Can you find the one I will like best?" – *accompanied by eight competitor sweets under a chromatic illumination.*

"To select a sweet for me to eat, click on it once. It will go big to show which sweet you have chosen. You can change your mind by clicking on another sweet." – *One of the sweets was enlarged to show how it would look when selected.*

"On each round, the game will look like this:" – *Immediately before showing the whole scene with Derek on it.*

"As you feed me, you will fill up this bar. When it is full you will earn a star!" - *With black screen containing energy bar in upper left hand corner*

"You are now ready to begin the game."

3.7.3 BIC of models

The Bayesian Information Criterion (BIC) for each of the four models (linear, jump, hockey stick, no change) fit to the CCIs for both children and adults in Experiment 4 are shown in Table 3.2, and are shown in Table 3.3 for Experiment 5.

Illumination	Linear	Jump	Hockey Stick	No Change
Blue	-59.181	**-67.897**	-60.047	-52.900
Yellow	**-100.314**	-94.255	-92.772	-97.251
Red	-91.759	**-112.897**	-105.318	-82.031
Green	**-119.474**	-114.169	-113.183	-113.760

Table 3.2: BIC for each of the four models fit to each of the four illuminations for Experiment 4. Bold numbers indicate the lowest BIC for the given illumination condition.

Illumination	Linear	Jump	Hockey Stick	No Change
Blue	**-62.547**	-61.952	-60.199	-62.538
Yellow	-102.374	-94.200	-94.138	**-103.119**
Red	**-97.444**	-95.433	-93.457	-79.923
Green	**-109.681**	-101.795	-100.696	-107.524

Table 3.3: BIC for each of the four models fit to each of the four illuminations for Experiment 5. Bold numbers indicate the lowest BIC for the given illumination condition.

3.7.4 Linear Mixed Effects Model Interpretation

For the interaction lme models in both experiments, adults, grey sweets and blue illumination were used as the reference factors. Therefore, the intercept is an estimate of the mean CCI for adults looking at grey sweets under the blue illumination. For any model estimates greater than 0, CCIs were higher in the comparison than the reference condition (intercept), and for any estimates below 0, CCIs were lower in the comparison condition. For the first seven effects in these models (labelled 1-factor), the Estimate is the difference in mean CCI between the comparison and reference conditions listed. For example, for the top row of the model in Experiment 4, *Children vs Adults (*$Grey_{sweet}$*,* $Blue_{illumination}$*)*, the estimate is the difference between the mean CCIs for children with grey sweets under the blue illumination (0.671) and the intercept (adults with grey sweets under the blue illumination; 0.549). For the two-way interactions shown in the next 15 rows (labelled 2-factors), significant estimates reflect a difference between two levels of one factor varying depending on the level of another factor. The third factor, shown in brackets, is fixed. These are to be added to the relevant 1-factor variables. For example, to determine an Estimate of the mean CCI for adults looking at green sweets under the green illumination in Experiment 4, one would use the following equation:

$$
\begin{aligned}
Estimate &= Intercept + Green_{illumination} + Green_{sweet} + Green_{sweet}|Green_{illumination} \\
&= 0.5490 + 0.0856 - 0.1837 + 0.1673 \\
&= 0.6182
\end{aligned}
$$

For the three-way interactions (labelled 3-factors), these Estimates are to be added to the relevant 2-factor and 1-factor variables in the same way as above.

3.7.5 Chromaticities of rendered sweets

The chromaticities, in u'v', of the rendered sweets after truncation in Experiment 5 are shown in Fig. 3.17. Each illumination condition is a separate column and each sweet colour is a separate row. The chromaticities from all 10 renderings (each containing the the eight competitor sweets in a different position) in each condition are plotted, although for the green and grey sweets there is little difference between renderings. The chromaticities of the inferred Reflectance (R) and Tristimulus (T) matches (see main text for details) are also plotted in black. As can be seen, the eight competitors typically fall on a straight line which contains T and R, but there are some deviations (e.g. for rose sweets under the red illumination).

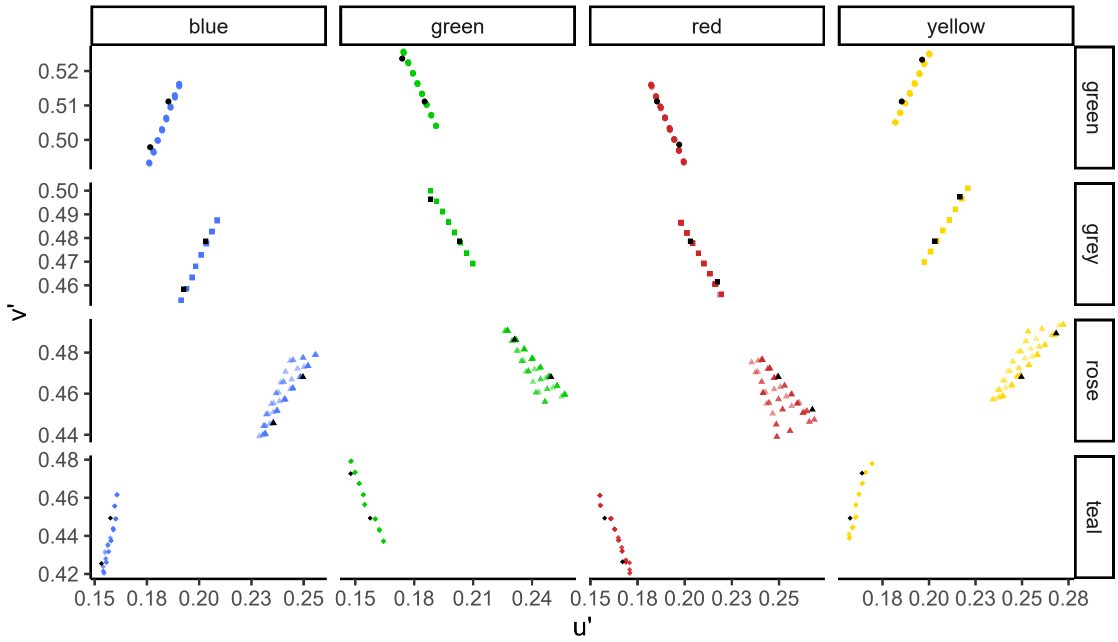


Figure 3.17: Chromaticity, in CIE u'v', of the rendered sweets. Each column is a different illumination and each row is a different sweet. Coloured symbols are the chromaticities of the eight competitor sweets; black symbols are the chromaticities of the Reflectance and Tristimulus matches

3.7.6 Experiment 5 CCI Calculation

As can be seen from the previous section, the chromaticity of the sweet designed to be the tristimulus match was not always the same as the chromaticity of the rendered target sweet. This occurred due to truncation of RGB values which were out of gamut of the monitor. In addition, there was a constant shift from the predicted chromaticities to the rendered chromaticities, which meant that when calculating colour constancy indices (CCIs), we were not able to use the predicted chromaticities calculated above.

To overcome this issue, we used the target sweet's rendered chromaticity as "T" in the CCI calculation. For "R", we used a theoretical chromaticity that the rendered target sweet would have under the chromatic illuminant. The theoretical spectrum of R was thus calculated as:

$$R(\lambda) = Target(\lambda) \times \frac{Chromatic\,illumination(\lambda)}{Neutral\,illumination(\lambda)} \tag{3.4}$$

where Target(λ) is the spectrum of the rendered target sweet; Chromatic illumination(λ)

	Illumination			
Sweet	Blue	Green	Red	Yellow
Green	1.016	1.053	1.065	0.994
Grey	0.970	0.893	1.000	0.996
Rose	1.058	0.985	0.946	0.932
Teal	0.977	0.990	1.018	1.038

Table 3.4: Optimal CCI possible for each illumination and sweet colour in Experiment 5.

is the spectrum of the Red, Green, Blue, or Yellow illumination; and Neutral illumination(λ) is the spectrum of the neutral illumination. The chromaticity of R(λ), in u'v', was then calculated for analysis.

As a result of using a theoretical R, there were no competitors that observers could select that had the exact same chromaticity as R in some scenes. The optimal CCIs that an observer would achieve if they had picked the sweet closest to R in each condition are shown in Table 3.4. It was possible to achieve close to perfect constancy in most conditions, with a maximum deviation from perfect constancy of less than 0.1, for all except Grey sweets under the Green illuminant.

Preface to chapter 4

In the previous chapter, a novel method of measuring colour constancy in children was developed. This task was a success, with children completing most of the trials, and many reporting that they found it fun. The data were able to be analysed in a way that gave us a clear measure of colour constancy, and the three-dimensional rendered scenes give us the freedom to manipulate many aspects. In chapter 2 we showed that adults could use specular highlights as a cue to improve their colour constancy. Bringing these two findings together, in the following chapter we altered the novel object-selection task developed in Chapter 3 to contain either matte or glossy sweets, to determine the development of the use of specular highlights.

Chapter 4

Development of specular highlight use in colour constancy

Previous research has found that adults are able to use specular highlights to improve colour constancy in a variety of tasks (Yang and Maloney, 2001; Wedge-Roberts et al., 2020; Lee and Smithson, 2016; Nagai et al., 2017). However, many of these tasks bear little resemblance to the real world. Here we aimed to further investigate the use of specular highlights using a more naturalistic object selection task. Another avenue of research and theory suggests that daylight priors can improve colour constancy for scenes illuminated by natural daylight illuminations (Delahunt and Brainard, 2004; Weiss et al., 2017; Aston et al., 2019), and that these may interact with other cues, such as specular highlights. Therefore, the second aim was to investigate the use of a daylight prior and whether it would interact with specular highlights. Finally, there has been very little research into the development of colour constancy during childhood, which is crucial to understand how people are able to develop the tools necessary to perceive colours as relatively stable under changing illuminations. Here we aimed to investigate how colour constancy develops over childhood, including the development of the use of specular highlights and daylight priors.

Using a child-friendly object selection task, we measured colour constancy for both matte and glossy objects under daylight and non-daylight illuminations. A group of naive adults and a small group of six to 10 year old children were tested. A larger sample of children was planned but due to the coronavirus pandemic, testing was halted.

Adults showed no benefit for scenes containing glossy, compared to matte shapes, suggesting that they did not use specular highlights in this task. They had the highest constancy under a yellowish daylight illumination, partially supporting the notion of a daylight prior, although this was dependent upon the surface reflectance of objects. In contrast, children's levels of colour constancy were negatively impacted by specular highlights, suggesting that they are not able to use them as a cue under 10 years of age. They showed no evidence of using a daylight prior. Overall, children had lower levels of colour constancy than adults in this task, suggesting that constancy is still developing up to at least 10 years old, although due to small sample sizes findings are to be taken with caution.

Further discussion of results, including directions for future research are discussed.

4.1 Introduction

Colour constancy is the ability to perceive the body reflectance of surfaces as relatively stable under a change in illumination. Although studies have found relatively high levels of constancy in adults (Brainard et al., 1997; Brainard, 1998), this is computationally difficult as the signal of light reaching the eye arises from a multiplication of the spectral reflectance of the surface and the spectral power distribution of the illumination. Furthermore, upon reaching the retina, this signal of light is received by three cone classes, making it even more difficult to disentangle the contribution of surface reflectance from illumination.

A number of cues, priors, and heuristics have been proposed to aid colour constancy, including assuming the brightest surface in a scene is white and perfectly reflects the illuminant (Brightest is white; Land, 1964); assuming the average reflectance over a scene is spectrally neutral, or grey (Grey World; Land, 1983; Buchsbaum, 1980); and assuming the local average reflectance is spectrally neutral (Valberg and Lange-Malecki, 1990). Kraft and Brainard (1999) tested whether observers can use these cues and found that all three cues improved colour constancy, but constancy was still imperfect with all present and valid. Amano and Foster (2004) used simulations to show that observers could use the global average photoreceptor activation in colour constancy, as proposed by the grey world hypothesis. Linnell and Foster (2002) found that observers can use both the global mean and brightest is white cues, with relative weights depending on the number of surfaces present. However, many of the early studies into cue-use in colour constancy focused primarily on matte, two-dimensional scenes. More complex cues have since been proposed, such as using prior knowledge of objects' canonical colours. Hansen et al. (2006) demonstrated that colour knowledge can influence colour perception, and Granzier and Gegenfurtner (2012) found that having objects with diagnostic colours present in a scene improved colour constancy when judging the colour of other surfaces.

A cue that has been researched relatively little is specular highlights. Specular highlights are present on glossy objects. Perfectly specular, smooth objects contain specular highlights which directly reflect the illumination, without absorbing it into the body material. Therefore, the highlights have the chromaticity of the illuminant. Partially specular objects, in contrast, reflect chromaticities lying on a line between the diffuse body reflectance of the object, and the illumination. Although it is possible that no individual highlight in such objects will reflect the chromaticity of the illumination, when multiple objects are present these lines will point towards the chromaticity of the illuminant and, if extended, would intersect at the chromaticity of the illuminant. This has been termed chromaticity convergence (Hurlbert, 1998).

Research into the use of specular highlights has typically found small benefits in colour constancy for scenes containing specular highlights, using achromatic or asymmetric matching with rendered stimuli (Snyder et al., 2005; Yang and Maloney, 2001; Yang and Shevell, 2003; Wedge-Roberts et al., 2020); colour matching with physical stimuli (Granzier et al., 2014); and in tests of operational colour constancy (Lee and Smithson, 2016; Nagai et al., 2017). However, the effect of highlights may depend on the type of illumination (Yang and Maloney, 2001) and the articulation of the background (Wedge-Roberts et al., 2020). Furthermore, many of the methods used in these studies bear little resemblance to real-world tasks. For example, in both Lee and Smithson (2016) and Nagai et al. (2017), the stimuli consisted of a single sphere or shape in a void. A more natural task to measure colour constancy is object selection, as used in Wedge-Roberts et al (submitted), Radonjić et al. (2015b), and Radonjić et al. (2015a). The first aim of the present study is to extend the research into the use of specular highlights for colour constancy to a more natural, object selection task.

As well as cues, priors may help improve colour constancy. One prior which has been proposed is a daylight prior, which should result in better colour constancy for scenes illuminated by daylights (ranging in appearance from blueish to yellowish) compared to non-daylights. In support of this, Pearce et al. (2014) and Aston et al. (2019) have found poorer illumination discrimination, implying better colour constancy, for illumination changes towards a blueish daylight compared to other directions. Furthermore, Weiss et al. (2017) and Delahunt and Brainard (2004) used achromatic matching and found the highest degree of colour constancy under blueish daylight illuminants. Brainard et al. (2006) modelled the data from Delahunt and Brainard (2004) using a Bayesian model with an illumination prior to help determine the most probable illumination-surface combination from a given set of photoreceptor responses. They found that a broad daylight illumination prior best fit the psychophysical data.

Some studies investigating the use of specular highlights have also manipulated the type of illumination to test for a daylight prior. Yang and Maloney (2001) found that specular highlights are only used when they signal a daylight illumination (D65) and not a non-daylight Tungsten illumination (A). In contrast, Wedge-Roberts et al. (2020) found no effect of illumination or interaction with highlights on colour constancy indices, but a bias towards a blueish daylight illumination in achromatic settings, similar to Weiss et al. (2017). Nagai et al. (2017) found no difference in operational colour constancy between shapes illuminated by daylight and non-daylight illuminations. Given the mixed findings regarding a daylight prior, the second aim of the current study is to determine whether adults use a daylight prior in a realistic object selection task.

None of the studies mentioned so far have investigated the development of colour constancy, cue use, or priors. In fact, very few studies have measured colour constancy in children. Some studies have used preferential looking with infants to determine that four-month-olds have a rudimentary form of colour constancy (Dannemiller, 1989; Dannemiller and Hanko, 1987; Yang et al., 2013a). Furthermore, Yang et al. (2010) found that four to eight month-old infants can perceive a Munker-White illusion, in which the perceived saturation depends on the local surround. This suggests that their perception is affected by simultaneous contrast, which is an important mechanism for colour constancy (Hurlbert and Wolf, 2004). Investigating colour constancy in older children, Rogers et al. (2020) found a reasonably high level of constancy in three to four year-old children, using a 2AFC task. Furthermore, Wedge-Roberts et al (submitted) found that six to 11 year-old children had higher constancy than adults in an object selection task. As only one of these studies investigated colour constancy in children over four years-old, the current study aims to build on developmental research to compare levels of colour constancy in children with adults.

As well as the development of colour constancy, we want to know at what age the ability to use priors and cues for colour constancy develops, in order to determine whether such information is innate or develops over time. Research into the development of prior-use in perception is mixed, with some studies suggesting that the light from above prior develops throughout childhood (Stone, 2011), and others finding that it is not age-dependent (Pickard-Jones et al., 2020). In a previous study (Wedge-Roberts et al, submitted), we measured colour constancy in children with daylight and non-daylight illuminants and found some evidence for better colour constancy under a blueish daylight illuminant in children. However, the effect of illumination was strongly dependent on the surface reflectance. The present study aims to extend this by asking whether children use a daylight prior with more saturated illuminations, and a different set of surface reflectances.

The ability to combine and appropriately weight cues and priors has also been found to develop with age (Chambers et al., 2018; Thomas et al., 2010; Alais and Burr, 2019). Therefore, we are interested in whether the ability to use specular highlights in colour constancy, in combination with other cues, develops with age over childhood. To use specular highlights, observers first need to be able to perceive gloss. Yang et al. (2011) found that seven month-old infants can already perceive gloss, in a preferential looking paradigm. However, this has not been investigated in older children, and although infants can discriminate between matte and glossy objects, it does not mean that they understand the physics of specular highlights, which is necessary to use them as a cue for colour constancy. Therefore, here we aim to investigate whether children can use specular highlights to improve their colour constancy.

In the present study, we aim to answer five research questions: (1) Can adults use specular highlights to improve colour constancy? (2) Will adults use a daylight prior in an object selection task, shown by better colour constancy for scenes illuminated by daylights compared to non-daylights? (3) Can children use specular highlights in colour constancy? (4) Can children use a daylight prior? (5) Is there a difference in (a) overall constancy, (b) use of specular highlights, or (c) use of a daylight prior between adults and children? To answer these five research questions, we adapted the methods of Wedge-Roberts et al (Submitted) to develop a child friendly object-selection task containing either glossy or matte objects, illuminated by either daylight or non-daylight illuminants. Due to the Coronavirus pandemic, fewer children were tested than planned meaning the developmental findings are to be taken with caution.

4.2 Experiment 6

4.2.1 Methods

Observers

It was planned that 30 adults and 60 children would be tested. As children only completed half the conditions while adults completed the full set, twice as many children were required. However, due to the Coronavirus pandemic, testing was halted part way through, and when it briefly resumed only adults could be tested. Therefore, 21 adults aged 19 to 45 years (mean =25.96, SD = 6.49) and 15 children aged six to 10 years (two 6, four 7, two 8, three 9, and three 10 year-olds) took part in this study. Adults were paid volunteers, and children were recruited from a database of parents interested in taking part in studies. Adults and parents of children gave informed consent. Children gave assent and the experimenter ensured that they were happy to continue between each round. All observers were tested for colour vision deficiencies using Ishihara plates (Ishihara, 2006), and none made more than two mistakes, meaning they were all included in the study. All observers had normal, or corrected to normal, visual acuity.

Materials and Apparatus

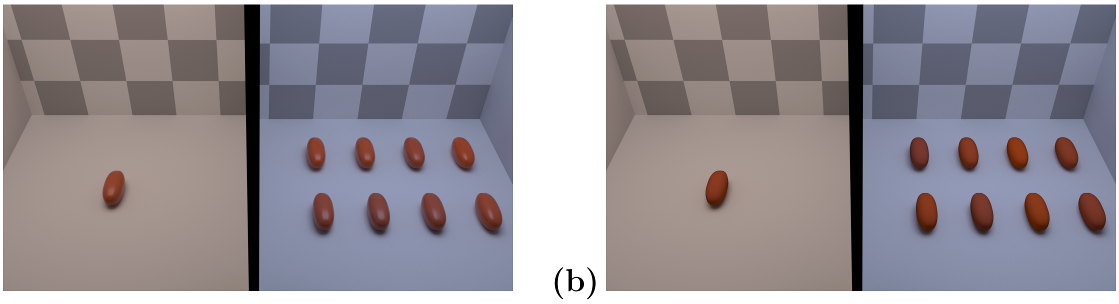
Testing took place in a dark, child friendly laboratory. A black cloth was placed over the desk to reduce the amount of light reflected from the surface of the desk. The stimuli were shown on an ASUS PA382Q 23" monitor, which has 10-bits per colour channel. An Nvidia quadro k600 graphics card presented the stimuli on the monitor. The monitor primaries were recorded using a Konica Minolta CS-2000 spectroradiometer. From these we generated a look up table to convert from XYZ to RGB, taking into account gamma non-linearities. Observers sat approximately 60cm from the monitor, with free head movement.

Stimuli

Three-dimensional scenes were generated in Blender (`https://www.blender.org/`) and hyperspectral images were rendered in Mitsuba (`http://www.mitsuba-renderer.org/`) compiled for spectral rendering with 30 spectral bands. After rendering, the images were converted to XYZ using the 2 degree CIE 1931 colour matching functions (CIE, 1932), and scaled to have a mean luminance of 50cd/m^2. These were then converted to RGB using the lookup table obtained from calibration, as described above. Out of gamut values were truncated such that any RGB values greater than 1 were clipped at 1 and values less than 0 were clipped at 0.

The rendered scenes consisted of two rooms side by side, both with light grey walls on the left and right sides and floor (reflectance = 0.75), and a checked back wall. Each room filled half the screen, extending 27 × 24 degrees of visual angle at a distance of 60cm. The checks on the back wall were made of white (reflectance = 1) and grey (reflectance = 0.5) alternating patches. Each patch extended approximately 4 degrees of visual angle. All walls were made using the rough plastic material in Mitsuba, and had a specular reflectance of 0, meaning they were perfectly matte. Both rooms had an area light encompassing their ceilings, which was not visible to observers. The camera was pointing down such that observers could see only the floor, the lower half of the back wall, and small portions of the outer side walls. An example of a scene is shown in Fig. 4.1.

The room on the left was always illuminated by a neutral illuminant – D57. This was chosen because Nascimento et al. (2016) measured the illumination in natural



(a) (b)

Figure 4.1: Example scene containing the orange sweet under the blue illuminant. (a) contains glossy sweets whereas (b) contains matte sweets.

scenes and found a peak in the distribution at a colour correlated temperature (CCT) of 5698K. The room on the right was illuminated by either a blueish or yellowish illumination on the daylight locus, or a reddish or greenish illumination which were perpendicular to the daylight locus in u'v' at 5700K. These will hereafter be referred to as *Blue*, *Yellow*, *Red*, and *Green*. The chromatic illuminations were all approximately $54\Delta E_{L*u*v*}$ away from neutral at a luminance of 50cd/m^2. To generate the illumination spectra, we first converted the desired u'v' chromaticities to Yxy. From these, we generated the illumination spectra using the mean and first two vectors of the Judd daylight basis functions (Judd et al., 1964), with the weights on the two vectors calculated as:

$$M1 = \frac{-1.3515 - 1.7703x + 5.9114y}{0.0241 + 0.2562x + 0.7341y} \tag{4.1}$$

$$M2 = \frac{0.03 - 31.4424x + 30.0717y}{0.0241 + 0.2562x + 0.7341y} \tag{4.2}$$

(as per Judd et al., 1964). For the Green illumination, some values of the spectrum obtained from these equations were negative so they were converted to 0. This shifted the chromaticity by less than $1\Delta E_{L*u*v*}$. The chromaticites corresponding to these spectra are shown in Table 4.1 under the columns entitled "predicted".

Pill-shaped sweets were generated in Blender by rounding the edges of a cube base shape. Within the room on the left, one sweet lay in the centre of the floor, which will hereafter be referred to as the target sweet. This extended 3.4 × 2 degrees of visual angle. In the right hand room, eight competitor sweets were placed on the floor, in two rows of four, as shown in Fig. 4.1. The material of these pills was generated using the rough plastic plugin in Mitsuba. The diffuse re-

Illumination	$x_{predicted}$	$y_{predicted}$	$x_{rendered}$	$y_{rendered}$	$u'_{rendered}$	$v'_{rendered}$
D57	0.3279	0.3441	0.3293	0.3456	0.2030	0.4794
Yellow	0.3955	0.3911	0.3958	0.3921	0.2290	0.5104
Blue	0.2764	0.2896	0.2779	0.2919	0.1869	0.4417
Red	0.3294	0.2861	0.3253	0.2858	0.2251	0.4451
Green	0.3256	0.4163	0.3336	0.4188	0.1813	0.5123

Table 4.1: Chromaticities of illuminations generated using Judd basis functions, in CIE Yxy and u'v'. Predicted columns are chromaticities of the illumination spectra; rendered columns are chromaticities reflected from perfectly reflective sweets in the scenes, after scaling.

flectance of the target was either Turquoise (Munsell chip 10BG-5/6), Orange (Munsell chip 2.5YR-5/10), or Purple (Munsell chip 5P-4/10). The Munsell reflectances were retrieved from `https://www.uef.fi/web/spectral/munsell-colors-matt-spectrofotometer-measured`. We originally included a fourth surface reflectance: lime (chip 5GY-8/6). However, during rendering the chromaticity shifted too much so these were not used in the experiment. These four surface reflectances were picked as, under the neutral illuminant (D57), they had hues approximately between the chromaticities of the four chromatic illuminations in CIE u'v'. The chromaticities of the four surface reflectances under D57, and the four illuminations, in u'v', are shown in Fig. 4.2, with the Munsell chips under D57 shown by Xs, and the colour corresponding to the colour of the chip (Turquoise, Purple, Lime and Orange), and the Os corresponding to the illuminations (Neutral, Green, Blue, Red, and Yellow).

When rendering scenes, the chromaticities sometimes shifted slightly, for example due to truncation. To determine the chromaticity of the illuminations after rendering, four scenes were rendered containing perfectly reflective (white), matte sweets under each of the chromatic illuminants on the right, and the neutral illuminant on the left. These scenes were scaled to have a mean luminance of 50cd/m^2. In addition, a mask was rendered which contained the nine white sweets with no floor under them. The mask was used to select pixels in the image which made up the sweets. To calculate the chromaticity of the four chromatic illuminants after rendering, we calculated the mean chromaticity reflected from the eight sweets (selected using the mask) in the right-hand box. The chromaticity of the neutral illuminant was the mean chromaticity reflected from the perfectly reflective sweet in the left-hand box. The chromaticities of these rendered illuminants are given in Table 4.1, under the columns entitled *rendered*.

The eight competitor sweets shown in the right hand room were designed to contain both a tristimulus match (T) indicating no colour constancy, and a reflectance match (R) indicating perfect constancy. The light reflected from T had the same chromaticity as the target sweet, despite being shown under a different illumination. In order for a real surface to reflect this light, it would have a different reflectance to the target sweet. The light reflected from R was the same as if the target sweet had been placed under the chromatic illuminant. The chromaticities reflected from three other competitors were equally spaced on a line joining T to R. In addition, one overconstant, and two underconstant competitors were generated with chromaticities beyond R (overconstant) or T (underconstant). All competitors were equally discriminable and lay on a straight line in u'v' space. A schematic of the eight competitors is shown in Fig. 4.3 with the blue square representing R and the red triangle representing T.

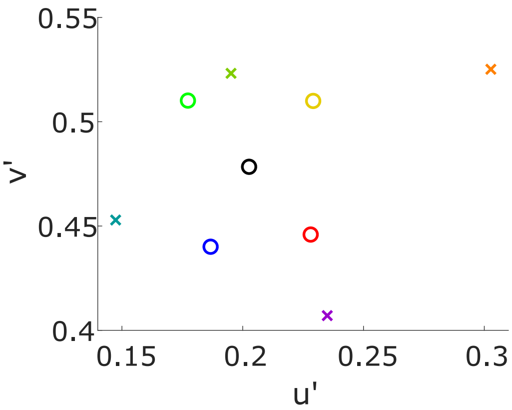


Figure 4.2: Chromaticities, in u'v', of the four surface reflectances (Orange, Purple, Turquoise, Lime) under D57 are shown by Xs. Chromaticities of the five illuminations (Neutral D57, Blue, Green, Red, Yellow) are shown by Os. The colour of each symbol corresponds to the reflectance or illumination.

The reflectances of the eight competitors were generated as follows. For each target reflectance and illumination, the spectrum of light to be reflected from R was calculated as the spectrum of the target Munsell chip multiplied by the (rendered) chromatic illumination's spectrum. The light to be reflected from T was calculated as the target Munsell spectrum multiplied by the spectrum of the neutral illumination. These spectra were then converted to L*u*v*, using the monitor white (XYZ

= 190.25, 205.24, 215.65) for the conversion. To generate the chromaticity of the remaining six competitors, the distance in L*, u*, and v* between T and R was calculated. These distances were divided by four to give the separation required between each competitor. Using this distance, three competitors were positioned equally between T and R (in L*u*v*). In addition, the chromaticity of an underconstant and two overconstant competitors were calculated by extending the line beyond T and R, by the same distance. The eight competitors were then converted to XYZ, and spectra generated with these chromaticities by least-squares fitting, using non-negative values only. The basis functions used for this were a set of 31 Gaussians, with peaks equally spaced between 400 and 700nm. As these spectra were those to be reflected from each competitor, to find the surface reflectance needed, they were divided by the relevant (rendered) chromatic illumination's spectra.

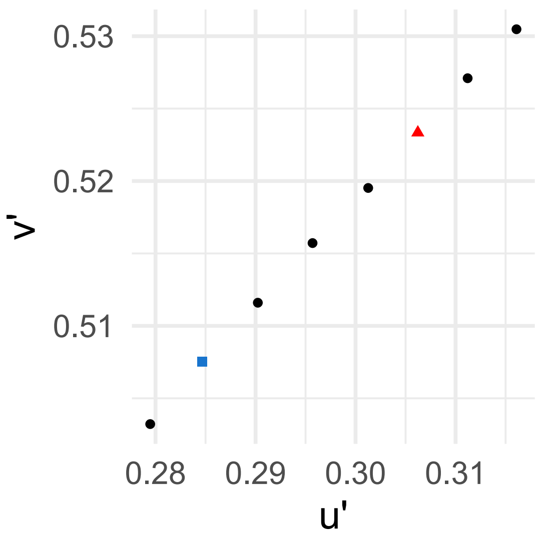


Figure 4.3: Example chromaticity of eight competitors in u'v', for Orange sweets under the Blue illumination. The red triangle represents T and the blue square represents R.

In addition to the four illumination conditions described above, we had two neutral illumination conditions to test for any inherent biases/ preferences. In these, both rooms were lit by the neutral illuminant. The same three target sweets were used in the left hand room as in the experimental conditions. In one control condition, the reflected competitor chromaticities were on a line between the Blue R and the Yellow R; in the other they laid between the Green R and Red R. To generate the first set of competitors, we used the spectrum to be reflected from the three middle competitors under Blue, T (always the same regardless of illuminant), the three middle competitors under Yellow, and the Yellow R. These were then divided by the neutral illuminant's spectrum to obtain sweet reflectances. To make the competitors for the Red-Green control condition, the same process was used with the competitors under Red and Green.

Matte and glossy versions of each scene were created. The rooms were identical in both. In the matte conditions, the sweets had a specular reflectance of 0, meaning they were perfectly matte. An example is shown in Fig. 4.1 (b). In the glossy conditions, the specular reflectance was 1, meaning they had visible specular highlights, as can be seen in Fig. 4.1 (a).

For each of the 36 conditions (six illuminations × three target sweet colours × two materials), ten unique scenes were rendered, each containing the eight competitor sweets in a random order. For the experiment, we required 13 repetitions of each. Therefore, three of then ten scenes were randomly chosen for each observer to be shown twice.

Task

Observers were introduced to a cartoon dragon, Derek (see Fig. 4.5), and were told that the target sweet, shown in the left hand box, was Derek's favourite sweet. They were instructed to pick which sweet, from the set on the right, was also Derek's favourite sweet. They selected this by clicking on it using a mouse. When a sweet was selected, a black arrow appeared above it (for the back row) or below it (for the front row) to indicate which had been selected. Observers then pressed the space bar to feed the sweet to Derek. They were free to change their mind as often as required before feeding Derek.

Procedure

Adult observers completed all the trials in two sessions whereas children only completed one session each. Each session lasted approximately one hour. For adults, each session contained trials for every sweet colour and every illumination, but only one material (matte or glossy). The order in which the two sessions were completed was counterbalanced between observers. As it would be difficult to test children on two different sessions, each child completed only half the conditions. All children completed trials with both matte and glossy sweets, to ensure material was a within subjects variable for all observers. Half the children completed the Red and Yellow illumination conditions while the other half completed the Blue and Green illumination conditions. This meant that every child saw both a daylight and a non-daylight illumination. Children also completed a matte and a glossy neutral control condition. Both controls contained six trials from both the Red-Green and Blue-Yellow adult controls.

All the children and the first 14 adults took part prior to the pandemic. The remaining seven adults took part during the pandemic. Therefore, the procedure was slightly different as the experimenter could not remain in the testing room for the final seven observers.

The experiment began with two minutes of dark adaptation. During this period, a jingle was played to keep observers entertained. Following this, a demonstration scene was presented to observers. This consisted of the same two boxes as in the main experiment, but instead of sweets, they contained glossy objects (torus, cone, sphere and cube). Five versions were rendered, each with a different illumination in the box on the right. An example is shown in Fig. 4.4 with the Blue illumination on the right. For the first 29 observers, the experimenter could control which illumination was shown. They read the following text out loud while changing the illumination at the relevant parts:

> In this game there are 2 boxes. The light in the box on the left will always stay the same colour. However, I can change the colour of the light in the box on the right. Watch me turn the blue light on. Now watch me turn the red, green, and yellow lights on. In each round, I will turn on a different coloured light in this box. On this round, it will be *insert appropriate illumination here*. Are you ready to meet Derek now?

For the remaining seven observers, the five illumination conditions were each shown for two seconds before automatically changing to the next one, until the observer pressed the space bar. The experimenter read aloud the instructions from behind a window. These were modified so the experimenter did not say they would turn on a different light; but instead a different light would be shown in each round. Following this, the observer pressed the space bar for the full instructions.

The remainder of the instructions were given by a monochrome animation of Derek the Dragon, as shown in Fig. 4.5. These included a random experimental scene. The full instructions were as follows:

> "Hi, I'm Derek! I like to eat colourful sweets! In the box on the left you will see the colour of sweet I want to eat." *(with left hand box)*
>
> "My friend has picked some sweets that he thinks I will like. These are in the box where the light changes colour. Can you find the one I will like best based off its colour? Some sweets will look bigger because they are nearer, but this is not important." *(with right hand box)*
>
> "To select a sweet for me to eat, click on it once. An arrow will appear to show which sweet you have chosen. You can change your mind by clicking on another sweet." *(with arrow shown above/below a random sweet in right hand box)*
>
> "On each round, the game will look like this:" *(followed by the whole scene)*
>
> "As you feed me, you will fill up this bar. When it is full you will earn a star!" *(with energy bar)*
>
> "You are now ready to begin the game."

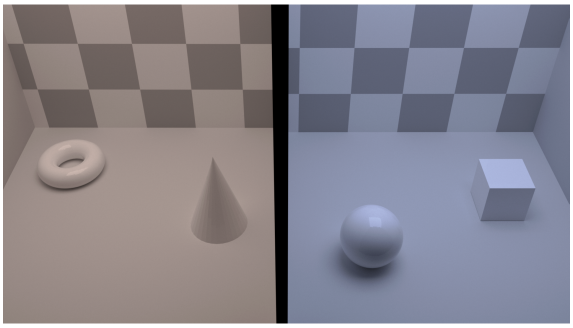


Figure 4.4: Blue demonstration scene used in instructions to show the illumination changing.

When happy with the instructions, observers began the first condition. Each block consisted of all 39 trials with the same illumination (13 repetitions of three sweet colours). The three sweet colours were randomly interleaved, and the order of the trials was randomised for each observer. Between each trial a black screen with an energy bar in the upper left corner was shown. The bar was filled incrementally, between 1/6 and 1/8 after each trial. When full, a star would appear on the screen alongside an animation of Derek talking and the text "Thank you for feeding me, you have earned a star!". Children were given a star chart at the beginning and given a silver star sticker whenever this occurred. Adults

Figure 4.5: Stills from the animation of Derek the dragon used to give instructions and a star reward.

were given the option to have a star chart, although most declined. Observers were dark adapted for two minutes between each block. Adults completed all six blocks (four experimental conditions + two neutral control conditions) for a single material within a session. The order of the blocks was randomised for each observer. Three observers completed the two sessions on the same day. One observer completed one session prior to lockdown and the other after, leaving a gap of 176 days. The remaining 17 adults completed the sessions on two days, separated by between one and 21 days. As noted above, children only completed one session consisting of six blocks (two illuminations with both materials, and a neutral control with both materials). The order of the blocks was randomised.

Data Analysis

After testing, we discovered that the rendering and truncation meant that the chromaticities of the Turquoise competitor sweets did not lie on a straight line in CIE u'v' (or L*u*v*). Therefore, these trials were not included in analysis. As a result, we had 26 trials remaining for each illumination/ material condition (13 from each of the Orange and Purple sweets).

To remove any conditions in which observers may have been responding at random, we developed two exclusion criteria. For this, we considered the distribution of responses (each a number between 1 and 8 to indicate which competitor had been selected) for each illumination/ material condition from each observer (26 responses). We first fit both a Gaussian and a Uniform distribution to these responses, and calculated a Bayesian Information Criterion (BIC) for each. Any set of responses which was better fit by the Uniform distribution (indicated by lower BIC), or in which the difference between the two BICs was less than 6, was deemed at least as well fit by a Uniform distribution and was thus removed from analysis. This was three conditions from two adults and eight conditions from children. In addition, we calculated the standard deviation of the 26 responses. We ran 100,000 simulations to determine that if observers had been randomly responding on each trial, they would have a SD of approximately 2.28. Therefore any set of responses with a SD larger than this were excluded. This was none from adults and six from children. Due to overlap between the two criteria, three adults' and eight children's conditions in total were excluded from analysis. Furthermore, one adult only completed one session due to the pandemic, and another adult completed the glossy session twice but did not complete the matte condition, due to computer error. Therefore, only 17 adults contributed full data sets which were analysed.

To calculate colour constancy indices (CCI), a measure of colour constancy, we required chromaticities indicating no constancy (T), perfect constancy (R), and the sweet selected by observers on a given trial. All analyses were conducted in CIE u'v' as chromaticity does not depend on luminance in this colour space. For the baseline measure, we would ideally use the chromaticity of T. However, during rendering the chromaticities shifted slightly such that the chromaticity of the target sweet (in the left hand box) was not identical to the chromaticity of the competitor designed to be T. Furthermore, the mean chromaticity reflected from glossy sweets lay between the chromaticity of the corresponding matte sweet and the illumination, and the chromaticity of the eight competitor sweets varied depending on position, due to mutual reflections. Therefore, the baseline chromaticity was defined as the mean chromaticity reflected from each pixel of the matte target sweet (shown in the left hand box). For ease, we will hereafter refer to this as $\mathrm{T}_{theoretical}$.

The reflectance match, R, was designed to have the chromaticity of the target sweet under the chromatic illumination. However, as with T, rendering issues meant that this was not always the case. Therefore, to calculate the chromaticity that observers would pick if they had perfect constancy, we rendered an additional set of scenes with the target sweet in the left hand box under each of the four chromatic illuminants. The mean chromaticity reflected from this sweet was defined as $\mathrm{R}_{theoretical}$.

To obtain the chromaticity of the sweet selected by observers, we calculated the mean chromaticity of each matte competitor across the 10 renderings. It should be noted that, although mutual reflections were present, the chromaticities did not differ much across scenes, as the sweets were perfectly matte. Observers' responses were converted to the corresponding chromaticity.

CCIs were calculated as:

$$CCI = 1 - \frac{a}{b} \tag{4.3}$$

where b is the Euclidean distance, in u'v', from $\mathrm{T}_{theoretical}$ to $\mathrm{R}_{theoretical}$, and a is the distance from the $\mathrm{R}_{theoretical}$ to the vector projection of the chromaticity of the sweet selected onto the line joining $\mathrm{T}_{theoretical}$ to $\mathrm{R}_{theoretical}$. The vector projection was used because, due to rendering, the three points did not always lie on a straight line in u'v'.

4.2.2 Results

Adults

We had two research questions to answer in adults: firstly, do observers have improved colour constancy for scenes containing specular highlights? And secondly, do they show evidence of using a daylight prior? The mean chromaticities selected for each sweet under each experimental illumination are shown in Fig. 4.6, with the circles showing the chromaticity selected for glossy sweets and the +s showing the chromaticity selected for matte sweets. Coloured triangles represent $\mathrm{R}_{theoretical}$ and black triangles represent the chromaticity of the target sweet under the neutral illuminant ($\mathrm{T}_{theoretical}$). Error bars represent ± 1 SEM. From this graph, it can be seen that the competitor picked varies depending on both the illumination and the sweet colour. There is no obvious overall difference between matte and glossy sweets.

Before analysing the CCIs, we wanted to ensure that the sweets were discriminable, and there was no internal bias towards a particular competitor (e.g. the most saturated) when the illumination was consistent across the two boxes. To this end, we analysed the two neutral control conditions. We defined the error as the Euclidean distance, in u'v', between the chromaticity of the competitor selected and the target sweet's chromaticity under neutral ($T_{theoretical}$). The average error across these control conditions, for the two sweets and two materials, was 0.00622 $\Delta E_{u'v'}$. This was smaller than the difference between two adjacent sweets, for all but two pairs of orange sweets, meaning that observers could typically discriminate between the sweets. An ANOVA on the errors found a significant main effect of condition, with a greater error for competitors ranging from the Blue to Yellow illuminations (mean = 0.00781) than Red to Green (mean = 0.00464). There was no significant difference between sweet colours or, crucially, between materials. This suggests that observers were equally able to discriminate matte and glossy competitors.

To answer our two research questions, we ran a 4 (illumination) $\times$ 2 (sweet colour) $\times$ 2 (material) ANOVA on the CCIs. The results of this are given in Table 4.2. In response to our first question - whether adults can use specular highlights for colour constancy - we found no significant main effect of material ($p = 0.546$). In response to our second question, which was whether adults use a daylight prior, there was a significant main effect of illumination ($p < .001$), with the highest CCIs under the Yellow daylight (mean = 0.435, SD = 0.334), followed by Red (mean = 0.371, SD = 0.236), Blue (mean = 0.270, SD = 0.336), and the lowest under Green (mean = 0.246, SD = 0.205). Bonferroni corrected t-tests showed there were significant differences between Blue and Red; Blue and Yellow; Green and Red; and Green and Yellow.

There was no significant main effect of sweet colour ($p = 0.713$). However, there was a significant interaction between illumination and sweet colour ($p < 0.001$). To further explore this interaction, we conducted pairwise Bonferroni-corrected t-tests to test for a difference between sweets for each illumination condition separately. There were significant differences between sweets under the Blue ($t(62.68) = -7.93$,

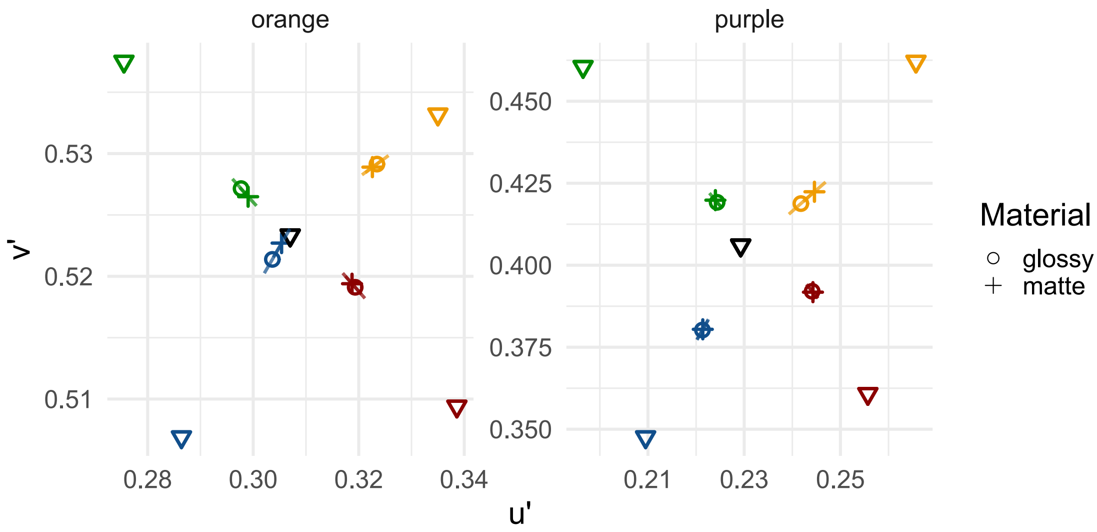


Figure 4.6: Mean chromaticities selected by adults for Orange (left) and Purple (right) target sweets in CIE u'v'. Black triangles represent the chromaticity of the target sweet under the neutral illuminant (T i.e. no constancy); chromatic triangles represent the chromaticity of the target sweet under each of the four illuminants (R i.e. perfect constancy). Os are the mean chromaticity selected by observers for glossy sweets, with the colour of the symbol corresponding to the illuminant. +s are the mean chromaticity selected for matte sweets. Error bars represent ±1 SEM along the line joining T to R.

Variable	DF	F	p	η_p^2	BF	Supporting
Illumination	3	19.837	<.001***	0.166	236307726	H1 (extreme)
Sweet	1	0.135	0.713	0.000	0.1265581	H0 (moderate)
Material	1	0.365	0.546	0.001	0.1522009	H0 (moderate)
Illumination × Sweet	3	39.399	<.001***	0.283	1.879289×10^{22}	H1 (extreme)
Illumination × Material	3	0.419	0.739	0.004	0.06945387	H0 (strong)
Sweet × Material	1	2.617	0.107	0.009	0.4108549	H0 (anecdotal)
Illumination × Sweet × Material	3	0.162	0.922	0.002	0.07693033	H0 (strong)

Table 4.2: Results of ANOVA on CCIs, with Illumination, Sweet, and Material as factors. * indicates significance at the <0.001 level. *BF* shows the Bayes Factors from the Bayesian ANOVA, and *Supporting* is the corresponding interpretation.**

$p < .001$) and Yellow ($t(73.152) = 5.498$, $p < .001$) but not the Red ($p = 0.118$) or Green ($p = 0.757$) illuminants. This interaction is shown in Fig. 4.7, with the CCIs for the Orange sweets shown on the left and the Purple sweets shown on the right. The colour of the violins correspond to the illuminations. From this we can see that for the Orange sweets, CCIs are highest under the Yellow illumination and lowest under the Blue illumination whereas for the Purple sweets, CCIs are highest under the Blue illumination and lowest under the Yellow and Green illuminations.

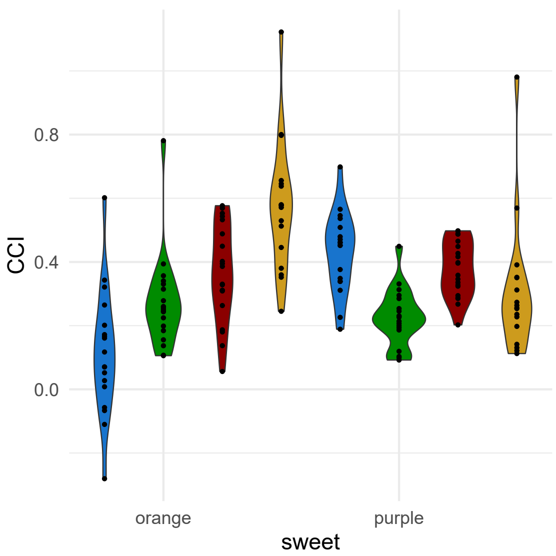


Figure 4.7: Violin plot showing CCIs for adults for Orange (left) and Purple (right) sweets under each of the four illuminants. The colours of the voilins correspond to the colour of the illuminant.

To check whether the null results in the ANOVA were not due to low statistical power, we also ran a Bayesian ANOVA with the same factors. The results of this are also given in Table. 4.2, under the columns *BF* and *Supporting*. This shows that there is sufficient evidence to accept the null hypothesis (H0) that there is no effect of material, with moderate certainty.

Children

The mean chromaticities of sweets selected by children, alongside $R_{theoretical}$ and $T_{theoretical}$ are shown in Fig. 4.8. From this, the pattern across sweets and illuminations looks similar to the pattern in adults. As children only completed six repeats of each sweet/ material in the neutral control conditions, we did not analyse these.

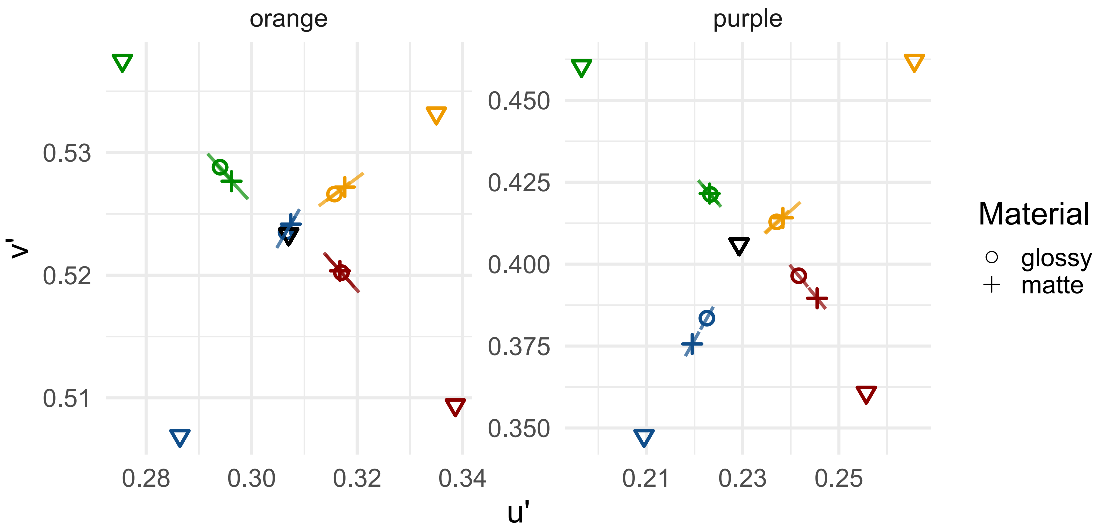


Figure 4.8: Mean chromaticities selected by children, in the same format as Fig. 4.6.

As with adults, we aimed to determine whether children would have higher CCIs when specular highlights and/or a daylight illuminant was present. However, as each child only completed half of the conditions, we could not run an ANOVA on the children's data. Therefore, we ran linear mixed effects models using the lme4 package in R (Bates et al., 2015). The first model was designed to test for main effects and included fixed factors of Illumination, Sweet colour and Material, with observer as a random factor. The results of this are shown in Fig. 4.9. Error bars show 95% confidence intervals, so any variables whose Estimates do not intersect with 0 are statistically significant at the 0.05 level, which are shown by bright red bars. Dark red bars indicate statistically non-significant variables. In response to our third research question, which was to determine whether children can use specular highlights, this model shows that there is a close to significant difference in CCIs between scenes containing matte, compared to glossy sweets ($p = 0.050$), with higher CCIs for matte (mean = 0.302, SD = 0.378) than glossy (mean = 0.265, SD = 0.337) sweets.

Our fourth question was whether children use a daylight prior on this task, and we found a significant difference between CCIs under the Green and Blue illuminants (p=0.001) with significantly higher CCIs under Green (mean = 0.315, SD = 0.305) than Blue (mean = 0.219, SD = 0.364). There were no other significant differences

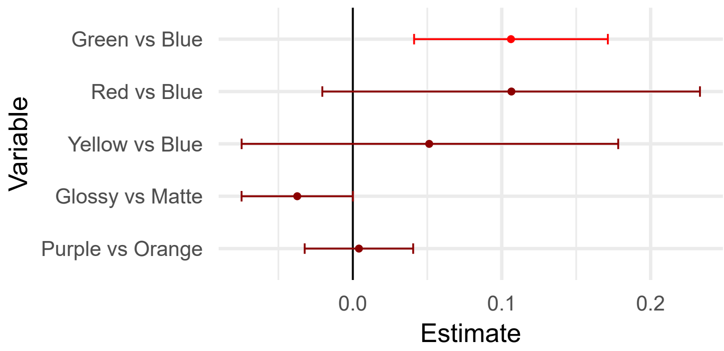


Figure 4.9: Results of lme main effects model on children's CCIs, with fixed effects of Illumination, Material, and Sweet, and random effect of observer. The reference conditions were Blue illumination, Matte material, and Orange sweets. Any estimates greater than 0 indicate that CCIs under the factor on the left of the Variable equation (e.g. Green in the first row) are greater than the reference (e.g. Blue in the first row). Estimates below 0 indicate that CCIs under the reference condition are higher. Bright red bars show significant effects whereas dark red bars show non-significant effects. Error bars are 95% confidence intervals.

between illuminations. This is in contrast to adults who had the lowest CCIs under Green.

As there was a significant interaction between illumination and sweet colour, and a close to significant interaction between material and sweet colour in adults, we ran a second linear mixed effects model on the children's data to test for interaction effects. We added Illumination × Sweet and Material × Sweet to this model:

$$CCI \sim (Illumination \times Sweet) + (Material \times Sweet) + (1|observer) \qquad (4.4)$$

The results of this are shown in Fig. 4.10. Further details on how to interpret this model are given in the Supplementary Materials of Chapter 3. There was a significant interaction between illumination and sweet colour, with CCIs lowest under the Blue illumination for Orange sweets, but highest under Blue for Purple sweets. This can also be seen in Fig. 4.11, although as only fifteen children took part, each completing only half the conditions, these violins appear either very thin or inverted. The pattern here is similar to that in adults, with higher CCIs under Yellow for Orange sweets, and under Blue for Purple sweets.

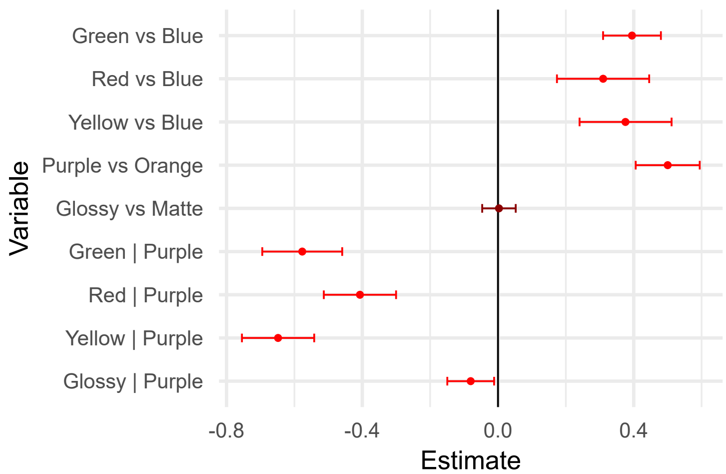


Figure 4.10: Results of lme interaction effects model on children's CCIs, in the same format as Fig. 4.9. The reference conditions were Blue illumination, Matte material, and Orange sweets. The first five variables only change one factor with the other factors fixed. For example, the first row is a comparison between CCIs for Orange, matte sweets under the Green illumination with CCIs for Orange, matte sweets under the Blue illumination. The final four rows show the two way interactions, with significant effects indicating that CCIs vary with one factor depending on the level of another factor. For example, Green | Purple indicates that the effect of the Green vs Blue illumination depends on the sweet colour.

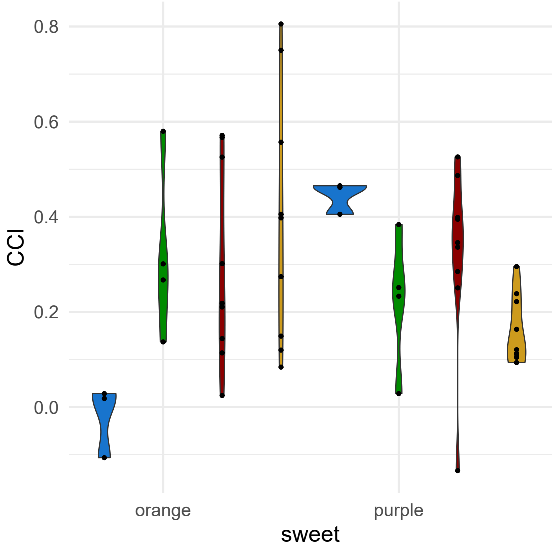


Figure 4.11: Violin plot showing children's CCIs for each illumination and sweet colour, in the same format as Fig. 4.7.

Furthermore, there was a significant interaction between material and illumination, with lower CCIs for glossy purple sweets than matte purple sweets, but no effect of material for orange sweets. This can also be seen in Fig. 4.8, which shows that the mean chromaticity selected for glossy purple sweets is closer to $T_{predicted}$ (black triangle) than for matte sweets.

Adults vs Children

As well as looking at the effects of specular highlights and illumination on adults and children separately, we aimed to determine whether there was a difference between the two age groups. To this end, we ran two more lme models. The first one was to investigate main effects of Age Group, Illumination, Sweet Colour and Material and the output is shown in Fig. 4.12. This is the same as the main effects model used to analyse children's data, with an additional factor of Age Group. It should be noted that there were considerably more adults than children in this analysis, so any main effects which are not between age groups are largely influenced by adults' data. In response to question 5a - whether there is a difference between children and adults' levels of colour constancy - children's CCIs were lower (mean = 0.283, SD = 0.358) than adults' (mean = 0.330, SD = 0.293) in this model, although the difference is not significant ($p = 0.0748$).

We were also interested in whether the effects of material (question 5b) or illumination (question 5c) varied by age group. To test this, we ran another lme model to look at interaction effects involving age group:

$$CCI \sim (Illumination \times AgeGroup) + (Sweet \times AgeGroup) +$$
$$(Material \times AgeGroup) + (1|observer)$$

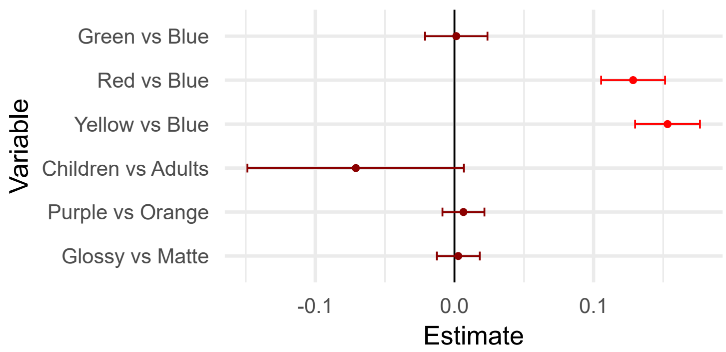


Figure 4.12: Results of lme main effects model with fixed factors of Ilumination, Age Group, Sweet, and Material, and random effect of observer, in the same format as Fig. 4.9.

The results of this model are shown in Fig. 4.13. There was a significant interaction between age group and material, with a negative estimate for Child | Glossy. This means that, while adults have (not significantly) higher CCIs for glossy, compared to matte, sweets, children do not. This interaction is shown in Fig. 4.14, with adults on the left and children on the right. Dark bars show mean CCIs with glossy sweets and light bars show mean CCIs with matte sweets. Bonferroni corrected t-tests showed a close to significant difference between adults and children with glossy sweets ($t(92.6) = 2.08$, $p = 0.08$), with higher CCIs for adults (mean = 0.336, SD = 0.294) than children (mean = 0.265, SD = 0.337). There was no significant difference between age groups with matte sweets ($t(79.28) = 0.61$, $p = 1$). In response to question 5a, this suggests that the effect of specular highlights does vary by age group, such that they only benefit adults.

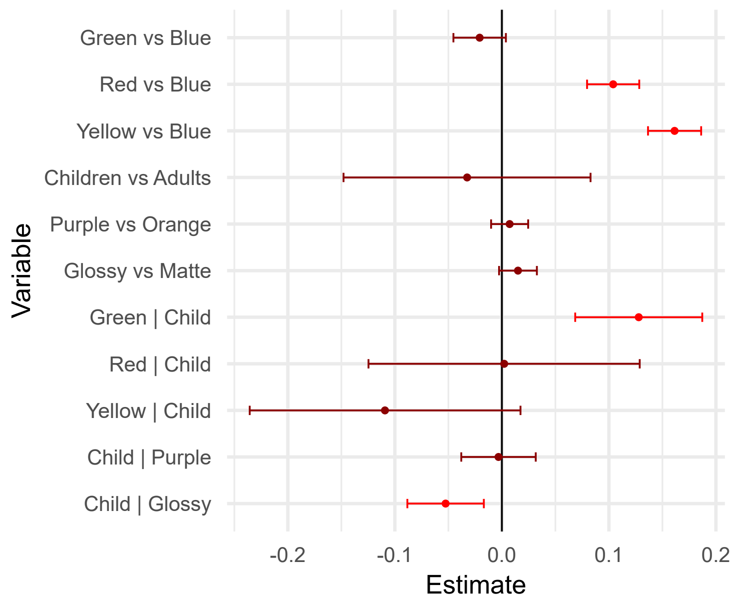


Figure 4.13: Results of lme model investigating interaction effects. See caption for Fig. 4.10 for interpretation of the model.

In addition, there was a significant interaction between illumination and age group. Whilst adults have lower CCIs under Green than Blue, and children have lower CCIs than adults under Blue, the Estimate for Green | Child is positive, and significant, meaning that children's CCIs under Green are relatively higher. To explore this further, we ran Bonferroni corrected t-tests to test for an effect of age group for each illumination condition separately. We found a significant difference between adults and children under Yellow only (t(71.35) = 3.66, p = 0.0019), with higher CCIs in adults (mean = 0.435, SD = 0.334) than children (mean = 0.256, SD = 0.395). There was no significant difference under the Blue (t(20.58) = 0.69 , p = 1); Green (t(23.32)= -1.48, p= 0.61); or Red (t(48.81)

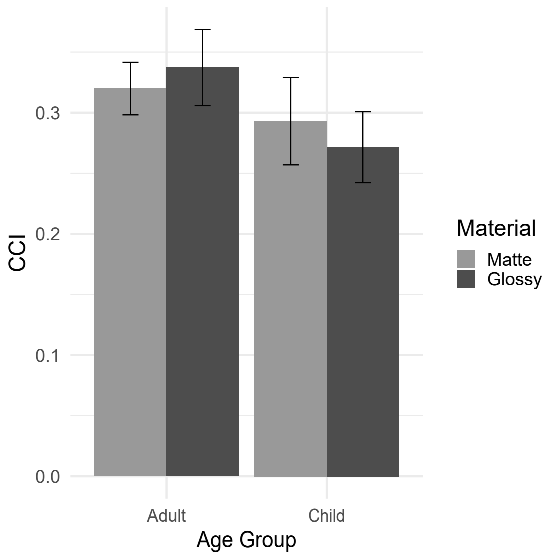


Figure 4.14: Mean CCIs for adults (left) and children (right) with matte (light grey) and glossy (dark grey) sweets. Error bars show ±1 SEM.

= 1.23, p = 0.90) illuminations. This suggests that the effect of illumination does vary by age group, in response to question 5c. Adults demonstrated slightly more use of a daylight prior than children, although even they have lower CCIs under a blue daylight than a red non-daylight. The mean CCIs for adults and children under each illuminant, are plotted in Fig. 4.15. There was no significant interaction between age group and sweet colour.

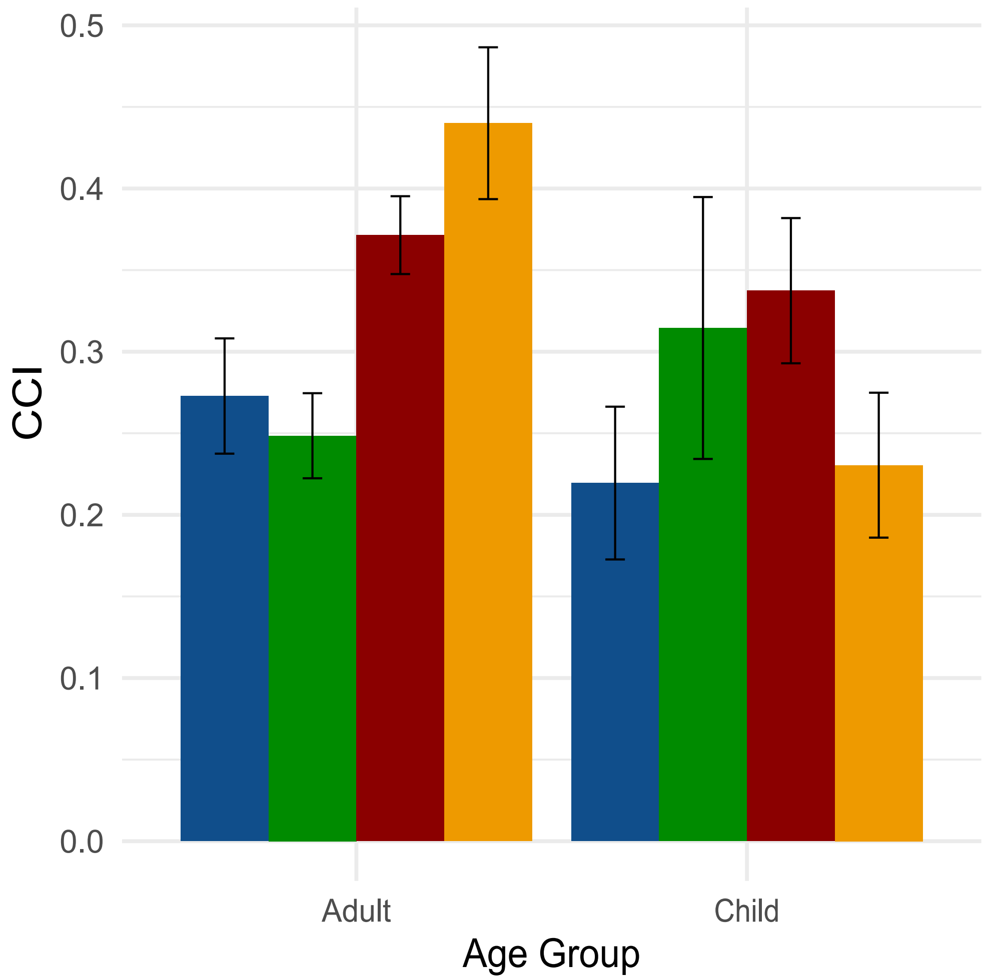


Figure 4.15: Mean CCI for adults (left) and children (right) under each of the four illuminants. The colour of the bars correspond to the illuminations. Error bars are ±1 SEM.

4.3 Discussion

In the present experiment, we adapted an object selection task to test whether adults and children could use a specular highlights cue, and a daylight prior, to improve their colour constancy. Due to the Coronavirus pandemic, fewer observers were tested than planned, particularly children, meaning we lacked power to draw any firm conclusions regarding development. Furthermore, due to issues with rendering stimuli, only two sweet colours were used rather than the four initially planned. Whilst this means that any results are to be taken with caution, we were able to tentatively answer five research questions related to the development of colour constancy.

Generally, colour constancy was quite low, with a mean CCI of 0.32 over all observers. A possible reason for this is that the illuminations we used were quite saturated, all over $50\Delta E_{L*u*v*}$ from neutral (D57). Gupta et al. (2020) showed that the rate of adaptation is the same regardless of the saturation of the illumination, meaning that it takes longer to adapt to more saturated illuminations. As there were only 39 trials per illumination condition and observers spent 6.8 seconds per trial, on average, it is unlikely that observers would be fully adapted to such saturated illuminants for most, if not all, of the trials. Furthermore, we were measuring simultaneous colour constancy, which meant that it was hard for observers to adapt to two illuminations with such different chromaticities. Nevertheless, in response to our five research questions, we found the following.

Our first research question was whether adults would use specular highlights in an object selection task. We found no significant effect of material in a frequestist ANOVA, and moderate evidence in support of the null hypothesis in a Bayesian ANOVA, suggesting that specular highlights did not improve colour constancy in this task. This is in contrast to previous studies finding a benefit of specular highlights (Yang and Maloney, 2001; Lee and Smithson, 2016; Snyder et al., 2005; Yang and Shevell, 2003; Nagai et al., 2017; Granzier et al., 2014; Wedge-Roberts et al., 2020). There are a number of possible reasons for this discrepancy. First, Wedge-Roberts et al. (2020) found no effect of specular highlights when using a uniform grey background, but with a Mondrian background made up of different surface reflectances, observers did use specular highlights. In fact, the effect size we previously found with a uniform background (.005) is comparable to that seen here (.001) - both are less than 0.1 which is considered a small effect. Therefore, it may have been difficult to detect an effect in this experiment due to the neutral background (grey and white) offering a strong cue, which causes observers to assign less weight to the specular highlights. Future studies could test this by using a similar methodology to that employed here, with a Mondrian background.

Another difference between the present study and previous research is that many of the previous studies have used stereoscopic viewing (Yang and Maloney, 2001; Yang and Shevell, 2003; Snyder et al., 2005) or real stimuli (Granzier et al., 2014). Yang and Shevell (2002) showed that binocular disparity, which results from stereoscopic viewing, improves colour constancy for scenes containing specular highlights. Therefore, it is possible that without binocular disparity, the realism of highlights is reduced, meaning observers are less likely to use them as a cue. Future studies using either stereoscopic viewing or real stimuli are needed to determine whether observers use highlights in a more realistic setup.

It could be argued that the fact that all objects had similar body reflectances reduced the effectiveness of specular highlights as the chromaticity convergence cue would be difficult to use. The chromaticities of the highlights did not have the same chromaticity of the illuminations, as can be seen in Fig. 4.16. In this figure, the black O is the chromaticity of the body colour; the blue O is the chromaticity of the Blue illumination, and the Xs (making a line between the Os) are the chromaticities of each pixel within a particular rendered sweet. As can be seen, the chromaticities point towards the illuminant but do not reach it. The plots for the other illuminations and sweet look similar. However, it is unlikely that the difficulty of the chromaticity convergence

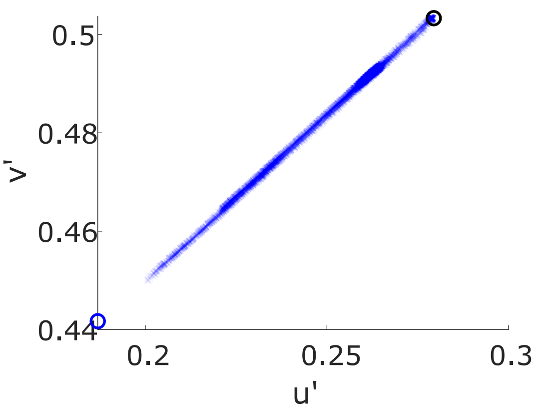


Figure 4.16: Chromaticity, in u'v', of each pixel in a glossy competitor sweet under the Blue illumination. The black O is the chromaticity of the sweet's diffuse reflectance; the blue O is the chromaticity of the illumination, and each blue X is the chromatcitiy of a given pixel. The chromaticities fall on a line between the diffuse and specular reflectances, but do not reach the illumination.

cue is causing the lack of effect of material, as Yang and Maloney (2001) showed that observers can use the specular highlights cue when all spheres had the same body reflectance but not when they had different body reflectances. Furthermore, both Lee and Smithson (2016) and Nagai et al. (2017) found that observers could use specular highlights on a single object in a void.

Our second research question was whether observers would use a daylight prior to improve colour constancy for scenes illuminated by daylights, compared to non-daylights. Whilst there was a significant main effect of illumination, this was not in the expected direction. Performance was best under the Yellow daylight illumination, but low under the Blue illumination. This is in contrast to previous research which has typically found a "blue bias" (Weiss et al., 2017; Pearce et al., 2014; Aston et al., 2019; Delahunt and Brainard, 2004). In our previous study (Wedge-Roberts et al, Submitted), using a similar methodology to that used here, we found highest overall constancy under the blueish illumination. However, a possible explanation for this discrepancy is in the surface reflectances. We found a very strong interaction between illumination and sweet colour, with better colour constancy under the Yellow illuminant for Orange sweets, and under the Red and Blue illuminants for Purple sweets. This effect is not an artefact of the discriminability of each competitor in each condition, as the $\Delta E_{u'v'}$ between each competitor is similar for each illumination condition of a given sweet colour. However, there is a large difference between the ΔEs of each sweet colour, with the chromaticities of the eight competitors closer for the orange than the purple sweets.

As noted in the methods, and shown in Fig. 4.2, the diffuse reflectances of the sweets under D57 were designed to be between the chromaticities of the illuminations. The Purple sweets had a reflectance designed to be closest to the Red and Blue illuminants, while the Orange sweets should have been closest to the Yellow and Red illuminations. However, due to the constraints of picking reflectances which appeared saturated but within gamut, the Orange sweets were closest to the Yellow illumination and the Purple sweets were closest to the Red illumination. The interaction between sweet colour and illumination, therefore, appears to be driven by higher CCIs when the surface reflectance and illumination have similar chromaticities (Orange sweets under Yellow illumination and Purple sweets under Red illumination), than when they have dissimilar chromaticities (Orange sweet under Blue illumination and Purple sweet under Green illumination). This pattern can also be seen in the mean chromaticities of sweets selected (Fig. 4.6). Furthermore, this is in agreement with the results of our previous study (Wedge-Roberts et al, Submitted), in which we found higher constancy when sweet colours and illuminations aligned. However, as we only used two reflectances here, and they did not align as planned, future studies are needed to explore this further, which use more surface reflectances, aligned with the illumination chromaticities.

Our next two research questions concerned colour constancy in children. Unfortunately, we were only able to test 15 children, rather than the planned 60, due to COVID-19. As each child only completed half the illumination conditions, this meant that five children completed the Green and Blue conditions, and 10 children completed the Yellow and Red conditions. Therefore, the results are to be taken with caution, as we lacked power. Nevertheless, from the data obtained there was no significant effect of specular highlights on colour constancy in children. There was a close to significant effect, in which children performed worse with specular highlights than without. This suggests that children may be impaired by specular highlights, rather than using them to improve constancy. A possible reason is that children may not be interpreting highlights correctly, and instead attributing them to the body colour of the sweets. If observers took the mean chromaticity over all pixels including the highlights, it would shift the perceived colour. Alternatively, the highlights may simply be a distraction for children, which cover up some of the sweet. To test this, future studies are needed to determine whether children can perceive gloss in rendered images. Although Yang et al. (2011) found that infants can discriminate glossy from matte objects, it is unclear how gloss is perceived and interpreted by children. The results here suggest that children may not perceive gloss in the same way adults do. Alternatively, the fact that images used here were rendered and not viewed stereoscopically may prevent children from perceiving the highlights as such, to a greater extent than adults. Therefore, this study should be replicated with real glossy objects to determine whether children can use highlights in such situations.

As with adults, we did not find evidence in support of a daylight prior with children. Although there was a significant effect of illumination, this was driven by better constancy under the Green non-daylight than the Blue daylight illuminant. However, the effect of illumination was mediated by the sweet colour, with the highest levels of colour constancy for Orange sweets under the Yellow illuminant, and for the Purple sweet under the Blue and Red illuminants. The pattern across sweets and illuminations for children (Fig. 4.8) and adults (Fig. 4.6), is almost identical, suggesting that this effect has already developed by six years. Very few previous studies investigating the use of a daylight prior in colour constancy have manipulated surface reflectance, with many using achromatic settings (e.g. Weiss et al., 2017; Delahunt and Brainard, 2004). Therefore, they would not have revealed the interactions between illumination and surface reflectance found here; simply showing a daylight prior for achromatic surfaces.

Our final aim was to determine whether there was a difference in colour constancy between adults and children. We found that adults were, overall, better than children on this task, although the difference was not significant. This is counter to findings from our previous study (Wedge-Roberts et al, Submitted), in which children had higher CCIs than adults. As we know there are large individual differences between observers, this could simply be due to the small sample size. However, an alternative explanation is the nature of the stimuli used. In the present experiment, the sweets were all sitting on the floor of the box, so it is clear that they are in the same illumination framework as the rest of the box. Furthermore, the checkered back walls makes it more clear that there is a different illumination, rather than different coloured walls, in the two rooms. In the previous study, adults may have performed worse as they were attempting to ignore the different room colours, whereas here they were more likely to perceive the rooms as differently illuminated. To explore this further, future studies investigating the development of colour constancy with physical, rather than rendered, stimuli are needed.

As discussed above, for adults there was no effect of material whereas children were (non significantly) worse with glossy, compared to matte, sweets. Combining the data we found a significant interaction, with no difference between adults and children with matte sweets but children significantly worse than adults with glossy sweets. This reinforces the notion that specular highlights have a negative effect on children, and suggests that the ability to use such a cue is still developing. Even though adults do not show a benefit of having specular highlights in this experiment, the fact that they do not have a detriment suggests that they are accurately perceiving the sweets as glossy.

Finally, we aimed to determine whether a daylight prior was still developing in six to 10 year old children. As noted above, the pattern across illuminations and sweets is consistent across age groups. However, there was a significant interaction between illumination and age group, with a different overall pattern across illuminations for the two age groups. This is not consistent with a developing daylight prior, as there was no significant difference between children and adults under the Blue illumination, but there were significant differences under both Red and Green non-daylights. Future studies with a larger sample of children, and a broader range of surface reflectances, can help make this pattern clearer.

Overall, we found that adults did not use specular highlights or a daylight prior to improve their colour constancy. Instead, the effect of illumination was mediated by the diffuse surface reflectance of the sweets. Furthermore, children had poorer constancy in the presence of specular highlights, suggesting that they are still developing the necessary skills to use such a cue - perhaps gloss perception. In addition, children did not use a daylight prior, but had a similar interaction between illumination and surface reflectance as adults. Children performed worse, overall, than adults, suggesting colour constancy is still developing at six to 10 years.

Due to the limitations of this study (low numbers of children and limited surface reflectances), further research is required to more concretely answer the questions regarding the development of the use of specular highlights and a daylight prior, which we have tentatively answered here. In addition, the findings lead to additional questions which can be investigated in future studies. Specifically, the interaction between illumination and surface reflectance suggests that colour constancy may be higher for surfaces which have a similar chromaticity to the illuminant, which can be tested by more controlled manipulation, and a wider range of illuminants and surface reflectances. Additionally, we found a detrimental effect of specular highlights in children, but further research is required to explore the causes behind this finding, such as by looking at the development of gloss perception. Furthermore, we found poorer constancy in children than adults in this study, but in our previous study (see Chapter 3), this finding was reversed. Future studies manipulating the realism of stimuli may help explain this discrepancy.

Chapter 5

General Discussion

Colour constancy is a computationally difficult problem to solve as an object's surface reflectance and the illumination falling on it are combined in the signal of light arriving at an observer's eye. Therefore, cues and priors have been suggested which observers can use to disentangle the contributions of surface reflectance and illuminant. As it can take time to build these priors and learn what cues are and how to use them, it is possible that it takes time during development to learn such information. Here we have investigated these aspects of colour constancy.

At the outset of this work, we aimed to answer six research questions relating to cue and prior-use, and development of colour constancy. To answer these questions, we ran six experiments. The first three experiments were run only with adults to determine whether they could use specular highlights and/or a daylight prior. Following this, it was determined that a novel task was needed to measure colour constancy in children. In the next two experiments, we developed and tested a novel object selection task with six to 11 year old children, using both simple two-dimensional stimuli, comparable to those used in many previous studies, and more realistic three-dimensional rendered stimuli. Having determined that this task was appropriate for measuring colour constancy in children in this age range, we ran a final experiment using the novel task to ask whether, and when, children can use specular highlights. Here, I will begin by discussing how the findings from each experiment helped to answer each research question outlined in the general introduction. I will then discuss any other findings of interest, outlining any outstanding questions and proposing studies needed to answer these, and ending with final conclusions.

5.1 Specular highlights and other cues

Based on previous literature (Yang and Maloney, 2001; Xiao et al., 2012), the first question we aimed to answer was whether observers would use a specular highlight cue differently depending on the presence or validity of other cues. To determine this, we first needed to find out whether adults could use specular highlights in any circumstances. In the first series of three experiments, observers were presented with stimuli containing either glossy shapes (with valid specular highlights) or matte shapes (with no highlights). When making achromatic settings, the matching patch was either against a uniform white wall (Experiment 1, wall condition); a teal shape (Experiments 1 and 2, shape condition); a Mondrian wall (Experiments 2 and 3 wall condition); or a shape of varying reflectances (Experiment 3 shape condition). If the use of specular highlights depended on the local surround cue, such that a greater weight was applied to the specular highlights when the local surround cue was harder to use, we would predict differing effects of highlights against each surround, with the least reliance on highlights against the uniform white wall. Furthermore, we would predict an interaction between specular highlights and patch position in all three experiments.

We found no difference in the weight applied to specular highlights (as measured by colour constancy indices) on the wall compared to on the shape in any experiment. Observers used specular highlights to improve constancy regardless of patch position in both experiments with a Mondrian wall (Experiments 2 and 3) but did not use specular highlights in either position when scenes contained a uniform white wall (Experiment 1). Given the stark difference in local surround between patch positions, this questions the idea that the weight applied to specular highlights does depend on the local surround cue. Furthermore, this finding is in contrast to Yang and Maloney (2001) who found that observers did not use specular highlights when a matching patch was on the back wall but did when it was on a glossy shape. A possible reason for this discrepancy is that Yang and Maloney (2001) used a uniform white background wall, similar to that used in Experiment 1 here, in which we found that observers did not use specular highlights. Furthermore, the back wall used by Yang and Maloney (2001) was perpendicular to the observers, without any indication of three-dimensionality, so it may have been less clear that the objects and back wall were part of the same illumination framework. This could mean that observers were less likely to apply information about the illuminant, obtained from specular highlights on the shapes, to judgements made on the back wall.

Whilst there were no interactions in any individual experiment, we found that observers only used the specular highlights (as shown by higher levels of colour constancy with glossy shapes) when there was a Mondrian background (Experiments 2 and 3), and not when there was a uniform white background (Experiment 1), lending support to the notion that the weight applied to specular highlights does depend on other cues. The difference in use of specular highlights between scenes containing a white wall and scenes containing a Mondrian wall may be due to a difference in global surround, rather than local surround which differs between patch positions.

In Experiment 6, we were able to test whether the use of specular highlights would extend to a different task. In this, observers performed an object selection task with either matte or glossy sweets. Adults did not use specular highlights in any condition, as shown by no difference in levels of measured colour constancy between scenes containing matte and glossy sweets. Whilst this is in contrast to Experiments 2 and 3, it is similar to the finding in Experiment 1. This is likely due to the fact that there was a spectrally uniform background wall in both Experiments 1 and 6. Therefore, this finding supports the notion that the use of specular highlights depends on other cues, such as global surround.

Additionally, specular highlights had a *negative* impact on colour constancy for children in Experiment 6, for purple sweets but not orange sweets. Whilst it is unclear what is driving this interaction, and there was no such significant interaction in adults, this may suggest that the use of specular highlights can depend on the surface they are on.

Overall, the differences in findings between experiments do suggest that the extent to which specular highlights improve colour constancy may interact with other cues such as the global surround. However, we found less evidence for an interaction between specular highlights and local surround, suggesting that the information obtained from highlights is not constrained to a local region near the highlights.

Although we did not directly measure the interactions between other cues, these findings suggest that cues to colour constancy may not be used independently; the use of each cue may depend on what other cues are available. Future studies manipulating different combinations of cues will be able to determine the nature of such interactions more precisely. In addition, differences between Experiment 6 (object selection) and Experiments 2 and 3 (achromatic setting) suggest that the effect of specular highlights may depend on the method of measuring colour constancy, which should be considered in future research.

5.2 Specular highlights and daylight priors

Previous research has suggested that observers may use specular highlights more when the scene is under a daylight than under a non-daylight illumination (Yang and Maloney, 2001). On the other hand, given that specular highlights improve colour constancy less when other cues are valid (see above), it was possible that observers would show less benefit of specular highlights on colour constancy when the illumination was a natural daylight, as they would have sufficient information from using their prior knowledge and other cues. Therefore, we aimed to determine whether the weight applied to specular highlights varied depending on the type of illumination.

In all experiments presented here, we tested observers with both daylight and non-daylight illuminations. In the first three experiments, we found that the use of specular highlights did not depend on the type of illumination, in terms of either colour constancy or variable error of estimates. Similarly, in Experiment 6, in which scenes were illuminated by more saturated illuminations, observers did not show any evidence of using specular highlights under any illumination condition. Although it is possible that the reason observers do not use specular highlights in Experiment 6 while they do in Experiments 2 and 3 is due to the illuminations being more saturated in the last experiment, there are more plausible explanations relating to the global mean, as noted above.

Taken together, these findings suggest that the extent to which specular highlights improve colour constancy does not vary depending on the validity of a daylight prior. This is counter to findings from Yang and Maloney (2001). A large difference between the experiments presented here and Yang and Maloney (2001) is that we did not perturb any cues. This meant that the specular highlights provided a valid cue regardless of the type of illumination. As Yang and Maloney (2001) were measuring the *weight* applied to specular highlights while we were determining whether highlights improved colour constancy, their measure was more sensitive, which could explain the seemingly contradictory findings. Furthermore, in general we found low levels of colour constancy, and weak evidence for a daylight prior (see following section). Given that observers did not have higher levels of colour constancy for scenes illuminated by daylights, we may not expect the use of specular highlights to be mediated by the illumination. Future studies in which observers show stronger evidence of using a daylight prior on colour constancy indices (e.g. using more immersive viewing, or varying the saturation of the illuminations) would be more likely to find this interaction, if it is present.

5.3 Daylight prior and cues

Probabilistic theories suggest that observers may be able to use a daylight prior - knowledge that scenes are more likely to be illuminated by natural daylights on the daylight locus - to improve colour constancy (Brainard and Freeman, 1997; Brainard et al., 2006). This has been partially supported by a small number of studies finding superior constancy under blueish illuminants (Delahunt and Brainard, 2004; Weiss et al., 2017; Aston et al., 2019; Pearce et al., 2014). In all six experiments presented here, the type of illumination was manipulated to test whether observers use a daylight prior to improve constancy under illuminations along the daylight locus compared to unnatural reddish and greenish illuminations. Furthermore, we were able to test whether the extent to which observers used a daylight prior was mediated by the presence of other cues, as in Delahunt and Brainard (2004).

In Experiment 1, using an achromatic settings task, there was no benefit, in terms of colour constancy or variability of estimates, for scenes illuminated by daylight, compared to non-daylight illuminants. However, achromatic settings were all biased towards the blueish direction of the daylight locus. This may suggest that observers assume illuminations are blueish meaning they attribute blueness in the matching patch to the illumination rather than the surface, thus perceiving a spectrally biased patch as achromatic. Furthermore, the extent to which observers' settings were biased towards blue differed across the patch positions, with a greater bias when the matching patch was against a turquoise shape than when it was against a white wall. This suggests that observers have a greater "blue bias" when a local surround cue is harder to use. This could reflect greater dependence on a daylight prior when other cues are weakened. However, there was no effect of specular highlights on this bias. Additionally, the effect of illumination on colour constancy indices and variability was not mediated by either patch position (local surround) or specular highlights.

In both Experiments 2 and 3, the results were similar to those in Experiment 1, with no benefit of daylight illuminants, on either colour constancy or variability of estimates. Additionally, settings were biased towards the blue illumination, to a greater extent than in Experiment 1. There was little difference in the bias between patch positions in these two experiments, although the use of a Mondrian background meant that the average chromaticity and luminance of the local surround was more closely equated between the back wall and the central shape in these experiments. The fact that there was a greater bias in these experiments than in Experiment 1, along with lower overall levels of colour constancy, suggests that observers may rely more on a daylight prior when the task is harder (e.g. when the global mean cue is harder to use due to background articulation). On the other hand, there was no significant interaction between illumination and patch position or specular highlights on levels of colour constancy or variability in either of these experiments.

In Experiments 4 and 5, observers completed an object selection task for four different sweet colours, under either daylight or non-daylight illuminants. This allowed us to determine whether observers use a daylight prior with a different task. In support of a daylight prior, we found higher constancy under the blueish daylight illuminant than any other illuminants. However, the lowest colour constancy was found under the yellow daylight illuminant tested. This is in line with many previous studies investigating the effect of a daylight prior, which tend to find an asymmetric effect. A possible explanation is that natural illuminations tend to be more saturated along the blue than the yellow direction (Nascimento et al., 2016; Hernández-Andrés et al., 1999), meaning the yellow used here may have been too saturated. In addition, whilst daylight illuminations range from blueish skylight to yellowish sunlight, skylight tends to be more diffuse while sunlight is punctate. Therefore, a diffuse yellow illumination may not be in line with a daylight prior.

An interesting finding in both experiments 4 and 5 was that the effect of the illumination was strongly dependent on the surface reflectance tested. Colour constancy was highest when the chromaticity of the sweet was similar to the chromaticity of the illumination. This was true across age groups and across experiments. This suggests that the overall benefit for scenes illuminated by a blueish daylight illuminant could have been an artefact of the surface reflectances used - in Experiment 4 colour constancy was highest under the blue illumination for the teal sweets but not the other three sweet colours tested. Whilst this does not answer the question of whether use of a daylight prior is mediated by cues, it does suggest that the use of a daylight prior is mediated by other factors (here surface reflectance).

The difference between Experiment 4 and Experiment 5 is that two-dimensional stimuli were used in Experiment 4 while three-dimensional rendered stimuli were used in Experiment 5. This means that there should have been more cues available in Experiment 5, for example scene geometry. If the use of a daylight prior depended on cues, we would expect to see less of an effect of illumination in Experiment 5, when more cues were present. However, the effect of illumination was similar across these experiments, suggesting that the additional cues did not reduce reliance on a daylight prior. It should be noted that constancy was actually *lower* with three-dimensional rendered stimuli than with two-dimensional stimuli. This suggests that the additional cues did not make the task easier - potentially due to the lack of realism in the stimuli, which contained sweets floating in a box. This may therefore explain why the cues did not affect the use of the prior, as they were not easy to use.

In contrast to all other experiments, in Experiment 6 (in adults) we found the highest constancy under a yellow-ish daylight illuminant, and the second lowest under blue. This is particularly surprising as the illuminations used in Experiment

6 were more saturated than in all previous experiments, which means the yellow illumination was even less natural. However, as in Experiments 4 and 5, the effect of illumination was mediated by the surface reflectance of the sweets. Only two reflectances were tested here: orange and purple. For orange sweets, constancy was highest under the yellow illumination whereas for purple sweets, observers performed best under blue. As noted above, this seems to be due to constancy appearing higher when the chromaticities of the illumination and surface reflectance align. It is possible that using more sweet colours which average at neutral, as was originally intended for this experiment, would result in higher overall constancy under a different illumination. Future studies using more surface reflectances are therefore needed to determine whether observers can use a daylight prior in this task.

There was no evidence of an interaction between a daylight prior and specular highlights in Experiment 6, which is in agreement with Experiments 1 to 3, suggesting that the strength of a daylight prior is not dependent upon the presence of glossy objects.

Overall, we found mixed evidence for a daylight prior in adults, with a blue bias in the first three experiments, but no benefit in terms of colour constancy; highest constancy under a blueish daylight in experiments 4 and 5, but a strong dependence on surface reflectance; and highest constancy under a yellowish daylight in the final experiment, but again interacting with surface reflectance.

There was some evidence for an interaction between a daylight prior and some cues, with the blue bias depending on local surround and articulation of the background. However, the influence of the illumination on levels of measured colour constancy was not affected by the presence of specular highlights, or the dimensionality of the stimuli.

5.4 Development of colour constancy

Whilst much was previously known about colour constancy in adults, prior to this work no studies had investigated when colour constancy develops during childhood. This is an important avenue of research as without this research, we do not know how, or when, the adult brain develops the skills to solve such a complex problem as colour constancy, which is computationally difficult. Furthermore, it is likely that common perceptual systems are used for different types of constancies (such as size or shape constancy), meaning that any findings on the development of colour constancy are useful for informing on the development of other constancies. Understanding how humans develop colour constancy will also be useful in designing artificial intelligence systems to learn how to perceive colours as invariant under changing illuminations. Additionally, if colour constancy does not develop at an early age, children may perceive colours differently to adults, meaning any coloured materials designed for children would not be perceived as expected. Furthermore, if children lacked colour constancy, their visual world would be much more variable that that of adults. In the last three experiments presented here, we aimed to determine the nature of the developmental trajectory between six and 11 years.

Counter to expectations, we found higher levels of measured colour constancy in children than adults in Experiments 4 and 5, with constancy decreasing with age over childhood. This was surprising, and a number of possible explanations are discussed in Chapter 3, including different interpretation of instructions leading to different strategy use, or a greater level of adaptation in younger children than in adults.

In Experiment 6, in contrast, children were (non-significantly) worse than adults, although this was mediated by both specular highlights, such that children were worse with glossy sweets only, and sweet colour, such that children were worse with purple sweets but not orange sweets. We were unable to track the developmental trajectory in this experiment due to small numbers of observers.

There are a number of possible explanations for the discrepancy between experiments. Firstly, it is important to remember that the data obtained in Experiment 6 were incomplete due to the impact of COVID-19, meaning that only a small sample of children were tested, with very small numbers in certain age bands. Therefore, it is impossible to say whether the tentative finding of poorer constancy in children would have been found with a larger sample. Future studies with complete samples are needed to determine the true pattern.

However, if the finding in the final experiment did reflect the true nature of development, there are other explanations for the difference between experiments.

Experiment 6 consisted of the most realistic stimuli, with three-dimensional rendered sweets containing mutual reflections and sitting on the floor of a room (rather than floating). In this experiment children's levels of colour constancy were lower than in either of the other experiments. The mutual reflections and specular highlights may be confusing younger observers who have less understanding of how light reacts with different types of surfaces. Alternatively, it could be that children do understand how mutual reflections and specular highlights work in the real world but are unable to apply this knowledge to rendered stimuli viewed on a monitor. To test this, future studies are needed to determine the development of colour constancy with real (rather than rendered) stimuli.

Another difference between these experiments is the illuminations used. In Experiment 6, the illuminations were more saturated than in Experiments 4 and 5. With two spatially separated, simultaneously presented illuminations, it is difficult to ensure close to complete adaptation to either without controlling the duration spent viewing each. The larger the difference in chromaticity between the illuminants, the harder it is to adapt to both. Furthermore, Gupta et al. (2020) showed that the rate of adaptation does not depend on the saturation; it simply takes longer to adapt to more saturated illuminants. We hypothesised that children may have greater adaptation than adults in Experiments 4 and 5, due to less divided attention between either side of the scene. Perhaps by increasing the saturation, it became harder to adapt to the chromatic illumination, meaning children did not have sufficient information for high levels of constancy, even if they were looking more towards the side illuminated by the chromatic illuminant than adults. Studies deliberately manipulating the saturation of the illuminants while keeping everything else constant and/or measuring and controlling for eye movements are needed to determine whether this can explain the difference in findings.

Other differences are that there were specular highlights present in Experiment 6 only (discussed below), and the surface reflectances were different. Given that there was no interaction between sweet colour and age group in Experiments 4 and 5, and there is no theoretical reason that children would behave differently to adults in response to different surface reflectances, it is unlikely that this can explain the difference.

Following Experiments 4 and 5, it was hypothesised that adults were using cognitive strategies to ignore the illumination and attempt to match on hue and saturation, thus leading to lower levels of colour constancy. This should be more difficult in the final experiment, where stimuli are more realistic, which should result in higher levels of constancy. However, adults were less constant in Experiment 6 (realistic three-dimensional rendering) than in Experiment 4 (two-dimensional).

Taken together, the results on the development of colour constancy are inconclusive with children having better colour constancy than adults with two-dimensional stimuli and simple (less realistic) rendered three-dimensional stimuli, but potentially worse than adults with more realistic three-dimensional stimuli (although this finding is preliminary and non-significant). Whilst these findings are important in understanding how perceptual constancies are acquired during development, future studies using real (physical) stimuli are needed in order to apply these findings to the development of real world constancy. Furthermore, there are many more aspects of the development which remain to be discovered, such as how development differs between different measures, and the development of cues and priors for colour constancy (see following sections). Nevertheless, the finding that children can perform well in an object selection task suggests they have at least developed some form of colour constancy by six years.

5.5 Development of daylight priors

Given previous research into the development of other perceptual priors, we were interested in determining at what age children are able to use a daylight prior to increase colour constancy for scenes illuminated by natural daylight illuminants (blues to yellows). To this end, illuminations both on and perpendicular to the daylight locus were used in all developmental experiments (4, 5 and 6).

In Experiment 4, using simple two-dimensional stimuli, we found no evidence in support of a developing daylight prior in children aged six to 11 years. Whilst children had overall higher levels of colour constancy than adults, both age groups had the highest constancy under the blue daylight illuminant and lowest under the yellow daylight. Regressions and model fitting showed a decrease in colour constancy with age for all four illuminations tested, with no difference between daylights and non-daylights. This suggests that any use of a daylight prior, shown by superior constancy under the blue daylight, has already developed by six years.

On the other hand, with three-dimensional rendered stimuli in Experiment 5, we did find an interaction between age group and illumination, with little difference between children and adults under the yellow illuminant, but better constancy in children under all other illuminants. If a daylight prior were developing with age, we would predict more equal levels of colour constancy across illuminations in children than adults. This finding is the opposite, with children having a greater asymmetry across illuminants than adults. However, the fact that children have higher constancy than adults under a blue daylight, as well and red and green non-daylight illuminants, is consistent with Experiment 4. Overall, this experiment does not provide strong evidence either in support of or counter to a developing daylight prior.

In Experiment 6, the illuminations used were more saturated than in Experiments 4 and 5, and the sweets' surface reflectances were different. In this experiment, there was an interaction between age group and illumination. Under a saturated yellow illuminant, adults had a higher level of colour constancy than children whereas there was no significant difference between age groups under all other illuminants. In contrast with the previous two experiments, this does support the notion of a daylight prior developing with age, such that children do not have the same benefit of a yellow daylight illumination that adults do. In addition, children were (non-significantly) worse than adults under a blue daylight, but better under a green non-daylight. Whilst the small amount of data collected on children in this experiment does mean this finding should be taken with caution, it does suggest that the ability to use a daylight prior with more saturated illuminants may still be developing up to at least 10 years.

Taken together, these three developmental studies have conflicting results regarding the development of a daylight prior: the first suggests that children have already developed a prior by six years of age; the second is inconclusive; and the third suggests that the use of a daylight prior is still developing in children over the age of six. A possible reason for this inconsistency is that there was not strong evidence of adults using a daylight prior in any of these studies: as noted in Section 5.3 above, the highest levels of constancy in adults were under blue in Experiments 4 and 5 but yellow in Experiment 6, and this varied depending on the sweet colour. Possible reasons for the lack of a daylight prior found in adults are discussed in Section 5.3. Therefore, future studies are needed to determine more conclusively whether adults truly have a daylight prior, before asking whether, and when, children do.

5.6 Development of specular highlights cue

Our final aim was to determine whether children can use a specular highlights cue to improve their colour constancy. In Experiments 1, 2, and 3, we determined that adults can use specular highlights to improve their levels of colour constancy, only when other cues were difficult to use. In Experiments 4 and 5, we developed a novel way to measure colour constancy in children using an object selection task. In Experiment 6, we aimed to answer the final research question by applying the novel method to scenes containing either matte or glossy sweets.

In the final experiment, we tentatively found that a small sample of children were negatively impacted by specular highlights, as shown by having poorer constancy in scenes containing glossy sweets. This is in contrast to adults whose levels of colour constancy were no different for scenes containing specular highlights compared to those without. This finding suggests that children aged six to 11 years may still be developing the ability to use specular highlights as a cue for colour constancy. However, given that adults did not use specular highlights to improve colour constancy in this experiment, and the sample of children tested was small, this finding should be taken with caution.

If further studies testing more children find that they truly do show a detriment of specular highlights, additional studies are needed to determine what is causing children to be impaired by specular highlights when adults are either not affected, or show greater levels of colour constancy when scenes contain highlights. A possible explanation is that children do not perceive the highlights as such, and assume they are part of the body colour. This would lead them to use them inappropriately, and possibly lead to poorer constancy. It may be that children particularly struggle to perceive highlights appropriately in rendered stimuli. Therefore, future studies using real glossy objects may yield different findings. Another future direction would be to introduce a Mondrian background. From the first set of experiments, we determined that adults only used specular highlights when the scene contained a Mondrian background. Therefore, it would be interesting to compare adults and children in a situation in which adults use highlights to improve their colour constancy.

5.7 Other findings

In addition to providing some answers to the six research questions outlined at the beginning, we found some additional findings of interest in these experiments. The first finding, as alluded to above, is that when looking for a daylight prior we found a strong interaction between the illumination and the surface reflectance. Whilst physically it may not be surprising that colour constancy depends on an interplay between the two spectra which are combined in the signal of light reaching the eye, this finding does have important implications for future research. Specifically, many studies investigating daylight priors in colour constancy have used achromatic setting (Delahunt and Brainard, 2004; Weiss et al., 2017; Bosten et al., 2015; Gupta et al., 2020). We have shown here in three separate experiments that use of a "daylight prior" is dependent upon the surfaces used, such that observers *appear* to use a blue daylight prior on blue-ish surfaces and a yellow daylight prior on yellow-ish surfaces. Therefore it is important for future studies to take this into account by manipulating the surface reflectances and determining whether observers use a daylight prior with all surfaces; not just for spectrally non-selective surfaces.

Another finding briefly discussed above is the effect of local surround in the first three experiments. For all three, we found a significant effect of patch position (against coloured shape vs plain or Mondrian wall) on both colour constancy and variability of settings. Whilst the focus of these experiments was on the less researched cue of specular highlights, it is interesting to note that there was an effect of local surround, in agreement with previous studies (Smithson and Zaidi, 2004; Kraft and Brainard, 1999; Hurlbert and Wolf, 2004). Furthermore, previous studies have considered the effect of local surround on colour constancy indices only, whereas here we also measured variable error. We found higher levels of variability on the turquoise shape than a plain white wall (Experiment 1), but less variability on the turquoise shape than a Mondrian wall (Experiment 2). This is in contrast to colour constancy which was always highest on the wall, regardless of the articulation of said wall. This shows the importance of using different measures depending on the question.

Related to this, measuring variable error in addition to colour constancy indices was a novel approach employed in these experiments. Whilst we did not find the expected effect of specular highlights on variable error, it would be interesting to determine whether variable error in colour constancy tasks is affected by other cues. As we simply measured variable error using an achromatic setting task, we have shown that it is easy for future experiments to test this whilst also using typical measures of colour constancy. The analysis techniques developed here are applicable to any future studies which require observers to make adjustments.

A finding throughout all six experiments was that levels of measured colour constancy were lower than have previously been reported. The highest colour constancy indices were found in Experiment 1 (0.519) whereas the lowest were in Experiment 2 (0.255), where 1 would be perfect constancy. A possible explanation for these low levels of constancy is that the experiments were not very immersive or realistic. In all six experiments, observers were viewing stimuli on a two-dimensional monitor, without a stereoscope, unlike in more immersive environments used previously (Radonjić et al., 2015b; Yang and Maloney, 2001; Granzier et al., 2014; Gupta et al., 2020; Delahunt and Brainard, 2004). Furthermore, in most of the experiments, the stimuli were not physically realistic, containing floating shapes (Experiments 1 to 3), or floating sweets (Experiments 4 and 5). Although stimuli were more physically plausible in Experiment 6, the increased saturation of the illuminations, coupled with the simultaneous viewing, made it hard for observers to adapt fully.

An additional factor influencing the low CCIs found is that in all experiments, all observers were naive (except one in Experiment 5 who was excluded from analysis). In many previous studies the observers have included the experimenters themselves (Lee and Smithson, 2016; Arend and Reeves, 1986; Delahunt and Brainard, 2004; Yang and Maloney, 2001). Testing solely observers who are unaware of the purpose of the study, and mainly unaware of the concept of colour constancy, allows us to measure colour constancy in a way that is more applicable to the general population, to determine how people truly perceive colours. Therefore, it may be that colour constancy has been overestimated in previous studies using non-naive observers.

Finally, we found that an object selection task can be adapted and gamified in a way that children aged six to 11 years enjoy, by using a cartoon dragon. Many of the children reported to the experimenter that they found this task fun, with one even stating that they wanted to play it at home. This means that future experiments aiming to investigate the development of particular aspects of colour constancy can adapt this task to answer different questions. However, given the results found in the present experiments, caution will need to be taken when selecting the surface reflectance of the sweets, and care is needed to ensure that observers are performing the task as expected, rather than choosing the nicest looking (which may be the most saturated) sweet for example.

5.8 Conclusions

Overall, we aimed to research three general areas of colour constancy: cues, priors, and development. With regards to cues, we tested observers' ability to use specular highlights, which had been little researched before. We found that observers can use highlights to improve colour constancy, but the extent to which they are used depends on the availability of other cues, including the global mean. The fact that observers do not benefit from specular highlights when other cues are valid and easy to use may be because specular highlights require more effort to use than some other cues. In addition, we found evidence of observers using a local mean cue.

Regarding priors, we did not find strong evidence of observers using a daylight prior for colour constancy, in either an achromatic settings task, or an object selection task. This suggests that observers are able to judge object colours as relatively unchanging under both natural daylights and unnatural illuminations. Whilst people tend to spend a large amount of their time indoors under unnatural illuminations, indoor lighting tends not to be as far from the daylight locus as those illuminations tested here. However it would be interesting to determine whether people who spend more time outside, and have less experience with unnatural illuminations, such as from non-western cultures, show greater use of a daylight prior.

Finally, in terms of development, we found that children outperformed adults in two out of three of the developmental experiments. The fact that there is a difference in measured colour constancy between adults and children suggests that children's object colour perception is still changing up to at least 11 years. Although there are a number of possible explanations for this, it seems unlikely that children have no colour constancy given how well they were able to pick a matching object under a different illumination. While further studies are needed using more realistic stimuli, this suggests that coloured teaching materials designed for children are appropriate, and will not appear to change colour vastly under changing illuminations for younger children.

References

Adams, W. J. (2007). A common light-prior for visual search, shape, and reflectance judgments. *Journal of Vision*, 7(11):11.

Alais, D. and Burr, D. (2019). Cue combination within a bayesian framework. In *Multisensory Processes*, pages 9–31. Springer.

Amano, K. and Foster, D. H. (2004). Colour constancy under simultaneous changes in surface position and illuminant. *Proceedings of the Royal Society B: Biological Sciences*, 271(1555):2319–2326.

Arend, L. and 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., Schirillo, J., and Goldstein, R. (1991). Simultaneous color constancy: papers with diverse Munsell values. *Journal of the Optical Society of America A*, 8(4):661.

Aston, S. and Hurlbert, A. (2017). What #theDress reveals about the role of illumination priors in color perception and color constancy. *Journal of Vision*, 17(9):4.

Aston, S., Radonjić, A., Brainard, D. H., and Hurlbert, A. C. (2019). Illumination discrimination for chromatically biased illuminations: Implications for color constancy. *Journal of Vision*, 19(3):1–23.

Balas, B., Auen, A., Thrash, J., and Lammers, S. (2020). Children's use of local and global visual features for material perception. *Journal of vision*, 20(2):10.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

Bäuml, K.-H. (2001). Increments and decrements in color constancy. *Journal of the Optical Society of America A*, 18(10):2419.

Beau Lotto, R. (2004). Visual development: Experience puts the colour in life. *Current Biology*, 14(15):619–621.

Beck, J. (1966). Age differences in lightness perception. *Psychonomic Science*, 4(6):201–202.

Bejjanki, V. R., Randrup, E. R., and Aslin, R. N. (2020). Young children combine sensory cues with learned information in a statistically efficient manner: But task complexity matters. *Developmental Science*, 23(3):1–13.

Blackwell, K. T. and Buchsbaum, G. (1988). The effect of spatial and chromatic parameters on chromatic induction. *Color Research & Application*, 13(3):166–173.

Bloj, M. G., Kersten, D., and Hurlbert, A. C. (1999). Perception of three-dimensional shape influences colour perception through mutual illumination. *Nature*, 402(1991):877–879.

Bone, R., Landrum, J., Fernandez, L., and Tarsis, S. (1988). Analysis of the macular pigment by hplc: retinal distribution and age study. *Investigative ophthalmology & visual science*, 29(6):843–849.

Bosten, J. M., Beer, R., and MacLeod, D. I. (2015). What is white? *Journal of Vision*, 15(16):1–19.

Boyaci, H., Doerschner, K., and Maloney, L. T. (2004). Perceived surface color in binocularly viewed scenes with two light sources differing in chromaticity. *Journal of vision*, 4(9):664–679.

Boyaci, H., Doerschner, K., and Maloney, L. T. (2006). Cues to an equivalent lighting model. *Journal of vision*, 6(2):106–118.

Brainard, D. (1998). Color constancy in the nearly natural image. 2. Achromatic loci. *Journal of the Optical Society of America A*, 15(2):307.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10:433–436.

Brainard, D. H., Brunt, W. A., and Spiegle, J. M. (1997). Color constancy in the nearly natural image. 1. Asymmetric matches. *Journal of the Optical Society of America A*, 14(9):2091–2110.

Brainard, D. H. and Freeman, W. T. (1997). Bayesian color constancy. *Journal of the Optical Society of America A*, 14(7):1393.

Brainard, D. H., Longére, P., Delahunt, P. B., Freeman, W. T., Kraft, J. M., and Xiao, B. (2006). Bayesian model of human color constancy. *J Vis*, 6(11):1267–1281.

Brainard, D. H., Pelli, D. G., and Robson, T. (2002). Display characterization. In Hornak, J. P., editor, *Encyclopedia of Imaging Science and Technology*, pages 172–188. Wiley.

Brill, M. H. (1995). The relation between the color of the illuminant and the color of the illuminated object. *Color Research & Application*, 20(1):70–76.

Brislin, R. W. and Leibowitz, H. W. (1970). The effect of separation between test and comparison objects on size constancy at various age-levels. *The American journal of psychology*, 83(3):372–376.

Brown, A. M. (1990). Development of visual sensitivity to light and color vision in human infants: A critical review. *Vision Research*, 30(8):1159–1188.

Buchsbaum, G. (1980). A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310(1):1–26.

Buldt, B. (2016). Johannes von kries: A bio-bibliography. *Journal for General Philosophy of Science*, 47(1):217–235.

Chambers, C., Sokhey, T., Gaebler-Spira, D., and Kording, K. P. (2018). The development of Bayesian integration in sensorimotor estimation. *Journal of Vision*, 18(12):1–16.

Chien, S. H. L., Bronson-Castain, K., Palmer, J., and Teller, D. Y. (2006). Lightness constancy in 4-month-old infants. *Vision Research*, 46(13):2139–2148.

CIE (1932). Commission internationale de l'eclairage proceedings. *Cambridge University, Cambridge.*

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

Dannemiller, J. L. (1989). A test of color constancy in 9- and 20-week-old human infants following simulated illuminant changes. *Developmental Psychology*, 25(2):171–184.

Dannemiller, J. L. and Hanko, S. A. (1987). A test of color constancy in 4-month-old human infants. *Journal of Experimental Child Psychology*, 44(2):255–267.

Day, R. H. and McKenzie, B. E. (1973). Perceptual shape constancy in early infancy. *Perception*, 2(3):315–320.

de Almeida, V. M. N., Fiadeiro, P. T., and Nascimento, S. M. C. (2010). Effect of scene dimensionality on colour constancy with real three-dimensional scenes and objects. *Perception*, 39(6):770–779.

Delahunt, P. B. and Brainard, D. H. (2004). Does human color constancy incorporate the statistical regularity of natural daylight? *Journal of Vision*, 4(2):1.

DiCarlo, J. M. and Wandell, B. A. (2000). Illuminant estimation: Beyond the bases. In *Color and Imaging Conference*, pages 91–96. Society for Imaging Science and Technology.

Doerschner, K., Boyaci, H., and Maloney, L. T. (2004). Human observers compensate for secondary illumination originating in nearby chromatic surfaces. *Journal of Vision*, 4(2):3.

Doherty, M. J., Campbell, N. M., Tsuji, H., and Phillips, W. A. (2010). The ebbinghaus illusion deceives adults but not young children. *Developmental science*, 13(5):714–721.

D'Zmura, M. and Lennie, P. (1986). Mechanisms of color constancy. *Journal of the Optical Society of America A*, 3(10):1662.

Emmerson, P. and Ross, H. (1987). Variation in colour constancy with visual information in the underwater environment. *Acta psychologica*, 65(2):101–113.

Ernst, M. O. (2006). A Bayesian view on multimodal cue integration. *Human body perception from the inside out*, 131:105–131.

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

Fairchild, M. D. and Lennie, P. (1992). Chromatic adaptation to natural and incandescent illuminants. *Vision Research*, 32(11):2077–2085.

Faul, F., Ekroll, V., and Wendt, G. (2008). Color appearance: The limited role of chromatic surround variance in the "gamut expansion effect". *Journal of Vision*, 8(3):1–20.

Field, J. and Collins, J. K. (1977). A developmental study of the effects of instructions on visual shape judgements. *British Journal of Psychology*, 68(3):297–304.

Forsyth, D. and Zisserman, A. (1989). Mutual illumination. In *Proceedings CVPR'89: IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 466–473. IEEE.

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

Franklin, A., Pilling, M., and Davies, I. (2005). The nature of infant color categorization: Evidence from eye movements on a target detection task. *Journal of experimental child psychology*, 91(3):227–248.

Gerardin, P., Kourtzi, Z., and Mamassian, P. (2010). Prior knowledge of illumination for 3d perception in the human brain. *Proceedings of the National Academy of Sciences*, 107(37):16309–16314.

Gilchrist, A. and Jacobsen, A. (1984). Perception of lightness and illumination in a world of one reflectance. *Perception*, 13(1):5–19.

Gilchrist, A. L. (1979). The perception of Surface Black and Whites. *Scientific American*, 24(3):88–97.

Gomez, J., Natu, V., Jeska, B., Barnett, M., and Grill-Spector, K. (2018). Development differentially sculpts receptive fields across early and high-level human visual cortex. *Nature communications*, 9(1):1–12.

Granrud, C. E. (2009). Development of size constancy in children: A test of the metacognitive theory. *Attention, perception & psychophysics*, 71(3):644–654.

Granrud, C. E. and Schmechel, T. T. N. (2006). Development of size constancy in children: A test of the proximal mode sensitivity hypothesis. *Perception and Psychophysics*, 68(8):1372–1381.

Granzier, J., Vergne, R., and Gegenfurtner, K. (2014). The effects of surface gloss and roughness on color constancy for real 3-D objects. *Journal of vision*, 14(2):1–20.

Granzier, J. J. and Gegenfurtner, K. R. (2012). Effects of memory colour on colour constancy for unknown coloured objects. *i-Perception*, 3(3):190–215.

Gupta, G., Gross, N., Pastilha, R., and Hurlbert, A. (2020). The time course of colour constancy by achromatic adjustment in immersive illumination: What looks white under coloured lights? *BioRxiv*, pages 1–32.

Hansen, T., Olkkonen, M., Walter, S., and Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience*, 9(11):1367–1368.

Hedrich, M., Bloj, M., and Ruppertsberg, A. I. (2009). Color constancy improves for real 3d objects. *Journal of Vision*, 9(4):16–16.

Helson, H. (1938). Fundamental problems in color vision. I. The principle governing changes in hue, saturation and lightness of non-selective samples in chromatic illumination. *Journal of Experimental Psychology*, 23(5):439–476.

Hernández-Andrés, J., Lee, R. L., and Romero, J. (1999). Calculating correlated color temperatures across the entire gamut of daylight and skylight chromaticities. *Applied optics*, 38(27):5703–9.

Hollants-Gilhuijs, M. A., Ruijter, J. M., and Spekreijse, H. (1998). Visual half-field development in children: Detection of colour-contrast-defined forms. *Vision Research*, 38(5):645–649.

Humphrey, G. K., Goodale, M. A., Jakobson, L. S., and Servos, P. (1994). The role of surface information in object recognition: studies of a visual form agnosic and normal subjects. *Perception*, 23(12):1457–1481.

Hurlbert, A. and Wolf, K. (2004). Color contrast: a contributory mechanism to color constancy. *Progress in brain research*, 144:145–160.

Hurlbert, A. C. (1998). Computational models of colour constancy. *Perceptual Constancy. Why things look as they do*, pages 283–322.

Ishihara, S. (2006). *Ishihara's tests for colour deficiency.* Kanehara Trading Inc., Tokyo, Japan.

Johnson, S. P. and Aslin, R. N. (2000). Infants' perception of transparency. *Developmental Psychology*, 36(6):808.

Judd, D. B., MacAdam, D. L., Wyszecki, G., Budde, H. W., Condit, H. R., Henderson, S. T., and Simonds, J. L. (1964). Spectral Distribution of Typical Daylight as a Function of Correlated Color Temperature. *Journal of the Optical Society of America*, 54(8):1031.

Kaess, D. W., Haynes, S. D., Craig, M., Pearson, S., and Greenwell, J. (1974). Effect of distance and size of standard object on the development of shape constancy. *Journal of experimental psychology*, 102(1):17.

Káldy, Z. and Kovács, I. (2003). Visual context integration is not fully developed in 4-year-old children. *Perception*, 32(6):657–666.

Katz, D. (2013/1935). *The world of colour*, volume 44. Routledge.

Kavšek, M. (2011). Infants' responsiveness to lightness changes on a dynamic three-dimensional surface. *Vision Research*, 51(23-24):2398–2404.

Kavšek, M. and Granrud, C. E. (2012). Children's and adults' size estimates at near and far distances: A test of the perceptual learning theory of size constancy development. *i-Perception*, 3(7):459–466.

Kimura, A., Wada, Y., Yang, J., Otsuka, Y., Dan, I., Masuda, T., Kanazawa, S., and Yamaguchi, M. K. (2010). Infants' recognition of objects using canonical color. *Journal of Experimental Child Psychology*, 105(3):256–263.

Kleiner, M., Brainard, D., and Pelli, D. (2007). What's new in psychtoolbox-3? *Perception*, 36.

Klenberg, L., Korkman, M., and Lahti-Nuuttila, P. (2001). Differential development of attention and executive functions in 3-to 12-year-old finnish children. *Developmental neuropsychology*, 20(1):407–428.

Knill, D. C. and Richards, W. (1996). *Perception as Bayesian inference*. Cambridge University Press.

Kraft, J. M. and Brainard, D. H. (1999). Mechanisms of color constancy under nearly natural viewing. *Proceedings of the National Academy of Sciences of the United States of America*, 96(1):307–12.

Kraft, J. M., Maloney, S. I., and Brainard, D. H. (2002). Surface-illuminant ambiguity and color constancy: Effects of scene complexity and depth cues. *Perception*, 31(2):247–263.

Kuriki, I. (2006). The loci of achromatic points in a real environment under various illuminant chromaticities. *Vision Research*, 46(19):3055–3066.

Land, E. H. (1964). The retinex. *American Scientist*, 52(2):247–264.

Land, E. H. (1983). Recent advances in retinex theory and some implications for cortical computations: color vision and the natural image. *Proceedings of the National Academy of Sciences*, 80(16):5163–5169.

Land, E. H. (1986). An alternative technique for the computation of the designator in the retinex theory of color vision. *Proceedings of the National Academy of Sciences*, 83(10):3078–3080.

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

Landy, M. S., Maloney, L. T., Johnston, E. B., and Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision research*, 35(3):389–412.

Langer, M. S. and Bülthoff, H. H. (2001). A prior for global convexity in local shape-from-shading. *Perception*, 30(4):403–410.

Lee, D. and Plataniotis, K. N. (2014). A taxonomy of color constancy and invariance algorithm. In *Advances in Low-Level Color Image Processing*, pages 55–94. Springer.

Lee, H.-C. (1986). Method for computing the scene-illuminant chromaticity from specular highlights. *Journal of the Optical Society of America A*, 3(10):1694.

Lee, R. J. and Smithson, H. E. (2016). Low levels of specularity support operational color constancy, particularly when surface and illumination geometry can be inferred. *Journal of the Optical Society of America A*, 33(3):A306.

Linnell, K. J. and Foster, D. H. (2002). Scene articulation: Dependence of illuminant estimates on number of surfaces. *Perception*, 31(2):151–159.

Liu, Y., Gevers, T., and Li, X. (2015). Color constancy by combining low-mid-high level image cues. *Computer Vision and Image Understanding*, 140:1–8.

Logvinenko, A. D., Funt, B., Mirzaei, H., and Tokunaga, R. (2015). Rethinking colour constancy. *PLoS ONE*, 10(9):1–29.

Maloney, L. T. and Landy, M. S. (1989). A statistical framework for robust fusion of depth information. In *Visual communications and image processing IV*, volume 1199, pages 1154–1163. International Society for Optics and Photonics.

Meneghini, K. A. and Leibowitz, H. W. (1967). Effect of Stimulus Distance and Age on Shape Constancy. *Journal of Experimental Psychology*, 74(2 part 1):241–248.

Minolta, K. (2008). Spectroradiometer cs-2000.

Morrone, M. C., Burr, D., and Fiorentini, A. (1990). Development of contrast sensitivity and acuity of the infant colour system. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 242(1304):134–139.

Myambo, K. (1972). Shape constancy as influenced by culture, western education, and agel. *Journal of Cross-cultural Psychology*, 3(3):221–232.

Nagai, T., Kaneko, S., Kawashima, Y., and Yamauchi, Y. (2017). Do specular highlights and the daylight locus act as cues for estimating illumination color from a single object? *Optical Review*, 24(1):47–61.

Nascimento, S. M., Amano, K., and Foster, D. H. (2016). Spatial distributions of local illumination color in natural scenes. *Vision Research*, 120:39–44.

Olkkonen, M., Hansen, T., and Gegenfurtner, K. R. (2009). Categorical color constancy for simulated surfaces. *Journal of vision*, 9(12):6–6.

Olkkonen, M., Saarela, T. P., and Allred, S. R. (2016). Perception-memory interactions reveal a computational strategy for perceptual constancy. *Journal of Vision*, 16(3):1–21.

Olkkonen, M., Witzel, C., Hansen, T., and Gegenfurtner, K. R. (2010). Categorical color constancy for real surfaces. *Journal of vision*, 10(9):16.

Parkkinen, J. P. S., Box, P. O., and Kuopio, S. (1989). Characteristic spectra of Munsell colors. *JOSA A*, 6(2):318–322.

Pearce, B., Crichton, S., Mackiewicz, M., Finlayson, G. D., and Hurlbert, A. (2014). Chromatic Illumination Discrimination Ability Reveals that Human Colour Constancy Is Optimised for Blue Daylight Illuminations. *PLoS ONE*, 9(2).

Peeples, D. and Teller, D. Y. (1975). Color vision and brightness discrimination in two-month-old human infants. *Science*, 189(4208):1102–1103.

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10:437–442.

Pickard-Jones, B., D'Avossa, G., and Sapir, A. (2020). 3D shape-from-shading relies on a light source prior that does not change with age. *Vision Research*, 177:88–96.

Poirel, N., Mellet, E., Houdé, O., and Pineau, A. (2008). First came the trees, then the forest: Developmental changes during childhood in the processing of visual local-global patterns according to the meaningfulness of the stimuli. *Developmental psychology*, 44(1):245.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Radonjić, A. and 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., and Brainard, D. H. (2015a). Color constancy in a naturalistic goal-directed task. *Journal of Vision*, 15(13):3.1–3.21.

Radonjić, A., Cottaris, N. P., and Brainard, D. H. (2015b). Color constancy supports cross-illumination color selection. *Journal of Vision*, 15(6):13.

Radonjić, A., Ding, X., Krieger, A., Aston, S., Hurlbert, A. C., and Brainard, D. H. (2018). Illumination discrimination in the absence of a fixed surface reflectance layout. *Journal of Vision*, 18(5):1–27.

Radonjić, A., Pearce, B., Aston, S., Krieger, A., Dubin, H., Cottaris, N. P., Brainard, D. H., and Hurlbert, A. C. (2016). Illumination discrimination in real and simulated scenes. *Journal of Vision*, 16(11):1–18.

Rapoport, J. (1967). Attitude and Size Judgment in School Age Children. *Child Development*, 38(4):1187–1192.

Rogers, M., Witzel, C., Rhodes, P., and Franklin, A. (2020). Color Constancy and Color Term Knowledge are Positively Related in Early Childhood. *Journal of Experimental Child Psychology*.

Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P., and Posner, M. I. (2004). Development of attentional networks in childhood. *Neuropsychologia*, 42(8):1029–1040.

Rushton, W. A. (1972). Review Lecture. Pigments and signals in colour vision. *The Journal of Physiology*, 220(3):1–31.

Saarela, T. (2018). Shapetoolbox: Creating 3d models for vision research. *Journal of Vision*, 18(10):229–229.

Seriès, P. and Seitz, A. (2013). Learning what to expect (in visual perception). *Frontiers in human neuroscience*, 7:668.

Shepp, B. E. and Barrett, S. E. (1991). The development of perceived structure and attention: Evidence from divided and selective attention tasks. *Journal of experimental child psychology*, 51(3):434–458.

Shevell, S. K. and Wei, J. (1998). Chromatic induction: Border contrast or adaptation to surrounding light? *Vision Research*, 38(11):1561–1566.

Smithson, H. and Zaidi, Q. (2004). Colour constancy in context: Roles for local adaptation and levels of reference. *Journal of Vision*, 4(9):693–710.

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.

Snyder, J. L., Doerschner, K., and Maloney, L. T. (2005). Illumination estimation in three-dimensional scenes with and without specular cues. *Journal of vision*, 5(10):863–877.

Sohoglu, E., Peelle, J. E., Carlyon, R. P., and Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *Journal of Neuroscience*, 32(25):8443–8453.

Speigle, J. M. and Brainard, D. H. (1999). Predicting color from gray: the relationship between achromatic adjustment and asymmetric matching. *Journal of the Optical Society of America. A, Optics, image science, and vision*, 16(10):2370–6.

Spitschan, M., Aguirre, G. K., Brainard, D. H., and Sweeney, A. M. (2017). Erratum: Variation of outdoor illumination as a function of solar elevation and light pollution. *Scientific reports*, 6(1):1–14.

Stone, J. V. (2011). Footprints sticking out of the sand. Part 2: Children's Bayesian priors for shape and lighting direction. *Perception*, 40(2):175–190.

Sugita, Y. (2004). Experience in Early Infancy Is Indispensable for Color Perception. *Current Biology*, 14:1267–1271.

Tanaka, J. W. and Presnell, L. M. (1999). Color diagnosticity in object recognition. *Perception and Psychophysics*, 61(6):1140–1153.

The Math Works, Inc. (2018). *MATLAB version 9.5.0,878302 (R2018b)*. Natick, Massachusetts.

Thomas, R., Nardini, M., and Mareschal, D. (2010). Interactions between "light-from-above" and convexity priors in visual development. *Journal of Vision*, 10(8):1–7.

Thouless, R. H. (1931). Phenomenal regression to the real object. *British journal of psychology*, 21(4):339.

Tronick, E. and Hershenson, M. (1979). Size-distance perception in preschool children. *Journal of Experimental Child Psychology*, 27(1):166–184.

Valberg, A. and Lange-Malecki, B. (1990). "Colour constancy" in Mondrian patterns: A partial cancellation of physical chromaticity shifts by simultaneous contrast. *Vision Research*, 30(3):371–380.

Van Beers, R. J., Sittig, A. C., and Gon, J. J. D. v. d. (1999). Integration of proprioceptive and visual position-information: An experimentally supported model. *Journal of neurophysiology*, 81(3):1355–1364.

Van De Weijer, J., Schmid, C., and Verbeek, J. (2007). Using high-level visual information for color constancy. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8. IEEE.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer, New York, fourth edition. ISBN 0-387-95457-0.

Von Kries, J. and MacAdam, D. (1878). Physiology of visual sensations. *Sources of Color Science*, pages 101–108.

Wagner, H. J. and Kröger, R. H. (2005). Adaptive plasticity during the development of colour vision. *Progress in Retinal and Eye Research*, 24(4):521–536.

Wang, Y., Zhang, J., Zou, J., Luo, H., and Ding, N. (2019). Prior knowledge guides speech segregation in human auditory cortex. *Cerebral Cortex*, 29(4):1561–1571.

Wedge-Roberts, R., Aston, S., Beierholm, U., Kentridge, R., Hurlbert, A., Nardini, M., and Olkkonen, M. (2020). Specular highlights improve color constancy when other cues are weakened. *Journal of vision*, 20(12):4.

Weiss, D., Witzel, C., and Gegenfurtner, K. (2017). Determinants of Colour Constancy and the Blue Bias. *i-Perception*, 8(6).

Werner, A. (2006). The influence of depth segmentation on colour constancy. *Perception*, 35(9):1171–1184.

Werner, A. (2007). Color constancy improves, when an object moves: High-level motion influences color perception. *J Vis*, 7(14):1–14.

Werner, A. (2014). Spatial and temporal aspects of chromatic adaptation and their functional significance for colour constancy. *Vision Research*, 104:80–89.

Werner, J. S. and Walraven, J. (1982). The effect of chromatic adaptation on the subjective neutral point; the role of luminance, contrast and background color. *Vision Research*, 22:929-944(1977):929–944.

Winkler, A. D., Spillmann, L., Werner, J. S., and 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., Flack, Z., and Franklin, A. (2013). Categorical colour constancy during colour term acquisition. In *AIC 2013 12th international AIC Congress, Newcastle upon Tyne, UK*.

Wurm, L. H., Legge, G. E., Isenberg, L. M., and Luebker, A. (1993). Color Improves Object Recognition in Normal and Low Vision. *Journal of Experimental Psychology: Human Perception and Performance*, 19(4):899–911.

Wyszecki, G. and Stiles, W. (1982). *Color science: concepts and methods, quantitative data and formulae*. Wiley classics library. Wiley.

Xiao, B. and Brainard, D. H. (2008). Surface gloss and color perception of 3D objects. *Visual Neuroscience*, 25(3):371–385.

Xiao, B., Hurst, B., MacIntyre, L., and Brainard, D. H. (2012). The color constancy of three-dimensional objects. *Journal of Vision*, 12:1–15.

Yang, J., Kanazawa, S., and Yamaguchi, M. K. (2010). Perception of Munker-White illusion in 4-8-month-old infants. *Infant Behavior and Development*, 33(4):589–595.

Yang, J., Kanazawa, S., and Yamaguchi, M. K. (2013a). Can Infants Tell the Difference between Gold and Yellow? *PLoS ONE*, 8(6):1–5.

Yang, J., Kanazawa, S., Yamaguchi, M. K., and Kuriki, I. (2013b). Investigation of color constancy in 4.5-month-old infants under a strict control of luminance contrast for individual participants. *Journal of Experimental Child Psychology*, 115(1):126–136.

Yang, J., Kanazawa, S., Yamaguchi, M. K., and Motoyoshi, I. (2015). Pre-constancy Vision in Infants. *Current Biology*, 25(24):3209–3212.

Yang, J., Otsuka, Y., Kanazawa, S., Yamaguchi, M. K., and Motoyoshi, I. (2011). Perception of surface glossiness by infants aged 5 to 8 months. *Perception*, 40(12):1491–1502.

Yang, J. N. and Maloney, L. T. (2001). Illuminant cues in surface color perception: Tests of three candidate cues. *Vision Research*, 41(20):2581–2600.

Yang, J. N. and Shevell, S. K. (2002). Stereo disparity improves color constancy. *Vision Research*, 42(16):1979–1989.

Yang, J. N. and Shevell, S. K. (2003). Surface color perception under two illuminants: the second illuminant reduces color constancy. *Journal of vision*, 3(5):369–379.

Yip, A. W. and Sinha, P. (2002). Contribution of color to face recognition. *Perception*, 31(8):995–1003.

Yonas, A. and Hagen, M. (1973). Effects of static and motion parallax depth information on perception of size in children and adults. *Journal of Experimental Child Psychology*, 15(2):254–265.

Zaidi, Q. (2001). Color constancy in a rough world. *Color Research and Application*, 26(S1):S192–S200.