JUSTUS-LIEBIG-
UNIVERSITÄT
GIESSEN

Determinants of colour constancy

Inauguraldissertation

zur Erlangung des Doktorgrades der
Naturwissenschaften (Dr. rer. nat.) der

Justus-Liebig-Universität Gießen
Fachbereich 06., Psychologie und Sportwissenschaften
Otto-Behagel-Straße 10F, 35394 Gießen

vorgelegt am 04. April 2017 von

David Weiß

geboren am 26. Mai 1986
in Siegen

Dekan:

Prof. Dr. Dr. Jürgen Hennig (Gießen, Psychologie)

Gutachter:

Prof. Karl R. Gegenfurtner, Ph.D. (Gießen, Psychologie)
Prof. Dr. Reinhard Lakes-Harlan (Gießen, Biologie)

Mitglieder der Prüfungskommission:

Prof. Roland W. Fleming, Ph.D. (Gießen, Psychologie)
Prof. Dr. Jörn Munzert (Gießen, Sportwissenschaft)

Acknowledgments

First of all, I would like to thank my supervisor Prof. Karl Gegenfurtner for his general support of this whole work. Without his encouragement, motivation and patience, this whole project would not have reached this point. His large pool of ideas, expertise and experience was of outstanding value to my progress.

No less I want to thank my collaborator Prof. Marina Bloj from Bradford University, who supported me during two of the studies presented here. Her spreading enthusiasm, motivation and personal support were uplifting and encouraging to me.

Moreover, I want to especially thank Christoph Witzel for his collaboration in two studies of this work, guiding me with his experience, never shy to introduce another analysis to explain the data in even more detail (and adding yet another page to the manuscript), but also supporting me and motivating me during the final phase of this work. Thanks also go to my other collaborators: Florian Bayer, Vivian Paulun and Elisabeth Baumgartner.

I want to thank all the people from the department in Gießen, especially the members of the colour group for their discussions, ideas, critique and comments on this work during the last four years: Florian Schiller, Zarko Milovic, the Italian guys Matteo Valsecchi and Matteo Toskani, Robert Ennis, Elisabeth Baumgartner, Thorsten Hansen, Alban Flachot, Arash Akbarina, Katja Doerschner and Hussein Boyaci. All of them enriched our weekly meetings with humour and their wide spread knowledge and understanding of the field of psychophysics, what I'm very grateful of. Special thanks go to Robert Ennis and Yaniv Morgenstern for helping with their native language skills and proof readings.

Especially I want to thank my family for always supporting me and backing me up. They believed in me and encouraged me when I lost hope. Most importantly, I want to thank Katja Giedigkeit, who gave me emotional support during the final year of my work, who lifted me up and believed in my abilities.

Abstract

Colour constancy describes the ability of our visual system to keep colour percepts stable through illumination changes. This is an outstanding feat given that in the retinal image surface and illuminant properties are conflated. Still, in our everyday lives we are able attribute stable colour-labels to objects to make communication economic and efficient. Past research shows colour constancy to be imperfect, compensating for 40% and 80% of the illumination change. While different constancy determinants are suggested, no carefully controlled study shows perfect constancy.

The first study presented here addresses the issue of imperfect constancy by investigating colour constancy in a cue rich environment, using a task that resembles our everyday experience with colours. Participants were asked to recall the colour of unique personal objects in natural environment under four chromatic illuminations. This approach yielded perfect colour constancy.

The second study investigated the relation between illumination discrimination and chromatic detection. Recent studies using an illumination discrimination paradigm suggest that colour constancy is optimized for bluish daylight illuminations. Because it is not clear if illumination discrimination is directly related to colour constancy or is instead explained by sensitivity to changes in chromaticity of different hues, thresholds for illumination discrimination and chromatic detection for the same 12 illumination hues were compared. While the reported blue bias could be replicated, thresholds for illumination discrimination and chromatic detection were highly related, indicating that lower sensibility towards bluish hues is not exclusive to illumination discrimination.

Accompanying the second study, the third study investigated the distribution of colour constancy for 40 chromatic illuminations of different hue using achromatic adjustments and colour naming. These measurements were compared to several determinants of colour constancy, including the daylight locus, colour categories, illumination discrimination, chromatic detection, relational colour constancy and metameric mismatching. In accordance with the observations in study 2, achromatic adjustments revealed a bias towards bluish daylight illumination. This blue bias and naming consistency explained most of the variance in achromatic adjustments, while illumination discrimination was not directly related to colour constancy.

The fourth study examined colour memory biases. Past research shows that colours of objects are remembered as being more saturated than they are perceived. These works often used natural objects that exist in a variety of colour and hue, such as grass or bananas. The approach presented here directly compared perceived and memorized colours for unique objects, used also in the first study, and confirmed the previous findings that on average, objects were remembered more saturated than they were perceived.

Zusammenfassung

Farbkonstanz beschreibt die Fähigkeit unseres visuellen Systems Farbeindrücke unter Beleuchtungsänderungen beständig zu halten. Dies ist eine außergewöhnliche Leistung, wenn man in Betracht zieht, dass in dem Lichtsignal welches das Auge erreicht Eigenschaften der Beleuchtung und der Oberflächen konfundiert sind. Trotz dieser Problematik sind wir in unserem alltäglichen Leben in der Lage Objekten stabile Farbnamen zuzuordnen, und damit unsere Kommunikation effizient und ökonomisch zu gestalten. Bisherige Studien zur Farbkonstanz berichten jedoch, dass Farbkonstanz nicht perfekt ist, Beleuchtungswechsel wurden nur zwischen 40-80% kompensiert. Während unterschiedliche Determinanten der Farbkonstanz vorgeschlagen wurden, konnte bisher keine sorgfältig kontrollierte Studie perfekte Farbkonstanz zeigen.

In der ersten Studie dieser Arbeit wurde dieser Aspekt untersucht, indem Farbkonstanz in einer hinweisreichen Umgebung unter Verwendung einer Aufgabe, die möglichst präzise unserer alltäglichen Erfahrung im Umgang mit Farben wiederspiegelt, gemessen wurde. Die Versuchsteilnehmer wurden aufgefordert die Farbe eines spezifischen persönlichen Gegenstandes unter vier farbigen Beleuchtungen aus dem Gedächtnis abzurufen. Unter Verwendung dieses Ansatzes konnte perfekte Farbkonstanz erreicht werden.

Die zweite Studie untersuchte die Beziehung zwischen Beleuchtungs-Diskrimination und chromatischer Detektion. Die Ergebnisse von kürzlich veröffentlichten Forschungsarbeiten, welche ein Beleuchtungs-Diskriminations-Paradigma verwendeten, zeigen das diese Diskrimination in Richtung bläulicher Beleuchtung verzerrt ist. Daraus wurde geschlossen, das Farbkonstanz für bläuliche Tageslicht-Beleuchtungen optimiert ist . Da es aber nicht klar ist, ob Beleuchtungs-Diskrimination in direkter Beziehung zur Farbkonstanz steht, oder aber vielmehr auf die Sensitivität für chromatische Veränderungen zurückführen ist, wurden Wahrnehmungsschwellen für Beleuchtungs-Diskrimination und chromatische Detektion für die selben 12 Beleuchtungsfarben gemessen und verglichen. Während die bereits berichtete Verzerrung in Richtung der bläulichen Tageslichtbeleuchtung repliziert werden konnte, wurde ebenfalls eine hoher Zusammenhang zwischen chromatischer Detektion und Beleuchtungs-Diskrimination gefunden, welcher darauf hinweist, dass die Verzerrung in Richtung bläulicher Farben keine exklusive Eigenschaft der Beleuchtung-Diskrimination ist.

Anknüpfend an die zweite Studie wurde in der dritten Studie die Verteilung von Farbkonstanz über 40 chromatische Beleuchtungen anhand von achromatischen Einstellungen und Farbbenennung untersucht. Farbkonstanz wurde auf ihren Zusammenhang zu mehreren Determinanten der Farbkonstanz überprüft, unter anderem mit Tageslichtvariationen, Farbkategorien, Beleuchtungs-Diskrimination, relationaler Farbkonstanz und metameric mismatching. In Übereinstimmung mit der zweiten Studie

wurde auch für achromatische Einstellungen eine Verzerrung in Richtung bläulicher Tageslichtbeleuchtungen gefunden. Diese Verzerrung und der Konsensus der Beleuchtungsbenennung erklärten den Großteil der Varianz der achromatischen Einstellungen, während Beleuchtungs-Diskrimination nicht in direkter Verbindung zur Farbkonstanz stand.

In der vierten Studie wurden Verzerrungen des Farbgedächtnisses untersucht. Frühere Studien berichten, dass Objektfarben häufig gesättigter erinnert werden als sie tatsächlich wahrgenommen werden. In diesen Studien wurden häufig natürliche Objekte verwendet, die in einer Vielzahl an Farbtönen und Sättigungen existieren, wie beispielsweise Gras oder Bananen. In dem hier präsentierten Ansatz wurden Farbwahlen aus dem Gedächtnis mit Farbwahlen der direkten Objektwahrnehmung für persönliche, spezifische Objekte, die auch schon in der ersten Studie verwendet wurden, verglichen. Die Ergebnisse der vorherigen Studien konnten für diese Objekte repliziert werden: Im Durchschnitt wurden Objektfarben gesättigter erinnert als das Objekt im direkten Vergleich wahrgenommen wurde.

Table of contents

Chapter

1 Introduction

One of the things that make our lives easier every day is the fact that we can attribute stable colours to objects. We take it for granted that an apple appears the same way in the supermarket where we buy it as it appears at home where we probably eat it. This phenomenon is in fact an outstanding performance of our brain: The light reflected from an object to the eye is always a combination of surface and illuminant properties and thus irreversibly confounds the reflectance distribution of the object surface and the spectral distribution of the illuminant. If we measure the light reflected from an apple under two different illuminations, similar to the situation described above, we find very different wavelength distributions (see **Figure 1.1**). To judge the objects' colours, the human visual system just has this confounded information available, but it is still able to achieve a constant object colour, despite the variable information coming from different illuminations. This phenomenon is known as "colour constancy", the ability of the human visual system to keep colour percepts stable through illuminant shifts (Arend & Reeves, 1986; Brainard, Brunt, & Speigle, 1997; Foster, 2011).

1.1 History

The work on colour constancy has quite a long history and it was already thought about in the late 18th century (Monge, 1789; Young, 1807). Monge (1789) as well as Young (1807) already realized that a chromatic illumination is not perceived as being chromatic or at least recognized less so, if the whole visual field is covered by that illumination. Both authors also point out that the percept of a chromatic illuminant on a neutral surface (white paper) is only really visible when a colour shadow is cast by another illuminant. In this case, parts of the surface appear in different colours according to the illuminants. In the case of just one illuminant, no matter of what colour, a white paper appears white.

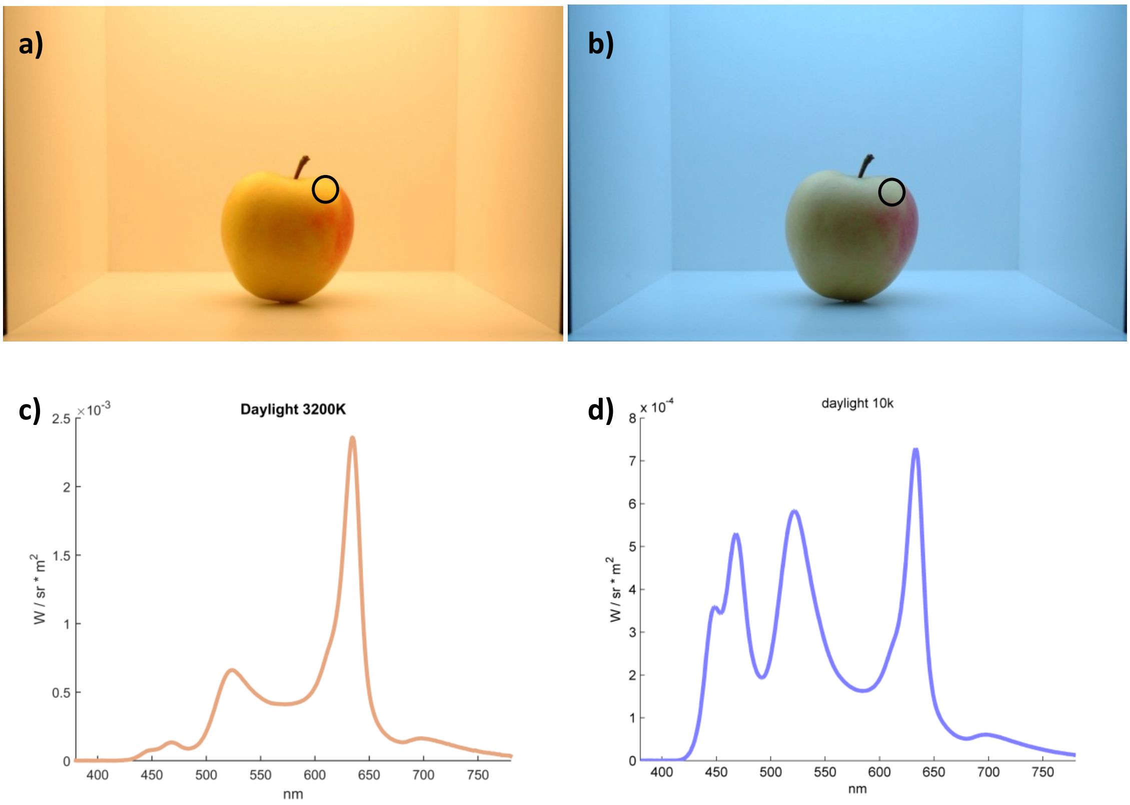


Figure 1.1: Pictures of the same apple taken under two illuminations that are from two points of the daylight locus. a) Daylight 3200K (evening, sunset). b) Daylight 10000K (cloudy day, dawn). c) & d) Spectra of surface measurements matching the illuminants above. The centre of the black circle in the photographs indicates the area from which the measurements have been obtained.

Helmholtz (von Helmholtz, 1867) formed the idea that the visual system achieves colour constancy by subtracting the illuminant from a viewed scene or object. Basing his theory about the constancy of object colours on memory and learning, his idea was that we gain knowledge of the appearance of an object under white light by seeing it under many different illuminations. He concluded that we use this knowledge to discount the illuminant from a given scene and reestablish stable object percepts.

A mathematical model of discounting the illuminant has been presented by von Kries (von Kries, 1902), known as "von Kries adaptation", by weighting of each of the three receptor classes to achieve "normal" object colours. It consists of a simple weight matrix for each receptor class, and works quite well for many illuminant changes, if one accepts negative weightings in some cases. Von Kries adaptation is still taken into account when constructing modern colour constancy algorithms.

Hering (Hering, 1920) proposed systems that correct for illuminant changes, as illuminant changes counteract the major task of the visual system: identifying objects. For most of our daily life, we are only really able to realize the impact of different illuminations when they are right next to each other or when they change rapidly. A striking demonstration of colour constancy is also given in Hering’s work: Two papers are selected (brown and blue) to appear the same, when placed on a prism that is seen through a tube from the top and are illuminated with differently coloured illuminants (blue daylight and orangeish light of an Edison candle in this case) on each side. But as soon as they are seen under just one of two illuminants, both papers again appear different and are easily identified as being of their real colour, independent of knowledge of the "real" colours of the papers. He explains that the human retina is able to change its sensitivity depending on the light source to correct for changes in the illumination. He proposes two different mechanisms: Simultaneous contrast and chromatic adaptation to the global mean. Like Helmholtz, he also emphasizes the role of memory colours as an additional factor to gain constancy.

Judd (Judd, 1940) pointed out that the perceived colour of a surface is a function of the reflectance of a given surface and the reflectance of the field in which it is placed, similar to a notion of Helmholtz (1876) that if a given scene is dominated by a certain hue, a pure white (i.e., an object which reflects all wavelengths equally) will appear as the complementary colour of the dominant hue.

Building a model for colour constancy, Land & McCann (Land & McCann, 1971) proposed a rather successful algorithm to achieve colour constancy for simple scenes, known as the "retinex algorithm", but Brainard & Wandell (Brainard & Wandell, 1986) could show that the algorithm is too dependent on the composition of the scene and does not work well for the wide variety of visual scenes we face every day.

1.2 Recent studies on Colour Constancy: Tasks and measurements

1.2.1 Asymmetric matching

One of the most important studies for the methodology of psychophysics to study colour constancy has been presented by Arend and Reeves (Arend & Reeves, 1986). They introduced the method of asymmetric matching, a paradigm where typically an observer sees two scenes (or one scene divided into two parts) illuminated by two chromatically different illuminants next to each other simultaneously. The task in this paradigm is to match a surface under one of the illuminants, to have the same appearance as a certain surface under the other (reference) illuminant. Their results pointed out that task instructions have a major impact on the amount of constancy that is observed. In appearance based matches (hue-saturation matches), observers are asked to match the hue, saturation and brightness of a surface. The reported performance is in general quite poor for these kind of instructions (Arend, Reeves, Schirillo, & Goldstein, 1991; Arend & Reeves, 1986; Bäuml, 1999; Radonjić & Brainard, 2016; Troost & De Weert, 1991). For paper matches instead, where observers are asked to match a surface under the other illuminant as if it was cut from the same piece of paper as a surface under the reference illuminant, performance was can be quite high (Bäuml, 1999; Radonjić & Brainard, 2016). Since it was one of the first methods to measure colour constancy, asymmetric matching faces some shortcomings: A major problem is uncertainty for the observer if he sees the same set of surfaces under different illuminations (left and right) or if the set of surfaces has also changed, what might lead to impoverished constancy. Another problem of asymmetric matching is partial and incomplete adaptation to either illuminant, as the observer is never fully immersed within a single illuminant.

1.2.2 Colour naming and successive asymmetric colour memory matching

Troost and de Weert (Troost & de Weert, 1991) came up with an alternative way to study the amount of colour constancy that observers achieve: colour naming. They replicated the study of Arend and Reeves (1986) and additionally asked observers to identify chips under both illumination conditions by naming a colour patch. They found that this method lead to reliable measurements of colour constancy. The major shortcoming of colour naming itself is, that it is not very precise, given the fact that depending on the language, an observer uses an average of just 4-6 basic categories (Berlin & Kay, 1969). Even the most versatile repertoire of colour names is in general pretty diminished, compared to the amount of colours the human visual system can distinguish.

In addition to the simultaneous presentation of illuminated scenes, they also used a successive presentation and found that observers tended to overestimate the illuminant shift instead. If a delay between presentation and matching is present, this measure is known as successive asymmetric colour memory matching (Ling & Hurlbert, 2008; Murray, Daugirdiene, Vaitkevicius, Kulikowski, & Stanikunas, 2006; Uchikawa, Kuriki, & Tone, 1998). Typically, an observer is asked to memorize a certain object colour under one illuminant, and then to recall it under another illuminant later. The advantage of these designs is the fact that adaptation is not disrupted, and in comparison to achromatic adjustments (see below; Brainard, 1998), it is possible to do research on all possible surface colours.

On the other hand, studies on colour memory come with a special problem: Colour memory effects. Colour memory effects or biases are known to occur for a variety of everyday objects like fruit, plants and materials(Pérez-Carpinell, Baldoví, de Fez, & Castro, 1998; Siple & Springer, 1983), and were often tested for objects which come in a variety of hues themselves ("the green of grass" or "the yellow of a banana”). Object and surface colours reproduced from memory are reported to be more saturated(Newhall, Burnham, & Clark, 1957; Siple & Springer, 1983) than the surfaces or objects they refer to are perceived. Ling and Hurlbert (2008) suggested constancy indices to control for those effects, but those again might lead to a overestimation of colour constancy and because of that are not really comparable to established measures for colour constancy.

1.2.3 Relational and operational colour constancy

A different approach to examine colour constancy is to look at the relations between surfaces in a given scene under illuminant changes on the basis of cone ratios, which is known as relational colour constancy (Foster & Nascimento, 1994; Nascimento & Foster, 1997). While relational colour constancy does not directly provide a measure for constancy, it can be used in operational colour constancy tasks (Craven & Foster, 1992) which aim to examine if an observer is able to detect if the illumination of a given scene has changed or if a surface has changed. For example, observers are asked if a surface remained the same in accordance with a given scene across an illuminant change, or if the surface itself changed relative to the rest of the scene.

1.2.4 Achromatic adjustment

A further measure of colour constancy was introduced by Brainard (1998). Making use of the fact that pure achromatic surfaces perfectly reflect the characteristics of the illuminant, he set up a task in which observers were asked to perform achromatic adjustments under different illuminations. To do so, observers were able to freely control the colour of a certain area in the scene. The resulting matches are informative about the amount of correction and adaptation that the visual system achieves due to an illuminant shift. This paradigm overcomes most of the shortcomings of asymmetric matching, but is only informative for achromatic surfaces; it is not directly possible to measure constancy for a collection of surfaces with different reflectances.

Using such an achromatic adjustment task, Kraft and Brainard (1999) were able to demonstrate three mechanisms (which already had been annotated by Helmholtz (1876), Hering (1920) and Judd (1940)) that the visual system uses to achieve colour constancy, by selectively silencing each mechanism in a well designed experiment. These were adaptation to the local surround (Walraven, Benzschawel, Rogowitz, & Lucassen, 1991), adaptation to the spatial average of the image (Buchsbaum, 1980) and the area of maximal reflectance, also known under the term “the brightest is white” (McCann, McKee, & Taylor, 1976). Selectively silencing each of these mechanisms, observers achieved constancy indices between .83 (full cue condition) and .33 (all three mechanisms silenced).

1.3 Cues to colour constancy

Besides the mechanisms shown by Kraft and Brainard (1999), there are some other cues that are thought to be used by the visual system to achieve colour constancy.

1.3.1 Three-dimensional scenes and mutual illumination

It has been predicted (Funt, Drew, & Ho, 1991) and proven (Bloj, Kersten, & Hurlbert, 1999) that the visual system is able to detect and to correct for mutual illuminations in three-dimensional scenes. Later Hedrich, Bloj and Ruppertsberg (2009) showed that constancy improves for three dimensional objects and scenes. This result was challenged by de Almeida, Fiadeiro, & Nascimento (2010), who did not find any difference between scenes containing three dimensional objects and two dimensional objects. A later study by Xiao, Hurst, MacIntyre, & Brainard (2012) again makes a clear conclusion difficult. Their results showed slightly higher constancy for two-dimensional objects placed into a three-dimensional scene, which they explain by the fact that the luminance distribution on the two-dimensional object is less complex, making it is easier to match.

1.3.2 Colour-diagnostic objects

Others suggested that colour constancy is enhanced by adding real (known) objects to a scene (Granzier & Gegenfurtner, 2012; Witzel, Valkova, Hansen, & Gegenfurtner, 2011), like fruit with colour diagnostic properties, such as a banana. As pointed out above, Hering (1920) as well as Helmholtz (1876) already thought that objects with a well known colour might be used by the visual system to estimate the illumination of a scene. However, this claim has been questioned by others (Kanematsu & Brainard, 2014), who did not find a clear enhancement when colour diagnostic objects were added to scenes.

1.3.3 Specular highlights

Lee (1986) created an algorithm that made use of specular reflectances, which directly reflect the chromaticity of the illuminant, to access information about the illumination and achieve constancy this way. Yang and Maloney (2001) examined if observers make use of specular highlights and found an asymmetric influence of the kind of illuminant they used: If the scene was illuminated with D65 and specular highlight cues where in the direction of illuminant A, an observers' estimation of the illuminant was less influenced compared to a scene illuminated with illuminant A but with specular highlight cues in the direction of D65. The authors' explanation is that we know from experience to tell global illumination (from daylight) and highlights from artificial local illuminants (like traffic lights) apart. Xiao et al. (2012) found just a small increase in constancy by adding highlights to a test object.

Contrary, Granzier, Brenner, & Smeets (2009) could not find increased constancy in scenes including glossy objects with specular highlights compared to scenes containing matte objects only. According to their explanation, a possible problem with specular highlights is the fact that they are much brighter than the rest of the scene, and as such, it might be problematic to perceive the illuminant colour at all because the hues are washed out quite fast.

1.3.4 Visual field size

One of the factors that is crucial to improved constancy is the size of the visual field that is immersed within an illuminant, because it influences the possibility of complete and uniform adaptation. In a natural environment, the complete visual field of an observer is immersed within a certain illumination. An improvement due to this factor has been shown by Murray et al. (2006), as well as by Hansen, Walter and Gegenfurtner (2007). Murray et al. (2006) reported close to perfect colour constancy in a large field (120°) adaptation (60s) experiment, while a smaller field (20°) and shorter adaptation times led to impoverished constancy. In the full-field condition of Hansen et al. (2007), nearly the whole visual field was exposed to an illumination, and constancy in a categorization task was close to perfect. In a reduced cue condition, constancy was clearly diminished.

A collection of studies from 1986 to 2015 that provided information about the actual size of the scene and distance from the observer (or the visual angle of the stimulus), is shown in **Figure 1.2** b. As many experiments did not aim to achieve perfect constancy, but to study the effect of certain scene properties, there are still lower constancy values with large visual fields present. In general, a trend towards high constancy achieved under full visual field conditions is clearly visible.

1.4 Meta trend over 30 years

In the past literature, colour constancy is known to be imperfect. In general, reported degrees of colour constancy vary between 40 and 80 percent (Foster, 2011). While one could argue that humans don’t need to be 100% colour constant to be able to pick ripe fruit and to stay alive, this is counterintuitive to our experience. If different studies which measure colour constancy are compared, there are certain conditions which lead to higher constancy than others, so it is tempting to reach perfect constancy in experimental settings.

One of the major problems of research on colour constancy is the use of computer based experiments, and with that the use of a restricted area of viewing field and rigid viewing positions relative to a given scene. The second problem is the use of abstract stimuli and tasks, as experiments with more natural environments and more cues available lead to higher constancy (Kraft & Brainard, 1999; Kraft, Maloney, & Brainard, 2002), compared to abstract computer displays (Arend & Reeves, 1986). **Figure 1.2** panel a) shows constancy values reported in studies from 1986 until 2015, colour coded by type of experiment, while studies from 1986 to 2012 are taken from the review of Foster (2011).

Most of the really low values (values around .02) indicate experimental conditions in which experimenters aimed to silence certain cues to constancy or specifically looked at the influence of certain scene properties.

While for some studies, the experimental features leading to high constancy values are hard to point out (Ling & Hurlbert, 2008; Xiao et al., 2012), other studies that achieved high constancy values used certain properties in their setups: Large visual field size and sufficient adaptation time (Hansen et al., 2007; Murray et al., 2006), as well as the use of real scenes and objects including many cues to the illuminant (de Almeida, Fiadeiro, & Nascimento, 2004; Kraft & Brainard, 1999; Kraft et al., 2002)).

As pointed out above, the amount of observed constancy highly depends on the task and the instructions (appearance vs. paper matches, see Arend & Reeves (1986)), and the naturalness of the task seems to be crucial as well (Bramwell & Hurlbert, 1996; Radonjić, Cottaris, & Brainard, 2015a). Also, the technical evolution seems to play a role. While early computer monitor based experiments led to lower constancy values, as depicted in **Figure 1.2** panel a), higher constancy was achieved with simulated scenes from the middle of the 90s onward.

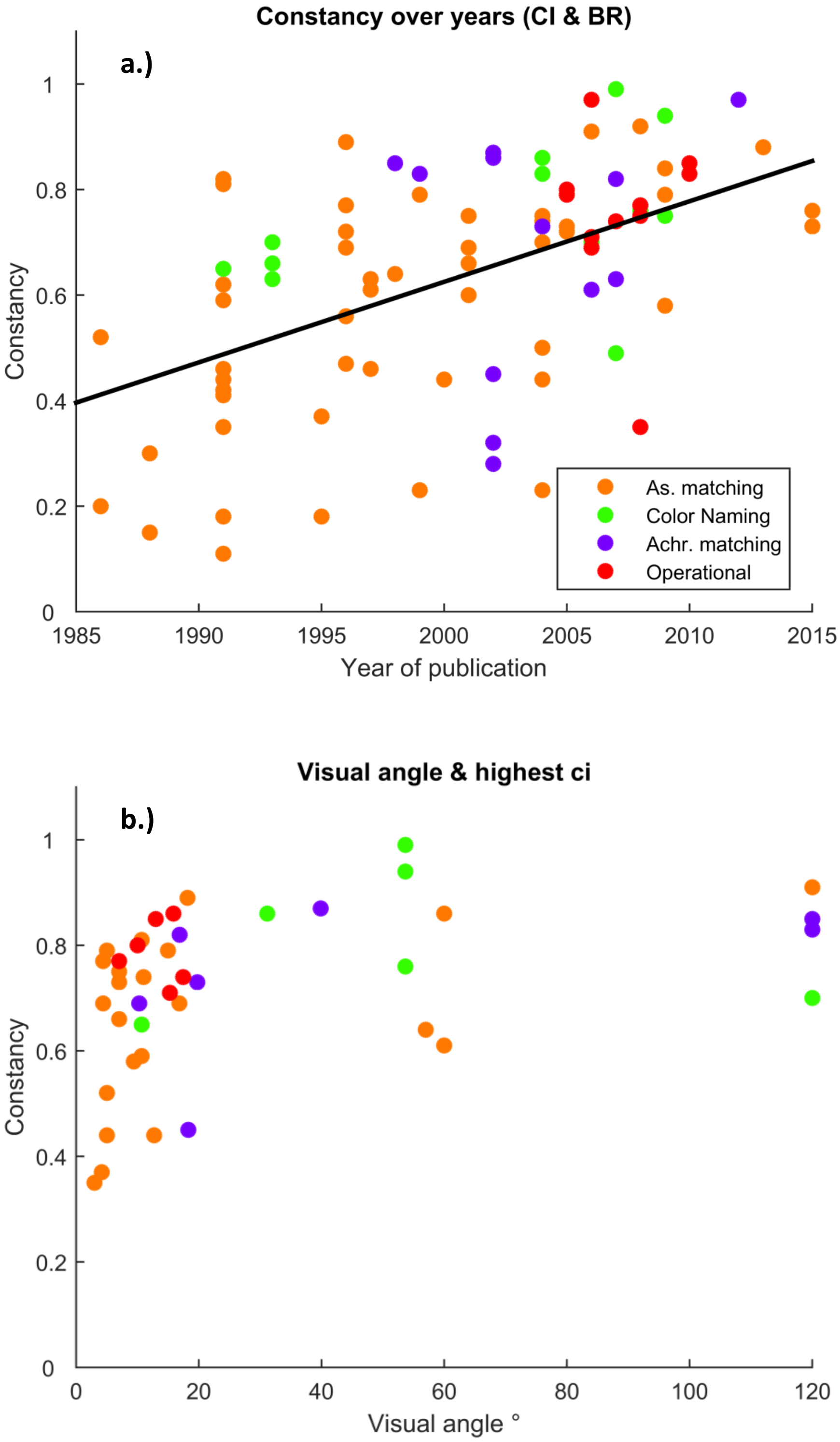


Figure 1.2: a.) Colour Constancy Indices reported in studies from 1986 to 2015. Studies until 2010 are shown as reported in Foster (2011). Orange dots indicate studies using asymmetric matching designs, green indicates colour naming and related designs, purple indicates achromatic matching and red indicates studies using relational colour constancy designs. The black line indicates a linear regression of the constancy indices. b.) Size of the visual field and constancy indices.

1.5 Constancy Indices

Figure 1.2 includes results from two different versions of constancy indices, which have been used in most of the studies since 1991. The Colour Constancy Index (CCI) as proposed by Arend et al. (1991), and the Brunswik ratio (BR) by Troost & de Weert (1991). The reasoning behind the two indices is to compare the magnitude of the illuminant shift with the match of the observer under this illuminant shift. Both express the amount of constancy shown by the observer, where a value of 0 indicates a failure of constancy and a value of 1 indicates perfect constancy. The Brunswik ratio directly compares distance of the illuminant shift between the coordinates of the neutral illuminant (A in **Figure 1.3**), (or the coordinates of an observer match under a neutral illuminant (A in **Figure 1.3**)) and the chromatic illuminant (or the coordinates of the surface under an illuminant change, B in **Figure 1.3**) with the distance between the neutral match of the observer and the colour match of the observer (see Equation 1) indicated by point C in **Figure 1.3**.

A common variation of the Brunswik ratio is to replace the distance between the neutral illuminant (or neutral match) and the match with the projection of that distance onto the vector between neutral match and test illuminant (see equation 1b in Olkkonen, Hansen, & Gegenfurtner, 2008). This is shown in **Figure 1.3** by the translucent red line between A and C'. A shortcoming of the Brunswik ratio is the assumption that all deviations in the adjustments are along the illumination shift, so it is sensitive to accuracy only. The Colour-Constancy Index (see equation 2) compares the distance of the illuminant shift (AB) with the distance between the adjustment error (distance between the chromatic illuminant and the match, (BC) in **Figure 1.3**). If the adjustment error is as large as the illuminant shift, complete absence of constancy is indicated by a ratio of one. To indicate constancy instead, this ratio is subtracted from one. Compared to the Brunswik ratio, an advantage of the Colour Constancy Index is that it is not based on the assumption that deviations in adjustments are along the illuminant shift. As it is sensitive to precision, the disadvantage is that it is not possible to reach perfect constancy (CCI = 1) with this index in empirical studies, which necessarily involve measurement noise.

Equation 1: $BR = \frac{|A-C|}{|A-B|}$

Equation 1b: $BR' = \frac{|A-C'|}{|A-B|}$

Equation 2: $CCI = 1 - \frac{|B-C|}{|A-B|}$

where A = Neutral illuminant or Neutral match; B = test illuminant; C = Match; C' = Projection of C to AB

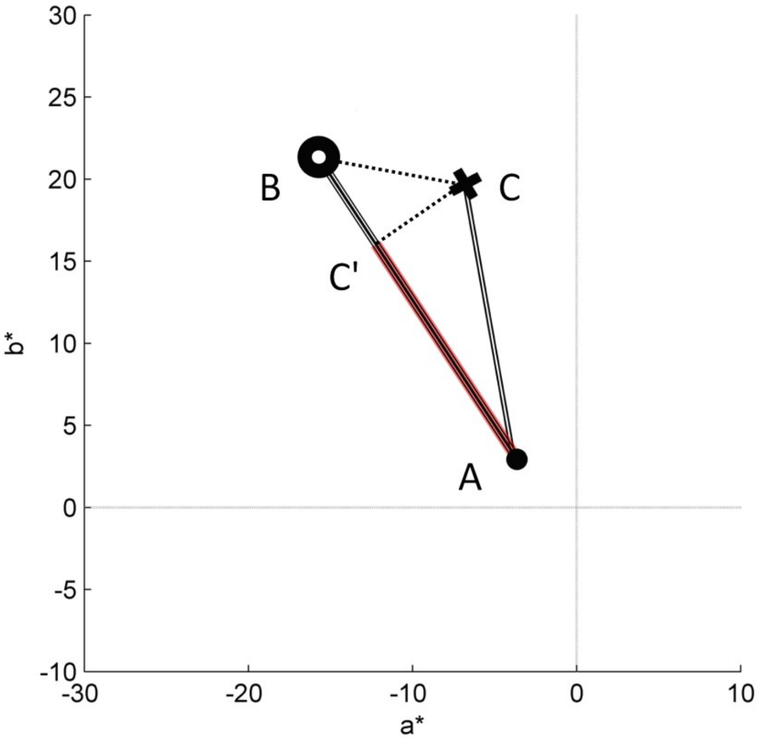


Figure 1.3: Constancy indices in the CIE-LAB colourspace. Point A represents the coordinates of the neutral illuminant, B represents the coordinates of the chromatic test illuminant. Point C indicates the match of the observer under the test-illuminant. AC′ indicates the vector projection of the distance between A and C onto the vector between A and B (shown in transparent red).

1.6 Colour categories vs. daylight regularities

1.6.1 Daylight regularities

As pointed out above, regularities in daylight changes are the most common illuminant change that we face throughout our lives. The light changes in colour over the course of the day, starting as dim blue at dawn, becoming orangeish yellow at sunrise, again turning to bluish hues around noon and again gets yellowish orangeish at sunset (DiCarlo & Wandell, 2000; Granzier & Valsecchi, 2014; Wyszecki & Stiles, 1982). Also, light in shadows is more bluish, as it contains light scattered from the blue sky (Churma, 1994; Troscianko, Benton, Lovell, Tolhurst, & Pizlo, 2009). Since our visual system has evolved under these illuminant changes, it's reasonable to assume that colour constancy is specialized to cope for illuminant shifts of these kinds. Typical changes due to daylight variations are depicted in **Figure 1.1** by illuminating an apple under orangeish and bluish daylight.

Recent studies by Pearce et al. (Pearce, Crichton, Mackiewicz, Finlayson, & Hurlbert, 2014), and Radonjić et al. (Radonjić et al., 2016), have highlighted the role of daylight illuminations, by showing that humans are worse in discriminating scenes illuminated with bluish daylight from scenes illuminated by a neutral light as compared to yellow, green and reddish lights. These studies define colour constancy as the inability to perceive the chromaticity of the illumination. It is still not clear how illumination discrimination is related to colour constancy, since surface colour percepts should still remain stable even if information about the chromaticity of the illumination is available. Murray et al. (2006) also reported higher constancy for the bluish daylight illuminant compared to illuminant A in their full field adaptation experiment. In a more recent study, Daugirdiene et al. (Daugirdiene, Kulikowski, Murray, & Kelly, 2016) reported elevated constancy for illuminants along the Planckian locus compared to illuminants orthogonal to the Planckian locus, with slightly higher constancy for bluish daylight hues.

Further asymmetries in constancy for different illuminant hues have been reported by Delahunt and Brainard (Delahunt & Brainard, 2004). Using an achromatic matching task, they found better constancy for bluish and greenish hues, while constancy for yellowish and reddish illuminants was diminished.

An earlier study did not find a difference for illuminations along the daylight locus compared to reddish and greenish illuminations, using real surfaces and illuminants in an achromatic matching task (Brainard, 1998). No differences for illumination hues were also reported by Hansen, Walter and Gegenfurtner (Hansen et al., 2007), for four full field illuminations that matched the cardinal axis of DKL-Space (Derrington, Krauskopf, & Lennie, 1984). Schultz, Doerschner, & Maloney (2006) also did not find differences between two daylight illuminants resembling blue and yellow, and two artificial illuminants resembling reddish and greenish illuminations. (Radonjić, Cottaris, & Brainard, 2015b) did not find differences between yellowish and bluish illuminant changes in a colour selection task as well as in an asymmetric matching task.

Others even found the opposite of Pearce et al. (2014), reporting instead that constancy decreased with rising correlated colour temperature, so constancy was worse for bluish compared to yellowish illuminations in an asymmetric matching task (de Almeida et al., 2004). It's important to note that in this study the fixed reference illumination was a bluish daylight illuminant (25K), and constancy increased with distance along the daylight locus to that illuminant, so it was highest at the yellowish daylight illuminant (4k).

More studies compared either different daylights or a single daylight illuminant to an artificial light, as the role of natural illuminants is a major topic for colour constancy. Using a successive colour constancy experiment, Ling and Hurlbert (2008) concluded that constancy for sets of surfaces varying in their saturation was slightly higher under yellowish and greenish illuminations, while sets varying in hue mainly showed better constancy under greenish and bluish illuminations.

A more recent phenomenon, a photograph known as #theDress, swarmed the field of colour constancy (Brainard & Hurlbert, 2015; Gegenfurtner, Bloj, & Toscani, 2015; Lafer-Sousa, Hermann, & Conway, 2015; Winkler, Spillmann, Werner, & Webster, 2015; Witzel, Racey, & O'Regan, 2017). The object of interest was a picture taken of a dress, which differs strongly in how it is perceived by different observers. One group sees the colours in the dress as being black and blue, while others see it as gold and white. Many of the published studies favor the connection of the colours in the Dress to daylight regularities as an explanation for the bistable interpretation of the picture by observers. Witzel et al. (2017) presented a sensible explanation for the phenomenon, by showing that placing the cut out picture into scenes that suggested a placement in shadows or in direct sunlight significantly changed the percept of the colours in the dress.

1.6.2 Colour categories

A possible alternative to the distribution of illuminant hues connected to daylight variations and colour constancy is the connection between language and colour perception. The topic of interest here is how we categorize different colours in language as being more or less equal to each other and give them colour names, known as colour categories. Research in this area aims to solve the question of how colour categories are formed, since low level visual features do not account for these categorizations: The hue favoured by colour opponency cells, so called second stage mechanisms, located in the the LGN (Derrington et al., 1984; Gegenfurtner, 2003) do not match the main colour categories and colour opponency we find in language.

Berlin and Kay (1969) hypothesized that basic colour terms (red, green, blue and yellow) are common in every language and share the same meaning, representing a cross-language match for colour categories. In their World Colour Survey (Berlin & Kay, 1969) they found that populations speaking over 110 different languages roughly categorize colours the same way, independent of their language. In psychophysics, there are also effects of colour categories. Regier, Kay and Khetarpal (2007) found that similarity judgments for hues were higher within categories compared to category boundaries. Witzel and Gegenfurtner (2013) measured discrimination thresholds between different hues and asked the observers to name the different hues to form colour categories. For some categories, they found that observers tend to show higher thresholds within the measured colour categories and lower at category borders, suggesting that the brain is especially precise when judging differences in physical input that is not semantically related. With respect to memory colours, Bae, Olkkonen, Allred and Flombaum (2015) reported that memory colour matches with and even without a delay tended to be shifted away from category boundaries, even if the recall was directly without a delay period.

For colour constancy, there are hardly any studies that investigated a relation to colour categories. For surfaces of different hues, Kulikowski & Vaitkevicius (1997), using Munsell chips of 40 different hues along the azimuth, showed that constancy peaked at category centres for red, yellow, green and blue, but not for violet hues under tungsten light or under a greenish illuminant. Recently, Daugirdiene et al. (2016) found a less specified pattern. Matches for 40 Munsell chips in two different saturations under an illuminant shift were placed into distinct groups that shifted clockwise or counterclockwise under illuminant changes. For an orange-yellowish Illuminant and bluish illuminant S matches were shifted in accordance with the Planckian locus, while matches under purplish illuminant P and greenish illuminant G were shifted orthogonal to the Planckian locus. Also, for the less saturated set of chips, constancy was higher for illuminant changes along the Planckian locus compared to illuminant changes orthogonal to that. Taken together, these findings suggest a relation towards daylight regularities instead of colour categories. If colour categories are formed by regularities in our natural surroundings, these could also have an impact on how colour constancy is achieved, leading to differences in constancy within categories and between categories, like the aforementioned studies suggest. Until now, no study has examined the connection between illuminant hues and colour constancy, since no study aiming for that has tested more than 4 different illuminant hues (five, if including the neutral reference illuminant), except for Brainard (1998). To test for 4 basic colour categories (red, green, blue & yellow), one needs to test at least eight different illuminations: One at each category centre and one for each category boundary.

1.7 Outlook

Drawn from the past studies described above, four studies will be presented here.

The first study presented here will try to enhance colour constancy to perfection. While previous studies, as pointed out above, reported colour constancy to be incomplete, some studies found scene and task features that enhance constancy. Here, evidence will be provided that colour constancy can be enhanced, if it is measured in a natural task in a cue-rich environment, using personal objects of observers in a memory matching paradigm.

A second study will examine the "blue bias" reported by Pearce et al. (2014). Since it is unclear if the illumination discrimination paradigm presented in their study really measures colour constancy, rather than expressing sensitivity to surface changes, the experiment will be replicated and compared to chromatic detection data for the same set of illumination hues.

The third study attempts to solve the mixed results of past studies regarding the distribution of colour constancy indices for different illuminant hues, one study that will be presented here will examine the distribution of colour constancy regarding illumination hues for 40 chromatic illuminations along the colour azimuth, and compare it to several determinants including colour categories, daylight regularities, metameric mismatching, sensory singularities and relational colour constancy.

The fourth study will examine colour memory effects for the objects used in the first study. Unique objects will be examined in a memory matching paradigm, and colour memory effects of the owners (long-term memory) will be compared to those of observers who see the objects just for a brief time (short term memory).

Chapter

2 Perfect colour constancy in real-world settings

A similar version of this manuscript has been submitted as:

Weiss, D., Bloj, M. & Gegenfurtner, K.R. (under review). Perfect colour constancy in real-world settings.

Colour constancy denotes the ability to assign a particular colour percept to an object, irrespective of its surroundings and illumination. The light reaching the eye confounds illumination and spectral reflectance of the object, making the recovery of constant object colour an ill-posed problem. How good the visual system is at solving this task is still a matter of heated debate, despite more than hundred years of research. Depending on the laboratory task and the specific cues available to observers, colour constancy was found to reach levels between 20% and 80%, which is incompatible with the relatively stable colour appearance of objects around us and the consistent usage of colour names in real life. Here we show that constancy is perfect using real objects in a natural task and environmental conditions. Our laboratory task and conditions were chosen to mimic the role of colour constancy in everyday life. Participants had to identify the colour of a (non-present) item familiar to them in a series of rooms under a variety of different experimental illuminations. In all conditions they mostly selected the same coloured chip as their match to the absent object, even though the light reaching the eye in each case differed considerably. Our results demonstrate that colour constancy in the real world can indeed be exceptionally good. We found it to be as good as visual term memory permits, and not generally compromised by sensory uncertainty. Previous measured insufficiencies can mainly be attributed to reduced laboratory settings and tasks.

2.1 Introduction

Perceptual constancies are the workhorse of our sensory abilities. In vision, the stimulation on the retina is extremely variable with respect to size, form, speed and wavelength. Yet, we do perceive a stable world where an object does not appear to change when we walk past it, for example, even though changes in distance, projection, eccentricity lead to a vastly different stimulation of our visual system. For many of these constancies, it is known that the compensation achieved by our visual system depends on the richness of the cues that are available, and that compensation is perfect under most natural settings, when numerous cues are available. This has been shown very elegantly for size constancy in classic experiments (Holway & Boring, 1941). Similar findings have emerged for other constancies (see (Troost, 1998)), with the notable exception of colour constancy. Some researchers argue that the problem of colour constancy is intractable mainly due to metamerism (Logvinenko, Funt, Mirzaei, & Tokunaga, 2015; Witzel et al., 2016), and initial empirical studies observed indeed rather low levels of compensation around 20% (Tiplitz Blackwell & Buchsbaum, 1988). More recent experiments have shown that as cues are added to the visual stimulus, colour constancy can increase to levels of about 80% (Kraft & Brainard, 1999). However, perfect constancy is elusive, which is at odds with our everyday experience.

In everyday life, we take it for granted that objects “have” a colour. From early childhood on we regularly use colour terms to describe objects (Bornstein, 1985; Franklin & Davies, 2004). We say that a shirt “is” green, for example, and not that the shirt “looks greenish under this particular lighting”. Our high, innate expectation of colour as a common and reliable object identifier became clear in February 2015 when what seemed to be the whole world become irritated when looking at the same picture but seeing the some object as different colours; the well-publicised blue-black/white-gold dress conundrum (Brainard & Hurlbert, 2015; Gegenfurtner et al., 2015; Lafer-Sousa et al., 2015; Winkler et al., 2015; Witzel et al., 2017).

Given the levels of colour constancy observed in previous experiments, and given the variability between different observers, one would expect these inconsistencies in colour naming across different observers to happen quite often, which does not seem to be the case. In order to be able to use colour terms consistently and persistently colours have to be remembered. However, solely remembering colours would not be enough to recognise our favourite shirt outside in bright sunlight and inside the office under artificial light. From a physical point of view the spectral distribution of the light that reaches our eyes from the shirt in these situations is very different as are the ensuing excitations of the cone photoreceptors in the retina and yet, indoors and outside, the shirt appears the same colour to us.

Our visual system seems able to compensate for such illumination and surround changes (Brainard & Radonjić, 2014; Foster, 2011; Hurlbert, 2007; Smithson, 2005) and enables us to perceive the environment as stable with respect to colour. In other words, we perceive objects as colour constant. **Figure 2.1** illustrates why this is a major achievement. The light reflected by a sample of the objects used in our study changes noticeably under daylight and the four other illuminants we employed. In each case the reflected wavelength composition and resulting cone excitations are very different. When viewed as isolated patches rather than full objects, the patches seem to change under the different illuminations and we would frequently even assign a different colour name to them. We investigated whether under these different conditions our participants were able to identify the same constant colour for their object.

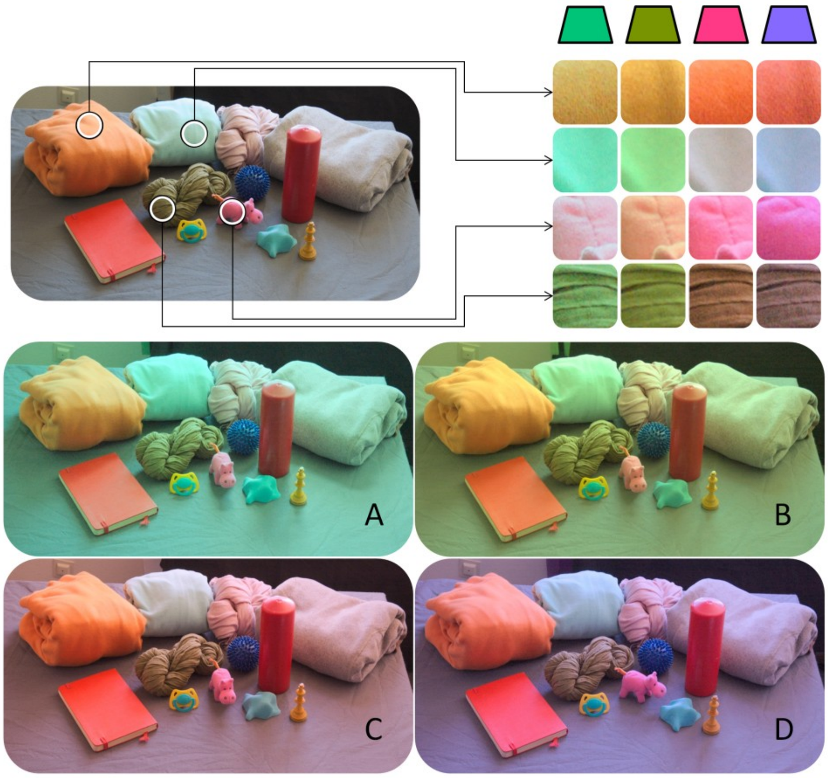


Figure 2.1: Personal objects brought for the study by participants shown under neutral daylight illumination in the top left panel. On the top right we show details of several objects under the four experimental illuminants to illustrate how different the reflected wavelength composition and cone excitations are under each illumination. In panels A to D we show the object collection under the different experimental illuminants.

2.2 Methods

2.2.1 Participants

16 subjects (5 females) participated in this study. All subjects were members of the Department of Psychology of the University of Gießen and provided informed consent before taking part. Experiments were performed in agreement with the Declaration of Helsinki and were approved by the local ethics committee (LEK 2013-0018). 12 participants were naïve to the purpose of the experiment. Mean age of the participants was 35 years, with a range between 26 and 61. All observers had normal or corrected to normal visual acuity and normal colour vision.

2.2.2 Materials and Methods

Subjects were asked to bring a personal object that is well known to them and that had a certain colour of which the participant was sure to have a good memory of. One subject provided 2 objects; another subject only performed matches under 2 out of the 4 filter conditions. Photographs of the 17 objects are shown in **Supplementary Figure S 2.1**.

In absence of the object (which had been taken away by the experimenter and had not since been seen by the participant) subjects selected a chip from the Munsell Book of Colours (Glossy Finish Collection) (Munsell Color, Baltimore,Md., 1976) arranged in 40 plastic bins by hue (1325 chips in total), as shown in **Supplementary Figure S 2.2** that best matched their recollection of the object colour.

Observers performed the chip selection task in rooms with white painted walls and grey floors filled with experimental equipment, office furniture and objects illuminated only by either neutral daylight or daylight modified by the use of one of four Lee filters that were attached in a inconspicuous way to the window in the room. Details of the filters used and their chromaticity are provided in **Supplementary Table S 2.1** and **Supplementary Figure S 2.3**, filters were obtained from Lee Filters (Lee Filters Ltd., Andover, UK). Before doing the chip selection, participants adapted for at least two minutes to the illuminant, while re-sorting the scrambled order of the bins by hue. This task allowed participants not only to adapt to the illuminant but also to become familiar with the Munsell chips. Each participant did two memory matches under neutral daylight and each filter, leading to 10 trials in total over separate days and without seeing their object. Before and after each participant completed their selection procedure, the ambient illumination was measured by using a RS-2 (51 mm diameter) PTFE reflectance standard (PhotoResearch Inc., Chatsworth, US) and a Konika Minolta CS2000 Spectroradiometer (Konica Minolta Sensing Inc., Singapore) and average chromaticities are reported in **Supplementary Table S 2.1** and **Supplementary Figure S 2.3**.

2.2.3 Data analysis

We chose to plot the chromaticity of selected chips in the approximately perceptually uniform colour space CIE_{1976} - L*a*b (Wyszecki & Stiles, 1982). For this we used measured reflectance of the selected Munsell chips provided by the University of Joensuu Spectral Colour Research Group (Orava, 2002) and average ambient illumination measured in our experimental rooms to calculate corresponding CIE-XYZ values using Judd-Vos (Judd, 1951; Vos, 1978) corrected colour matching functions, provided by the CVRL-database ("Colour & Vision Research Laboratory (CVRL)," 1995)and converted them to CIE 1976 - L*a*b* (Wyszecki & Stiles, 1982).

For some object/illumination combinations the chromaticity of the chip selected by the owner under daylight is no longer represented in the Munsell collection under a different illuminant. Strictly speaking, zero colour constancy cannot be obtained under these conditions, because any selected chip will be shifted towards the new illuminant (i.e. will tend towards constancy). In each of these cases we calculated for the corresponding object/illumination combination the lower bound of colour constancy using the chromaticity of the chip that under that illuminant would provide the lowest possible colour constancy index. The calculated lower bounds for constancy are shown in the bar charts in **Figure 2.4** as dark shaded regions and **Supplementary Figure S 2.4** illustrates this problem for a particular object. The problem emerges only for few object/illumination combinations and is not systematically related to the overall high level of constancy we obtained. All analysis has been done using Matlab 2012b (The MathWorks Inc., 2007).

2.2.4 Metameric mismatch volume

Metameric mismatch volumes were calculated using the approach suggested by Logvinenko et al. (Logvinenko, Funt, & Godau, 2014), as described recently (Witzel et al., 2016). For each reflectance of the chips selected by participants under neutral daylight, metameric mismatch volumes were calculated for the four illuminant changes from neutral daylight to the chromatic illumination in CIE-L*a*b.

2.2.5 Metamer analysis:

To calculate potential metamers of the 17 Munsell chips selected by participants in our study as memory matches under neutral illumination, we compared the spectral distribution of each of them to 11,302 surface spectra gathered from different online databases (Arnold, Savolainen, & Chittka, 2008; Barnard, Martin, Funt, & Coath, 2002; Berns, n.d.; Haanpalo, n.d.; Hiltunen, n.d.; Jaaskelainen, Silvennoinen, Hiltunen, & Parkkinen, 1994; Marszalec, n.d.; Matsumoto et al., 2014; Orava, 2002; Parkkinen, Jaaskelainen, & Kuittinen, 1988; Regan et al., 1998; Westland, Shaw, & Owens, 2000) including the Munsell Database (which incorporates all chips used in this study), leading to 192,117 possible comparisons. Each reflectance was rendered under each of the 5 illuminant spectra used in our experiment. The resulting spectral product was converted to CIE_{1931}-XYZ values using CIE_{1931}-2°-observer colour matching functions. To make the illuminants comparable for this analysis, the luminance according to an ideal reflector was normalized to Y = 1 before calculating the spectral product. Further, to be able to examine perceived differences between two surface reflectances, XYZ tristimulus values were converted to CIE-L*a*b using the whitepoint of the given illumination. Metamer pairs were defined using a threshold limit in perceived differences (CIE-ΔE2000 (M. R. Luo, Cui, & Rigg, 2001)) in CIE-L*a*b. If two surfaces were below that threshold under one illumination and above it under another illumination, they were regarded as metamers. Since there is no official agreement on how CIE-ΔE2000 is related to JNDs (just notable differences), we chose a threshold that was slightly above JNDs as defined by MacAdam-ellipses (CIE-ΔE2000 = .5 with a range of .23 to .89) of CIE-ΔE2000 = 1.5.

2.3 Results

We asked participants to bring along a personal object of colour they knew well (see **Figure 2.1** and **Supplementary Figure S 2.1**). Without having further sight of the object, participants were asked to select a coloured chip that best represented the colour of their object from a collection of 1325 chips, displayed in small plastic bins in an office environment illuminated only by natural daylight (see **Supplementary Figure S 2.2**). In later days, they had to repeat the selection in four other rooms each with a very different illumination (see Methods and **Supplementary Figure S 2.3** and **Supplementary Table S 2.1**). Again, the task was performed without seeing their object.

These measurements allowed us to calculate a constancy index (Olkkonen et al., 2008; Troost & de Weert, 1991) by comparing the colour of the chip selected as a memory match under daylight with the colour of the chip selected under each other illuminant. The colour constancy index indicates the degree of colour constancy achieved by an observer for a particular object, with 100% indicating that the same chip was selected under both illuminations, and 0% indicating that the selection of the chips was based on cone excitations only, without any compensation for the different illuminations. **Figure 2.2** illustrates this using as example the green scarf. For this object, under neutral daylight illuminant, the owner selected chip 5GY5/4 (represented by point A in **Figure 2.2**) as a match to their object. The matches, as well as those under the other illuminants, were made from memory – the actual object was not present. If this participant were fully colour constant, then she would also select chip 5GY5/4 under all four experimental illuminants (represented here, under the purplish illuminant by point B in **Figure 2.2**). If they were less constant they might select an alternative chip such as 2.5GY5/4 (represented by point C in **Figure 2.2**). We compute the corresponding colour constancy index by projecting vector AC onto AB (indicated by AC'). If C and B are identical then the index equals 1 (expressed as 100%), indicating perfect colour constancy. Zero constancy would instead be represented if the observer would pick a chip that has an equal colour signal under the purplish illuminant as the memory match under neutral daylight, in this case 5GY7/8 (represented by point D in **Figure 2.2**).

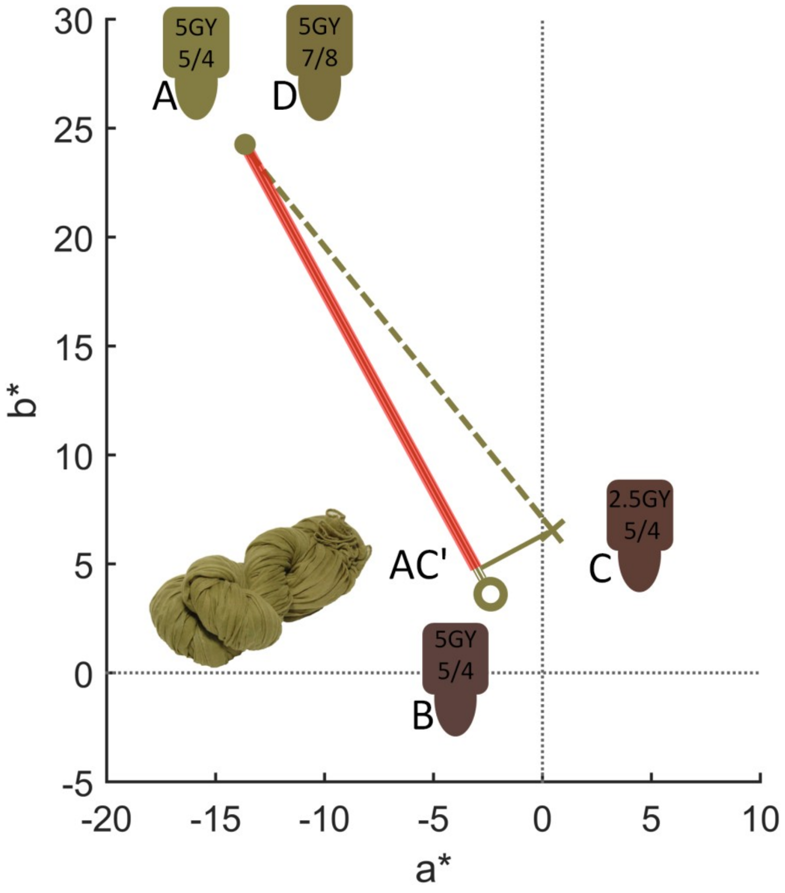


Figure 2.2: Representation in CIE (a*, b*) plane of the colour co-ordinates of the chip selected by the owner of the green scarf under neutral daylight illuminant (dot, labelled A), the coordinates of that same chip under one of the test illuminants, represented by the circle (B) and of an alternative chip under the same illuminant, the cross in (C). AC' , indicated by a red line gives the vector projection of AC onto AB. D represents a zero constancy choice for this object-illuminant combination.

Figure 2.3 shows the results for all objects under the four experimental illuminants, average of 2 sessions. For each object, there are two lines emerging from the memory colour selection under daylight. One line connects the colour coordinates of the selected chip under daylight to the colour coordinates of the same chip under the new illuminant (represented by circle). The second line extends to the colour coordinate of the chip the observer selected from memory under the new illuminant (represented by a cross). **Figure**

2.3 shows that for most objects these lines stay close together and frequently overlap. Constancy is close to perfect under all conditions, with a mean value of 93.9% (+/- 20% S.D.) and a median of 99.2% across all objects and illuminants. There was small variation between objects, as shown in **Figure 2.4**A.

Another way to evaluate constancy is to look at the differences between the perfect match and the selected chip. In CIE 1976 L*a*b* colour space, the Euclidean distance ΔE to a first approximation represents a perceptual just noticeable difference, based on measurements by MacAdam (Brainard, 2003). From a previous memory study (Bloj, Weiß, & Gegenfurtner, 2016) we know that the reliability of visual short-term memory for our participants and objects is of the order of one Munsell step in Hue, one and a half in Chroma and half a step in Value (see table 2 left hand column of (Bloj et al., 2016)). This roughly corresponds to the distance between 2 neighbouring Munsell chips, which is equivalent to 5.4 ΔE units for the 17 chips selected under neutral daylight in this study. The observed deviations between the selected chip and the chip representing perfect colour constancy were small and of the same order as this memory limit [mean ΔE: 7.3, median ΔE: 7.0 (+/- 5.6 S.D,+/- 1.4 S.E), as shown in **Figure 2.5**. In 23 out of 132 cases, participants even selected the very same chip under the test illuminant as under natural daylight. The bimodal shape of the histogram arises due to the discreteness of the collection of chips. The orange line in **Figure 2.5** models this effect and predicts the observers' choices based on random variations of about one Munsell Hue, one and a half in Chroma steps, and half a Value step, as we had observed in our short term visual memory (Bloj et al., 2016). This means that our observers were as good as their visual memory allowed them, with just a handful of exceptions.

One long-standing question in colour constancy research has been whether there are differences in the degree of constancy for different illumination conditions. In particular, the idea that the visual system has adapted to the type of illumination changes occurring naturally has been often tested with, so far, inconclusive results (Delahunt & Brainard, 2004; Hedrich et al., 2009). In our settings, the illumination changes along the bluish and yellowish directions are more similar to daylight changes than the reddish and greenish illumination changes. The results in **Figure 2.4**, right hand panel, show that there was little difference between the four illumination changes. There is a tendency for better constancy for the bluish and yellowish filters, but it is not statistically significant, and might be hidden by ceiling effects.

A repeated measure ANOVA showed a significant effect of the individual objects F(16,16)= 2.346, p=.049. We did find a few combinations of objects and illuminants that consistently led to relatively poor values of constancy, such as the red candle under the reddish illuminant and the light brown chess piece under the yellowish illuminant. In our previous study on memory colour for the objects used in this experiment (Bloj et al., 2016), we found highest variance in Chroma for the chess piece compared to all other objects even in object-present matches, so we suggest that the object itself is difficult to match due to its surface properties.

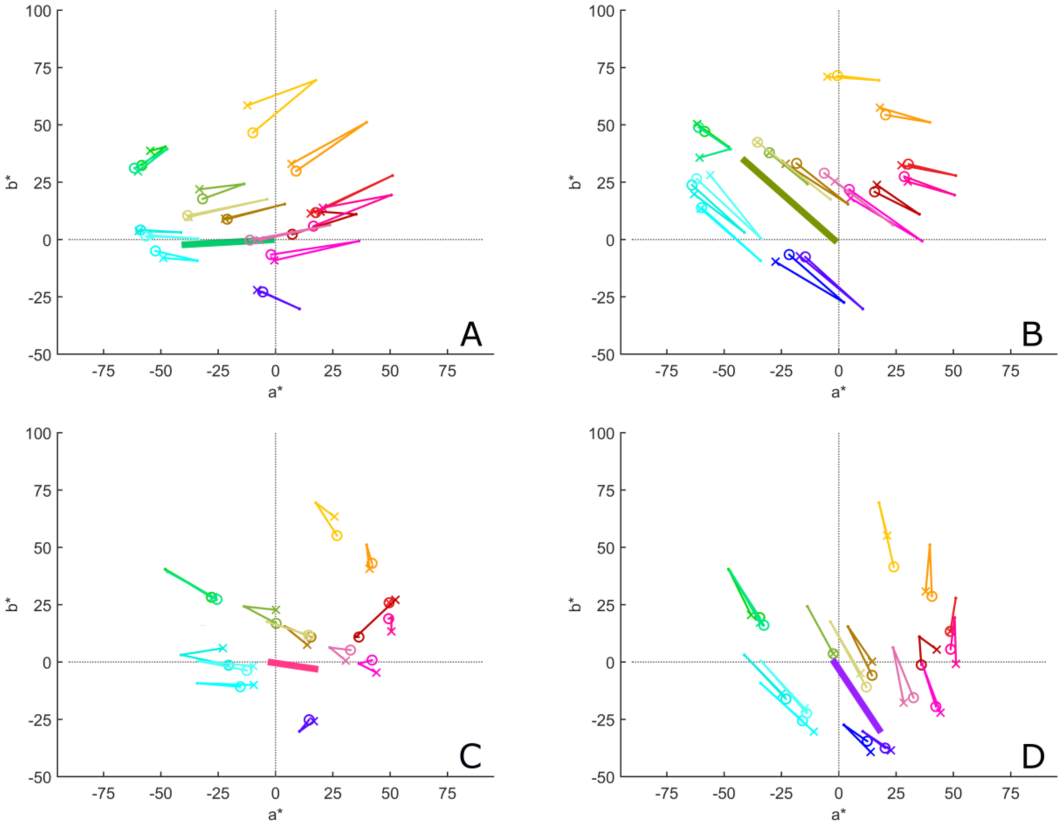


Figure 2.3: Diagrams showing in CIE a*, b* plane the direction and magnitude of the four illuminant shifts (A, B, C, D) with a heavy line that starts at the chromaticity of neutral daylight, near (0,0) and ends at the chromaticity of the test illumination. As in **Figure 2.2**, the lines for each object start at the chromaticity of their memory match under neutral daylight and extend to their memory match under the new illuminant, represented by crosses. The line ending in a circle represents the chromaticity shift of the memory selection for a given object from neutral daylight to the test illuminant. Under illuminants B and D we tested 17 objects, 16 under A and C. Plots are average from 2 sessions.

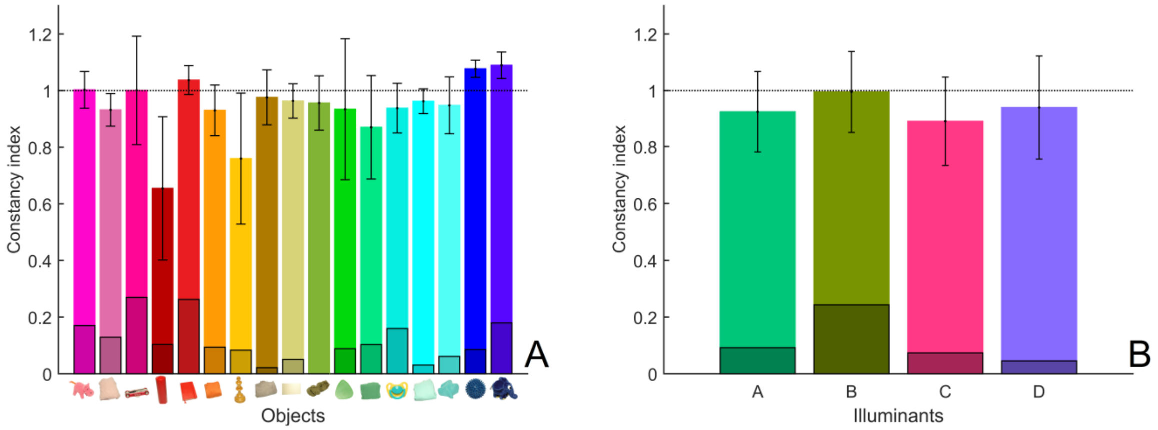


Figure 2.4: (A) colour constancy indices averaged over all illuminant conditions for each object. A value of 1 indicates perfect colour constancy and error bars represent standard deviation from the mean. Shaded regions indicate lowers bounds for colour constancy as described in Methods and **Supplementary Figure S 2.4**. (B) constancy indices averaged over all objects for each test illuminant.

Recently, it has been argued that colour constancy is impossible to achieve because of the metamer mismatch problem. In colour vision, reflectance spectra varying continuously in wavelength get reduced to three numbers, the excitation of long-, middle-, and short-wavelength-sensitive cones (Stockman & Sharpe, 2000). Consequently, there must be many physically different spectra that lead to the same cone excitations under a particular illumination, but which may differ under another illumination. In theory, the volume of colour space under one illumination that gets mapped into a single point under a second illumination has been shown to be quite large, and a recent study suggested that the size of that volume is negatively correlated to measures of colour constancy (Witzel et al., 2016). We therefore computed the size of the metameric mismatch volume for the colours of the objects used in our study (see **Supplementary Table S 2.2**). There was no correlation with the colour constancy index (all r(15) between -.32 and -.01, all p>.22, see **Supplementary Table S 2.3**).

We also computed the frequency of metamers to the selected Munsell chips within a large set of 11,302 natural reflectance spectra. Metamers were infrequent (5.1 x 10-04) for the illumination changes we used. On average, there were 5.76 +/- 10.32 metamers per object (see **Supplementary Table S 2.4**), but only the green scarf (42 metamers) and the red book (16 metamers) had a larger number of metamers. For 8 objects, there was no metameric surface within the whole set of 11.302 natural reflectance functions. Based on these calculations, we do not think that metameric mismatch represents a big problem for colour constancy under natural conditions.

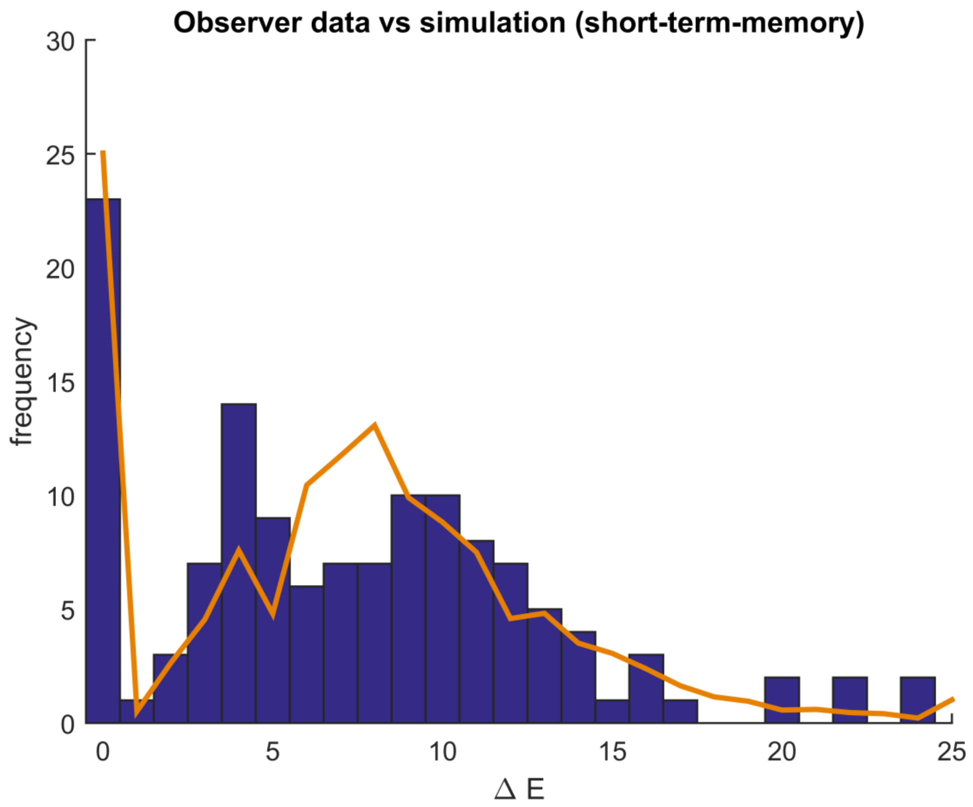


Figure 2.5: The bars depict the frequency of deviations from a perfect memory match in Euclidian distances in CIE-LAB colour space. Total number of matches was 132, the largest group of ΔE deviations is zero, meaning observers picked the very same chip as under neutral daylight. The orange line shows predicted frequencies when assuming that observers follow a Gaussian distribution when picking chips close to the originally selected chip (SD (in Munsell steps): Hue = 1.0521, Value = 0.53, Chroma = 1.37).

2.4 Discussion

The colour constancy values we have found in our study are higher than those previously reported in earlier studies that have relied on different methods. In the classic experiments using asymmetric matching, participants were asked to adjust the colour of an object or surface seen under one illuminant to another shown under a different illuminant. In a simultaneous version of this task when both illumination conditions were presented at the same time participants achieved an average colour constancy index of just 20% to 50% (Arend & Reeves, 1986; Arend et al., 1991). The results depended on the instructions given to the observer. When asked to make the two patches under different illuminations appear the same, constancy was quite low. When asked to adjust the two patches as if they were made from the same paper, constancy was higher. In both cases participants' adaptation to either illuminant is compromised due to limited eye movements while in our study participants had at least 2 minutes to adapt to the single illuminant in the room with a complete field of view and real objects. Still, there are examples of studies using asymmetric matching paradigms that achieved higher degrees of constancy (Bäuml, 1999; Troost & de Weert, 1991), but on average these designs lead to constancy values of about 49% (Foster, 2011).

Kraft & Brainard (1999) used an achromatic matching task (Brainard, 1998) with a single illuminant to which participants were well adapted to elegantly show that constancy can vary between 20% and 80%, depending on the richness of the cues available to the observers. They found that three factors – global scene colour, local contrast, and the colour of the brightest region – had an effect on colour constancy. Two problems remain, though. First, the use of this technique where participants are asked to adjust the patch until it appears grey is distinct from how we experience colour constancy in our daily lives and thus cannot provide a general test since it is limited to the single point in colour space void of any colour. In our experiments we used participants' internal representations of many colours as shown by the good coverage of the colour space in **Figure 2.3** and the variety of objects in **Supplementary Figure S 2.1**. Second, even when all cues are provided, constancy remains significantly below perfection. This is distinct from other constancies. For size constancy, for example, the availability of different cues also plays a big role, but when sufficient cues are provided, constancy is perfect (Holway & Boring, 1941).

A few studies have obtained even higher values of constancy. These studies have used either colour naming (Smithson & Zaidi, 2004; Troost & de Weert, 1991) or colour categorization (Hansen et al., 2007; Olkkonen, Hansen, & Gegenfurtner, 2009) to approximate the hypothesized function of colour as an identity tag in our daily lives. Their results show that the appearance of objects under different illuminations typically does not cross into a different colour category. Although both naming and categorization remain very much intact under different illuminants they only provide a coarse sampling of our sensory and perceptual colour space due to linguistics limitations. In our current study we show that colour constancy is better than just categorical and can represent the richness of the millions of colours we can discriminate.

A crucial aspect of our work is, as in real life, the use of memory to make the measurement of constancy for absent coloured objects possible. In some previous work (Allred & Olkkonen, 2013; Hedrich et al., 2009; Ling & Hurlbert, 2008), memory performance was taken into account when calculating the degree of constancy. These authors found for real 2D paper patches (Ling & Hurlbert, 2008), real 3D paper cubes (Allred & Olkkonen, 2013) or both (Hedrich et al., 2009) that constancy was as good as memory allowed for. Our experiments differ in that we used objects that were already highly familiar to our participants. We can exclude a possible role of memory biases (Bloj et al., 2016) in the present study by only comparing memory matches across different illuminations. It has also been shown recently that constancy is slightly improved in a natural selection task (Radonjić et al., 2015a, 2015b). Our results show that colour constancy is indeed perfect when a natural task, natural illumination conditions, and natural stimuli are used in combination.

Our results are exceptional, since this is the first study to show that performance in a colour constancy task can be perfect. Previous studies have often used more impoverished paradigms to get performance into a range where the effect of different cues could be investigated. Still, in all of these studies a ceiling on the very best performance emerged at levels of about 80% constancy. Our results proof that no such hard ceiling exists and that the limitations of earlier studies are mainly due to the limitations of the experimental paradigms. Indeed, there are three good reasons why our results should not come as a surprise. First, they do agree with our experience of colour in the natural world and with the use of colour names to label objects. Second, even relatively simple computational models can achieve high levels of colour constancy. This is evidenced by the fact that photographs taken under quite varying illumination conditions typically look correct when viewed in a different illumination context. The white balance algorithms built into most cameras achieve this feat. The magnitude of the correction easily becomes visible by turning off the automatic correction (if possible), as illustrated in **Figure 2.1**. Third, the neural mechanisms for colour constancy are well known. There are relatively local adaptation mechanisms throughout the visual system, starting in the cone photoreceptors. There are global adaptation mechanisms emerging in higher cortical areas. And there are mechanisms for computing colour constrast across edges as early as in primary visual cortex (V1) (Shapley & Hawken, 2011).

Our study resolves a riddle that has long puzzled vision scientists. In everyday life, we take colour constancy for granted, while in the lab constancy turned out to be between mediocre and incomplete. We show that in a natural environment, using a natural task where the visual system has all possible cues available, we can indeed achieve perfection, as has been shown decades ago for most other perceptual constancies, e.g. size constancy (Holway & Boring, 1941). This is reassuring in the light of a recent widely publicized example showing a lack of constancy and large individual variation when insufficient cues about the illuminant and scene are available (Brainard & Hurlbert, 2015; Gegenfurtner et al., 2015; Lafer-Sousa et al., 2015; Winkler et al., 2015; Witzel et al., 2017).

Chapter

3

Sensitivity to hue explains “blue bias” in colour constancy

A similar version of this manuscript has been submitted as:

Weiss, D., Witzel, C. & Gegenfurtner, K. R. (under review). Sensitivity to hue explains the "blue bias" in color constancy.

In order to clarify the determinants of colour constancy, this study investigates the role of the daylight axis and the “blue bias” in the perception of scene illuminations. A first experiment reproduced the blue bias with an illumination naming task: Observers described a blue illumination as white while naming red, yellow, and green illuminations of comparable saturation by the respective chromatic colour terms. The second experiment was aimed at determining the specificity of the blue bias. Detection thresholds for the chroma of the illuminations were measured in 12 hue directions, including the yellow and blue direction of the daylight locus. Results confirmed that observers were most insensitive for illuminations with bluish hues. However, the sensitivity was not lowest for the blue direction of the daylight axis, but for greenish blue colours. In order to examine whether the blue bias is particular to scene illuminations or a more general effect of colour perception, a third experiment measured simple detection thresholds for chroma in the 12 hue directions of the second experiment. These detection thresholds were highly correlated with illumination discrimination, suggesting that the blue bias is not particular to changes in illumination, but a more general result of the anisotropy of colour space. At the same time, there was still a bias in illumination discrimination when accounting for the sensitivity to hue. However, this bias was towards green-blue, rather than towards the blue pole of the daylight axis as suggested previously.

3.1 Introduction

Colour vision allows for identifying of objects by their colour. This would not be possible without *colour constancy*. Colour constancy is the ability to identify the colour of an object or surface across changes in illumination. For a given object or surface colour, changes in illumination alters the light that reaches the eye. The ability to perform colour constancy requires that the human observer somehow accesses information about surface colour that is independent of that change. How this is possible is partly known, but important aspects of colour constancy cannot be explained by known illumination cues and perceptual mechanisms (Brainard & Radonjić, 2014; Foster, 2011; Smithson, 2005).

Recent studies suggested that the magnitude of colour constancy depends on the colour of the illumination (Brainard, 2016; Pearce et al., 2014; Radonjić et al., 2016; Winkler et al., 2015; Wuerger, Hurlbert, & Witzel, 2015). The general idea is that human observers are most experienced with changes in illumination due to variations of daylight. Daylight varies along a curve between blue and yellow, the *daylight locus* (Granzier & Valsecchi, 2014; Judd et al., 1964; Mollon, 2006; Taylor & Kerr, 1941). Moreover, shadow and shading produce colour shifts in the bluish direction (Churma, 1994; Troscianko et al., 2009). If human colour vision is adapted to these illumination changes, colour constancy should be best along the daylight axis, and in particular towards the blue direction. In other words, changes towards blue and maybe yellow illuminations should least affect how human observers perceive colours. The idea that human colour vision is most robust against changes towards bluish illuminations is called a *blue bias* in colour constancy (Aston, Le Couteur Bisson, Jordan, & Hurlbert, 2016; Pearce et al., 2014; Winkler et al., 2015; Wuerger et al., 2015).

Some studies support the idea of a blue bias. Delahunt and Brainard (2004) compared colour constancy under daylight illuminations and under red and green illuminations in an achromatic adjustment task, and found higher constancy for bluish daylight and greenish illuminations. Moreover, a blue bias was found when comparing colour constancy under bluish daylight illuminations and under yellowish illuminant A (Murray et al., 2006).

However, other studies did not find higher colour constancy either for blue illuminations or for illuminations along the daylight axis in general. One study observed higher colour constancy for illuminant hues orthogonal to the daylight locus (Rüttiger, Mayser, Sérey, & Sharpe, 2001). Another study, obtained higher colour constancy for the yellowish than the bluish part of the daylight locus (de Almeida et al., 2004). Still other studies did not find any significant difference in colour constancy between illuminations with different hues (Brainard, 1998; Hansen et al., 2007; Olkkonen et al., 2009, fig. 6; Olkkonen, Witzel, Hansen, & Gegenfurtner, 2010, figs. 12-13; Radonjić et al., 2015a; Schultz et al., 2006). Finally, there was some evidence that the effect of illumination hue on colour constancy depends on the hue and saturation of the surface colours (Ling &

Hurlbert, 2008). Taken together, it is yet unclear whether and how colour constancy depends on the illumination.

Another approach suggested that observers are less sensitive to illumination changes across the daylight axis. First evidence for this idea came from achromatic (grey-scale adjustments. Achromatic surfaces reflect the chromaticity of the illumination, and thus are known as equivalent illuminants (Brainard, 1998). It has been shown that observers are uncertain in the estimation of grey along a curve that links blue and yellow and is very similar to the daylight locus (Bosten, Beer, & MacLeod, 2015; Witzel et al., 2011). In addition it has been observed that the estimation of grey is shifted towards blue, indicating that observers see slightly bluish colours as grey (see also Olkkonen, Hansen, & Gegenfurtner, 2008; Winkler et al., 2015; Wuerger et al., 2015). These findings indicate that colour identification without changes in illumination are least reliable along the daylight locus and towards the blue hue direction. It has been suggested that this is due to the fact that the colour signal is least reliable in the natural environment (Witzel et al., 2011). The variation of daylight implies that the sensory colour signal corresponding to a given object or surface in the natural environment varies along the daylight locus. Due to this variation, observers may be less certain about the identity of colours along the daylight locus.

Recent studies have further substantiated this idea. They found that observers were less able to see illumination changes towards blue (Pearce et al., 2014; Radonjić et al., 2016). The insensitivity to the effects of blue illumination has been taken as evidence for higher colour constancy blue illuminations: The less the change in illumination affects perception, the more stable is the perception of object colours. This effect strongly supported the idea of a “blue bias” in colour constancy.

It has been speculated that the colour-switching dress, having colours along the daylight locus (Gegenfurtner et al., 2015), could be explained by the uncertainty about surface colours along the daylight locus (Brainard & Hurlbert, 2015; Gegenfurtner et al., 2015; Lafer-Sousa et al., 2015) or by the blue bias (Winkler et al., 2015). If this is true, the phenomenon of the dress illustrates that the uncertainty about colours along the daylight axis and the blue bias can have striking effects on how human observers perceive colour.

However, differences in colour perception as a function of hue are not specific to the perception of illumination changes. If we consider that observers are completely adapted to the colour of a given illumination, the change to the colour of another illumination corresponds to a perceived change in saturation: It is a change from the adapted and hence white illumination to an illumination with chroma. The sensitivity to changes in saturation varies strongly as a function of hue (Giesel, Hansen, & Gegenfurtner, 2009; Hansen, Giesel, & Gegenfurtner, 2008; Krauskopf & Gegenfurtner, 1992; Witzel, Maule, & Franklin, under revision). While some evidence has been provided that the blue bias is independent of the discriminability of uniform colours (Pearce et al., 2014), this issue has not yet been tested specifically and must be considered as open.

Moreover, changes in illuminant spectra can have complex effects on the reflected light and the resulting colour signal (Logvinenko, 2009; Logvinenko et al., 2015; Witzel et al., 2016; Witzel, Cinotti, & O'Regan, 2015). Depending on the illuminant, colours in a scene may be changed to different degrees. In particular, certain illuminations may strongly change the relationship between colours, i.e. differences between colours under one illumination may considerably change under another illumination. An extreme example is a monochromatic light under which many colour differences become zero (namely when they result in the same lightness under the monochromatic illumination). Changes in colour differences strongly affect relational colour constancy. In sum, the degree to which illuminations affect colour differences may strongly impact on colour constancy and the perception of the illumination. If these effects depend on the hue of the illumination, this could potentially explain different effects of hue on colour constancy and the perception of illuminations, such as particular effects along the daylight axes or the blue bias.

Here, we investigated the perception of illumination changes through different methods, and we tested whether effects of illumination hue may be explained by sensitivity to hue and overall colour changes in the scenes. In a first preliminary experiment we tested the subjective appearance of the illumination with a simple task, in which observers have to describe the illumination through colour names. In the second experiment, we measured observers' ability to detect an illumination change for many hues and using illuminant spectra that are directly comparable across hues. Finally, we measured simple detection thresholds in the hue directions of the second experiment and compare the detection thresholds to the illumination estimation thresholds from that second experiment.

3.2 Experiment 1: Illumination colour naming

This experiment was meant to explore interesting questions and the design for Experiments 2 and 3. In this preliminary experiment, we simply re-examined the results found with the illumination discrimination paradigm (Aston et al., 2016; Brainard, 2016; Pearce et al., 2014; Radonjić et al., 2016) with a simpler paradigm that does not involve direct comparisons between successively presented scenes. In this way, we were hoping to avoid effects of memory and discrimination. To do this, we simply asked observers to describe the colour of the illuminations of the scenes through colour terms.

Moreover, different illuminations may affect surface colours differently. In particular, the relation between colours may be changed under chromatic illuminations because highly saturated surface colours with peaky reflectance spectra are less affected by illumination changes than desaturated surface colours with comparatively flat reflectance spectra. Here, we examined whether those effects on surface colours depend on the hue of the illumination. The question of how the relations between colours change under illumination change may be understood as a form of relational colour constancy (Foster & Nascimento, 1994; Nascimento & Foster, 2000). Originally relational colour constancy is assessed through cone rations, which express the constancy of the sensory signal. However, we were more interested in the differences as perceived by observers, and hence focused on measures of perceptual differences rather than cone ratios.

3.2.1 Method

3.2.1.1 Observers

13 observers (8 female, 18-29 years old) of the Justus-Liebig University participated in the experiment. Participants were tested for normal colour vision using Ishihara plates (Ishihara, 2004). All experiments were carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) and informed consent was obtained for experimentation with human subjects.

3.2.1.2 Apparatus

To produce the stimuli, scenes were set-up in a LED-Box (Just Normlicht LED Colour Viewing Light), which allowed to control illuminations through the adJust LED control professional Software (ver.:1.1.21). Photos of the scenes under the different illuminations (see Stimuli) were taken from a 45 degree angle, using a Nikon D70 camera. The white-balance of the camera was switched off and saved in raw-format so that images were recorded without white-point adjustments.

Stimuli were presented on an EIZO Colour Edge CG223-BK Wide Gamut LCD monitor with a diagonal of 22“, a resolution of 1680*1050 pixels, and a refreshrate of 60hz. The monitor was driven by an Nvidia Quadro FX1800 graphic card with 8 bit per channel colour resolution. The monitor was calibrated and gamma-corrected (without bit loss). The Judd-corrected CIE-xyY values (Judd, 1951) of the primaries were: R = [0.652, 0.332, 33.5], G = [0.202, 0.678, 65.4], and B = [0.151, 0.066, 8.2]. Experiments were programmed in Matlab (The MathWorks Inc., 2007) with the Psychtoolbox extensions (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997).

The experiment was conducted in a completely dark experimental room with the only light source being the monitor. The participants were seated 42 cm in front of the screen (controlled with a chin rest), and looked through a black viewing tunnel.

3.2.1.3 Stimuli

We used two kinds of scenes, one containing colour-diagnostic objects, the other only coloured papers (**Figure 3.1**). The scene with the objects contained fruits and vegetables (e.g. banana, carrot, apple etc.) and man-made objects with typical colours (e.g. coca cola and Nivea can), arranged on a grey table and background (leftmost column in **Figure 3.1**). The non-diagnostic scene consisted of the Macbeth ColourChecker chart (McCamy, Marcus, & Davidson, 1976).

These scenes were photographed under five different illuminations, a neutral grey, a blue, a green, a yellow and a red one (**Figure 3.2)**. Illuminants were chosen to have hues that coarsely correspond to the cardinal directions of Derrington-Krauskopf-Lennie-space (Derrington et al., 1984) to allow for potential comparisons with earlier studies (Hansen et al., 2007; Olkkonen et al., 2009, 2010). Moreover, illuminants had equal luminance (56.35-56.66 cd/m^2) and similar chroma in DKL space (and hence coarsely similar chroma in CIELUV space, cf. **Figure 3.2)**. Since the scene under the blue illumination in the light box appeared very similar to the one under the neutral illumination, we used a slightly higher saturation for the blue illumination.

Two version of each stimulus type was used, one version with and the other version without achromatic colours (columns in **Figure 3.2)**. Images of colour-diagnostic scenes were presented at 20.5 x 11.2 degree visual angle, those with the ColorChecker at 17.9 x 11.9 degree.

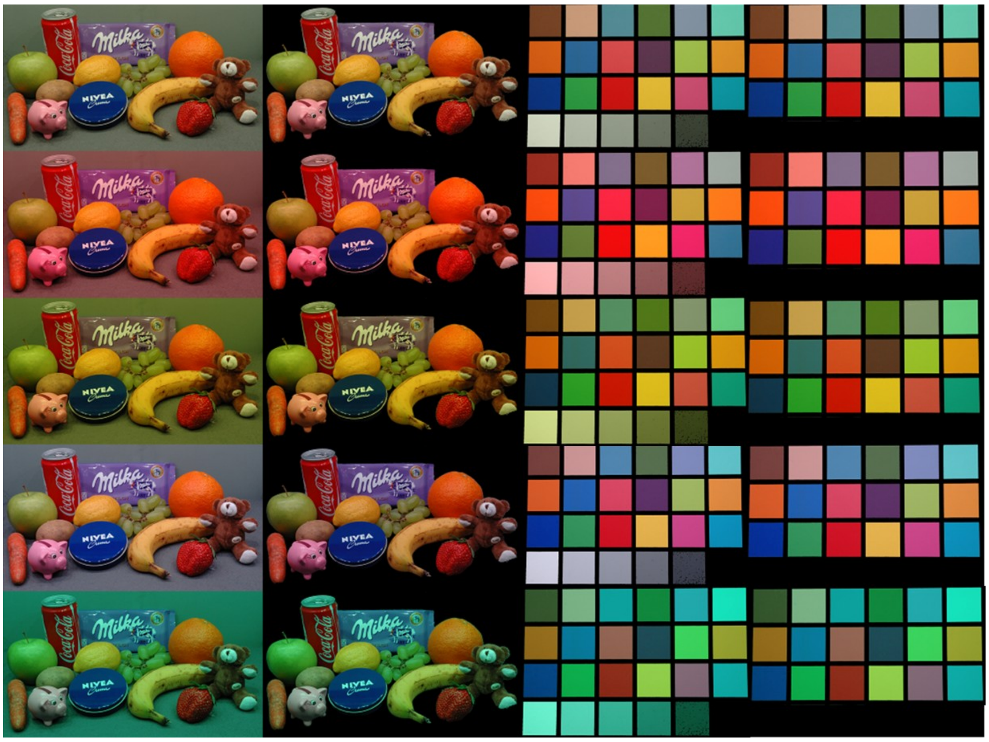


Figure 3.1. Stimulus images. Rows correspond to the different illumination colours (neutral, red, yellow, blue, and green). The first and the second column show the colour diagnostic scene, the third and fourth the ColorChecker chart. The second and fourth column show the versions without achromatic colours.

3.2.2 Results and Discussion

3.2.2.1 Illumination naming

Figure 3.3.a illustrates the accuracy of responses. Observers identified the white and the red illumination correctly in almost all cases. Accuracy for yellow and green were lower than for white and red; and accuracy for blue was close to zero. We calculated a three-way repeated measurements analysis of variance with the factors “illumination colour”, “scene type” and “presence of grey colours”. There was a highly significant main effect of illumination colour (F(4,48)= 39.1, p<0.001). There were no other main effects or interactions, except for a three-way interaction (F(4,48)=3.2, p=0.02), which is of minor importance for the overall blue bias.

Figure 3.3.b illustrates the kind of mistakes observers made for each illumination colour (pooled for scene type and presence of grey). Most errors for the green and yellow illuminations consisted in naming the green illumination as blue, and the yellow as green. These errors may be explained by the fact that the “green” pole of the L-M axis and the “yellow” pole of the S-(L+M) axis are very close to the green-blue and yellow-green category boundaries, respectively (see Witzel & Gegenfurtner, 2013, fig. 9.a,c).

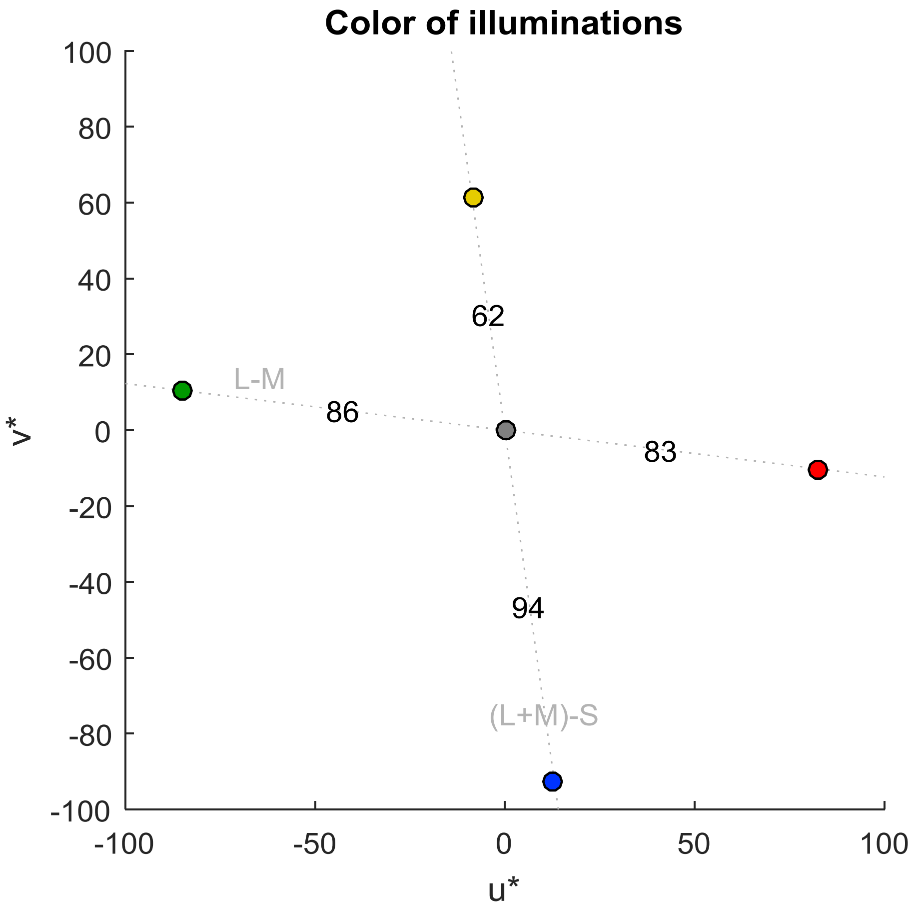


Figure 3.2. Colour of illuminations. Colours of five illuminations in the light box based on measurements with a spectroradiometer (PR650) and represented in CIELUV. The x-axis corresponds to the green-red (u*), the y-axis to the blue-yellow axis (u*) in CIELUV space, when using the neutral illumination as the reference white point for CIELUV. The disks are coloured in correspondence with the illumination colours they represent, i.e. grey, blue, red, yellow, and green. Numbers report Euclidean differences of coloured illuminations from the neutral grey illumination. Dotted lines indicate DKL axes, which were used to determine the hue directions of the illumination colours.

However, the large rate of errors for naming the blue illumination cannot be explained by an ambiguity of colour terms. Blues was called white in about 80% of all answers, indicating that blue was almost always mixed up with white. This was the case, even though the saturation of the blue illumination was higher for the blue and for the other chromatic illuminations (**Figure 3.2**), and even though the difference between the blue and the white illumination was clearly discriminable (supra-threshold). This result indicates that blue appears like a neutral illumination, in line with the idea of a blue bias.

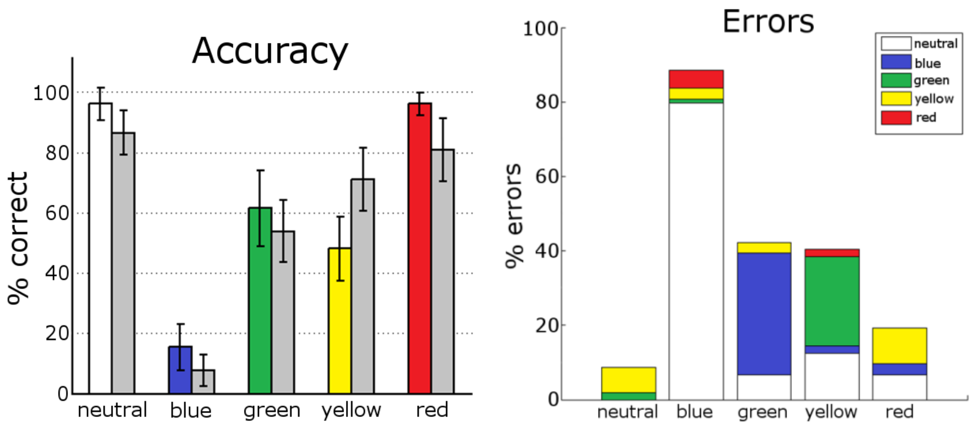


Figure 3.3. Results from Experiment 1. Panel a shows the accuracy rates for the scene (coloured bars) and the colour checker chart (grey bars) under each of the five illuminations (from left to right: white, blue, green, yellow, and red). Panel b illustrates which other colour terms observers chose when they gave a wrong response. Each of the five bars corresponds to one illumination in the same order as in Panel a. Note that participants mostly described the blue illumination as white (panel b).

3.2.2.2 Statistics of colour changes

To clarify why blue appears to be white, we determined the colour shifts across illuminations. For this purpose we used CIELUV and CIELAB because they coarsely represent discriminable distances between colours. **Figure 3.4** illustrates the colour shifts in CIELUV between the white (grey dots) and the chromatic illuminations (coloured dots in the respective panels).

The average difference between colour distributions indicates how much colours were shifted due to the illumination change: The higher the average difference, the more colours were shifted due to the illumination change. The standard deviation of those differences shows how homogeneously these shifts were and gives an idea of relational colour constancy: The lower the standard deviation, the more homogeneously the shifts and the higher “relational colour constancy” (in our sense). These calculations were done for all ten pairwise comparisons of illuminations (for illustration purposes **Figure 3.4** only shows change from white to another colour).

The change between the white and the blue illumination provided by far the lowest average shift and standard deviation ($\Delta E_{Luv} = 19.1 \pm 9.6$ and $\Delta E_{Lab} = 18.7 \pm 7.7$) compared to other illumination changes (at least $\Delta E_{Luv} = 27.4 \pm 15.1$ and $\Delta E_{Lab} = 26.2 \pm 14.4$ for white to yellow). Similar results were obtained with ColorChecker chart (**Figure S 3.2** and **Figure S 3.3**): Surface colour shifts were lowest and varied least when illumination changed between white and blue ($\Delta E_{Luv} = 24.3 \pm 9.8$ and $\Delta E_{Lab} = 22.0 \pm 3.9$) than when it changed between illuminations of other colours. These results suggest that colours were most similar and that the relationship between colours was most stable under changes between the white and the blue illumination.

Since the blue illumination had a higher chroma than the other chromatic illuminations, the observation of smaller colour shifts may not simply be explained by the chroma of the blue illumination. Instead, the reason for the comparatively small surface colour shifts with our blue illuminant may reside in our particular illuminant spectra. It is an open question whether low surface colour shifts and high relational colour constancy are a property of blue illuminations in general, or whether it is a particularity of the LED illuminants used here. It is also possible that our approach of photographing the real scenes might have altered colours and their relations. It is unlikely that this explains the blue bias since the real (non-photographed) scenes under blue illumination clearly appeared as if they were illuminated by a white light.

In any case, the present results raise two important questions that warrant further investigation. First, are low surface colour shifts a general property of blue illuminations, in particular of those that occur in the natural environment? Second, does the blue bias in the perception of illumination depend on the magnitude and consistency of surface colour shifts or does it also occur when colour shifts are controlled across illuminations with different hues. In the following experiments, we focus on this latter question.

3.3 Experiment 2: Illumination discrimination

The illumination naming (Experiment 1) confirmed a blue bias in the perception of illuminations found in previous studies (Aston et al., 2016; Brainard, 2016; Pearce et al., 2014; Radonjić et al., 2016). However, our experiment 1 and those previous studies used only a few illumination colours (i.e. not more than 4 chromatic illuminations). With so small a sample of colours, it is difficult to establish whether the blue bias is specific to a particular bluish hue direction (e.g. daylight or tritanopic confusion line) or whether this bias is a more global phenomenon, involving larger regions of colour space. The main goal of this second experiment was to determine whether the blue bias is specific to a particular bluish hue direction. For this, we measured illumination discrimination for 12 hue directions with the task proposed by Pearce et al. (2014). Moreover, experiment 1 suggested that the blue bias is possibly related to smaller effects of blue illuminant spectra on colour changes in the scene. In previous studies and in our experiment 1 involved illuminant spectra of LED lights (Pearce et al., 2014; Radonjić et al., 2016). These illuminant spectra tend to be “peaky” (i.e. to have sharp local maxima and minima). Peaky illuminations may have particular effects on illumination-induced surface colour shifts (including changes in the relation between surface colours). In experiment 2, we wanted to avoid this in order to examine whether the blue bias occurs for illuminants independent of peakiness and differences in colour shifts. For this purpose, we used smooth spectra (based on Gaussian fits) spread across the visual spectrum to obtain the twelve illumination colours.

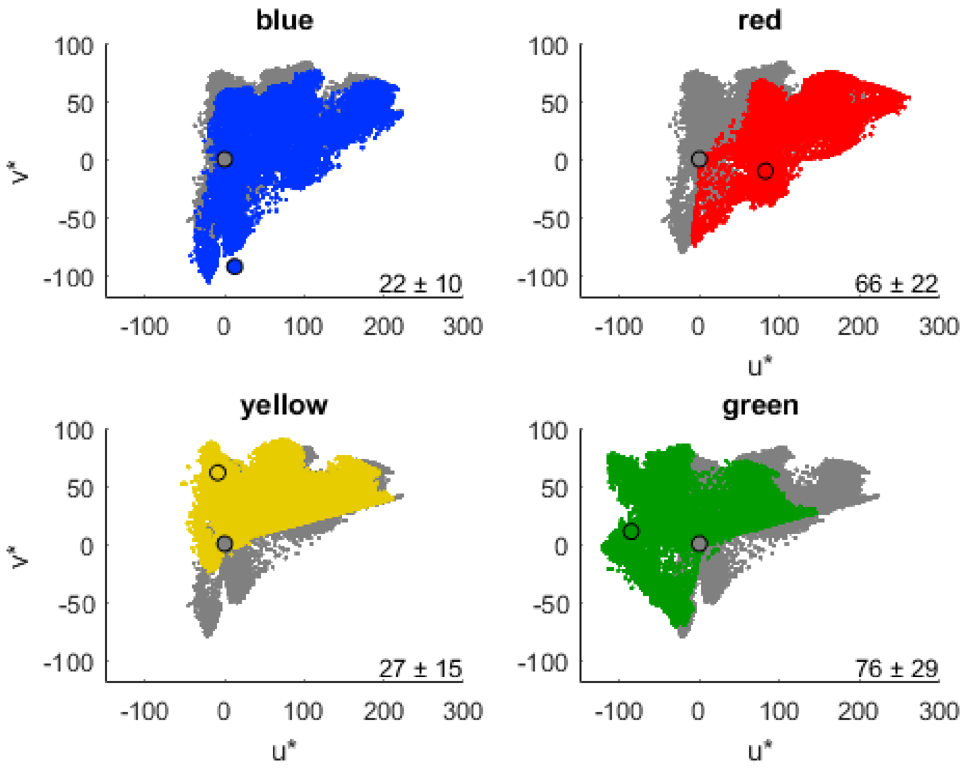


Figure 3.4. Colour shifts of the scene under different illuminations. Diagrams show colour distributions in CIELUV colour space with the neutral illumination as the white-point. The x-axis corresponds to the green-blue dimensions (u*), the y-axis to the blue-yellow dimensions (v*). In all diagrams, grey dots correspond to the colours in the scenes under the neutral illumination. Circles show the colour of the illumination. For a representation in CIELAB see **Figure S 3.1**

In this experiment we used cone-opponent DKL-colour space for two reasons. First, we wanted to examine the role of second stage mechanisms in colour discrimination and more generally we wanted to compare results to previous studies on colour discrimination (e.g. Giesel et al., 2009; Hansen et al., 2008; Krauskopf & Gegenfurtner, 1992; Witzel & Gegenfurtner, 2013). Second, our stimulus displays were based on colours induced through colour contrast (see Methods), and cone-opponent mechanisms seemed to be most appropriate to model the induced colours along opponent directions. To some extent, this idea is supported by the fact that induced colours in afterimages can be modelled through cone-opponent mechanisms (Zaidi, Ennis, Cao, & Lee, 2012).

3.3.1 Method

In the illumination discrimination task, observers were asked to identify which one of two illuminations (the *comparison illuminations*) was different to a *reference illumination*. In this experiment, illuminations were presented by rendering scenes under the different illuminations on the computer monitor (**Figure 3.5**).

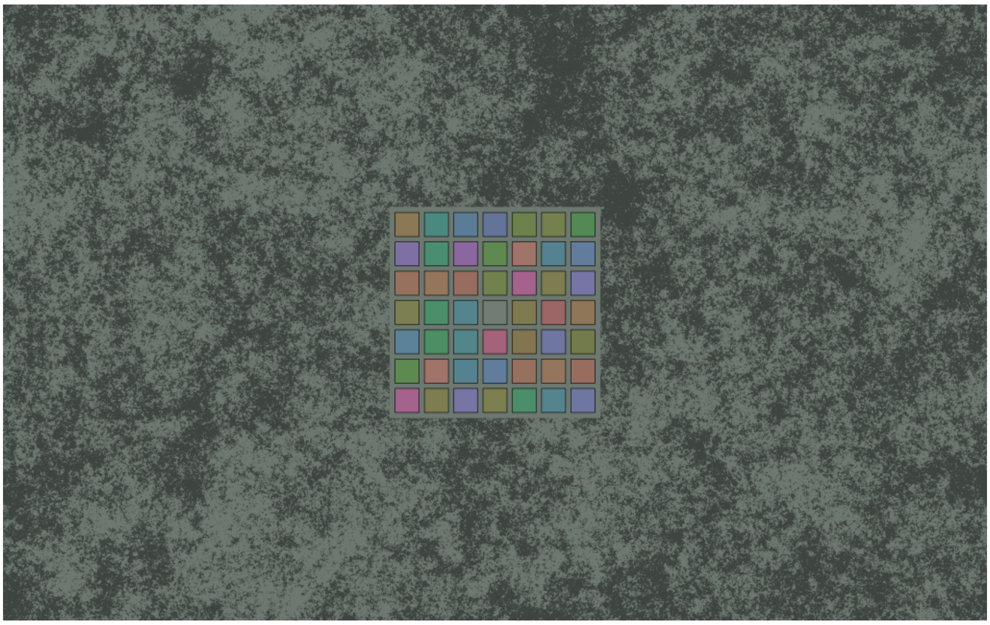


Figure 3.5. Stimulus display for the neutral reference illumination. The square in the centre shows the scene composed of Munsell-like reflectance spectra and illuminated by the neutral illuminant. The scene is surrounded by dark and light grey spatially distributed by a brown-noise function.

3.3.1.1 Observers

16 naive students (10 females, 21-31 years old) of the Justus-Liebig-University participated in the experiment. All participants were tested for normal vision using Ishihara plates (Ishihara, 2004).

3.3.1.2 Apparatus:

All experiments were presented on an EIZO CG223W monitor driven by an AMD FirePro V4900 with a resolution of 1680 x 1050 pixels, and a colour resolution of 10 bit per channel. The Monitor was calibrated using a Konika Minolta CS2000 Spectroradiometer (Konica Minolta Sensing Inc., Singapore), CIE-xyY specifications of the channels were: R = [0.655, 0.332, 34.6]; G = [0.207, 0.678, 64.2]; B = [0.15, 0.065, 7.8]. All stimuli used in the experiment have been gamma corrected (without bit loss). The Monitor was placed in a black painted tunnel, 40 cm away from the participant. The screen had a visual angle of 61.3° x 40.6°. The distance was fixed by a chin rest mounted to the table. The experiments were programmed in Matlab 2012b (The MathWorks Inc., 2007), using the psychophysics toolbox 3 extensions (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997).

3.3.1.3 Stimuli

Figure 3.5 illustrates the stimulus display use in this experiment. The scene in the centre of the display consists of 7 x 7 squares with 40 different surface colours (a few colours occur twice). The background of the scene consisted of a neutral grey surface. The surround of the scene showed a brown noise distribution of the background grey and a darker grey surface. The stimulus display was rendered under the different illuminations.

Overall, we determined 13 illuminants with different hues (cf. **Figure 3.6**). One of these was the neutral white illumination. Four illuminants were designed to test the role of the daylight axis. They corresponded to the blue and yellow hue direction of the daylight axis, and to the greenish and reddish hue directions orthogonal to the daylight axis. The other eight illuminants were matched with the four cardinal and four diagonal directions of DKL-space. This allowed us to take effects of cone-opponency into consideration. In order to obtain smooth illumination spectra we created artificial illumination spectra based on Gaussian functions. Moreover, our illuminations were designed to match the surface colours in the scene in a particular way: For each illumination there should be one surface colour that is shifted towards the colour signal of the neutral surface colour under the neutral illumination (“absolute” or “colourimetric” grey). In this way, the surface colour shifts due to illumination changes were comparable to the distribution of surface colours under the neutral illumination (cf. black and red dots in **Figure 3.7**). Finally, we wanted realistic surface colours. Hence, we created surface colours based on the reflectance spectra for matte Munsell spectra (Munsell Color Services, 2007).

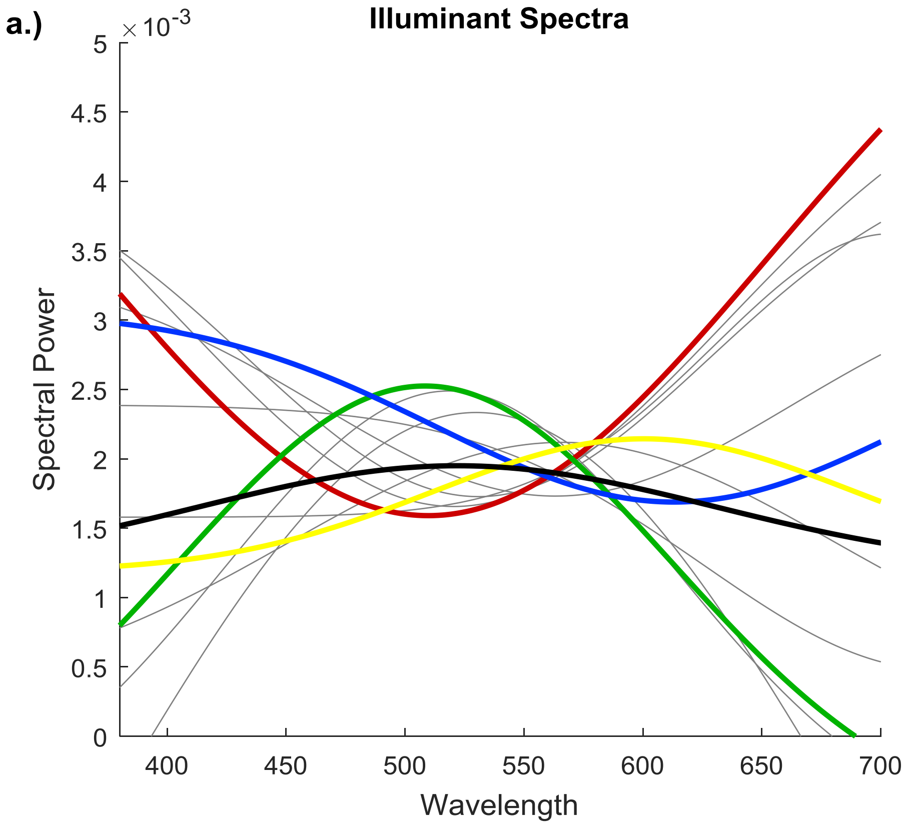
In order to obtain illuminant and reflectance spectra that fulfil all these criteria, we started with a neutral illuminant composed of a broad Gaussian distribution that spread across the whole visible spectrum and had chromaticities that are typical for neutral illuminations (xyY_{Judd} = [0.3265, 0.3419, 136.0]). We then selected 12 Munsell Hues that corresponded to the opponent hue directions of the wanted illuminations (Note: Due to the limited sample of Munsell spectra, hue directions of the Munsell spectra under the neutral illumination and the hue directions of the illuminants were close but not exactly hue directions). For this, we retrieved the Munsell spectra from the Jeonsuu colour group (Kohonen, Parkkinen, & Jääskeläinen, 2006; Parkkinen, Hallikainen, & Jaaskelainen, 1989), which are now available via the University of Eastern Finland and is now available at: http://www.uef.fi/en/web/spectral/munsell-colors-matt-spectrofotometer-measured). To obtain surface colours with comparable hues, we interpolated reflectances for Munsell chips with different Munsell chroma so that all surface colours (under the neutral illumination) had the same chroma (radius) in DKL-space (cf. black circle in **Figure 3.7**).

The equal chroma of the Munsell-like reflectances corresponded to a Munsell chroma between 4 and 6.

Illuminants were produced by a minimization algorithm that fitted the parameters of Gaussian functions to cancel the colour of one of the Munsell-like surfaces under that illuminant (**Figure 3.6**). As a result, one of the colours under each illumination was colourimetrically the same as the grey surface under the neutral illumination. This is evident by the fact that the red ellipses in **Figure 3.7** always the centre of the black circle. This approach also allowed us to control the effect of illumination hues on surface colour shifts: The average and the standard deviation of surface colour shifts were similar for all illumination hues: In DKL-space averages varied between 0.258 and 0.272, and standard deviations between 0.045 and 0.075. This corresponds to 26.3 to 27.7 for averages and 4.4 to 5.9 for standard deviations in CIELUV. Hence, this approach allowed for measuring the blue bias independent on the magnitude and consistency of surface colour shifts.

For neutral reflectances in the background and surround, the spectra for Neutral 6.5 (background and surround) and Neutral 5 (surround only) from the MacBeth ColorChecker (McCamy et al., 1976) were used (since the Munsell reflectance database did not provide any neutral reflectance spectra).

For the measurement of discrimination thresholds it was necessary to modulate the “intensity of the illumination change” that is the effect of the respective illuminations on the surface colours in the scene. For this purpose, surface colour shifts were rendered at nine chroma levels. Chroma levels were determined by the distance between the neutral grey surface under the neutral and under the respective chromatic illumination. Chroma levels were defined as the radius in DKL-space and varied between 0 and 0.08 in steps of 0.01. Scenes at different chroma levels were achieved by interpolating between the neutrally illuminated scene and the chromatically illuminated scene.


a.)
Illuminant Spectra
× 10^{-3}
Spectral Power
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
400
450
500
550
600
650
700
Wavelength

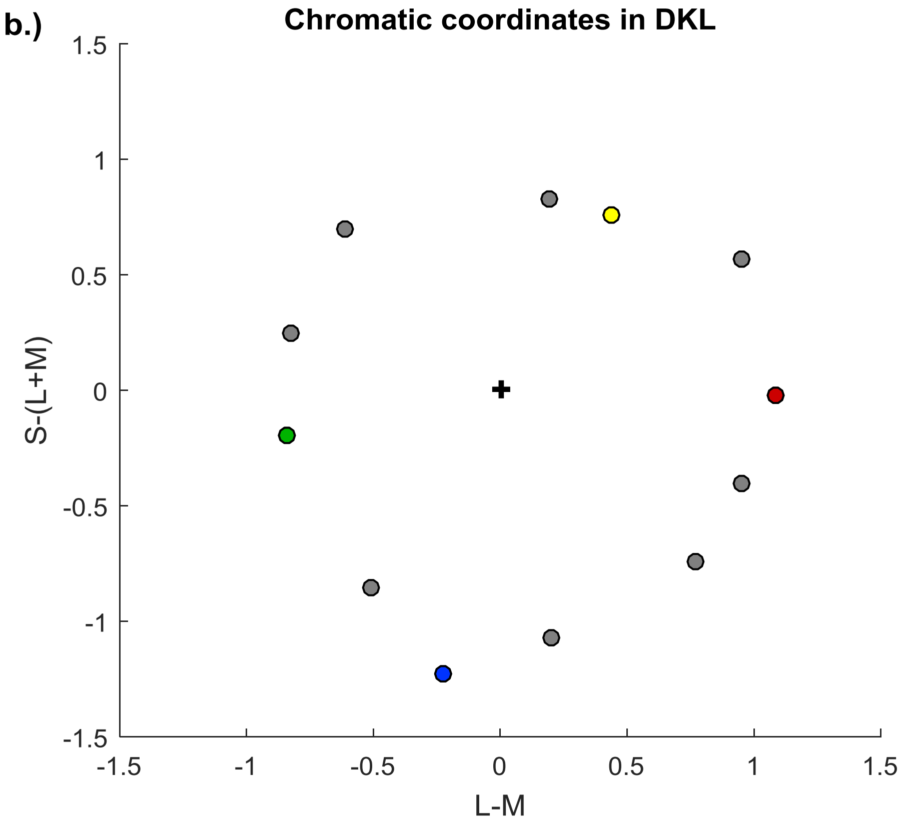

b.)
Chromatic coordinates in DKL
S-(L+M)
1.5
1
0.5
0
-0.5
-1
-1.5
-1.5
-1
-0.5
0
0.5
1
1.5
L-M

Figure 3.6. Gaussian illuminants. Panel a shows illuminant spectra. The thick black curve corresponds to the neutral illuminant and the grey curves to the chromatic illuminants. Four curves are highlighted in colour to illustrate the correspondence with panel b. Panel b illustrates the coordinates of the illumination in DKL space. The cross corresponds to the neutral illumination, the grey disks to the chromatic illuminations. For illustration the red, the yellow, the green and the blue disks correspond to the respectively coloured spectra in panel a.

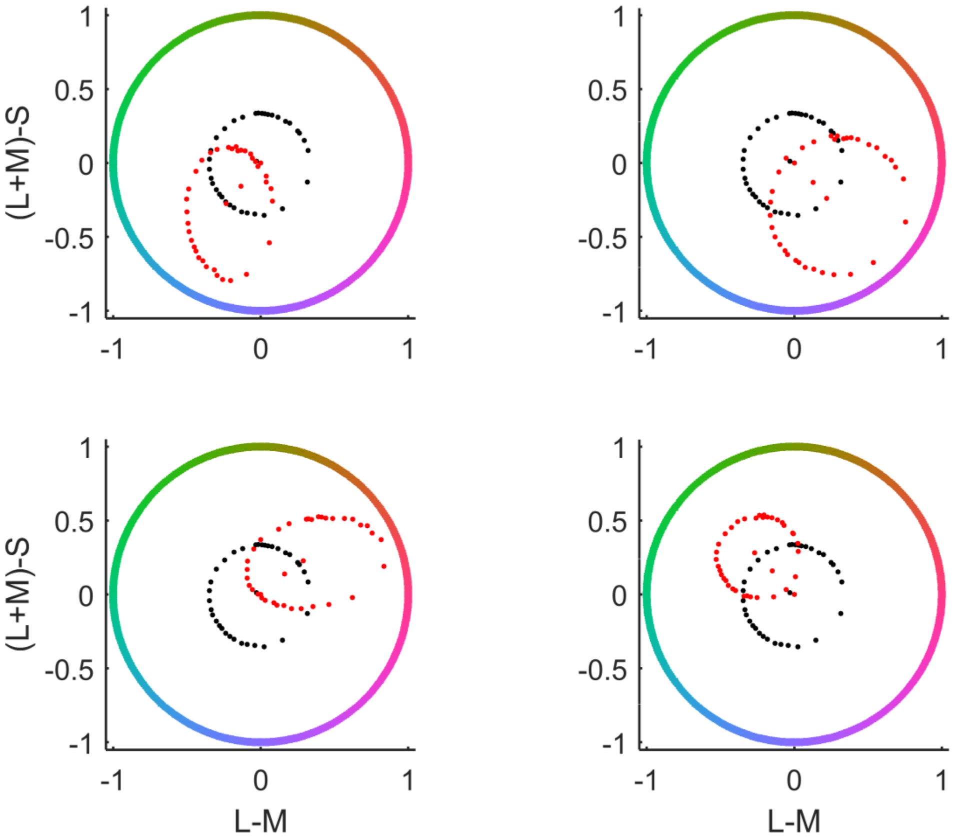


Figure 3.7. Changes of surface colours in DKL-space for four different illuminations.

3.3.1.4 Procedure

Figure 3.8 illustrates a trial of our task. A trial started with the presentation of the stimulus display under the reference illumination (*reference displays*) followed by the presentation of the scene under the comparison illuminations (*comparison displays*). The reference illumination was always the neutral illumination. One of the comparison displays was the same as the neutral referent display and the other comparison display was a scene under a chromatic illumination.

The reference image was shown for 2 seconds, the comparison images for 1 seconds each. Between the presentations of two images a blank black screen was displayed for 400ms. After the second comparison image observers were prompted to choose between the first and the second comparison image to match the test image. This was done by showing a black background with a grey square in the centre. Observer pressed one of two keys to indicate which one of the two comparison illuminations differed from the test illumination.

Measurements were done in separate blocks for each of the 12 hue directions. Within each block, the comparison images with the nine chroma levels were presented 20 times in random order, resulting in 180 trial in each block (*constant stimuli technique*). One such block took about 20min. To avoid fatigue, observers completed only four such blocks in one session, with 30min break after two blocks. Consequently, observers participated in overall 3 session of about 2h duration to complete the experiment.

In order to control the state of adaptation across observers, an experimental session begun with two minutes of dark adaptation before starting with experimental trials. Moreover, before the first experimental session, observers completed two practice sessions to make sure they understood the task. The first practice sessions was much shorter than the experimental sessions, with only one clearly visible chroma level for one hue direction. The second practice session was similar to the experimental sessions regarding chroma levels, but much shorter.

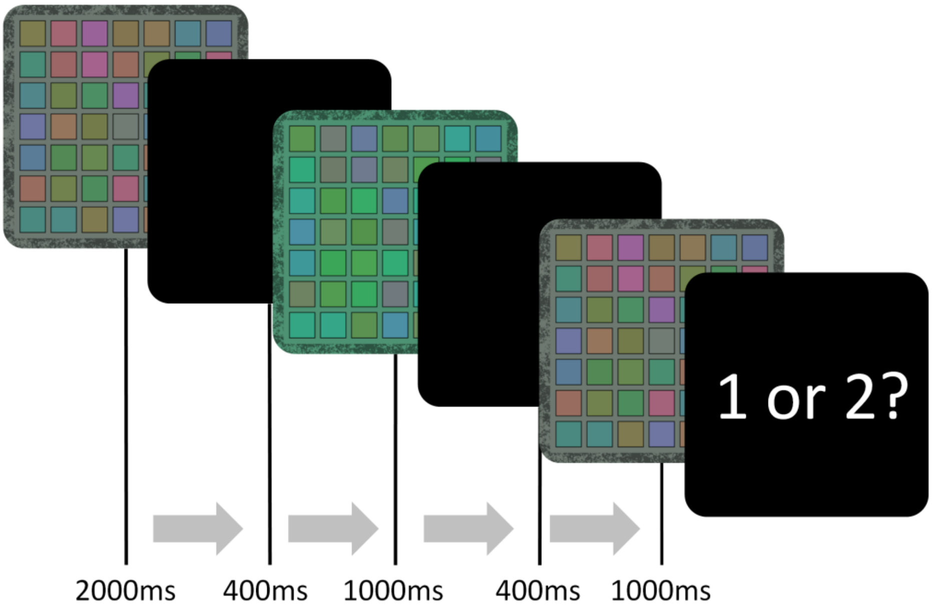


Figure 3.8. Procedure for illumination discrimination.

3.3.2 Results

Figure 3.9.a illustrates how we determined discrimination thresholds based on psychometric functions. Thresholds were calculated for each illuminant hue direction. For this, we used *psignifit 3* toolbox (Fründ, Haenel, & Wichmann, 2011; Schütt, Harmeling, Macke, & Wichmann, 2016) to estimate psychometric functions. For this, a cumulative Gaussian distribution was fitted to the data. Thresholds were determined at a probability of 0.5 for perceiving the stimulus. Since the probability of a correct answer by chance is 0.5, the threshold corresponds to a percent correct level of 75%.

When the chroma level was zero, there was no difference between the comparison images and observers must respond at chance level (0.5). As expected, the grant average (across hues and observers) of measured accuracies for this condition was close to 0.5 (M = 50.50%) and did not differ significantly from chance level (0.5) in a paired t-test across observers (t(15) = 0.5, p= 0.64). For this condition, differences of individual accuracies from chance level must necessarily be unwanted noise that distorts the estimation of psychometric functions. For this reason, we replaced individual accuracies by chance probability (0.5) for the data of each individual observer before we fitted the psychometric function.

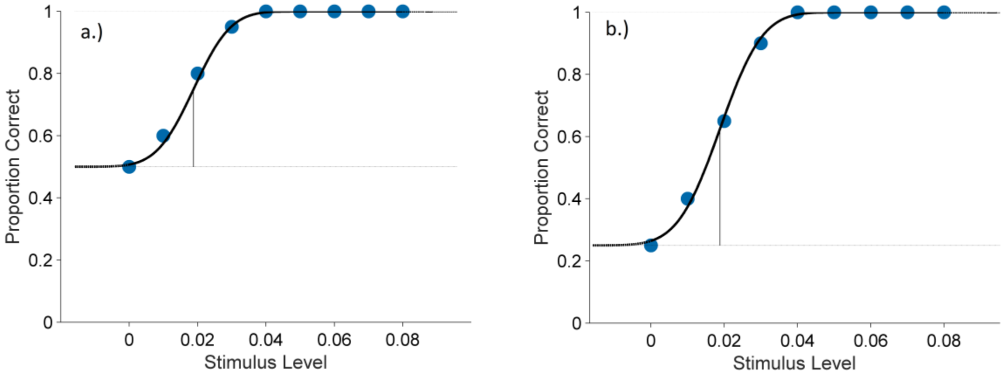


Figure 3.9. Psychometric functions for threshold estimation. Panel a: illumination discrimination (2AFC). Panel b: and chromatic detection (4AFC). *Note that the thresholds are different for proportions of correct responses, but correspond to the same probability of seeing the difference (0.5).*

3.3.2.1 Thresholds across hue direction

Averaged discrimination thresholds and their standard deviations of thresholds are shown in **Figure 3.10**. Thresholds were higher along the daylight illuminations and lowest along the orthogonal axis. A repeated measures ANOVA showed that Mauchly's test for sphericity was significant (X^2=139.2, p<.001), applied Huynh-Feldt corrected measures lead to a significant main effect for Hue (F(7.4,110.5)= 5.8, p<.001), indicating significant differences in thresholds for the illuminant directions compared in this experiment.

To examine whether thresholds were symmetrical in opponent hue directions, we compared thresholds in paired t-tests across participants for each of the 6 opponent hue directions. We applied a Bonferroni correction for the 6 tests, resulting in an alpha level of 0.0083. Although the average threshold was higher for the blue direction (0.042) than for the yellow direction (0.028) of the daylight axis, this difference was not significant (t(15) = -1.9, p = 0.08). Instead, thresholds were significantly higher for green-blue than for the reddish colours along the opponent-hue direction close to the L-M mechanism (t(15) = -3.2, p = 0.006) and close to the first diagonal of DKL-space (t(15) = -4.0, p = 0.001). This suggests that the asymmetry mainly occurs along the first diagonal and discrimination thresholds are generally higher in the third compared to the first quadrant of DKL-space (**Figure 3.10**.b).

To test whether discrimination thresholds were higher along the daylight axis than in the other directions, we calculated the average thresholds across the two directions of the daylight axis (blue and yellow), and first tested whether they were higher than the average of all other directions in a paired t-test. This was indeed the case (t(14) = 2.4, p = 0.03). To examine whether each other orientation yielded lower thresholds than the daylight axis, we compared the average along the daylight axis to the average of each single other orientation through paired t-tests. If the thresholds were elevated specifically in the direction of the daylight axis, all t-tests should be significant (i.e. no Bonferoni correction needed). However, only thresholds along the (L+M)-S axis (t(15) = 2.4, 0.03), the second diagonal (t(14) = 3.4, p = 0.004) and the direction orthogonal to the daylight axis (t(15) = 3.0, p = 0.009) were significantly lower than those along the daylight axis; but the comparison with the L-M (t(15) = 1.3, p = 0.21) and the first diagonal (t(15) = 1.2, p = 0.23) were not significant. These results suggest that illumination discrimination is generally worse along orientations across the first and third quadrant (**Figure 3.10**.b). While these include the daylight axis lower sensitivity is not specific to the daylight axis.

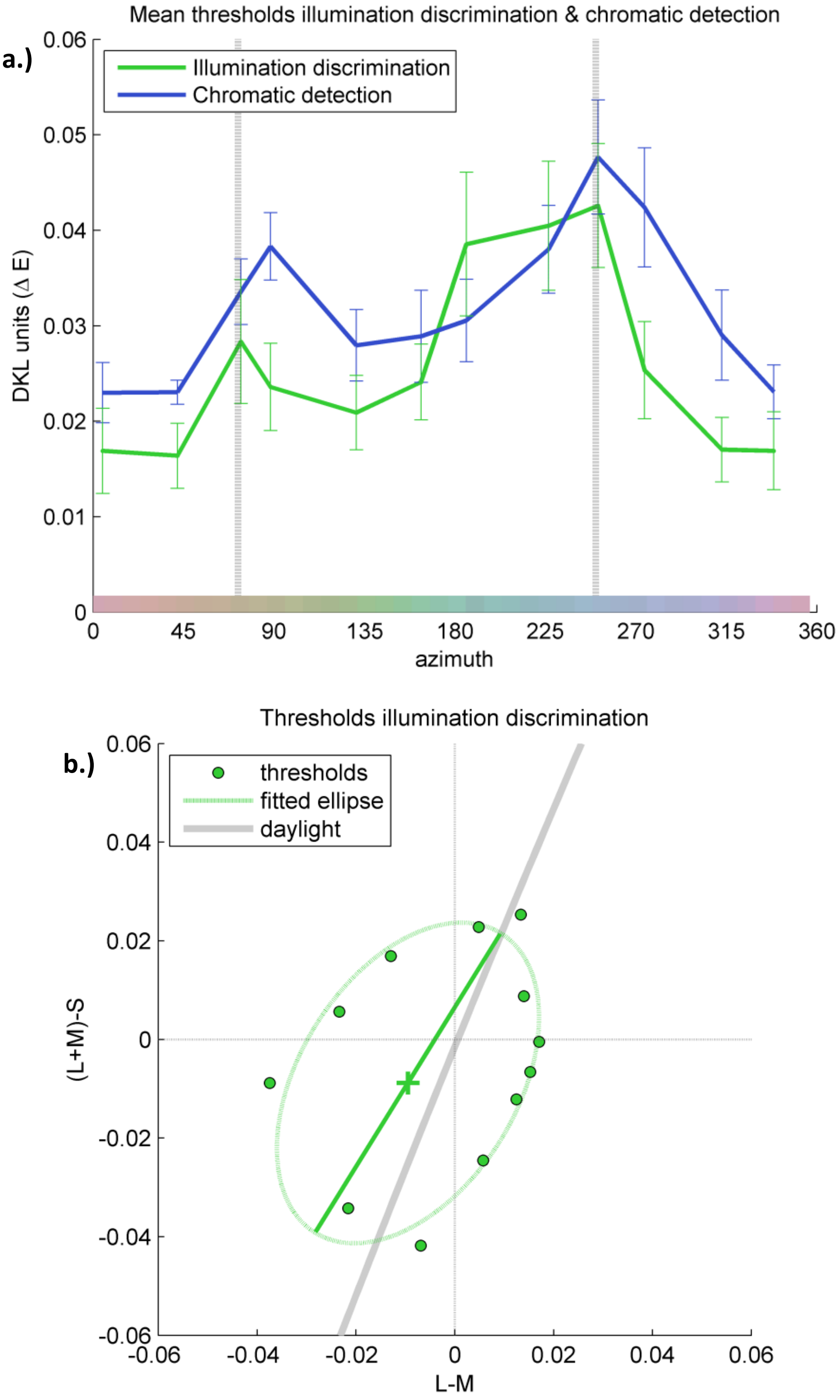


Figure 3.10. a) Discrimination thresholds for the 12 illuminations tested in the illuminant discrimination experiment given in DKL-units. Error bars indicate standard errors, dotted lines mark the daylight locus. b) Polar plot for illumination discrimination thresholds. Green dots represent discrimination thresholds. Eccentricity corresponds to the magnitude of the thresholds. The red curve shows an ellipse fitted to the thresholds and the cross corresponds to the centre of the fitted ellipse. The daylight locus is illustrated by the light grey curve in the background. Note that hues of our daylight illumination correspond to the saturated colours (blue and yellow) of the daylight locus. Since the daylight locus is bent this implies that our daylight illumination is slightly off (to the right) compared to the part of the daylight locus in this diagram.

3.3.2.2 Relative orientation of thresholds

It might be argued that the distribution of detection thresholds (green curve in Figure 3.10.b) is shifted away from the origin. In this case, the distances and hue directions should be calculated from the centroid of the distribution rather than from the origin of DKL-space. A shift of the distribution towards blue would be in line with a blue bias. We tested this idea by inspecting the shift of the centre of the distribution away from the origin. Moreover, an effect of the daylight axis might be understood as variation parallel to the daylight axis, independent of any shift. To examine this idea, we computed the direction of the shift of the centre and the orientation of the tilt of the distribution relative to the x-axis.

For this purpose, we fitted an ellipse to the data based on the direct least squares algorithm (Fitzgibbon, Pilu, & Fisher, 1999). The centre of the resulting ellipse (green cross and curve in **Figure 3.10**.b) is shifted in the 215.0 deg direction and the major axis of the ellipse is tilted by 54.0 deg. Both, the direction of the shift and the direction of the major axis are oriented more along a greenish blue to reddish yellow direction than the daylight axis (69 deg when rectified for the shift between daylight axis and origin).

3.3.2.3 Colour shifts

Contrary to the colour shifts in Experiment 1, averages (overall magnitude) and standard deviations (consistency) of surface colour shifts did not vary strongly across illumination of different hue. To test for a potential relationship between surface colour shifts and blue bias, we calculated correlations across the 12 illumination hues between the discrimination thresholds in **Figure 3.10** and the averages and standard deviations of surface colour shifts in DKL-space. However, thresholds were neither correlated with average ($r(10) = 0.24$, $p = 0.45$) nor with the standard deviations of the surface colour shifts ($r(10) = 0.07$, $p = 0.80$). This results indicates that the blue bias observed in this experiment is not due to differences in surface colour shifts across illuminant colours.

3.3.3 Discussion

Illumination discrimination thresholds were higher along cone-opponent orientations that contrast greenish blue and reddish yellow (first and third quadrant) colours. Moreover, there was an asymmetry between the greenish blue and the reddish yellow hue direction, with greenish blue illuminations leading to higher thresholds (third quadrant) than reddish yellow (first quadrant) illuminations. In contrast to experiment 1, the magnitude and the variation of colour shifts due to the illuminant changes were barely varied across the different illuminants in this experiment and did not correlate with the illumination discrimination thresholds. Biases towards the bluish hue direction were nevertheless present, indicating that these biases were not due to differences in colour shifts across illuminants of different hue. At the same time, illuminations with hue directions along the cardinal axes of DKL-space did neither yield consistently high nor low thresholds.

Although the results of this second experiment generally confirm the idea of a blue bias suggested by previous studies (Pearce et al., 2014; Radonjić et al., 2016) and the results in experiment 1; there are also considerable differences to those results. In our experiment one, the observed blue bias went in the blue direction (+S) of the cone-opponent (L+M)-S axis, and there was no bias towards the green direction (+M) of the cone-opponent L-M axis. In this second experiment we observed a bias in the +M direction of the L-M axis, but not towards the +S direction of the (L+M)-S axis. Those cone-opponent directions do not correspond to pure green and blue, but rather to a greenish blue and a purple hue direction (Malkoc, Kay, & Webster, 2005; Webster, Miyahara, Malkoc, & Raker, 2000; Witzel & Gegenfurtner, 2013).

In contrast, those previous studies (Pearce et al., 2014; Radonjić et al., 2016) compared the blue and the yellow directions of the daylight axis to the red and green orthogonal directions. These hue directions coincide with blue and yellow hue directions, and are closer to (but not precisely coincident with) typical green and red hues (Malkoc et al., 2005; Webster et al., 2000; Witzel & Gegenfurtner, 2013)(Malkoc et al., 2005; Webster et al., 2000; Witzel & Gegenfurtner, 2013). Our results are in line with the main observations of previous studies (Pearce et al., 2014; Radonjić et al., 2016) in that the blue daylight direction yielded lower sensitivity to illumination changes than the opposite and orthogonal directions (**Figure 3.10**.b). A difference to the results of Pearce et al. (2014) is that they observed particularly low discrimination thresholds in the green direction, while the thresholds in our second experiment were low in both the green and the red direction (orthogonal to the daylight axis).

However, detailed relationships between hue directions also varied between the study of Pearce et al. (2014) and the one of Radonjić et al. (2016), and even across the single experiments and conditions of Radonjić et al. (2016). This was the case even though all those measurements were designed to match one another in terms of the surface colour arrangements in the scenes and the precise illuminants. Like us, Radonjić et al. (2016) did not find lower thresholds in the green direction in many of their measurements (Experiment 2 and Experiment A). Moreover, under some conditions Radonjić et al. (2016) found comparatively low discrimination thresholds for the red direction and no blue bias when they changed the average surface colour in their scene (their Experiment 2 and Experiment A). Despite several control measurements by Radonjić et al. (2016), it remained unclear where these differences came from.

An important difference between our and previous studies is the use of cone-opponent DKL-space instead of CIELUV space to sample and represent colours (Pearce et al., 2014; Radonjić et al., 2016). A first point is that the relative scaling of the cone-opponent axes in DKL space is arbitrary and could explain differences in thresholds between the orientations of these axis. However, since the observed pattern was not specific to the comparison between the axes and since we observed pronounced asymmetries in opponent directions, we can exclude that the observed (greenish) blue bias was due to the relative scaling of the cone-opponent axes.

Moreover, while DKL-space captures signals of sensory cone-opponent mechanisms, CIELUV is better in representing perceivable colour differences. Since illumination discrimination involves the comparison of colours, it is natural to consider that the general ability to see colour differences could affect illumination discrimination thresholds. In line with this idea, the asymmetries of the illumination discrimination thresholds (**Figure 3.10**.b) look roughly similar to the detection and discrimination ellipses measured previously for uniformly coloured disks (Giesel et al., 2009, fig. 1 and 4; Krauskopf & Gegenfurtner, 1992, fig. 14). This suggests that illumination discrimination thresholds may be related to the way sensitivity changes across DKL-space. CIELUV or other existing colour appearance spaces (including S-CIELAB as in Radonjić et al., 2016) are not very accurate in estimating the sensitivity to small differences between colours. In order to compare our illumination discrimination thresholds from experiment 2 with the more general sensitivity to simple isolated colour patches, we measured empirical detection thresholds in a third experiment.

3.4 Experiment 3: Detection thresholds

Experiment 3 investigated whether there is a relationship between detection thresholds for simple uniform colours and the thresholds in tasks involving complex stimuli as in our experiment 2. For this purpose, we measured chromatic detection thresholds for isolated colour patches with the hues used for the illuminations in experiment 2.

3.4.1 Method

Detection thresholds were measured with the 4-Alternative Forced Choice task used in previous experiments (e.g. Giesel et al., 2009; Krauskopf & Gegenfurtner, 1992). In this task, a test colour is presented at one of four locations on the computer screen (cf. **Figure 3.11**), and the observer indicated the location of the colour by pressing one of four keys that correspond to the locations of the four candidate stimulus positions.

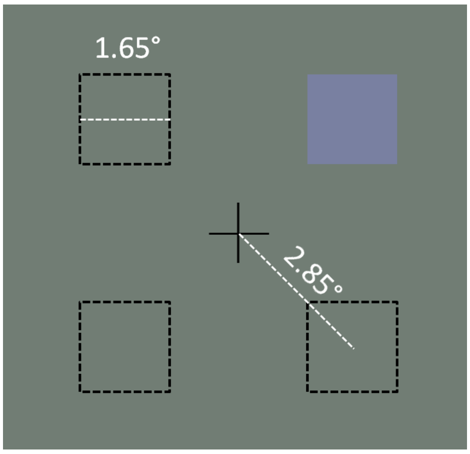


Figure 3.11. Stimulus display in the chromatic detection task of experiment 2. A fixation cross was presented in the centre of the display. In this example, the test colour is shown in the top-right location. The other candidate stimulus locations are illustrated by the dashed squares. Size and distances are indicated by the dashed lines and reported visual angles (dashed squares and lines, and specifications were not part of the actual stimulus display). Observers had to indicate the location with the test colour. The test colour was presented with different levels of chroma to determine Just-Noticeable Differences (JND) with a constant stimulus technique.

3.4.1.1 Observers and Apparatus

The same 16 observers that participated in the illumination discrimination experiment also participated in the present measurements of detection thresholds. The apparatus was the same as in the illumination discrimination experiment.

3.4.1.2 Stimuli

The stimulus display is illustrated in **Figure 3.11**. Test colours were shown on a square patch with the same size as the patches in the scene used in experiment 1 (1.65 deg visual angle). The four candidate stimulus positions were top-left, top-right, bottom-left and bottom-right, and the centre of the square was 2.85 deg visual angle from the centre of the display.

Unlike the scene in experiment 2, test colours were shown on a uniform grey background. The chromaticity and luminance of the background was the same as the lighter of the two grey colours used for the patterned background of the neutral scene in experiment 2 (xyY = [0.33, 0.34, 49cd/m^2]; cf. **Figure 3.5** and

Table S 3.1). Test colours were isoluminant with the background (49 cd/m^2).

The hues of the test colours were defined by the DKL-coordinates of the 12 illumination hues in experiment 2. More precisely, they corresponded to the hue of the lighter grey of the patterned grey background in experiment 2 under each of the 12 illuminations (cf.

Table S 3.1). Stimuli were determined with the same nine chroma levels (0:0.1:0.8) as in experiment 2 to be used in the constant stimuli method.

3.4.1.3 Procedure

One trial started with the 1000ms presentation of a fixation cross at the centre of the screen. Then the test colour was presented for 500ms at one of the four possible locations (**Figure 3.11**). To implement a smooth stimulus on- and offset, it was with a temporal Gaussian envelope. This was done because an abrupt stimulus onset might allow observers to complete the task by responding to an undefined flicker without seeing any colour. Note that the onset positions were not visible to the observers, but were learned quickly during the training trials. After stimulus presentation, the observer pressed one of four keys on a tailor-made keypad. There was no time limit for the answer.

Each test colour was presented five times at each of the four location. Presentation of the 12 hues, 9 chroma levels and four positions was randomized. The completion of the overall 2160 trials (12 hues x 9 chroma levels x 4 positions x 5 repetitions) took less than two hours since stimulus presentation was much faster than in experiment 2. The experiment was divided into two parts, each about one hour, with a break in between. Before the main experiment, observers completed two short training sessions to become familiar with the task, with one of them having clearly visible targets.

3.4.2 Results

Figure 3.9.b illustrates the determination of detection thresholds. As for the illumination discrimination data, we fitted cumulative Gaussian distributions as psychometric functions and determined thresholds with a probability of 0.5 for perceiving the difference. Since chance probability for a correct answer is 0.25, the threshold corresponds to 62.5% percent correct level in the 4-Alternative Forced Choice task. As expected, the grant average (M = 24.82%) for the chroma level of zero was very close to and did not differ significantly (t(15) = -0.2, p = 0.82) from chance (50%). As for illumination discrimination, accuracies for this condition have been replaced by chance probabilities for better estimates of psychometric functions.

3.4.2.1 Detection thresholds across hues

The blue curve in **Figure 3.10** shows the chroma detection thresholds across the 12 hue directions. We conducted a 1-way repeated measures ANOVA to test for significant differences across the twelve hues. We used Huynh-Feldt corrections as the sphericity assumption was violated (X^2=176.18, p<.001). There was a highly significant effect of hue (F(4,60.6)=6.4, p <.001) indicating that detection thresholds differed systematically across hues. As for the illumination discrimination thresholds, the highest thresholds were obtained for the blue direction of the daylight locus.

However, as for the illumination discrimination thresholds the asymmetry along opponent axes occurred rather to the greenish blue than towards the blue direction of the daylight axis. In particular, detection thresholds were not significantly higher towards the blue than towards the yellow direction of the daylight axis (t(15) = -3.0, p = 0.009) when test probability is Bonferroni corrected (alpha = 0.0083). Instead, they were significantly higher for the greenish blue direction of the first diagonal (t(15) = -4.2, p = 0.0008) and of the L-M axis (t(15) = -4.6, p = 0.0004). These results suggest a bias for detection thresholds towards the greenish blue direction similar to the one found for illumination discrimination.

For the sake of completeness, we also conducted the tests that examined whether thresholds were elevated in the direction of the daylight axis. This was true when average thresholds in the two directions of the daylight axis were compared to the average of all other directions (t(15) = 2.8, p = 0.01). When compared to each of the other five orientations, thresholds along the daylight axis were higher than those along the L-M axis (t(15) = 2.9, p = 0.01), the first (t(15) = 3.0, p = 0.008) and the second diagonal (t(15) = 2.7, p = 0.02), and the orientation orthogonal to the daylight axis t(15) = 3.1, p = 0.008. However, thresholds along the (L+M)-S axis were larger than those along the daylight axis, though this difference was not significant (t(15) = 0.07, p = 0.94). These results suggest that the variation of thresholds is not specific to the daylight axis, in that the effect cannot be disentangle from a potential effect of the (L+M)-S-axis.

3.4.2.2 Comparison with illumination discrimination thresholds

The profile of detection thresholds across hues (blue curve in **Figure 3.10**.a and **Figure 3.12**.a) looks fairly similar to the one of the illuminant discrimination thresholds from experiment 2 (green curve in **Figure 3.10**.a and **Figure 3.12**.a). To assess this similarity, we calculated a correlation across the 12 hues between the two kinds of thresholds (**Figure 3.12**.b). The correlation was positive, highly significant and explained more than 50% (51.8%) of the total variance (r(10) = 0.72, p = 0.008). To examine whether the correlation may be due to single values, we calculated robust correlations with the Robust correlation toolbox (Pernet, Wilcox, & Rousselet, 2012). All measures of robust correlations were still higher than the simple Pearson correlation. We also inspected correlations for individual datasets. Except for two observers, all correlation coefficients were positive, and four of these correlations were significant. These individual correlation coefficients were significantly larger than zero with a t-test based on Fisher's z-transports (t(15) = 4.1, p = 0.0009; mean r = 0.39). These correlations show that the illumination discrimination and detection thresholds (**Figure 3.10**.a and **Figure 3.12**.a) are very similar.

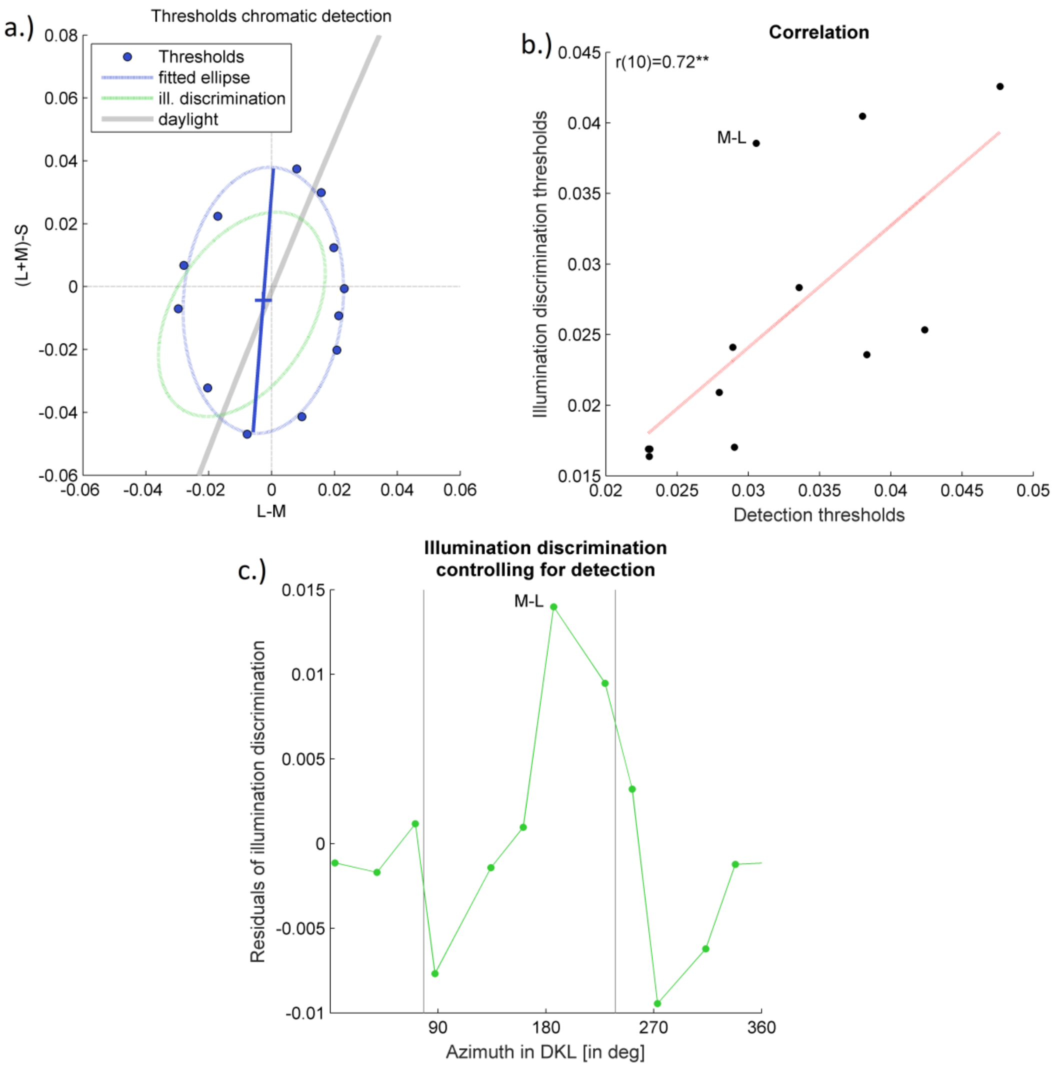


Figure 3.12. Comparisons between detection and illumination discrimination thresholds. Panel a shows detection thresholds in the format of **Figure 3.10**.b. For comparison, the average illumination discrimination thresholds (green curve) from **Figure 3.10**.b are reproduced. Panel b provides a scatter plot that illustrates the correlation between detection thresholds (x-axis) and illumination discrimination thresholds (y-axis). The red line corresponds to the regression line. Panel c plots the residuals of the illumination discrimination thresholds (y-axis) when controlling for the variation of the detection thresholds. The x-axis corresponds to azimuth in DKL-space as in **Figure 3.10**.b. To highlight the link between panel b and c we identified the data point that corresponds to the M-L pole of the L-M axis by the annotation in both panels. *Note the high degree of similarity between illumination discrimination and detection thresholds (panels a and b) and the fact that residuals of illumination discrimination peak at the M pole of the L-M dimension rather than the daylight axis (vertical grey lines).*

The question arises what kind of variation is specific to illumination discrimination and cannot be predicted by detection thresholds. This is illustrated by **Figure 3.12**.c, which shows the residual variation across hue of the illumination discrimination thresholds when controlling for the variation captured by the detection thresholds. The residuals peak at the green-blue M-pole of the L-M axis, which reflects the difference in overall orientation of the two distributions of thresholds (**Figure 3.12**.a): while detection thresholds (blue curve) are rather oriented towards the S-pole of the (L+M)-S axis, illumination discrimination thresholds are oriented in a direction between both the S-pole of the (L+M)-S axis and the M-pole of the L-M axis. According to these observations, illumination discrimination is biased towards green-blue when accounting for the general ability to detect colours.

In order to compare the orientation of thresholds for illumination discrimination and detection, we fitted ellipses to the datasets of each individual observers. We used a paired t-test across observers to examine whether there is a difference between the orientations of the ellipses for illumination discrimination and those for hue detection. However, the difference was not significant (t(15) = 1.2, p= 0.26).

3.4.2.3 Comparison of relative orientation of thresholds

The average of detection thresholds is shifted away from the origin towards the blue direction. As for the illumination discrimination thresholds, we fitted ellipse to the detection thresholds (blue curve in **Figure 3.12**.a). The centre of this ellipse is shifted away from the origin towards the blue direction, namely towards 234.9 deg and the major axis of the ellipse was tilted towards 83.3 deg. These measures show that the shift and the rotation of the detection thresholds were much closer aligned with the daylight axis than it was the case for the illumination discrimination thresholds.

The centres of detection thresholds and illumination thresholds are shifted to different degrees and in slightly different directions. To account for differences in these shifts, we recalculated the two kinds of thresholds as the distances from the respective centres, and the correlations between the resulting thresholds. There was again a positive correlation across hues (r(10) = 0.59, p = 0.04), but it explained less variance than the one observed with the original thresholds (r(10) = 0.72, see above). The fact that this correlation explains less variance than the one observed for the original thresholds suggests that the remaining differences between the distributions of the two kinds of thresholds may not be explained by differences in shifts.

So, we went one step further and accounted for both shifts and rotations together. To do so, we shifted and rotated the two kinds of thresholds so that the centre and the orientation of the two corresponding fitted ellipses (green curve in **Figure 3.10**.b and and blue and green curve **Figure 3.12**.a) were the same. We then calculated the thresholds as the distances from the (common) centre of the ellipses. The correlations between the recalculated illumination discrimination and detection thresholds explained 73% of the variance and was highly significant (r(10) = 0.85, p = 0.0004). This correlation indicates the similarity of the shapes of the two distributions when accounting for shifts and rotation. The observed correlation explains about 21% more than the correlation between the original thresholds (51.8%), indicating that this amount of variance is due to differences in shifts and rotation. At the same time, the 27% of unexplained variance (100%-73%) must be attributed to the difference in shape (**Figure 3.12**.a).

3.4.3 Discussion

In sum, there are some differences in orientation and in shape between the distributions of the two kinds of thresholds across hue directions. These differences notwithstanding, there is a high degree of similarity between detection and illumination discrimination thresholds.

3.4.3.1 Detection thresholds

We compared detection thresholds to those obtained in previous studies. Overall, our thresholds for the eight hue directions measured by Giesel et al. (2009) had on average (0.032) a similar size as those from panel a (0.028) and panel b (0.032) of Giesel et al. (2009)'s **Figure 1** and did not differ significantly (both $p > 0.29$). However, our detection thresholds were elongated along the (L+M)-S axis (**Figure 3.12**.a), whereas those of Giesel et al. (2009) were mainly elongated along the first diagonal. As a result, there was no significant correlation between our and their measurements even after converting them into a DKL space with the same white-point and axis scaling ($r(6) = 0.57$, $p = 0.14$ and $r(6) = 0.19$, $p = 0.65$).

At the same time, our measurements were completely in line with those obtained by Krauskopf and Gegenfurtner (1992, fig. 14). The apparent difference between their and our data is merely due to axes scaling. Their axes were scaled by detection thresholds, resulting in a circular distribution of detection thresholds instead of an ellipse. Our detection thresholds were higher along the S-axis. Scaling the axes by the average thresholds along the axes compensates for elongations along the axis. As a result of this, the distribution of our detection thresholds is also approximately circular (blue line and disks in **Figure 13**), and hence similar to the one found by Krauskopf and Gegenfurtner (1992, fig. 14). But rescaling cannot compensate for the elongation along the diagonal in the Giesel et al. (2009)'s data (squares and diamonds in **Figure 13**).

There were only three differences between the implementation of our task and the one of Giesel et al. (2009). First, the luminance of their stimuli was about 10 cd/m^2 lower than ours; but that was also true for the ones used by Krauskopf and Gegenfurtner (1992). Second, the stimulus displays in our study and in Krauskopf and Gegenfurtner (1992) were half as large. However, this should not be of great importance because large parts of the stimuli of Giesel et al. (2009) were within the range of foveal presentation. Finally, we and Krauskopf and Gegenfurtner (1992) presented our stimuli with a smooth stimulus on and offset (i.e. a temporal Gaussian envelope) while Giesel et al. (2009) used an abrupt on and offset. We do not see any apparent reason why this should systematically influence the variation of thresholds across hues; however, in principal it could be possible that timing has such an effect because perceptual mechanisms may differ in their latencies and speed (e.g. Lee, Mollon, Zaidi, & Smithson, 2009; Stromeyer, Eskew, Kronauer, & Spillmann, 1991).

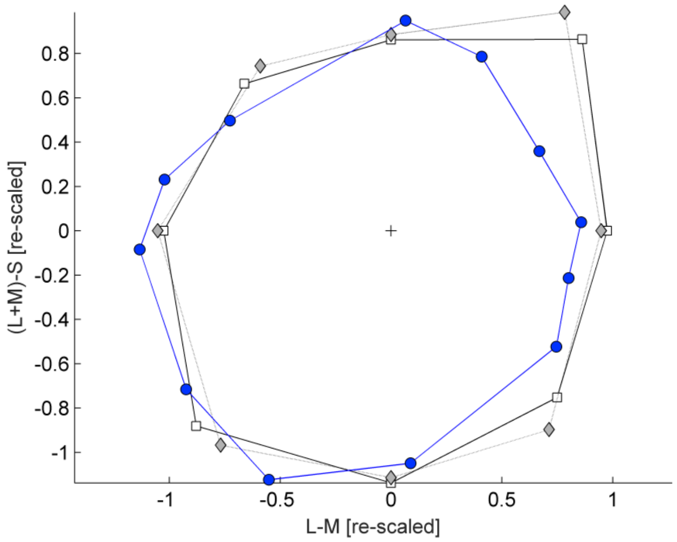


Figure 13. Comparison of detection thresholds. X- and Y-axes correspond to L-M and (L+M)-S as in **Figure 3.12**, but axis scaling is relative to the detection threshold along the respective axis. The white squares and the solid curve shows detection thresholds from panel a, the grey diamonds and the dotted line those from panel b of Figure 12 of Giesel et al. (2009). The blue disks and curves represent the data from the present studies as in Figure 12, but rescaled. Note that the present data is approximately circular after rescaling, while the data from Giesel et al. (2009) is elongated along the first diagonal.

3.4.3.2 Detection and illumination discrimination

Differences across studies notwithstanding, the relationship between detection thresholds and illumination discrimination thresholds observed in our study seem not to depend on a close match between the stimuli and tasks for the two kinds of measurements. The illumination discrimination experiment involved complex stimulus displays with coloured and patterned backgrounds, whereas the detection experiment used simple colour patches on a uniform grey background. Moreover, the illumination discrimination experiment used a successive 2AFC recognition task. This task requires some memory performance because of the delay between reference and comparison stimuli. In contrast, detection thresholds were measured with a 4AFC task with no implication of memory. Despite all these differences we observed a strong correlation between the two kinds of thresholds. This relation might be even higher, if the tasks of experiments 2 and 3 were more similar.

This correlation suggests that the illumination discrimination task reflects, to a large extent, how well observers can detect simple colour changes in different hue directions that are not specific to illumination changes. The illumination discrimination task is meant to measure the extent to which the influence of illuminations on a scene is discriminable (Aston et al., 2016; Pearce et al., 2014; Radonjić et al., 2016). However, before it is possible to conclude about the detectability of an illuminant change, it has to be proven that that the task actually measures illumination, and not just sensitivity towards changes in saturation of different hue directions. As thresholds of the illumination discrimination experiment highly correlate with the pattern of thresholds in the chromatic detection task in our study, it seems likely that a major part of the illumination discrimination task can be performed by just taking the change of a single patch in the scene into account.

In our experiments, sensitivity to chroma across hue directions explains the blue bias in illumination discrimination by 50% of the variance. This suggests that the blue bias is not specific to illumination changes, but is inbuilt in the way colour information is processed. When accounting for the effect of sensitivity to chroma illumination discrimination in Experiment 2 showed a bias towards the greenish blue pole of the L-M mechanism (**Figure 3.12**.c) instead of the blue direction of the daylight axis. According to these results, it seems unlikely that this bias is due to the observers' experience with daylight variations (Winkler et al., 2015).

3.4.3.3 Comparison to other studies on illumination discrimination

In contrast to our findings previous studies reported a blue bias specifically for illumination discrimination (Aston et al., 2016; Pearce et al., 2014; Radonjić et al., 2016). Three main sources of variation across studies may explain the different findings.

One possible source of difference across studies is the use of different colour spaces. While we used DKL-space to inspect the role of low-level perceptual differences, those previous studies (Aston et al., 2016; Pearce et al., 2014; Radonjić et al., 2016) sampled and represented colours in colour appearance spaces, such as CIELUV. CIELUV may provide a better control of sensitivity across hues than DKL-space. However, these spaces as very general and only roughly approximate in their or control of sensitivity and sensitivity is not fully controlled in these spaces either. In contrast, we measured detection thresholds empirically to examine the role of sensitivity across hue directions. Moreover, the differences in space imply that illuminations have been sampled in slightly different direction.

Another source of difference may be differences in the scenes used in our experiment 2 and in those previous studies. In those previous studies (Aston et al., 2016; Pearce et al., 2014; Radonjić et al., 2016) colour patches were irregularly arranged across the scene and involved a large range of chroma. In contrast, the patches in our scenes were regularly arranged and involved desaturated colours, including grey. It might be that the presence of desaturated colours furthers a correlation between detection and illumination discrimination, in particular since desaturated colours shift more evenly under a different of illuminations. The idea that the distribution of surface colours matters for the illumination discrimination across hue is further supported by the findings of (Radonjić et al., 2016). However, the mere fact that surface colours modulate illumination discrimination thresholds suggests that observed patterns are not specific to illumination changes, and hence further support our conclusion.

Moreover, differences in the results across studies may be due to different sets of illuminations. In particular, previous studies sampled illuminations with at least partly different hues than ours (Aston et al., 2016; Pearce et al., 2014; Radonjić et al., 2016). However, those studies used small sets of maximally five illumination colours and such small sets only allow for very coarse comparisons across hues. In contrast, we measured illumination discrimination for a whole range of hues, including the hue directions of the daylight axis. These fine-grained measurements allowed us more general insights into how illumination discrimination is modulated across the hue circle. It is well possible that previous studies missed the greenish blue tendency of the putative blue bias simply because they did not sample illuminations in that hue direction.

Still more important than the precise hues of the illumination samples are differences in illuminant spectra. Previous studies (Aston et al., 2016; Pearce et al., 2014; Radonjić et al., 2016) and also our experiment 1 used illuminant spectra of LED light sources, whereas our simulated spectra were artificial and tailor-made to be as smooth as possible. The way colour signals from surface colours are shifted under illumination change depends on the spectra of the illuminants. Maybe, the LED spectra are more prone to produce a blue bias, while our spectra tend towards a more greenish bias. It is an important question to clarify the precise role of the illuminant spectra for illumination discrimination, and more generally the perception of illumination changes. With respect to the role of illuminant spectra, it seems most important to investigate the effects of illumination spectra similar to those that occur in the natural environment.

In any case, detection thresholds do not involve any illuminant spectra. Hence, our main observation of a relationship between sensitivity and illumination discrimination cannot be specific to our particular set of illuminant spectra. The observation of that relationship highlights that observed differences in illumination discrimination might well be due to differences in the sensitivity to hue. We think it is important that the role of sensitivity in illumination discrimination needs to be accounted for, before conclusions about biases in the perception of illumination or in colour constancy may be drawn.

3.5 Conclusion

This study investigated the ability to perceive illumination changes depending on the hue direction of the illumination changes. Results obtained with different methods showed that observers are comparatively insensitive to illumination changes towards bluish hues, hence confirming the “blue bias” found in previous studies (Aston et al., 2016; Brainard, 2016; Pearce et al., 2014; Radonjić et al., 2016; Winkler et al., 2015; Wuerger et al., 2015). The observed blue bias, however, was not specific to illumination changes, but could also be found in detection thresholds. In fact, detection thresholds were highly correlated to illumination discrimination thresholds. These findings suggest that measures of illumination discrimination are strongly, but not exclusively, determined by the general sensitivity to hue. There was still a bias in illumination discrimination when accounting for the sensitivity to hue in our data. However, this bias was towards the green-blue hue direction, rather than towards the blue pole of the daylight axis as suggested previously.

Chapter

4 Determinants of colour constancy and the blue bias

A similar version of this manuscript has been submitted as:

Weiss, D., Witzel, C., & Gegenfurtner, K. (under review). Determinants of colour constancy and the blue bias.

We investigated several sensory and cognitive determinants of colour constancy across 40 illumination hues. In a first experiment, we measured colour naming for the illumination and for the colour induced by the illumination on the colorimetric grey. Results showed that the induced colours are approximately complementary to the colour of the illumination. In a second experiment, we measured colour constancy using achromatic adjustments. Average colour constancy was perfect under the blue daylight illumination and decreased in colour directions away from the blue daylight illumination due to undershooting and a strong blue bias. Apart from this blue bias, colour constancy was not related to illumination discrimination and to chromatic detection measured previously with the same set-up and stimuli. We also observed a strong negative relationship between the degree of colour constancy and the consensus of naming the illumination colour. Constancy coincided with a low naming consensus, in particular because bluish illumination colours were sometimes seen as achromatic. Blue bias and category consensus alone explained >65%, and all determinants together >75% of the variance of achromatic adjustments. These findings suggest that colour constancy is optimized for blue daylight and that the perception of the illumination does not determine colour constancy

4.1 Introduction

While several important mechanisms and cues to colour constancy are known, it is far from being fully understood (Foster, 2011; Kraft & Brainard, 1999; Smithson, 2005). For example, it is still a matter of debate how colour constancy varies across different illumination colours and whether constancy is particularly tuned to certain illuminations. Here, we evaluated potential determinants that modulate colour constancy in scenes with many densely sampled hue directions.

4.1.1Background

It is frequently hypothesized that colour constancy is higher for illuminations varying along the daylight locus, where hue direction of natural illumination varies between yellow and blue over the course of the day. Thus, due to common experience with such colour variation, constancy is expected to be most proficient with changes along the yellow-blue direction (Delahunt & Brainard, 2004; Shepard, 1992).

An alternate hypothesis proposes that the exposure to the frequent variation of colours along the daylight axis produces uncertainty of colour appearance (Beer, Dinca, & MacLeod, 2006; Bosten et al., 2015; Gegenfurtner et al., 2015; Lafer-Sousa et al., 2015; Witzel et al., 2011). Moreover, several studies observed that bluish illuminations tend to be perceived as neutral, indicating an asymmetry of colour constancy towards the blue direction of the daylight axis (blue bias) (Aston et al., 2016; Pearce et al., 2014; Radonjić et al., 2016; Weiss, Witzel, & Gegenfurtner, under review; Winkler et al., 2015; Wuerger et al., 2015). It has been suggested that bluish illumination might be mistaken for shadows (Winkler et al., 2015), which can be bluish in the natural environment due to Rayleigh scattering (Churma, 1994; Troscianko et al., 2009).

Previous findings concerning the relationship between colour constancy and daylight were contradictory. Some studies found evidence for higher colour constancy for blue illumination colours (Daugirdiene et al., 2016; Delahunt & Brainard, 2004), while others did not find differences across illumination hues (Brainard, 1998; Hansen et al., 2007; Olkkonen et al., 2009, 2010; Schultz et al., 2006) and some found even better constancy for illumination hues other than blue (de Almeida et al., 2004). Moreover, studies investigating the perception of illumination found that observers had difficulties detecting changes towards bluish illuminations (Aston et al., 2016; Pearce et al., 2014; Radonjić et al., 2016). They suggested that the inability to see an illumination change is an indication of colour constancy. However, one might also make the opposite case and claim that the inability to see an illumination change implies an insensitivity to colour differences, which would undermine colour constancy. Taken together, it is still unclear how colour constancy relates to the variation of daylight.

A third hypothesis to the variation of colour constancy across colours, called *categorical colour constancy,* suggests that colour constancy is related to colour categories. Colour categories are the ensembles of colours designated by colour terms, such as “red”, “purple” or “blue”. It is yet unknown what determines colour categories and whether and how they are related to colour perception.

Previous research suggests that colour constancy might be highest around the centres of colour categories (Olkkonen et al., 2009, 2010). However, Olkkonen and colleagues (2009, 2010) did not measure constancy for individual points in colour space, but the constancy of colour categories and their borders across illuminations. Kulikowski and Vaitkevicius (1997) measured colour constancy with an asymmetric matching technique. They found local peaks of colour constancy for typical red, yellow, and blue, and to a lesser extent for green. This finding is substantiated by the observation that surfaces with the prototypical colours of categories have particular physical properties (*sensory singularities*) that make the sensory colour signal more predictable across illumination changes (Philipona & O’Regan, 2006; Vazquez-Corral, O’Regan, Vanrell, & Finlayson, 2012; Witzel et al., 2015).

These findings suggest that colour categories developed around the colours that are most stable under illumination changes and hence could serve as “perceptual anchors” under changing illumination (Kulikowski & Vaitkevicius, 1997; Witzel et al., 2015; Witzel, Maule, & Franklin, 2013). The idea of perceptual anchors also fits to a recent observation according to which memorised colours are shifted towards category prototypes (Bae et al., 2015).

Relational colour constancy (Foster et al., 1997; Foster & Nascimento, 1994; Nascimento, de Almeida, Fiadeiro, & Foster, 2004) is another important hypothesis to explain both the variation colour constancy across surface and across illumination colours. According to relational colour constancy, observers use cone-ratios to accomplish colour constancy, because these ratios are largely invariant across illumination changes. Predictions based on cone-ratios vary depending on the surface colours in a stimulus display and the illuminations, and might explain variation in colour constancy.

More recently it was suggested that colour constancy is related to *metamer mismatching* (Logvinenko et al., 2015; Witzel et al., 2016). Metamer mismatches describe the phenomenon that surfaces that are metameric under one illumination can result in different colours under another illumination (Burns, Cohen, & Kuznetsov, 1989; Cohen & Kappauf, 1982; Logvinenko et al., 2014; Wyszecki, 1958).

Witzel and colleagues (2016) claimed that higher volume of metamer mismatches (*metamer mismatch volume*) leads to higher uncertainty about a colour under illumination change, and thus weaker colour constancy. They measured colour constancy through asymmetric matching and found a strong relationship between colour constancy and metamer mismatching.

4.1.2 Objective

Taken together, different studies suggest very different determinants of colour constancy across colours, and the question arises how these diverse findings are related. One issue that makes it problematic to compare different studies is that a very limited number of illumination hues were used (mainly four; eight in Brainard, 1998) and that illuminations differed across studies. Some studies investigated illumination hues along and orthogonal to the daylight axis (de Almeida et al., 2004; Delahunt & Brainard, 2004). Others investigated illuminations with colours along the DKL-axes (Hansen et al., 2007; Olkkonen et al., 2009, 2010), which are oriented towards distinctly different hue directions. In particular, the +S endpoint of the so-called blue axes appears purple rather than blue (Malkoc et al., 2005; Webster et al., 2000; Witzel & Gegenfurtner, 2013, 2015). Another problem that makes comparisons of colour constancy across illumination colours difficult is that the shift of the sensory colour signal due to illumination changes does not just depend on the colour of the illumination, but on the actual spectra of the reflectances and the illuminants. The precise magnitude of the shift due to the illumination is particularly important when evaluating colour constancy through colour constancy indices.

In this study, we tested the candidate determinants of colour constancy across a large number of illumination hues. We used an achromatic adjustment method, which allowed us a high degree of control of experimental conditions. We created two-dimensional variants of the configuration introduced by Purves & Lotto (2002) with illuminations simulated in 40 hue directions. This configuration Purves-Lotto provide a striking illustration of the strong effects of colour induction in rendered scenes under simulated illuminations. The large number of directions allowed us to investigate how colour constancy changes as a function of hue, while also enabling statistical comparisons across conditions of interest. To control the effect of the illumination colours on the sensory colour signal, parameters of reflectance and illuminant spectra were matched to produce colorimetric grey in each of the 40 displays. Moreover, this design made it possible to investigate the role of illumination colours and the role of the induced colours on the colorimetric grey patches. Finally, we designed the study to match the set-up used for the measurement of perceived illuminations in a companion study on the perception of illumination colours (Weiss et al., under review). This allowed us to test in how far the variation of constancy across hue can be explained by how observers perceive the illumination.

In a first experiment, we compared colour categories for illumination colours and the complementary colours induced by the illumination on the colorimetric grey patch. In the second experiment we measured achromatic adjustments, and tested the role of the daylight locus, the blue bias, categorical colour constancy, metamer mismatching, sensory singularities, and relational colour constancy.

4.2 Colour Naming

This experiment provided the colour categories for stimuli in the achromatic adjustment experiment (see below), which allowed us to examine the relationship between colour categories and colour constancy. Moreover, this experiment allowed testing the idea that the colours induced in colour constancy are opponent to the inducing colours of the illumination as predicted by second-stage mechanisms.

4.2.1 Method

4.2.1.1 Observers

Colour naming was measured for 30 German observers (27 women, 22±2years). Observers were students at the Justus-Liebig-University as part of an experimental course. All participants were tested for normal colour vision using Ishihara plates (Ishihara, 2004). All experiments were carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) and were approved by the local ethics commission (LEK 2015-0015). Informed consent was obtained from our participants.

4.2.1.2 Apparatus

Stimuli were presented on an EIZO CG2420 monitor driven by an AMD FirePro V4900 with a resolution of 1.920 x 1.200 pixels, and a colour resolution of 8 bit per channel. The Monitor was calibrated using a Konika Minolta CS2000 Spectroradiometer (***Konica Minolta*** Sensing Inc., Singapore), CIE-xyY specifications of the channels were: R = [0.685, 0.311, 23.4]; G = [0.216, 0.725, 67.8]; B = [0.151, 0.046, 5.7]. All stimuli used in the experiment have been gamma corrected. The Monitor was placed in a black painted tunnel, 50cm away from the participant.

The numpad of the keyboard was used for entering responses. The respective keys were marked by the initials of the colour terms, and a printed scheme was also available displaying the complete colour terms in the spatial arrangement of the response keys.

Experiments were programmed in Matlab 2012b (The MathWorks Inc., 2007), using the psychophysics toolbox 3 extensions (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997).

4.2.1.3 Stimuli

Figure 4.14 illustrates our stimulus display that was inspired by the Purves-Lotto cubes (Lotto & Purves, 2002, fig. 9). It consisted of a large square (the “scene”) composed of 7x7 small coloured squares (the “patches”) embedded in a background with naturalistic luminance noise (i.e. “brown” noise with an amplitude of $1/f^{1.7}$). This display was rendered with a neutral achromatic (**Figure 4.14**.a) and 40 chromatic illuminants (**Figure 4.14**.a).

The particularity of this display is that the test patch in the very centre had the same colorimetric colour signal [x = 0.327, y = 0.342, Y = 48.70 cd/m^2] under all 41 illuminations. Embedded in a scene under the (simulated) neutral illumination this colour signal appears grey; but when the (simulated) illumination is chromatic this same colour signal appears chromatic due to colour induction and colour constancy.

The challenge in the creation of these displays consists of determining pairs of reflectance and illuminant spectra that all result in the same colorimetric grey, while controlling perceptual parameters of the colours, such as hue and chroma. Moreover, we wanted realistic reflectance spectra and smooth illuminant spectra.

To obtain realistic reflectance spectra we determined the spectra based on the reflectance spectra for matte Munsell spectra (Munsell Color Services, 2007). For this, we retrieved the Munsell spectra from the Jeonsuu color group (Kohonen et al., 2006; Parkkinen et al., 1989), which are now available via the University of Eastern Finland: http://www.uef.fi/en/web/spectral/munsell-colors-matt-spectrofotometer-measured).
Since these reflectances do not include achromatic reflectance spectra we used the spectra for Neutral 6.5 and Neutral 5 from the MacBeth ColorChecker (McCamy et al., 1976). These achromatic reflectances were used for the surround with the noise pattern (**Figure 4.14**.a), and the lighter of the two (Neutral 6.5) also defined the colour of the test patch under the neutral illumination. The other 40 reflectances were defined by the 40 Munsell Hues and Munsell Value 7. The latter matched the lightness of the light grey test patch (Colour Checker Neutral 6.5). To control perceived chroma, we linearly interpolated the reflectances for each Munsell hue so that the colour signal resulting of all chromatic reflectances under the neutral illumination (xyY_{Judd} = [0.3265, 0.3419, 136.0]) formed a hue circle in DKL-colour space. The size of this hue circle was defined by the criterion that the colour signals of all reflectances had to fit into the monitor gamut under all 41 illuminations. Note that the illuminations that shift the colour signal of the equally saturated reflectances to colorimetric grey do not have equal chroma; we come back to this in the experiment on achromatic adjustments (see below).

In order to obtain smooth illuminant spectra, we created the 41 illuminants based on Gaussian functions. We used a minimisation algorithm to fit the parameters of the Gaussian functions so that the resulting illuminants cancel the colour signal of the complementary Munsell-like reflectance and hence yielded the colorimetrically grey colour signal for that reflectance (xyY_{Judd} -coordinates of the illuminants are given in **Table S4.1**).

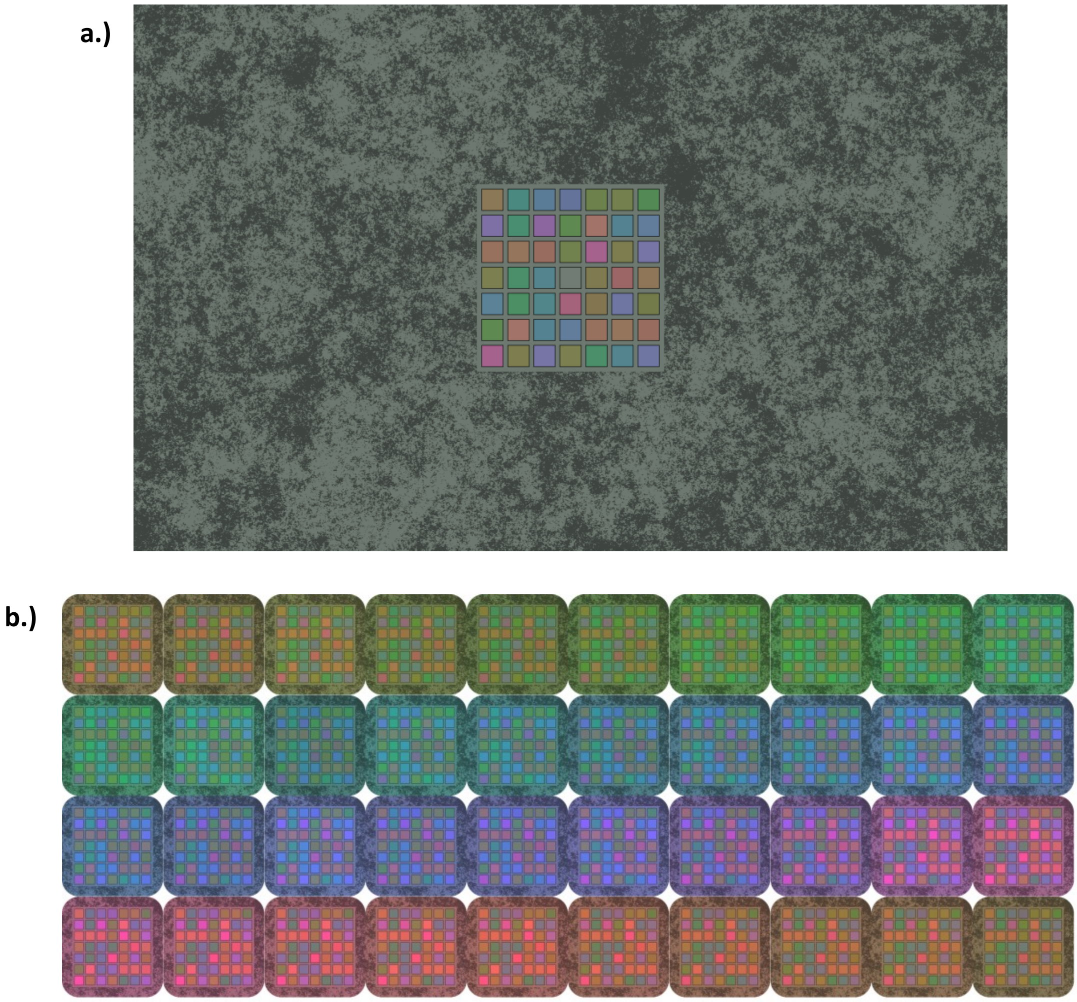


Figure 4.14. Stimulus display. a.) Scene with background under the neutral illumination. b.) Isolated scene (without background) under each of the 40 chromatic illuminations. For a better overview only the central checkerboard of the scene is depicted in panel b, but in the experiment, all scenes were presented with complete illuminated background as depicted in panel A.

4.2.1.4 Procedure

There were two versions of colour naming. In one version, observers were asked to name the colour of the colorimetrically grey patch in the centre of the display. This task provided data on how observers categorise the colour appearance induced by context and background based on the simulated illumination. In the second version, observers were asked to name the colour of the background that reflects the colour of the illumination. The presentation of version order was determined randomly.

In both versions, the 41 images were presented one at a time in a random order. A trial began with the presentation of a fixation point for 500ms, followed by the presentation of the scene until a response was given. Observers could enter a response by pressing one of 11 keys, corresponding to the German Basic Colour Terms: *Rosa* (pink), *Rot* (red), *Orange* (orange), *Gelb* (yellow), *Grün* (green), *Blau* (blue), *Lila* (purple), *Braun* (brown), *Schwarz* (black), *Grau* (grey), and *Weiß* (white). For each version of the naming task, the complete set of images was presented three times in three consecutive blocks, separated by a short-break. Overall, the measurements for both versions took about 15min.

4.2.2 Results and Discussion

Figure 4. illustrates the aggregated colour categories obtained from the two versions of the colour naming task. To calculate the azimuth, the grey of the background and test surface were used as the origin. For further details, the corresponding individual naming data may be found in **Figure S 4.4** of the Supplementary Material. The data in **Figure 4.** has been aggregated by determining the mode colour term for each stimulus display. Category membership is uncertain at the boundaries and category boundaries are not sharp and clear-cut (e.g. Olkkonen et al., 2010, fig. 8; Witzel & Gegenfurtner, 2013, fig. 6; Witzel, Hansen, & Gegenfurtner, 2008). Hence, we determined the boundaries at the hue that had a probability of 50% of being included in one or the adjacent colour category (as in Witzel & Gegenfurtner, 2013, fig. 7).

4.2.2.1 Category membership

Figure 4..a illustrates category consensus and average response times for naming the illumination colour reflected off the background. Consistency and average response times are measures of the uncertainty of category membership (Olkkonen et al., 2010; Witzel et al., 2008). With higher uncertainty towards category centres consistencies increase, and response times decrease, and vice versa towards category boundaries. As a result, these measures are negatively correlated (Witzel et al., 2008). Consistency and response times for the background naming in our study are highly correlated across colours ($r(39) = -.73$, $p < 0.001$).

Figure 4..b shows the category consensus (consistency of naming across observers) and response times for naming the induced colour of the test patch in the centre of the display. As a hue coordinate for the induced colour, the hue opponent to the illumination hue is shown along the x-axis. For the induced colours, consistency and average response times were also highly correlated across the 41 colours ($r(39) = -.69$, $p < 0.001$), indicating that there was a clear consensus of category membership across observers.

Moreover, consistency for naming the background was significantly lower than consistency for naming the induced colours of the test patch ($t(80) = 2.8$, $p = 0.006$). This is noteworthy because the colours in the background were really chromatic while those of the patch are induced colours. One could have expected induced colours to be more elusive and less consistent, but these results show that this is actually not the case. At the same time, lower consistency for naming of the illumination colours in the background can be explained by partial adaptation to the background colour. Adaptation desaturates the colours and desaturated colours are named with lower consistency (Olkkonen et al., 2010; Witzel, 2016; Witzel et al., 2015).

Another point is noteworthy about the lower consistency of background naming. Consider **Figure S 4.4**.b. Blue colours are comparatively often described by achromatic colour terms (grey or white). Other hues are never categorised as achromatic. This is in line with the idea of a blue bias in the perception of illuminations as observed previously (Aston et al., 2016; Pearce et al., 2014; Radonjić et al., 2016; Weiss et al., under review). At the same time, blue colours are never described by achromatic colour names for the induced colours of the test patch (**Figure S 4.4**.a), suggesting that the blue bias is specific to the perception of the illumination.

4.2.2.2 Opponency of induced colours

We examined whether colour categories for induced colours are rotated by 180 degrees in DKL-space compared to the categories of inducing colours. **Figure 4.**.c allows for comparing the colour categories obtained for the induced and for the inducing colours of the test patch and the background, respectively. In general, categories for induced and inducing colours closely correspond to each other in the upper and lower part of **Figure 4.**.c.

However, there were also differences. The main difference occurred for the yellow category, which is much smaller for patch than for background naming. For each observer we calculated differences between the boundaries of the two kinds of naming, and calculated t-test across observers to establish whether the differences were significant. It must be noted that this test is subject to additional noise due to the fact that different observers employed different sets of categories, resulting in different kinds of boundaries, e.g. brown-green and yellow-green (cf. **Figure S 4.4**.a-b). We only consider results that are consistent across the two tests for adjacent boundaries (e.g. yellow-green vs. green-yellow). The yellow-orange boundary was significantly different in both tests (both $p < 0.001$). The yellow-green, the green-blue, and the purple-pink boundaries were only significant in one of the tests due to occasional grey, brown, and red naming (see **Figure S 4.4**.a-b for details).

To assess how close these naming sets are to categorisation of simple uniformly coloured patches, we compared them to the categories obtained in a previous study (see Witzel & Gegenfurtner, 2013, fig. 9). Despite slight differences in the white-point (origin of DKL-space), category boundaries for uniformly coloured patches (dotted vertical lines) were close to those for patch and background naming.

In sum, induced colours are consistently named and show clear category memberships. There was also some evidence for a blue bias for perceived illuminations (Aston et al., 2016; Pearce et al., 2014; Radonjić et al., 2016; Weiss et al., under review). Moreover, the hues of induced colours can be roughly approximated by the opposite hue direction in

DKL-space. As a result, the hues opponent to the illumination hues may be used as a measure of hue for the appearance of induced colours on the patches.

Finally, the comparatively small yellow category for induced colours may be explained by the idea that blue illuminants are less saturated and weaker inducers. This idea is supported by two instances of grey naming for the orange-yellow colours that are opponent to bluish illuminations (**Figure S 4.4**.a). This is in line with the observation that the blue bias in illumination discrimination is related to the anisotropy of colour space that is reflected in the variation of sensitivity across hues (Weiss et al., under review).

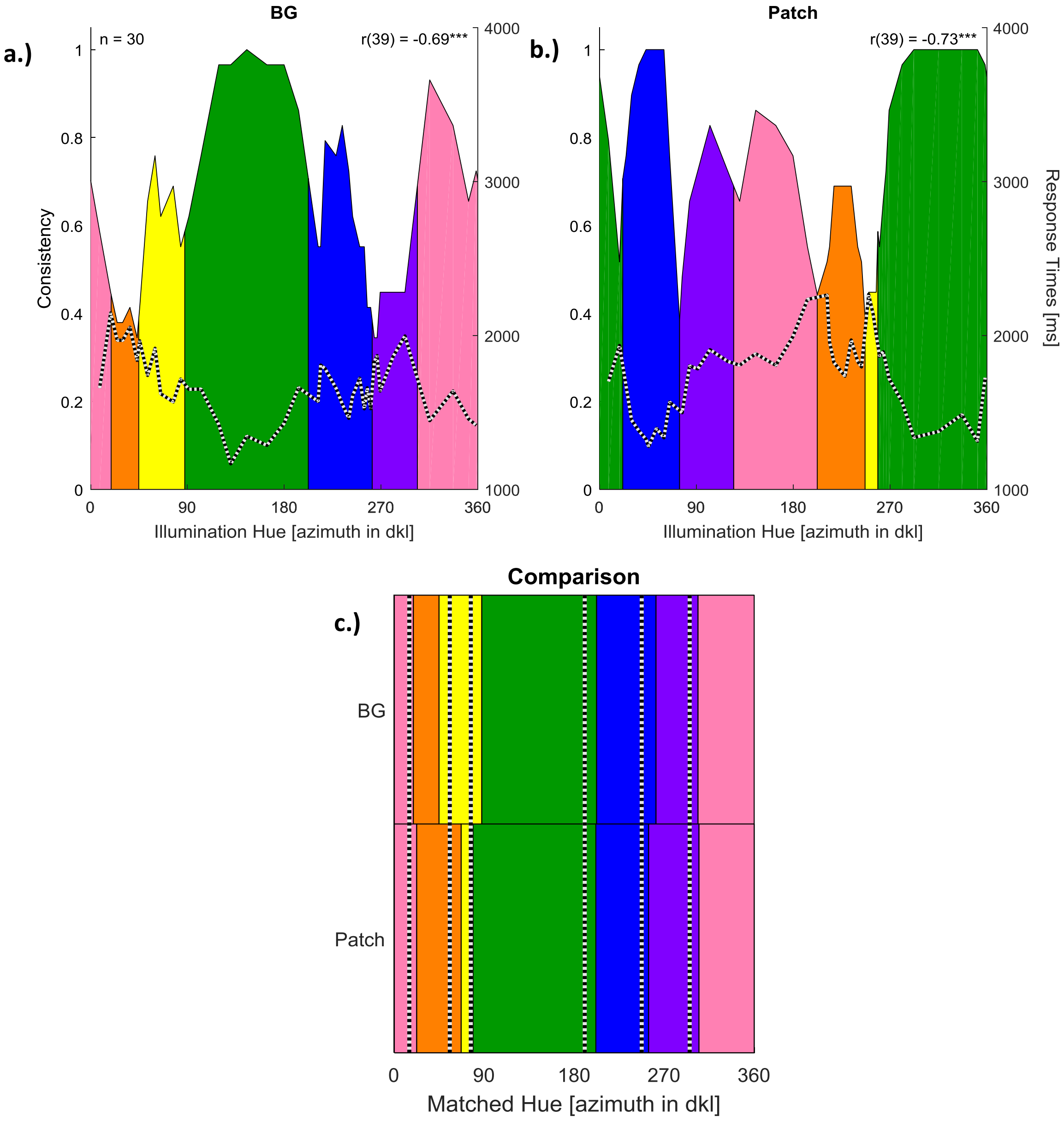


Figure 4.2. Results from colour naming. Panel a.) illustrates colour naming for the colour of the background, panel b.) the results for naming the induced colours on the test patch in the centre. Panel c.) compares colour naming of illumination in the background (panel a), the patch with the induced colour (panel b), and colour naming for simple coloured patches as obtained in a

previous study Witzel and Gegenfurtner (2013). In all panels, the x-axis represents hue of the illumination as determined by azimuth in degree in DKL-space, and coloured areas and vertical lines indicate the mode colour terms and their category boundaries. In the lower part of panel c, the azimuth of the illumination has been shifted by 180 degree in order to approximate the induced hue of the patch so as to match the hue of the illumination and of simple colour patches. In panels a and b the left y-axis represents the consensus of colour naming (i.e. the consistency across observers) and the right y-axis the average response times in colour naming. The thin solid curve above the coloured areas and the thin dotted curves in panels a and b show the variation of consensus and response times across hues. The correlation between consensus and response times is given in the upper right corner. The dotted vertical lines in panel c show the category boundaries for simple colour patches obtained in the previous study (same as in figure 9.a in Witzel & Gegenfurtner, 2013). Note the correlation between consensus and response times in both, induced colour (a) and background (b) naming, the high degree of similarity between the categories obtained for induced colours, background colours, and uniform colours (c).

4.3 Achromatic Adjustments

In this main experiment we measured colour constancy through achromatic adjustments for illumination changes in 40 hue directions. We examined how colour constancy changes depending on the hue of the illumination and tested the role of candidate determinates of the variation of colour constancy across hues.

4.3.1 Method

To compare the results of this experiments to those of the former study (Weiss et al., under review), the same participants were measured and the same apparatus was used. Moreover, the 40 stimuli used here included the 12 from that previous study.

4.3.1.1 Observers

Another 16 naïve observers (10 females, 21-31 years old) participated in the achromatic adjustments. All observers were students of the Justus-Liebig University, tested for normal vision using Ishihara plates (Ishihara, 2004) and gave informed consent before participating.

4.3.1.2 Apparatus

Stimuli were presented on an EIZO CG223W monitor driven by an AMD FirePro V4900 with a resolution of 1680 x 1050 pixels, and a colour resolution of 10 bit per channel. The Monitor was calibrated using a Konika Minolta CS2000 Spectroradiometer (***Konica Minolta*** Sensing Inc., Singapore), CIE-xyY specifications of the channels were: R = [0.655, 0.332, 34.6]; G = [0.207, 0.678, 64.2]; B = [0.15, 0.065, 7.8]. All experimental stimuli were gamma corrected. The Monitor was placed in a black painted tunnel, 40 cm away from the participant. From this distance, the screen subtended a visual angle of 61.3° x 40.6°. The distance was fixed by a chin rest mounted to the table. The experiments were programmed in Matlab 2012b (The MathWorks Inc., 2007), using the psychophysics toolbox 3 extensions (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997).

4.3.1.3 Stimuli & Procedure

The same stimuli were used as in the colour naming experiment (cf. **Figure 4.14**). In each trial, observers were presented one of the 41 stimuli, and were asked to adjust the test patch in the centre until it appeared achromatic to them. In the instructions it was emphasized that the test patch should not appear reddish, yellowish, greenish, bluish, or otherwise colourful.

Initially, the test patch had the same colorimetric grey (cf. **Figure 4.14** and **Table S 4.1**: Neutral) for all 40 chromatic illumination colours. However, to see the test patch as achromatic, observers had to compensate for the induced colour and adjust the patch towards the hue of the illumination, which is opponent to the induced hue (see colour naming experiment above). Observers were not told that the test patch was physically identical across stimuli to avoid cognitive efforts to counteract induction effects. For the control display with the achromatic illumination, the test patch was shown in a random initial colour.

To adjust the colour of the test patch, observers could press one of four keys. The keys corresponded to the four opponent directions of DKL-space and were spatially arranged accordingly. Luminance was fixed to the maximum luminance of the background. There were two different step sizes available, so that the observers could first approximate the colour region they aimed for and then fine-tune their match.

After confirming the adjustments, a sequence of colour noise patterns was presented in order to prevent after images in the following trial (cf. **Figure S 4.5**). The noise in these sequences changed with every frame and the sequence lasted 3 seconds.

Each of the 41 scenes were adjusted twice in interleaved order, resulting in overall 82 trials. Before starting the experimental trials, participants performed practice trials until they felt comfortable with the task. A session of adjustment took about 50 minutes.

4.3.2 Results

4.3.2.1 Patterns of Adjustments

Figure 4..a compares the achromatic adjustments averaged across the 16 observers (black triangles) to the illumination colour (coloured disks). Here, the illumination colour is the colour of the illumination reflected off the grey surface. Individual data may be found in **Figure S 4.6**.

In contrast to previous studies (Bosten et al., 2015; Chauhan et al., 2014; Witzel et al., 2011), adjustments in the control condition with the neutral illumination did not vary along the daylight axis. Three observers provided strong shifts to the reddish hue direction, while providing sensible measures in the colour constancy conditions. Even when excluding these observers a variation along the daylight axis was not clearly visible (**Figure S 4.6**.a). In line with previous observations (Witzel et al., 2017; Wuerger et al., 2015), there was a small shift towards blue (grey disk in the centre of **Figure 4.**.a).

For adjustments with chromatic illuminations, there were undershoots (i.e. adjustments with lower chroma than the illumination) for all hue directions except for the blue direction (i.e. in the third quadrant in **Figure S 4.6**.b). As a result, average adjustments were less saturated. Only average adjustments for illuminations with a bluish hue coincided almost exactly with the illumination colour.

Another feature of the average adjustments (black triangles in **Figure 4.**.a) was that they are all shifted towards blue. There is no such effect in the yellow direction of the daylight locus. To capture this overall shift towards blue, we identified the blue with the smallest shift, where the average adjustment was almost exactly the same as the illumination colour. This was the case for the blue colour located directly under the daylight locus in **Figure 4.**.a, henceforth *daylight blue*.

To quantify the shift of adjustments towards daylight blue, we calculated the distance of each illumination colour and each adjustment to daylight blue. Then, we determined the blue shift as the difference between the distance of an adjustment and of the corresponding illumination colour from daylight blue. A positive blue shift means that the adjustment was closer to daylight blue than the illumination colour. These blue shifts are illustrated by **Figure 4.**.b. Almost all blue shifts were above zero, indicating a shift towards daylight blue. A t-test across colours indicated that blue shifts were significantly above zero (t(40)= 9.5, p<0.001). For further specification of the blue shift, **Figure S 4.7** illustrates the rotations of average adjustments towards daylight blue. Moreover, the further the illumination

colour was away from daylight blue, the more the adjustment was shifted towards daylight blue. Consequently, there was a highly significant correlation between the blue shift and the distance of the illumination colour from daylight blue (r(39) = -0.82, p < 0.001).

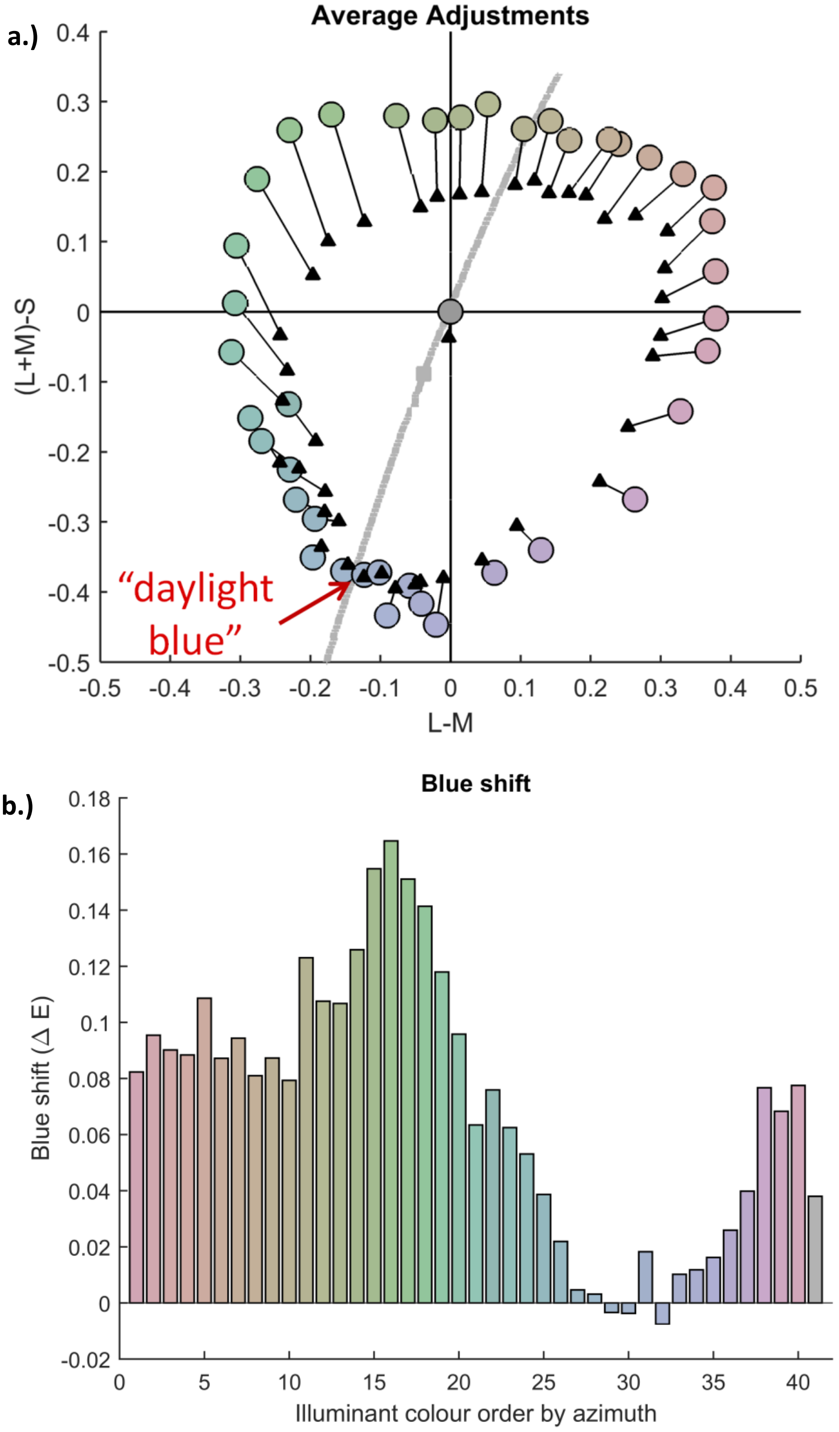


Figure 4.3. Achromatic adjustments. Panel a shows illumination colours (coloured disks) and average adjustments (black triangle) in DKL-space, with the L-M contrast along the x- and the (L+M)-S contrast along the y-axis. The grey curve in the background represents the daylight locus; the grey square on the daylight locus corresponds to d65. The red arrow identifies “daylight blue”. Panel b illustrates the blue shifts (y-axis) as a function of hue (azimuth along the x-axis). The blue shift quantifies how much closer achromatic matches were to daylight blue as compared to illumination colours. The last bar corresponds to the adjustment in the control condition. Note that almost all achromatic adjustments were shifted towards blue.

4.3.2.2 Colour constancy

Colour constancy is perfect when achromatic matches (black triangles in **Figure 4.**.a) coincide with the colour of the achromatic reflectance under the respective chromatic illumination (coloured disks in **Figure 4.**.a), and lower the further away the adjustments are from the colour of the achromatic reflectance (length of black lines in **Figure 4.**.a). We consider this distance as a raw measure of colour constancy, or rather of colour "inconstancy", and will refer to it as the adjustment error. The adjustment error is plotted as a function of azimuth in **Figure** .b (black curve).

Based on the adjustment error we calculated two more specific measures of colour constancy, the *Colour Constancy Index* (Arend et al., 1991) and an adaptation of the *Brunswick ratio* (Olkkonen et al., 2008; Troost & de Weert, 1991). The calculation of these measures is illustrated by **Figure** .a (see also Foster, 2011). The Colour Constancy Index (CCI) is based on the "relative adjustment error". To calculate the Colour Constancy Index, the adjustment error (black curve in **Figure** .b) is divided by the illumination shift, i.e. the distance between the achromatic reflectance (Neutral 6.5) under a chromatic illumination (coloured disk in **Figure 4.**.a) and under the neutral illumination (grey disk in **Figure 4.**.a). This ratio is one if the adjustment error is as large as the illumination shift. This indicates a complete absence of colour constancy. In order to obtain an index that reflects colour constancy, this value is subtracted from one, so that one corresponds to complete colour constancy:

eq.1: $CCI = 1 - \frac{|C-B|}{|A-B|}$

Where A, B and C correspond to the points in **Figure** .a.

The Brunswick ratio assumes that the only systematic variation in colour constancy occurs along the direction of the illumination shift and all other variation of adjustments is due to noise. For this reason, it projects the adjustments on the direction of the illumination

shift (i.e. the direction from the grey disk to the respective coloured disk in **Figure 4.**.a), resulting in the distance AC' according to the nomenclature of **Figure** .a. As for the Colour Constancy Index, this distance is expressed relative to the illumination shift (AB):

eq.2: $BR = \frac{|A-C'|}{|A-B|}$

The advantage of the Colour Constancy index is that it does not need to assume that all deviations of adjustments from the target colour defined by the illumination shift are noise. Moreover, while the Colour Constancy Index is sensitive to precision, the

Brunswick ratio is exclusively based on accuracy. The disadvantage of the Colour Constancy Index compared to the Brunswick ratio is that it does not disentangle systematic biases in adjustments and noise, implying that it cannot reach a value of 1 (perfect constancy) in empirical measurements which necessarily involve measurement noise. These conceptual differences are visible in **Figure** .b. The Brunswick ratio is higher than the Colour Constancy Index in bluish regions of colour space, in which there is overshoot in the adjustments (third and fourth quadrant in **Figure S 4.6**.b).

Finally, following Witzel et al. (2016) we determined the interindividual variation of adjustments as an index of the precision independent of the congruence with a target colour (black-white dotted curve in **Figure** .b). The interindividual variation is calculated as the mean differences of each individual observer's adjustment from the average across observers (similar but not the same as the standard deviation, i.e. the grey shade in **Figure** .b). This measure is particularly interesting when examining the relationship between achromatic adjustments and uncertainty.

Despite the conceptual differences between the four measures all curves indicate that colour constancy is comparatively high in the blue region of colour space and maximal for daylight blue (see red arrow in **Figure 4.**.a and highest peak of colour constancy close to the blue-purple boundary in **Figure** .b). This is due to the fact that the adjustment error (black curve in **Figure** .b) and the interindividual variation (dotted curve and grey shade in **Figure** .b) are minimal for daylight blue. Hence, adjustments are not only closer to the illumination colour, but there is also less uncertainty about the appearance of the achromatic colour when the illumination is daylight blue.

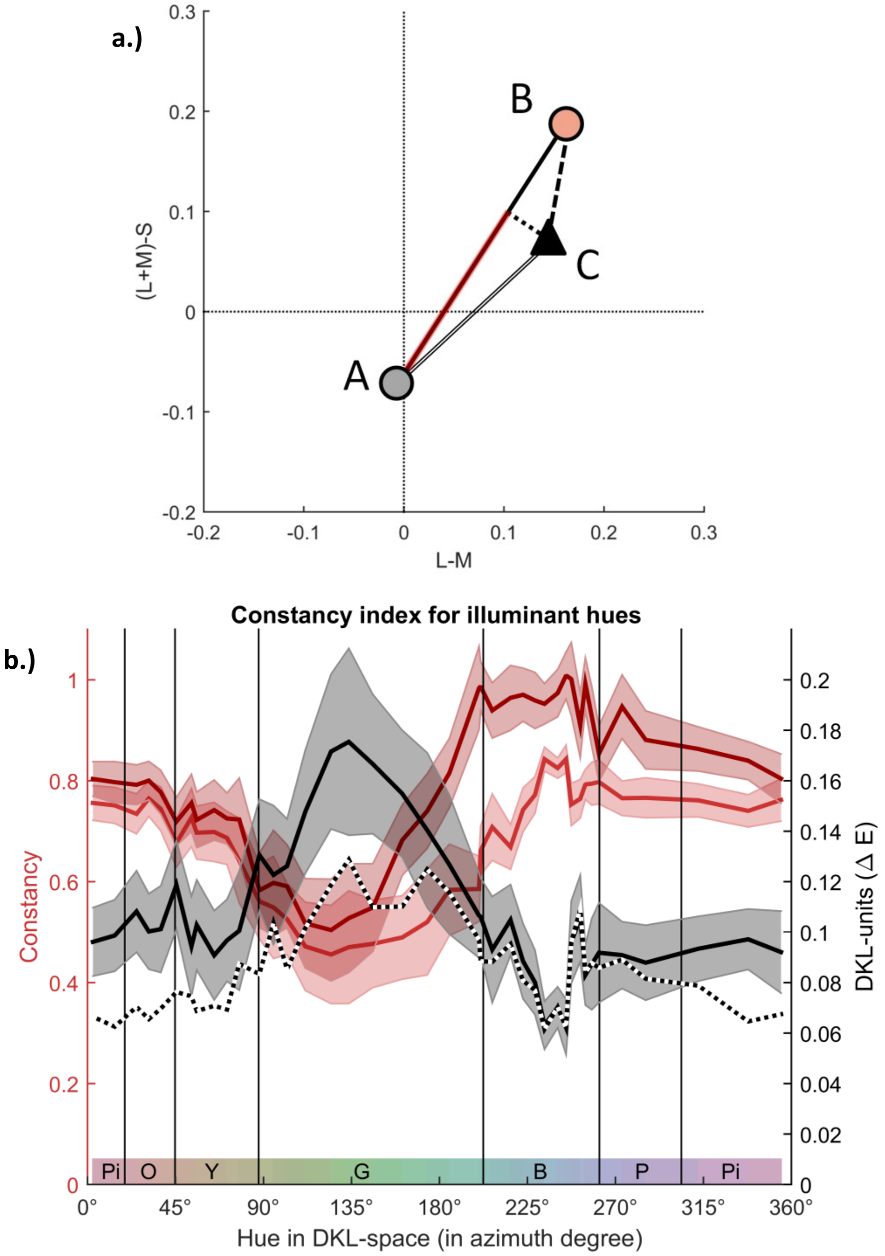


Figure 4.4. Colour constancy. Panel a illustrates the calculation of colour constancy measures. The grey disk (A) depicts the neutral illumination colour, the orange disk corresponds to the respective chromatic illuminant colour (B) and the black triangle to the achromatic match. The colour constancy index (CCI) is the distance BC divided by AB. The Brunswick ratio (BR) consists of the projection from AC to AB (red line) divided by AB. Panel b shows the colour constancy measures obtained in our study (red y-axis on the left) as a function of azimuth (x-axis). The dark red curve shows the Brunswick ratio, the light red curve shows the Colour Constancy Index. The black curve and the black-and-white dotted curve correspond to the adjustment error (i.e. BC in panel a) and to interobserver variation (see text for explanation). The corresponding black y-axis on the right represents Euclidean distances in DKL-space. Note that the data along the solid black curve correspond to the length of the black lines in **Figure 4.**.a and form the basis of the CCI (red curve).

4.3.2.3 Comparison with illumination discrimination and chromatic detection

We then tested the idea that observers discounted for the illuminant to accomplish the achromatic adjustments. For this purpose, we compared colour constancy of achromatic adjustments in this study with the illumination discrimination thresholds measured in the companion study (Weiss et al., under review). In case illumination estimation matters for achromatic adjustments, colour constancy should be higher for illuminations that are easy to perceive (low illumination discrimination thresholds). Hence, illumination discrimination thresholds measured in the companion study should be negatively correlated with the Colour Constancy Index and the Brunswick ratio, and positively correlated with the adjustment error and the interobserver variation across the twelve illumination colours (see **Table S 4.2** for details). The correlation between Brunswick Ratio and illumination discrimination thresholds was close to significance ($r(10) = 0.53$, $p = 0.08$), but was positive and hence contradicted the hypothesis. None of the other measures were correlated with illumination discrimination thresholds (all $p > 0.71$).

In general, any adjustment depends on the ability to perceive colour differences, and hence on discrimination thresholds. In particular, the standard deviations of adjustments may be translated into just-noticeable differences. In achromatic adjustments, these differences are presumably differences to the adapting white-point and mainly concern detection thresholds. To test for a relationship between achromatic adjustments and colour detection, we calculated correlations between the detection thresholds measured in the companion study and the above four measures of colour constancy. However, there was no significant correlation (all $p > 0.26$), indicating that achromatic adjustments are not related to detection thresholds in a simple way.

Moreover, in the companion paper we reported a blue shift for illumination discrimination that could partly be explained by a blue bias in hue detection (Weiss et al., under review). To compare the overall blue bias in achromatic adjustments to the bias obtained for illumination discrimination and detection thresholds, we integrated all shifts of achromatic adjustments away from the illumination colour. To do so, we subtracted the respective illumination colour (coloured disks in Figure 4..a) from the respective achromatic adjustment (black triangles in **Figure 4.**.a). In this way, the deviations between achromatic adjustments and illumination colours (black lines in **Figure 4.**.a) are relative to the origin. We will call these measures adjustment shifts. We fitted an ellipse to the adjustment shifts to capture their overall tendencies (black ellipse in **Figure 4.**) and compared the centre and orientation of the ellipse to the centres and orientations of the ellipses fitted to illumination discrimination (green ellipse) and detection thresholds (blue ellipse) from the companion study.

As for chromatic detection thresholds (blue cross in **Figure 4.**), the centre of the adjustment shifts (black cross in **Figure 4.**) is shifted towards the S-pole of the (L+M)-S axis, but the shift is much larger for achromatic adjustments than for detection thresholds.

However, while the ellipse for detection thresholds is aligned with the (L+M)-S axis, the orientation of the ellipse for adjustment shifts is oblique to the axes. Other than the blue shift, there does not seem to be any other commonality between achromatic adjustments and the other two measures.

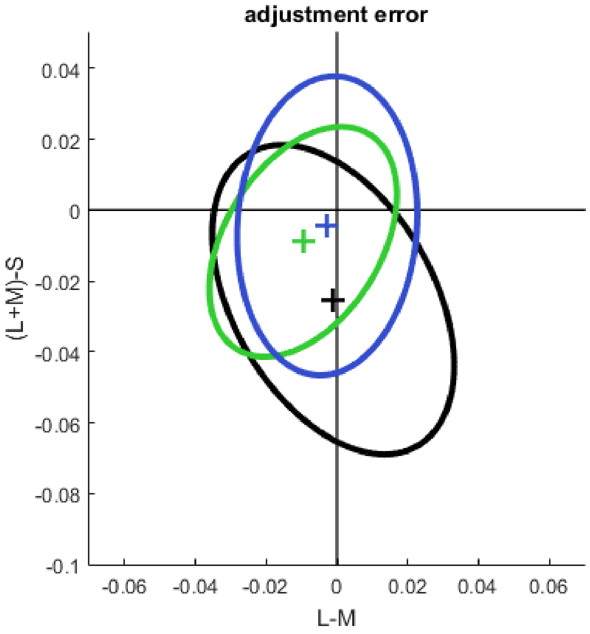


Figure 4.5. Comparison with chromatic detection and discrimination thresholds. Ellipses are fitted to the adjustment error (black ellipse), illumination discrimination thresholds (green), and chromatic detection thresholds (blue). Note the strong blue shift for achromatic adjustments.

4.3.2.4 Candidate determinants

We then investigated several other determinants that potentially explain the variation of achromatic adjustments and colour constancy. We focused on the Colour Constancy Index as a measure of colour constancy because we observed systematic shifts of adjustments towards the blue direction and the Brunswick ratio is insensitive to these shifts due to the projection on the illumination shift.

To assess the variation of adjustments perceptually, we recalculated adjustments errors (black curve in **Figure** .b), interindividual variation of adjustments (black-dotted curve in **Figure** .b) and the Colour Constancy Index (light red curve in **Figure** .b) in CIELAB space. We assumed that the observer's adapting white-point was the illumination of each scene and used the respective chromatic illuminations as white-points for the CIELAB calculations. Because of the strong variation of the white-point we did not use CIELUV because it provides an unreliable chromatic adaptation and hence Euclidean differences strongly change across colour space for different white-points. The colour constancy index is largely the same in DKL and CIELAB and colour space (**Figure S 4.8**) and hence allows tests independent of colour spaces.

As candidate determinants, we examined illumination shifts, blue bias, colour categories, metamer mismatch areas, sensory singularities, and cone ratios. As an overview, **Figure 4.** illustrates the variance of the adjustment errors explained by each determinant.

Illumination shifts. As can be seen from Figure 4., the illumination shift is not the same for all hue directions (this is also true in CIELAB; see Figure S 4.8). The larger the shift, the more colour constancy must be performed by the observer. Hence, failures of colour constancy might be expected to be higher for larger shifts. This was indeed the case (**Table S 4.3**): illumination shifts were correlated with adjustment errors (r(38) = .47, p = .002, cf. "illum. Shift" in Figure 4.) and interindividual variation in CIELAB (r(38) = .43, p = .005).

Blue bias. Above we observed a blue bias, according to which adjustments were shifted towards blue in DKL-space (see Figure 4..b). To assess the importance of the blue bias for our measures of colour constancy, we determined the distance between the blue daylight illumination (see arrow in **Figure 4.**.a) and the colour of each other illumination (see coloured disks in **Figure 4.**.a) in CIELAB. We then calculated correlations between those distances from daylight blue and our three measures of colour constancy (**Table S 4.3**). The distance from blue daylight was negatively correlated with the CCI (r(39) = -.52, p < .001) and positively with the adjustment error (r(39) = .67, p < .001, cf. "blue bias" in Figure 4.) and the interindividual variation (r(39) = .57, p < .001). These results indicate that constancy decreases with distance to the blue daylight illumination.

Colour categories. Figure .b shows that colour constancy changes rather smoothly across hues. This contradicts the idea of category effects on colour constancy, which would imply abrupt changes at category borders and/or at category prototypes. Further analyses also showed that colour constancy did not systematically differ between colours at the boundaries and colours in the centre of the categories. This is further illustrated by **Figure S 4.10** in the Supplementary Material.

However, the boundaries shown in Figure 4. and **Figure** .b are merely hue boundaries. Since colours are not very saturated category membership is not always maximal at the centres of those hue boundaries (see observer consistency and response times in **Figure 4.**.a-b). According to the idea of categorical colour constancy, category membership and colour constancy should be positively related (see also Witzel et al., 2016). We used category consensus as a measure of category consistency (**Figure 4.**.a-b) and correlated it with each of our three measures of colour constancy. Since response times provided an alternative measure of category membership, we also calculated correlations for response times. We calculated these correlations for both, category membership of induced patch colours and of inducing background colours (see **Table S 4.3** for details). Note that the colour names for the induced colours of the patches are shifted by approximately 180 degrees compared to the background naming (**Figure 4.**.b).

Consensus and response times for naming the patch colours were not correlated to any of our three measure of colour constancy (all p > .07; cf. “Patch categories” in **Figure 4.**). In contrast, category consensus for naming background colours was significantly negatively correlated with the colour constancy index (r(38) = -.50, p = .001), and positively with the adjustment error (r(38) = .68, p < .001; cf. “BG categories” in **Figure 4.**), and interindividual variation (r(38) = .68, p < .001). Response times were also significantly correlated with all of these three measures (all p < .05, cf. **Table S 4.3**). These correlations imply that colour constancy is lower for illumination hues with high naming consensus. This result contradicts the idea of categorical colour constancy, which predicts the inverse.

Metamer mismatching. We calculated metamer mismatch volumes in CIELAB for the light achromatic reflectance (Neutral 6.5) that reflects the illumination colour (disks in **Figure 4.**.a) for the 40 changes from neutral to each of the chromatic illuminations. The calculation of metamer mismatch volumes has been described previously (Logvinenko et al., 2014, 2015; Witzel et al., 2016). We focus on the two-dimensional projections of the volumes on the chromatic plane (i.e. metamer mismatch areas) because observers could not adjust lightness (Witzel et al., 2016). However, results were similar with the three dimensional volumes.

If the uncertainty represented by the metamer mismatch areas were related with the uncertainty of achromatic adjustments, there should be a positive correlation with our measures of adjustments errors and interindividual variation, and a negative correlation with the colour constancy index (see **Table S 4.3**). However, there was no significant correlation between metamer mismatch areas and our three measures of colour constancy (all p > .22). This result provides little support for a relationship between metamer mismatching and achromatic adjustments.

Sensory singularities. We determined sensory singularities for the Munsell-like reflectances that cancel the chromatic illumination so as to produce colorimetric grey. For the calculations we used the programs provided by Witzel et al. (2015). The idea is that adjustments should be more accurate and precise if these reflectances are singular because singularity makes the colour signal of these reflectances more predictable. However, correlations between sensory singularities and measures of colour constancy did not support this idea (see **Table S 4.3**). Sensory singularities were not correlated to the CCI (p > .35), but there were positive correlations between adjustment error and interindividual variation (r(38) = .35 p = .026) and (r(38) = .41, p = .008). Positive correlations contradicted the idea that sensory singularities reduce adjustment errors and variation.

Moreover, the observed correlations may be explained by the role of chroma since sensory singularities are strongly related to differences in chroma (Witzel et al., 2015). When controlling for Chroma (which is equivalent to the illumination shift in the present study), none of the measures was correlated with sensory singularities anymore (all r(38) < .22, p > .17).

Cone ratios. Finally, we also calculated the cone ratios for the 39 distractor and 2 background colours for each change from the neutral to each of the 40 illuminations. First of all, we observed that cone excitations for all three types of cones and all 40 illumination changes were almost perfectly correlated (min. r(39) = .88, max. r(39) = 1), implying that cone rations are almost constant (cf. Foster & Nascimento, 1994, fig. 2). Following the approach of Nascimento et al. (2004), we calculated predictions of the adjusted colour under the respective chromatic illuminations based on the 41 cone ratios of the surrounding colours. If colour constancy was determined by cone ratios, observers' adjustments should be closer to the predictions based on cone-ratios than to the actual colour of the surface under each chromatic illumination. We determined the distance between the cone-ratio prediction (averaged across the 41 estimations) and the average achromatic adjustment for each illumination, and compared them to the adjustment error across the 40 illuminations in a paired t-test (analoguous to Figure 10 in Witzel et al., 2016). Results showed that adjustments were further away from the cone-ratio predictions than from the predictions based on the light reflected of the grey reflectance (Neutral 6.5) under the chromatic illuminations (t(39) = 25.2, p < .001).

Then we determined the distance between the 41 predictions and the actual illumination colour (on the test patch) according to the Munsell-like reflectances and the Gaussian illuminations. The average of these distances provides the prediction error of the cone ratio predictions for each illumination. We calculated the correlations between cone ratio prediction error and the Colour Constancy Index, the adjustments error, and interindividual variability (see **Table S 4.3**). The prediction error was correlated with the adjustment error (r(38) = .38, p = .02) and with the interindividual variation (r(39) > .41, p < .008), but not with the Colour Constancy Index (r(38) = -.11, p =.48).

These results suggest that the more the illumination colour deviates from the cone ratio prediction, the more adjustments deviate from the illumination colour and the more variable adjustments are across individual observers. Since the Colour Constancy Index accounts for the illumination shift, the absence of a correlation between cone-ratio predictions and the Colour Constancy Index indicates that the correlations with the raw distance measures (adjustment error and interindividual variation) are due to the variation of illumination shifts (cf. **Figure S 4.9**). In fact, cone-ratio prediction errors were correlated to the size of illumination shifts (r(38) = .54, p < .001), and correlations between cone ratio prediction errors and constancy measures disappeared when controlling for illumination shifts in partial correlations (see **Table S 4.3**).

Combination of determinants. To assess the variance explained by the combination of all determinants, we calculated a multiple regression (last bar in **Figure 4.**). All determinants together explained R^2 = 75.5% of the variance of the adjustment error (F(7,

32) = 14.1, p < .001). The two most important determinants were the distance from daylight blue (blue bias: R² = 44.9%) and the consistency in naming the illumination colour (BG categories: R² = 45.8%). These two determinants were correlated with each other (r(38) = .39, p = .01), but each of them were still correlated with the adjustment error when controlling for the respective other determinant (BG categories: r(38) = .61, p < .001, blue bias: r(38) = .60, p < .001). These two factors together explained R² = 65.5% of the variance of adjustment errors (F(2, 32) = 35.1, p < .001).

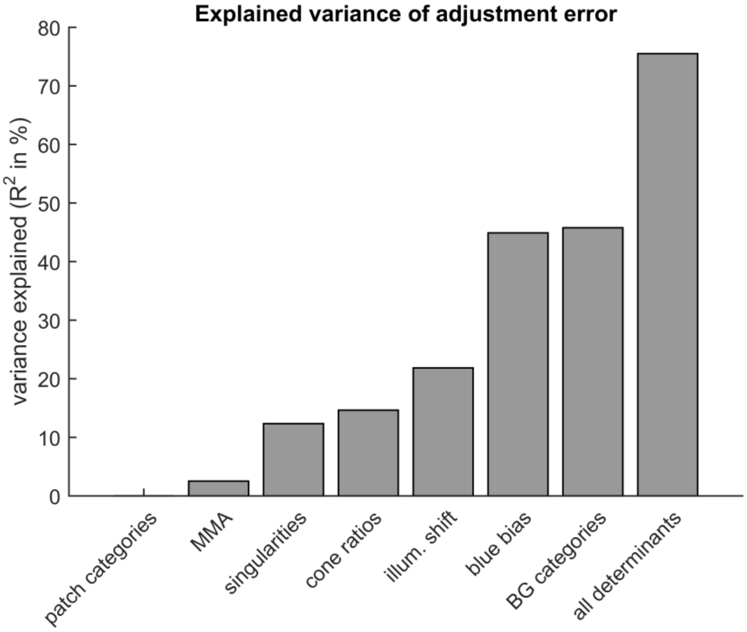


Figure 4.6. Variance of adjustment errors explained by each determinant. The x-axis lists different determinants ordered by the variance they explain. The y-axis represents the variance in percent explained by these determinants. All determinants and the adjustment error were computed in CIELAB. The first seven bars correspond to the correlations of the adjustment errors with naming consistencies for the induced colour on the patch (categories: patch), metamer mismatch areas (MMA), sensory singularities (singularities), cone ratios, illumination shift (illum. shift), the distance of the illumination hue to daylight blue (blue bias), and the naming consistency for illumination in the background of the stimulus display (categories: BG). The right-most bar illustrates the variance explained by a multiple regression with all determinants together as predictors.

4.4 Discussion

4.4.1 Daylight and Blue bias

Our results showed a systematic shift of achromatic adjustments towards the blue direction of the daylight locus (**Figure 4.**.b and **Figure 4.**). The closer the colour of an illumination was to daylight blue, the closer adjustments were to the colour of the grey reflectance under the other illumination. Moreover, adjustments tended towards undershooting (i.e. shifts towards the colorimetric grey) when colours were away from daylight blue (**Figure 4.**.a). The distance of each illumination from daylight blue was one of the two most important determinants of adjustment errors, explaining a large proportion of its variance (44.9%).

In general, undershooting may be explained by incomplete adaptation. Adaptation in this set-up was only controlled through the patterned background in the stimulus display. This may be too little to guarantee full adaptation as it is obtained by adapting to the illumination colour across the full visual field (see e.g. Hansen et al., 2007). For this reason, undershooting itself is not surprising.

What is particular is that the undershoot disappears almost completely under the blue illumination. Due to the way we designed the stimuli for this experiments, illumination shifts are not completely equal for every hue direction (**Figure 4.**.a). Although the 40 reflectances yielded the same chroma under the neutral illuminant in DKL-space, the chromatic illuminations needed slightly different levels of chroma to cancel the chroma of those reflectances and produce colorimetric grey under every illumination. In DKL-space (**Figure 4.**.a) the daylight blue illuminant yielded one of the largest illumination shift (difference between grey and coloured disk), and illumination shifts were negatively correlated with the blue bias in **Figure 4.**.b ($r(38) = -.60$, $p < .001$). This implies that higher blue shifts appear with smaller illumination shifts. The contrary would be expected if illumination shifts increased the blue bias by furthering undershoots. To double-check, we also inspected illumination shifts in CIELAB (**Figure S 4.9**) and found that illumination shifts seem to be small for blue illuminations; at the same time they are also small for yellow illuminations and yet there is only a shift towards blue, but not towards yellow. Consequently, the blue bias cannot be explained by illumination shifts.

Moreover, the blue bias contradicts the idea that blue illuminations were less saturated and weaker inducers, as discussed for the colour naming experiment. If induction was weakest for blue illuminations adjustments should be less shifted away from colorimetric grey. Instead, observers' adjustments were most strongly shifted, resulting in almost perfect coincidence of adjustments with the illumination shift and highest colour constancy under blue illuminations. For this reason, the blue bias may not be explained by the anisotropy of colour space or weaker induction by bluish illuminations.

As summarized in the Introduction, previous studies found inconsistent results concerning the variation of colour constancy across illumination hues (Brainard, 1998; Daugirdiene et al., 2016; de Almeida et al., 2004; Delahunt & Brainard, 2004; Hansen et al., 2007; Olkkonen et al., 2009, 2010; Rüttiger et al., 2001; Schultz et al., 2006). To control effects of hues, illumination and reflectance spectra, we used a large range of 40 illumination hues, smooth illumination spectra to avoid unpredictable effects of spectral discontinuities, and carefully controlled surface colour shifts. It is still possible that results are affected by the fact that we used Munsell-like reflectances and artificial Gaussian illuminant spectra rather than naturally occurring surface and illumination spectra.

However, it is difficult to control parameters of natural spectra across colour space, in particular since certain spectra, such as turquoise illuminants, barely exist in the natural environment. More importantly, natural reflectances and illuminants typically have rather smooth spectra and should be well approximated by our technique. Moreover, the blue bias was not particular to single illumination directions, but occurred across the ensemble of 40 illuminant spectra. Hence, even if there were single spectra in our set of illuminants that might have unnatural spectral properties, they could not explain the observed blue bias. For these reasons, we expect that the blue bias for achromatic adjustments in this experiment is likely to occur in the natural environment. Hence, our observation that colour constancy is highest for daylight blue illuminations suggests that colour constancy is optimized for such blue daylight illuminations.

The question arises of where this blue bias comes from. One possibility is that it is built into the visual system, and in particular is a feature of adaptation to different hue directions. However, the fact that the bias is towards daylight blue rather than the S-pole of the second-stage mechanisms suggests that the effect is not due to asymmetric adaptation along the (L+M)-S axis (Delahunt & Brainard, 2004; Foster, Amano, & Nascimento, 2003). Another possibility is that observers have more experience with illumination shifts along the daylight axis (Pearce et al., 2014), but this is inconsistent with our results that show the effect does not occur in the yellow direction of the daylight axis.

4.4.2 Perceived illumination and colour constancy

Moreover, our results revealed clear differences between achromatic adjustments and perceived illumination as measured through illumination discrimination thresholds (**Figure 4.**). Apart from the blue bias, we did not find any relationship between achromatic adjustments and illumination discrimination thresholds. This suggests that the ability to discriminate illuminations is of minor importance for colour constancy, at least when measured by achromatic adjustments.

The observation that colour constancy is unrelated to perceived illumination is in line with previous studies according to which observers are surprisingly bad in estimating illumination (Granzier et al., 2009). Taken together, these findings put into question the idea that observers consciously discount the illuminant.

At the same time, recent studies (Aston et al., 2016; Pearce et al., 2014; Radonjić et al., 2016) suggest that illumination discrimination may be considered as a measure for colour constancy because implicit mechanisms compensate for the effects of the illuminant change. In the companion study (Weiss et al., under review) we observed a strong correlation between sensitivity to hue and illumination discrimination ($r(10) = .72$, $p = .009$), indicating that illumination discrimination may be explained to a large degree by asymmetries in hue sensitivity.

In contrast, the present study showed that colour constancy (as measured through achromatic adjustments) is neither related to illumination discrimination nor to the sensitivity to hue. This was the case even though, the present measurements included exactly the same stimulus displays as the companion study (Weiss et al., under review). Since achromatic adjustments measure colour constancy, the observation that they are unrelated to illumination discrimination casts doubt on the idea that illumination discrimination directly translates into colour constancy. Moreover, in contrast to illumination discrimination achromatic adjustments cannot be explained by the sensitivity to hue. This suggests that the large blue bias observed for achromatic adjustments might be qualitatively different from the blue bias in the sensitivity to hue and illumination discrimination.

If this is so, our findings also inform us about the idea that colour appearance is uncertain along the daylight locus (Beer et al., 2006; Bosten et al., 2015; Gegenfurtner et al., 2015; Lafer-Sousa et al., 2015; Winkler et al., 2015; Witzel et al., 2011). Previous experiments found that achromatic adjustments under neutral illumination and adaptation vary most strongly along the daylight axis (Beer et al., 2006; Bosten et al., 2015; Witzel et al., 2011) and are shifted on average towards blue (Winkler et al., 2015; Wuerger et al., 2015). Our experiments extend these observations by showing that this shift towards blue is stronger the further the illumination colour is away from blue and it disappears when the illumination is blue. This asymmetry speaks against a general effect along the daylight axis. Moreover, our observation that achromatic adjustments are not related to illumination discrimination suggests that this blue bias is not due to uncertainty about the illumination.

According to Winkler et al. (2015) the asymmetry along the blue-yellow direction is due to observers' tendency to attribute bluish colours to the illumination and yellowish colours to objects and surfaces. This is in line with the observation that colour constancy in our achromatic adjustments is highest for blue, because this shows that observers completely attribute the blue colour to the illumination. It might also be visible in our colour naming data where the blue category is larger for naming the illumination in the background than for naming the induced colour of the patch (**Figure 4.**). High colour constancy for daylight blue can be particularly helpful to recognize objects under shadow because shadows shed blue lights on objects (Churma, 1994; Troscianko et al., 2009). Hence, the blue bias could be an adaptation to the natural environment.

4.4.3 Categorical colour constancy

Our findings contradicted the patterns of categorical colour constancy. First of all, our results contest the idea that adjusted colours are shifted towards prototypes in colour constancy, as they do in colour memory (Bae et al., 2015). In our experiment achromatic adjustments were not shifted towards category centres. Instead, all adjustments were shifted towards blue (see Results section Patterns of Adjustments).

Moreover, our findings undermined the idea that colour constancy is highest around category centres and prototypes and decreases towards the boundaries of colour categories, as has been suggested by a study using asymmetric matching (Kulikowski & Vaitkevicius, 1997) and by studies on category constancy (Olkkonen et al., 2009, 2010). In contrast to those studies, a previous study, using asymmetric matching, did not find evidence for higher constancy within the categories or any other relationship between colour constancy and colour categories (Witzel et al., 2016). Using achromatic matching, the present study found a negative relationship between colour constancy and category membership, which completely contradicts categorical colour constancy.

These diverging results across studies indicate that the relationship between colour constancy performance and colour categories depends on the methods and set-ups used to measure colour constancy, rather than being a general feature of colour constancy. In particular, the results of our study can be explained by different degrees of adaptation depending on the illumination hue. If observers adapt most easily to the blue daylight illumination, their colour constancy is high. At the same time, the colours of this illumination look more desaturated due to adaptation and hence are named less consistently. The inverse is true for illumination colours far off daylight blue, if observers only achieve lower levels of adaptation for those illumination hues.

The idea that both the blue bias and category consistency are due to a common source, namely different levels of adaptation, is supported by the fact that both measures are correlated. At the same time, each of these two determinant contributes to the explanation of achromatic adjustments and colour constancy, when controlling for the respective other determinant. These results make sense if we consider that the blue bias and the category consensus capture different aspects of the variation of adaptation across illumination hues. If this is true, future experiments that specifically measure the variation of adaptation across illumination hues might reveal the origin of the patterns we observed for colour constancy with respect to the blue bias and category consensus.

4.4.4 Other determinants

Contrary to previous observations (Witzel et al., 2016), we did not find consistent evidence for a relationship between colour constancy and metamer mismatching. One reason for the absence of the strong relationship observed previously might be that this study investigated effects of illumination hues, rather than surfaces colours. Moreover, observers might be unfamiliar with illuminants used here, while Witzel et al. (2016) used real daylight spectra. In any case, the present study highlights the limits of estimating colour constancy based on metamer mismatching.

The absence of clear effects of sensory singularities and the observation that cone-ratios are bad predictors of average adjustments is in line with previous observations (Witzel et al., 2016). According to our results, errors in the prediction of cone ratios increase with the illumination shift, and this might be related to adjustment errors and interindividual variation. Such a relationship notwithstanding, the fact that observers outperform cone-ratio predictions speaks against the idea that colour is driven by cone ratios and relational colour constancy.

4.5 Conclusion

Using achromatic adjustments, we investigated colour constancy for chromatic illuminations along 40 different colour directions. We also measured colour categories for the induced colour of the test patch and of the illumination colour.

Although we found some small but systematic differences between the different colour categories, the results of the naming experiment generally confirm the idea that colours induced by colour contrast are shifted to the opponent hue direction predicted by the second-stage mechanisms. Results also provided some evidence that bluish illuminations are seen as less saturated, which is in line with the blue bias for illumination estimation observed previously (Aston et al., 2016; Pearce et al., 2014; Radonjić et al., 2016; Weiss et al., under review).

In the achromatic adjustment experiment, we observed a strong blue bias: Independent of the illumination colour, adjustments were strongly shifted towards the blue pole of the daylight axis. Average colour constancy was perfect under the blue daylight illumination, but not in the other hue directions, due to undershooting and the blue shift. Our findings support the idea that colour constancy is optimised for bluish, but not for yellowish daylight.

Moreover, we observed a negative relationship between colour constancy and consistency of naming the illumination colour. This relationship is in conflict with the idea of categorical colour constancy. Instead, it suggests that observers more easily adapt to bluish illuminations.

According to our findings, colour constancy was not related to illumination discrimination, chromatic detection, metamer mismatching, sensory singularities, and relational colour constancy. Rather, the blue bias and the consistency of the illumination categories explained most of the variance of the achromatic adjustments.

Chapter

5 Bias effects of short- and long-term colour memory for unique objects

---

A similar version of this manuscript has been published as:

Bloj M., Weiss D., Gegenfurtner K. R. (2016). Bias effects of short- and long-term color memory for unique objects. Journal of the Optical Society of America A: Optics Image Science and Vision, 33(4), 492-500. doi: 10.1364/JOSAA.33.000492

Are objects remembered with a more saturated colour? Some of the evidence supporting this statement comes from research using 'memory colours' – the typical colours of particular objects, for example the green of grass. The problematic aspect of these findings is that many different exemplars exist, some of which might exhibit a higher saturation than the one measured by the experimenter. Here we avoid this problem by using unique personal items and comparing long- and short-term colour memory matches (in Hue, Value and Chroma) with those obtained with the object present. Our results, on average, confirm that objects are remembered as more saturated than they are.

5.1 Introduction

In our everyday life colour is a property strongly associated with objects and we have all heard and maybe even seen with our mind's eye the "red rose" of Robert Burn's poem or the "green grass of home" sung by Tom Jones. This idea of linking certain objects with a defined colour was formalized by Hering and Katz in the early 20 century, as discussed in Adams (1923) and Hanawalt & Post (1942). Subsequently Koffka (1935) and others put forward the specific hypothesis that colours are remembered more saturated, in line with the Gestalt hypothesis of change towards better or the ideal; sometimes identified as a 'positive time error' (Tate & Springer, 1971). A modern formulation of this can now be found as a statement of fact in some image processing (Sangwine & Horne, 1998) and photography books (Langford & Bilissi, 2011). However in more than 80 years of research into colour and memory the evidence to support this has been inconclusive at best.

Hanawalt & Post (1942) report 4 studies using different methods to explicitly test for an increase in saturation of remembered colours. They used abstract colour stimuli, different task and timings; under none of their tested conditions they find that remembered colours are more saturated.

In Newhall et al. (1957) the main experiment is a study using coloured lights (2 deg visual field) presented against a neutral background where the authors compared simultaneous colour matches with successive. The latter being the memory task, where after a gap of 5 sec. participants were asked to set the colourimeter to the colour that had been previously presented also for 5 sec. A comparison between the matches in the simultaneous and the successive condition indicates a consistent increase in Munsell Chroma (and Value to a lesser extent) for the memory matches. In this paper they also describe an older study (their Supplementary Experiment 3) in which participants are asked to select Munsell samples that best represent the recalled colours of a series of unseen objects of highly diagnostic colour (brick, sand, grass and dry grass, skin, concrete, pine trees and weathered wood). For some of these objects the memory matches seem to be of higher Munsell Chroma and Value than the 'standards' they compare them to. This latter experiment was re-created by Bartleson (1960) using essentially the same methods and arriving at similar results described as remembered colours shifting in the "…direction of the typical or dominant hues commonly associated with the actual objects." A comprehensive review of these and other early studies can be found in Tate & Springer (1971).

Siple & Springer (1983) used a colorimeter and controlled photographs of six fruits and vegetables to study the memory colour and preference for them when presented with and without context. Although they also obtained simultaneous matches using the photographs and real fruit as references the memory matches were obtained before participants had seen this particular version of the objects so they were based on some previous internal representation. Across objects and participants they report that memory for Munsell Hue and Value is quite accurate but that Chroma is preferred and remembered as higher.

Another paper often cited as providing evidence of the increased saturation for remembered colour is Pérez-Carpinell et al., (1998). In this study participants' recollections of the colour of 8 common fruit and vegetables were assessed under two illuminants and compared to measurements taken from real examples of the produce. As in other studies participants never saw these measured objects and their matches were reliant on each individuals own interpretation of the colour of a given fruit or vegetable that they externalized by selecting one of the 10 possible alternative NCS papers provided by the experimenter for each object. The author's report their findings in a seldom used colour representation space (SVF) and conclude that the colour difference found between participants' selection and measurements is due mostly to variation in chroma, rather than hue or lightness.

A part of Jin and Shevell's colour constancy study (Jin & Shevell, 1996) using computer simulated Munsell patches in a simple geometric arrangement the authors also report than when there was no illuminant change between the learning and testing phase (i.e. a colour memory task) participants matches deviated, for both durations tested, from the original colour but the nature of this variation was not systematic. They indicate that variations in Chroma and Hue for blue and yellow colours were larger than those for red and green ones and not in the same direction. They also report finding larger shifts for the condition with complex background.

In a more recent and sophisticated study (Vurro, Ling, & Hurlbert, 2013) that incorporates the polychromaticity of natural objects participants adjusted only the mean hue of 3-dimensional objects that represented fruits and vegetables to settings that were "redder" and "bluer" than the measured values of the unseen exemplars. A common element between this study and other recent ones (e.g. Bae et al., 2015) focusing on colour memory, many of them using coloured patches displayed on monitors, is the decision to reduce the dimensionality of the adjustment or choices available to participants only to the hue dimension of colour.

From the summaries above it is clear that a common problem undermines the results of previous studies involving objects with highly diagnostic colours or so called memory-colours. In all cases the experimenters chose **not** to show to the participants the exemplars of the target objects that were later measured to provide a comparison point for the memory matches. In other words there is no way for them to resolve the confound: are discrepancies between remembered and actual colours due to the choice of standard (which might be different to the individual experience each person has of that object) or is there a true memory effect indicative of how we encode the properties of familiar objects.

In our study we avoid this problem by using unique personal items for which the owner has a strong internalized colour representation, i.e. they can produce a long-term memory match to this particular object that is very familiar to them. But crucially we also ask them to provide us with a perceptual match when the object is present which provides us with a baseline to compare with the match done from memory. By also presenting the objects to participants other than their owners we are able to study short-term memory effects and compare them to simultaneous matching across all three colour dimensions (Hue, Value and Chroma).

5.2 Materials and Methods

5.2.1 Participants

In total 16 participants (5 females) participated in this study. Mean age was 35 years, with a range between 26 and 61. All but one were members of the Department of Psychology of the University of Gießen and all provided informed consent before taking part. 14 participants were naïve to the purpose of the experiment; two were authors (MB and KRG). All had normal or corrected to normal visual acuity and normal colour vision, verified via pseudoisochromatic plates (Ishihara, 2004). In the second, short-term memory part of the study, 12 of these participants (4 females) took part, none of them authors. The average age of this sub-group was 31 years with a range between 25 and 42.

5.2.2 Objects

Participants brought a personal object (two in the case of KRG) that was well known to them and had a colour of which they were confident to have a well-established memory of. Photographs of the 17 objects are shown in Figure 5.1. Objects had been owned for a median of 3 years with a range of 10 months to 21 years and in most cases seen daily or at least once a week. For the second part of the study we eliminated 5 objects from the collection; 3 because they were too well know by all participants (blue elephant, orange and brown jumper), 1 was no longer available (multi-tool) and 1 was of an overrepresented colour (turquoise jumper).

Figure 5.1. Photographs of the 17 objects brought by our participants. Each object is shown next to the Munsell chip most often selected as a simultaneous (with object present) match. Notice the variety of materials, size and colour of objects used in this study. Objects with a black border were only used in the first part of the study.

5.2.3 Experimental set-ups and procedures

5.2.3.1 Owners long-term memory (O-LTM) and object (O-OM) matches

In a standard office room with grey walls and floor and just one window, the only source of illumination used during the experiments, a table was covered with a grey cloth and 1325 chips from the Munsell Book of Colors (Glossy Collection) (Munsell Color Services, 1990) were displayed in 40 plastic bins. As part of the 2 minute enforced adaption period participants were asked to order the scrambled bins by hue (see **Figure 5.2**). This task allowed participants not only to adapt to the daylight but also to become familiar with the Munsell chips.

After adaptation and in absence of the object (which had been taken away by the experimenter and was not returned to the participant until all experiments were completed) participants selected chips from the Munsell collection that best represented the recalled colour of their object. For this purpose, they were allowed to spread the Munsell chips on the grey cloth to compare them and make their final selection. Participants were allowed to select a maximum of three chips, allowing for the possible fact that a single chip in the collection did not precisely represent the exact colour of a given object; some owners selected a single chip. All owners performed two long-term memory matches (O-LTM) for their object on two different days. On the second day, after completing their memory match, their object was brought back out and with it present they selected the Munsell chip that best matched the perceived colour of their object (O-OM).

Figure 5.2. Photograph of the 1325 chips from the Munsell Book of Colors (Glossy Collection) displayed in 40 plastic bins, ordered by hue as participants saw them after completing the adaptation task and before making their selection.

5.2.3.2. Participants short-term memory (P-STM) and object (P-OM) matches

For this task we used a sub-set of 12 objects and 12 participants; participants only evaluated objects brought by others (i.e. not their own objects). This second task was completed in a seminar room that was approximately three times the size of the first room, also only illuminated by daylight that came through three windows and contained six tables, which offered enough space for the short-term memory experiment. Like in the first room, the bins containing the Munsell collection where placed on a table covered with a grey cloth in front of a window. Next to it, another table, covered with the same cloth, was used to present objects, select chips and perform object matches.

After participants adapted while performing the previously described bin-sorting task they turned their back to the bins and the experimenter presented them with one of the objects brought by another participant. For 30 seconds the subject was allowed to handle the object and asked to memorize its colour, but not permitted to look at the bins containing Munsell chips. After the 30-second memorization phase, the object was taken away and the subject selected the Munsell chip that best represented the colour of the object they had just seen. Each participant performed one short-term memory match (P-STM) per object. Participants were also asked to provide a confidence rating for their match using a 5-point scale, with 5 representing absolute confidence. At the end of the session the experimenter brought out all the objects and participants were asked to select the Munsell chip that best matched the objects' colour, these constituted the participants object matches (P-OM). As before, for all matches, participants also provided a confidence rating on a five-point scale.

All experiments were completed between 21 November 2014 and 12 of December 2014 in Gießen, Germany. During this time of the year there is limited daylight hours and experiments were only run between 10 am and 3 pm; half of them during cloudy days and half during sunny days. Regardless of conditions, chips and objects (when present) were never placed in direct sunlight. At the beginning and end of each participant's matching session we measured with a Konica Minolta CS-2000 Spectroradiometer the luminance of a white standard (Photo Research RS2) placed 80 cm along and orientated at 45 degrees from the window normal at a height of 20 cm from the table surface. For the owner matching sessions the luminance measured ranged form 73 to 316 cd/m^2 (mean 185 SD =52 cd/m^2). For participants' sessions mean measured luminance was 261 cd/m^2 (SD = 250 cd/m^2), with a minimum value of 31 and a maximum of 979 cd/m^2.

5.3 Results

As we used Munsell chips to collect our participants' responses we can directly report our results using the Munsell system (Munsell Color Services, 1990). This will have the advantage of allowing straightforward comparison with earlier studies. Although conversion to other systems; for example the CIE 1976 - L*a*b* colour space (Wyszecki & Stiles, 1982), are possible as we have both Munsell chip reflectance data and measurement of average illumination during our experimental sessions this is an unnecessary step that does not add information to the analysis or change the overall conclusions.

In **Table 5.1** we report all matches from object owners; each object shown in the left column corresponds to a different participant. The Munsell chips they selected as long term memory matches (O-LTM) are show in the central columns, one column for each session. The column on the right shows the chips that were picked by the owners as the best match to their object when the object was present (O-OM). For some sessions/objects participants selected more than one chip (up to three) to indicate that the desired match was in between the selected chips. In those cases, and for representations in **Figures 5.3** and **5.4**, the mean is calculated, by averaging each of the three attributes (Hue, Value and Chroma) separately.

Table 5.1. Owners' long-term memory (O-LTM, 2 sessions) and object (O-OM) matches

Object	O-LTM (1)	O-LTM (2)	O-OM
	2.5BG 6/10 2.5BG 5/10	10G 7/8 10G 6/10	5BG 6/10
	5BG7/8	2.5BG8/6	5BG8/4
	7.5BG 5/10	7.5BG 6/8	10BG 6/6
	5PB 2/8	5PB 2/8	5PB 2/8
	7.5PB2/10	7.5PB2/10 5PB2/8	7.5PB2/10
	5RP 6/7 5RP 7/10	7.5RP 6/12	5RP 6/12
	2.5R 7/6 2.5R 7/8	10RP 7/6	10RP 7/4
	7.5R4/16 5R4/14	2.5R4/14 5R4/12	2.5R4/14
	5R 3/10	5R 3/10	7.5R 3/12
	7.5R 4/16	7.5R 4/14	7.5R 4/14
	2.5YR6/16	2.5YR5/14	2.5YR6/12
	10YR 6/10 10YR 6/12	10YR 7/14 7.5YR 7/14	10YR 7/10 10YR 7/12
	10YR6/4 10YR5/4	7.5YR6/2	2.5Y6/2
	5Y 9/2	5Y 9/2	5Y 8.5/2
	5GY 5/4	5GY 5/4	2.5GY 5/4
	10GY 5/12	10GY 5/10	10GY 5/10
	10GY 6/12 10GY 6/10 10GY 7/10	10GY 6/12 10GY 6/10 10GY 5/10	2.5G 5/8

Figure 5.3 presents the colour attributes of Chroma, indicated by distance from the centre with each concentric circle corresponding to a variation of two Chroma steps and Hue which varies along the circumference, with each radial line corresponding to one Hue step as available in the Munsell Book of Colors (Glossy Collection). Each line represents an object, with the origin (dot) located on the average of the object's owner' match with the object present (O-OM) and the end of the line on the average long-term memory match for that object. A cutout of the object is placed at the dot (O-OM) end of the line. In this way the length and orientation of the line indicates the variation in chromaticity between the object and memory matches. If there is no difference between object and memory matches then there is only a dot representing that object. If objects

were, for example, consistently remembered more saturated then we would expect all memory matches to be further from the centre than their corresponding object match.

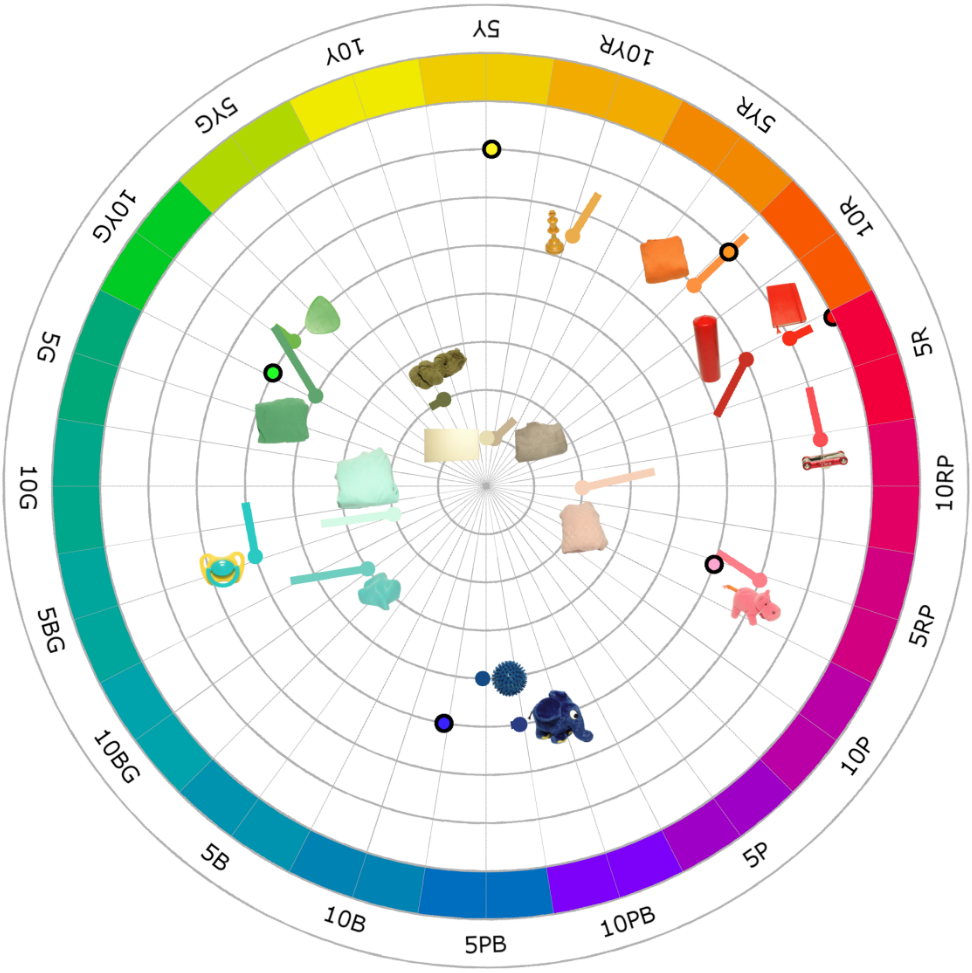


Figure 5.3. Radial plot showing Chroma and Hue of owners matches when the object is present (small dot next to object) and average owners long term memory match (end of line). If there is no shift, i.e. object and memory match coincide they are shown as a dot as in the case of the blue spiky ball. Each concentric circle represents 2 Chroma steps and each radial line represents one of the 40 available Hues in the Munsell Book of Colors (Glossy Collection). See **Table 5.1** for list of actual Munsell papers selected in each case. Dots with solid black outline depict colour category prototypes [17, 18].

In **Figure 5.4** we show the remaining colour attribute of Value along the vertical axis with each division corresponding to one Value step as represented in the Munsell Book of Colors (Glossy Collection) and Hue along the horizontal axis. In this case we have converted Hue notation to degrees by selecting an arbitrary zero point. As before, a cutout of the object is placed at the dot (O-OM) end of the line representing each object and the end of the line is located on the average long-term memory (O-LTM) match for that object. If only a dot is visible for a given object that means that memory and object match are identical.

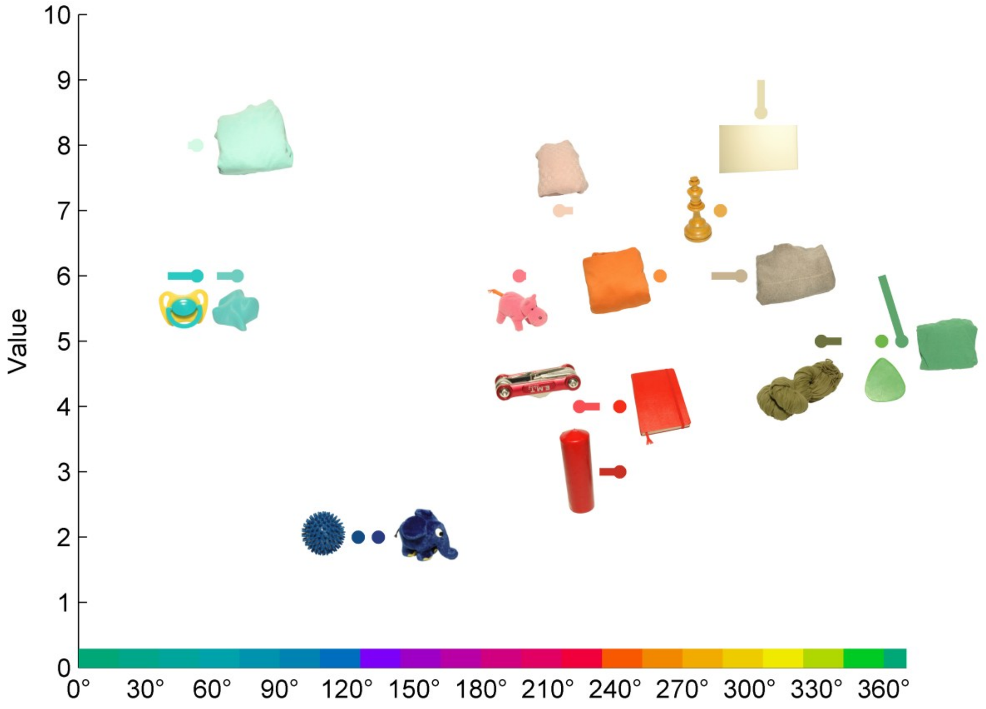


Figure 5.4. Plot showing Value and Hue of owners' object matches (small dots next to objects) and average owners long-term memory match (end of line). If object and memory match coincide they are shown as a dot with no line. Each available step in the Value dimension is represented in the vertical axis and on the horizontal axis we show all the 40 available Hues in the Munsell Book of Colors (Glossy Collection) converted to degrees with and arbitrarily chosen zero point. See **Table 5.1** for list of actual Munsell papers selected in each case.

In **Table 5.2** we report average and standard deviation of participants matches for objects other than their own. For each object shown on the left column the central column shows the corresponding Hue, Value and Chroma of the short-term memory (P-STM) matches averaged over 11 participants. The column on the right indicates the mean matches made by participants when the object was present (P-OM). As before, each of the three attributes (Hue, Value and Chroma) is averaged separately and the Hue notation was converted to its corresponding angular value (degrees) for calculations.

Table 5.2. Participants' (N=12) short-term memory (P-STM) and object (P-OM) matches

Object	Mean P-STM ± SD	Mean P-OM ± SD
	H = 69 ± 11 V = 5.2 ± 0.4 C = 9.5 ±0.9	H = 58 ± 5 V = 5.2 ± 0.4 C = 9.8 ± 0.6
	H = 73 ± 13 V = 5.9 ± 0.7 C = 8.5 ±1.5	H = 69 ± 7 V = 6.3 ± 0.5 C = 6.7 ± 0.9
	H = 124 ± 5 V = 2.7 ± 0.6 C = 9.3 ±1.6	H = 125 ± 0 V = 2.6 ± 0.5 C = 8.8 ± 1.3
	H = 208 ± 8 V = 5.8 ± 0.5 C = 11.7 ±1.4	H = 204 ± 7 V = 5.7 ± 0.5 C = 12.4 ± 0.8
	H = 213 ± 16 V = 7.2 ± 0.5 C = 4.8 ±1.0	H = 212 ± 15 V = 7.5 ± 0.5 C = 4.0 ± 0.0
	H = 237 ± 4 V = 3.9 ± 0.2 C = 14.3 ±0.8	H = 238 ± 5 V = 3.7 ± 0.6 C = 13.2 ± 1.0
	H = 242 ± 0 V = 4.2 ± 0.4 C = 15.5 ± 0.9	H = 242 ± 0 V = 4.0 ± 0.0 C = 14.5 ± 0.8
	H = 283 ± 16 V = 6.5 ± 0.9 C = 10.0 ± 3	H = 287 ± 0 V = 6.3 ± 0.8 C = 10.4 ± 2.8
	H = 301 ± 14 V = 8.9 ± 0.3 C = 3.3 ± 1.0	H = 307 ± 15 V = 8.8 ± 0.3 C = 2.7 ± 1.0
	H = 328 ± 10 V = 5.3 ± 0.6 C = 5.2 ± 1.0	H = 320 ± 10 V = 4.7 ± 0.5 C = 4.2 ± 0.6
	H = 356 ± 6 V = 5.5 ± 0.5 C = 11.0 ±1.3	H = 357 ± 4 V = 5.1 ± 0.3 C = 9.5 ± 0.9
	H = 359 ± 8 V = 6.1 ± 0.7 C = 9.0 ±1.8	H = 366* ± 4 V = 5.6 ± 0.7 C = 8.4 ± 0.8

*Due to where the zero value of the hue circle was arbitrarily set this value should be 006 degrees but for ease of plotting in **Figure 5.6** we have used 366 degrees instead.

Figures 5.5 and **5.6** are analogous to **Figures 5.3** and **5.4** and illustrate participants' short-term memory and object matches to the objects that were not their own. In the case of participants' object and short-term memory matches we also obtained for each match a confidence rating (1 to 5, with 5 being completely confident on their match). When averaged over all objects and participants we find that the average confidence rating for a memory match is 3.4 ± 0.9 and for a match with object present (or simultaneous match) it increases to 4.1 ± 0.9.

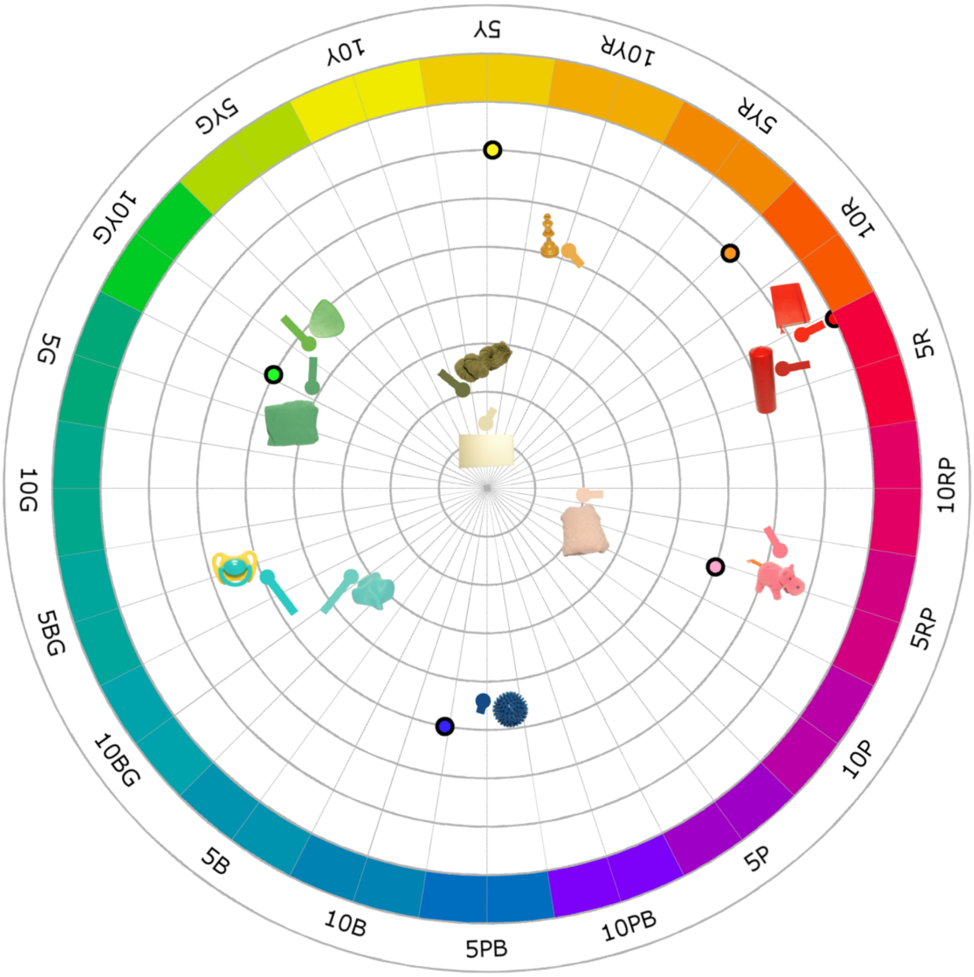


Figure 5.5. Radial plot showing Chroma and Hue of average participants object match (small dot next to object) and average participant short-term memory match (end of line). Each concentric circle represents 2 Chroma steps and each radial line represent one of the 40 available Hues in the Munsell Book of Colors (Glossy Collection). Values and SD are reported in **Table 5.2**, error bars omitted from plot for clarity. Dots with solid black outline depict colour category prototypes (Olkkonen et al., 2010; Witzel & Gegenfurtner, 2013).

5.4 Analysis

The three panels on the left hand column of **Figure 5.7** show how in the case of simultaneous matches (i.e. with object present) there is no significant difference between how long term owners or participants that have just recently acquainted themselves with them match the object colour. For all three colour dimensions the points lie along the 45 degree line. Two –sided t-tests comparing owners and participants' simultaneous matches show no significant term for Hue: t(11) = .68, p = .51; Value: t(11) = -.55, p = .596 or Chroma: t(11) = -1.72, p= .11. The right hand column of **Figure 5.7** compares the long-term memory matches of object owners with the short-term memory matches of participants for the same object, and the corresponding two-sided t-tests show no significant term for any of the three colour dimensions; Hue: t(11) = -.54, p = .598;

Value: t(11) = -.89, p = .39 or Chroma: t(11) = -.51, p=.62 confirming that there is no significant difference on how owners or participants remember the colour of the objects.

To test if there was a difference in any of the Munsell dimensions between long–term memory and simultaneous matches completed by the objects' owners, we conducted one-sided t-tests, comparing owners' mean long-term memory matches with their corresponding object match. For Hue the test showed no significant term: t (16) = 1.18, p = .873, neither did it for Value: t(16) = .16, p = .561. For Chroma, there was a significant effect: t(16) = -2.05, p = .029, indicating that owners long-term memory matches are more saturated than matches made with the objects present.

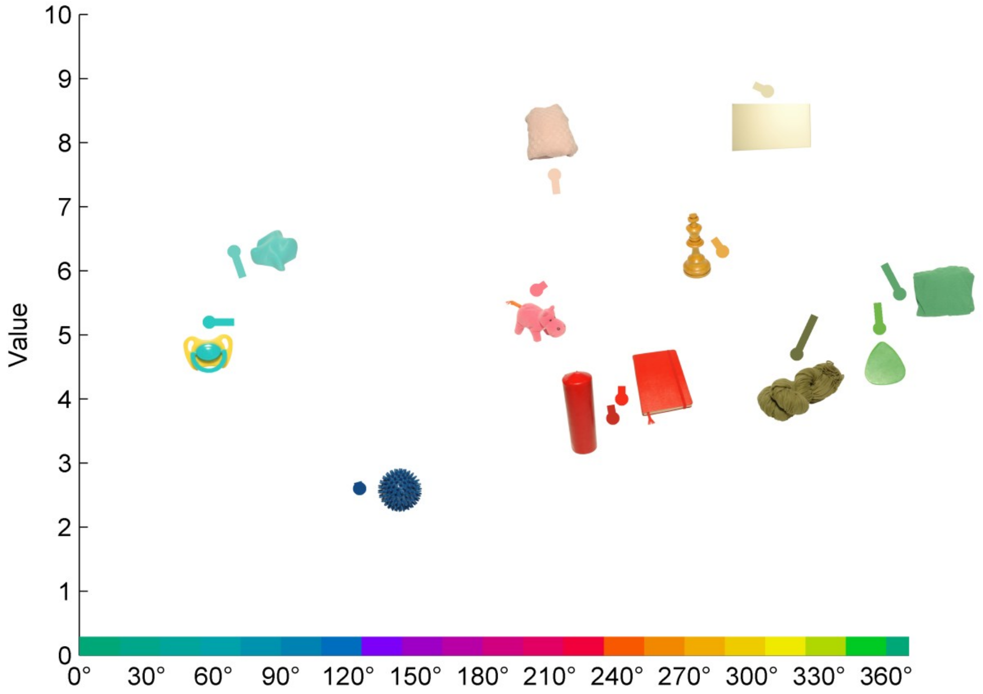


Figure 5.6. Plot showing average Value and Hue of participants' object matches (small dot next to object) and average participant short-term memory matches (end of line). If object and memory match coincide they are shown as a dot with no line. Each available step in the Value dimension is represented in the vertical axis and on the horizontal axis we show all the 40 available Hues in the Munsell Book of Colors (Glossy Collection) converted to degrees with and arbitrarily chosen zero point. Values and SD are reported in **Table 5.2**, error bars omitted from plot for clarity.

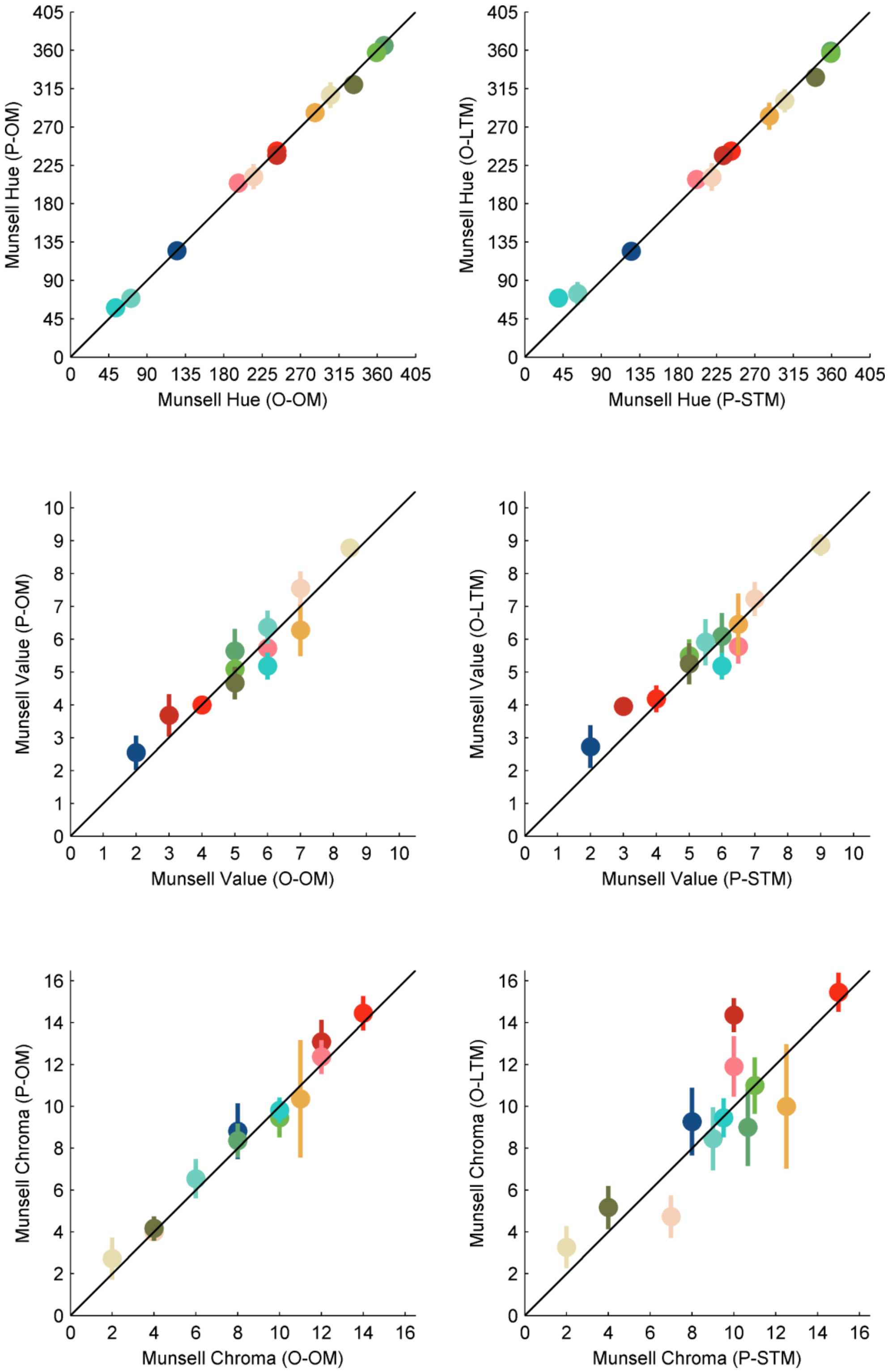


Figure 5.7. The three panels in the left column compare owners' (horizontal axes) with participants' matches (vertical axes) in the object present (simultaneous) condition for Hue (top), Value (middle) and Chroma (bottom). The panels on the right hand column show the same comparisons for the memory match condition. Vertical error bars represent SD of participant's matches. The 45-degree line indicates that participants and owners matches are identical. Only the 12 objects used in both studies are illustrated.

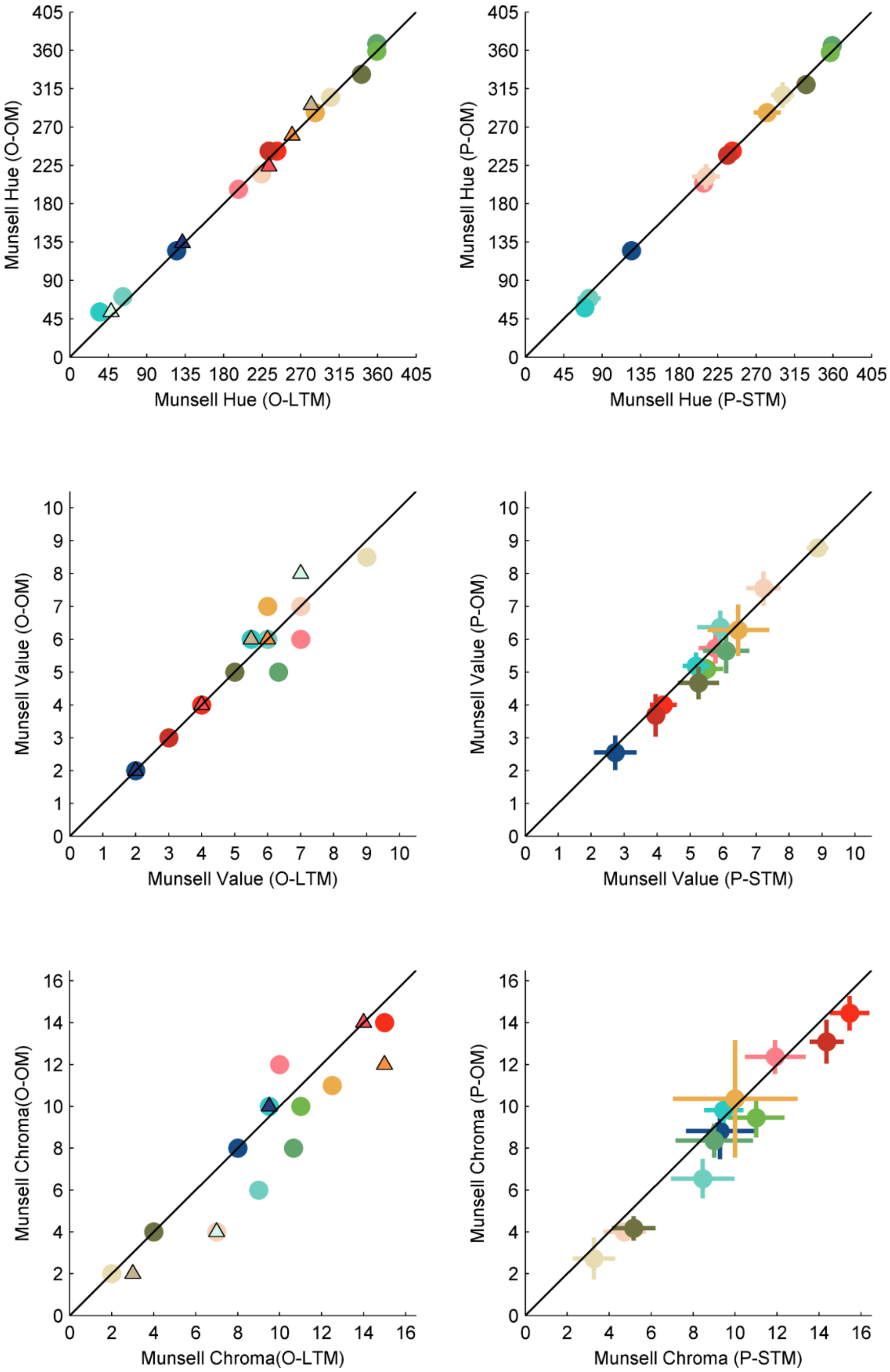


Figure 5.8. Left column shows owners' matches for Hue (top), Value (middle) and Chroma (bottom). The triangles represent the 5 objects not used in the short-term memory task, coloured disks objects used in both parts of the study. Right column shows mean and SD of participant's (N=12) Hue (top), Value (middle) and Chroma (bottom) matches by colour coded crosses. Object matches are shown along the vertical axis and memory matches along the horizontal. The 45-degree line indicates when memory and object match are identical.

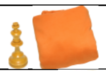
For effects between the short-term memory and simultaneous matches we conducted repeated measures ANOVA, comparing participants' short-term memory matches with those done with object present. As in the case of the owners' matches we find no effect for Hue: F(1, 11) = .82, p = .384, or Value: F(1, 11) = 3.79, p = .078. As before there was significant main effect for Chroma: F(1, 11) = 13.65, p = .004, suggesting that participants selected more saturated chips in the short-term memory condition when compared to the object present match.

These effects can be seen in **Figure 5.8** where we show owners' matches (left column) and participants' average matches and SD (right column) for Hue (top), Value (middle) and Chroma (bottom). If memory and simultaneous (object) matches were identical they would lie along the 45-degree line, and this is the case for the Hue (top row) and Value (middle row). In the case of Chroma (bottom row) the majority of points are under the diagonal line indicating that the matches from memory tend to be to chips with a higher Chroma than those chosen in the simultaneous (object present) condition.

There is a possibility that due to the finite nature of the Munsell collection our study, for some objects, underestimates the increase in saturation. In the case of owners simultaneous matches; four of the objects (candle, animal, ball and dummy) were matched to the chip with highest Chroma meaning that in the memory condition there was no chip available with a higher Chroma for that Hue/Value combination. For the two remaining objects (green scarf, kitchen cabinet) that do not show an increase of saturation with memory, however chips of higher Chroma were available. In the case of the participants matches all three objects for which memory matches were not of higher Chroma than simultaneous ones chips of higher Chroma were available.

In our current study we did not ask our participants to sort the Munsell chips into colour categories or to choose the best example for a category (known as prototype or focal colour), however Olkkonen et al. (2010) as well as Witzel and Gegenfurtner (Witzel & Gegenfurtner, 2013) did exactly this in previous studies. Seven participants across two studies sorted a subset consisting of 320 Munsell chips with maximal Chroma across all Hues and several Value levels into eleven categories that correspond to the basic colour terms (red, orange, yellow, green, brown, blue, purple, pink, white, gray, and black) and selected category prototypes in a similar experimental set-up to ours, i.e. office environment with natural daylight from windows. In **Table 5.3** we list the Munsell chip most frequently selected as the category prototype in these studies, and indicate to which category our objects belong to based on the location of the chip most often identified as an object match within the category boundaries from (Olkkonen et al., 2010; Witzel & Gegenfurtner, 2013).

Table 5.3: Munsell chip categories and prototypical chips.

Category	Prototype chip	Object
Pink	5RP7/10	
Red	7.5R4/16	
Orange	2.5YR6/14	
Yellow	5Y8/14	
Green	2.5G4/10	
Blue	2.5PB4/10	

The locations of the prototypical Munsell chips are depicted in **Figure 5.3** for owners' matches and in **Figure 5.5** for participants' matches as a dot with a solid black outline. The colour of the dots roughly represents the colour category of a given chip.

As participants in (Olkkonen et al., 2010; Witzel & Gegenfurtner, 2013) only sorted chips with a maximal Chroma, the following analysis only takes Hue of our participants matches into account.

From observation of **Figure 5.3** we can see that for two objects (red book, orange sweater) the Hue of the owners' memory and object matches coincide with the category prototype, and for three other objects (ball, kitchen door, guitar pic) there was no change between the memory match and the simultaneous match. These five objects cannot be analyzed further, because there is no shift between memory and object present match. Seven of the remaining objects were remembered closer to the prototypical colours (turquoise sweater, dummy, potato, elephant, bike tool, green scarf, chess piece), while five others were recalled away from the prototype (brown sweater, pink scarf, candle, animal, t-shirt). Over all 17 objects we found that the Hue of owners long-term memory matches, were no closer to the prototypical colours for each colour category than to the simultaneous matches: t(16) = .95, p = .35, and did not differ significantly from the simultaneous matches: t(16)= 1.33, p = .20.

In one case the Hue of the participants' short-term memory matches, simultaneous matches and prototypical chip overlapped for one object (red book, see **Figure 5.5**). For the remaining 11 objects, short-term memory matches were shifted towards the prototypical colours for 7 objects (candle, pink scarf, ball, green scarf, pic, chess piece), while the remaining 5 were remembered away from the focal colour of their category. The Hue of short-term memory matches were no closer to the prototypical colours for each colour category than to the simultaneous matches: t(11) = .68, p = .68 nor did they significantly differ from their simultaneous matches: t(11) = .51, p = .62.

5.5 Discussion

Undeniably our ability to remember colours plays an important role in our everyday life. It aids our recognition and identification of objects such as cars in the parking lot and socks in our drawers and contributes to our decision making during shopping for food and clothes. Overall, photographs of natural scenes seen in colour are recognized quicker and remembered better than when grayscale is used (Gegenfurtner & Rieger, 2000; Gegenfurtner, Wichmann, & Sharpe, 1998) and even in the case of simple lights, chromatic components are better remembered than brightness ones (Sachtler & Zaidi, 1992). Our ability to remember colours has been put to practical use in applications such as colour quality metrics for solid-state light sources (Smet, Ryckaert, Pointer, Deconinck, & Hanselaer, 2012) and to enhance digital images (Xue, Tan, McNamara, Dorsey, & Rushmeier, 2014) and plays a significant role in colour constancy (Allen, Beilock, & Shevell, 2012; Granzier & Gegenfurtner, 2012; Hedrich et al., 2009; Ling & Hurlbert, 2008).

Given this important role of colour in visual memory, large biases would be somewhat surprising. However, our results do confirm previous studies in finding an increase in saturation in the memory construct compared to direct viewing. Our study is the first one that uses particular exemplars engrained in the observer's long-term memory, thus getting past potential artifacts of previous studies. We are also able to extend this finding to short-term memory, for which we found an analogous bias towards more saturation for the objects represented in memory. It is interesting to speculate about a potential functional role for such a bias. If the memory trace literally fades with time, then any kind of bias towards more saturation would counteract this tendency.

A notable bias along the saturation axis has also been reported in some experiments measuring the effect of memory colours on perception (Hansen, Olkkonen, Walter, & Gegenfurtner, 2006; Witzel et al., 2011). Note that in this case it is not the memory that is distorted. Rather, perception itself gets distorted when participants are asked to adjust an object with a typical colour to a neutral gray. They adjust the neutral point in the direction opposite to the typical colour of the object, when compared to a neutral setting for a neutral object. The image of a banana, for example, would be adjusted more bluish than a random noise patch or an image of a pencil. The most straightforward explanation of this effect is that the memory colour is added to the sensory signal in cases of high uncertainty. These results would not predict any bias when viewing objects in their typical colour. In this case, there is a strong sensory signal, and the memory colour would be quite similar to the sensory signal, too.

There is evidence from carefully controlled computer displayed studies involving 2D patches and a variety of surrounds that patches appear much more vivid and richly coloured against low-contrast, gray surrounds than against high- contrast, multicoloured surrounds (Brown & MacLeod, 1997). Systematic exploration of this effect (Faul, Ekroll, & Wendt, 2008) indicates that this is a local effect that can almost be completely eliminated by the introduction of a thin black line between the patch and surround and that its effect is maximal in the case in which the centre and surround have the same luminance. Even though in our study participants were able to place the Munsell chips against a uniform grey background these was in the process of manipulating real 3D objects in a rich context that we believed disrupted any contrast or gamut expansion effects.

For the Hue dimension, there does not seem to be any advantage in a memory bias, and in our study we did not find any. A recent study reported a shift in hue towards colour category prototypes (Bae et al., 2015), and earlier studies emphasized the important role of focal colours for colour memory, in particular being remembered better (Bartleson, 1960) and more precisely (Heider, 1972). In our current study participants had either a long-term exposure to real tangible coloured objects (as in the case of the owners) or an opportunity to familiarize themselves with them over a 30 second period. We collected not only memory matches but also recorded simultaneous matches with the object present. We find no bias effect of either long- or short-term memory for the Hue dimension.

Bae et al. (2015), uses computer controlled and displayed stimuli to systematically explore the Hue dimension (while keeping Chroma and Value fixed) in a way that is not possible with our methods and find evidence that memory for Hue of simple coloured patches is significantly biased towards prototypical or focal colours. As the authors indicate, this effect would not be noticeable in a sparse sampling of the Hue dimension, such as the one in our study. Our own results show that participants' memory matches of colours associated to an objects do not deviate significantly from their simultaneous (object matches) and do not seem to be biased towards category prototypes.

Different from Newhall et al. (1957) we do not find a bias in our memory matches towards higher Value although as those authors we do find an increase in Chroma and in the variability of participants memory matches when compared to the object present (simultaneous) condition.

5.6 Conclusion

In the case of unique singular objects we find, in average, across both the long- and short-term memory tasks a tendency for objects to be remembered as more saturated (higher Chroma) than they are actually perceived (simultaneous match with object present).

We do not find a systematic bias for the other two dimensions in Munsell space, Hue and Value, nor is there evidence for a systematic bias towards category prototypes.

Chapter

6 General Conclusion

In our everyday lives, our visual system keeps colour percepts stable through illuminant changes. Subjectively, as far as we can tell from daily experience, failure of these constant percepts are quite rare. If frequent failures would occur, applying colour-labels to objects would be in vain. Still, previous research shows colour constancy is incomplete (Foster (2011)).

The first study of this dissertation shows colour constancy is not just perfect in our daily experience, but also approaches perfection in experimental conditions. The key factor to measure perfect colour constancy in experimental settings lies in the naturalness of the task, in creating an experimental situation that closely approximates the we way we experience colour constancy in our natural environment. By using real unique objects, observers were able to precisely recall their respective colour under different illuminations in a cue-rich environment. The results of this study provide evidence that colour constancy can be perfect if investigated under the right conditions and raises the question in how far past research on colour constancy was able to precisely describe the phenomenon in experimental settings. Nevertheless, the results presented here close the gap which has been present in research over the last three decades.

The second study examined whether illumination discrimination offers a solid measure for colour constancy or if it is instead driven by differences in sensitivity to saturation of colours of different hue. There was a common finding between the results presented here and the initial studies on illumination discrimination (Pearce et al. (2014); Radonjić et al. (2016)). Compared to other illumination hues, observers where less sensitive to changes towards bluish illuminations: A bias towards bluish daylight. But in the study presented here, illumination discrimination thresholds were highly correlated with detection thresholds, so this insensitivity to chromatic changes for bluish hues was not exclusive to the perception of illumination changes, but rather a general property of the visual system. This finding highlights the importance of controlling underlying sensory mechanisms involved in higher order tasks which are in the focus of research.

The third study compared colour constancy measured by achromatic adjustments under 40 different chromatic illuminations to several suggested determinants of colour constancy. The results show that colour constancy varies with illuminant colour and is especially optimized for bluish daylight illuminations, a finding replicated in the second study. Bluish daylight is the most frequent daytime natural illumination that we encounter, so it's not surprising to find that our visual system evolved with this bias. It matches our experience that we nearly never perceive bluish illuminations, but we are able to perceive the warm orangish light that is present during sunset. The reason for these differences might lay in the implicit importance of the situations that are accompanied by these illuminations. While bluish daylight appears during the middle of the day, yellowish daylight carries the important information that the day is going to end soon and that we will have to prepare for the night. Further, we often encounter objects that are partly lit by sunlight and partly covered by shadows cast by objects. As surfaces lying in shadows receive more bluish light scattered from the sky, it can be speculated that these would also be perceived more bluish if our visual system would not compensate better for bluish illuminations than it does for others. As our experiments did not test these hypothesis, this issues have to be addressed by future research. Another focus of the second study was to investigate the relation of colour constancy and colour categories. While there was no effect for colour categories, the consistency in naming of the illumination hue was negatively related to colour constancy. This finding might again be explained by a more efficient adaptation towards bluish illuminations, for which consistency was lowest.

It has been claimed that colour constancy might not be possible because of the occurrence of metamers (Logvinenko et al., 2015; Wyszecki, 1958), surfaces which look the same when seen under one illumination and different under another illumination. Examining the relationship of colour constacy and metamers in study 1 and 3 presented here lead to the conclusion that colour constancy observed here was not affected by metamers. This finding is backed up by the observation that the frequency of metameric surfaces in the natural environment is quite low, as shown by Foster, Amano, Nascimento and Foster (2006).

In the fourth study, we found a colour memory bias for real unique objects. Contrary to our expectations, it did not matter how familiar the object was. Seen frequently over months or just briefly for a few seconds, observers tended to remember the objects colours to be more saturated than they were actually perceived.

References

Adams, G. K. (1923). An Experimental Study of Memory Color and Related Phenomena. *The American Journal of Psychology*, *34*(3), 359–407.

Allen, E. C., Beilock, S. L., & Shevell, S. K. (2012). Individual differences in simultaneous color constancy are related to working memory. *Journal of the Optical Society of America A*, *29*(2), A52–A59. https://doi.org/10.1364/JOSAA.29.000A52

Allred, S. R., & Olkkonen, M. (2013). The effect of background and illumination on color identification of real, 3D objects. *Frontiers in Psychology*, *4*(NOV), 1–14. https://doi.org/10.3389/fpsyg.2013.00821

Arend, L. E., & Reeves, A. (1986). Simultaneous color constancy. *Journal of the Optical Society of America. A, Optics and Image Science*, *3*(10), 1743–1751. https://doi.org/10.1364/JOSAA.3.001743

Arend, L. E., Reeves, A., Schirillo, J., & Goldstein, R. (1991). Simultaneous color constancy: paper with diverse Munsell values. *Journal of the Optical Society of America. A, Optics and Image Science*, *8*(4), 661–672. https://doi.org/10.1364/JOSAA.8.000661

Arnold, S., Savolainen, V., & Chittka, L. (2008). FReD: The floral reflectance spectra database. *Nature Precedings*. https://doi.org/10.1038/npre.2008.1846.1

Aston, S., Le Couteur Bisson, T., Jordan, G., & Hurlbert, A. C. (2016). Better colour constancy or worse discrimination? Illumination discrimination in colour anomalous observers. *Perception*, *45*(S2), 207. https://doi.org/10.1177/0301006616671273

Bae, G.-Y., Olkkonen, M., Allred, S. R., & Flombaum, J. I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *Journal of Experimental Psychology: General*, *144*(4), 744–763. https://doi.org/10.1037/xge0000076

Barnard, K., Martin, L., Funt, B., & Coath, A. (2002). A data set for color research. *Color Research & Application*, *27*(3), 147–151. https://doi.org/10.1002/col.10049

Bartleson, C. J. (1960). Memory Colors of Familiar Objects*. *Journal of the Optical*

Society of America, *50*(1), 73–77. https://doi.org/10.1364/JOSA.50.000073

Bäuml, K. H. (1999). Simultaneous color constancy: How surface color perception varies with the illuminant. *Vision Research*, *39*(8), 1531–1550. https://doi.org/10.1016/S0042-6989(98)00192-8

Beer, R. D., Dinca, A., & MacLeod, D. I. A. (2006). Ideal white can be yellowish or bluish, but not reddish or greenish. *Journal of Vision*, *6*(6), 417–417. https://doi.org/10.1167/6.6.417

Berlin, B., & Kay, P. (1969). *Basic Color Terms: their Universality and Evolution*. Berkeley and Los Angeles: University of California Press. Retrieved from http://wals.info/refdb/record/Berlin-and-Kay-1969

Berns, R. (n.d.). Artist paint spectral database. Retrieved May 3, 2017, from https://www.rit.edu/cos/colorscience/mellon/Publications/Artist_Spectral_Database_CIC2016.pdf

Bloj, M., Kersten, D., & Hurlbert, A. C. (1999). Perception of three-dimensional shape influences colour perception through mutual illumination. *Nature*, *402*(6764), 877–879. https://doi.org/10.1038/47245

Bloj, M., Weiß, D., & Gegenfurtner, K. R. (2016). Bias effects of short- and long-term color memory for unique objects. *Journal of the Optical Society of America A: Optics Image Science and Vision*, *33*(4), 492–500. https://doi.org/10.1364/JOSAA.33.000492

Bornstein, M. H. (1985). On the development of color naming in young children: Data and theory. *Brain and Language*, *26*(1), 72–93. https://doi.org/10.1016/0093-934X(85)90029-X

Bosten, J. M., Beer, R. D., & MacLeod, D. I. A. (2015). What is white? *Journal of Vision*, *15*(16), 5. https://doi.org/10.1167/15.16.5

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

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

https://doi.org/10.1364/JOSAA.15.000307

Brainard, D. H. (2003). Color Appearance and Color Difference Specification. In S. K. Shevell (Ed.), *The Science of Color* (Vol. 116, pp. 191–216). Washington DC: Optical Society of America. https://doi.org/10.1016/B978-044451251-2/50006-4

Brainard, D. H. (2016). Color Psychophysics in the Distal Stimulus. *PERCEPTION*, *45*(S2), 174.

Brainard, D. H., Brunt, W. A., & Speigle, J. M. (1997). Color constancy in the nearly natural image .1. Asymmetric matches. *Journal of the Optical Society of America a-Optics Image Science and Vision*, *14*(9), 2091–2110. https://doi.org/10.1364/josaa.14.002091

Brainard, D. H., & Hurlbert, A. C. (2015). Colour Vision: Understanding #TheDress. *Current Biology*, *25*(13), R551–R554. https://doi.org/10.1016/j.cub.2015.05.020

Brainard, D. H., & Radonjić, A. (2014). Color Constancy. In J. Werner & L. Chalupa (Eds.), *The new visual neurosciences* (pp. 545–556). Cambridge, MA: MIT Press.

Brainard, D. H., & Wandell, B. A. (1986). Analysis of the retinex theory of color vision. *Journal of the Optical Society of America A*, *3*(10), 1651–1661. https://doi.org/10.1364/JOSAA.3.001651

Bramwell, D. I., & Hurlbert, A. C. (1996). Measurements of colour constancy by using a forced-choice matching technique. *Perception*, *25*(2), 229–241. https://doi.org/10.1068/p250229

Brown, R. O., & MacLeod, D. I. A. (1997). Color appearance depends on the variance of surround colors. *Current Biology : CB*, *7*(11), 844–849. https://doi.org/10.1016/S0960-9822(06)00372-1

Buchsbaum, G. (1980). A spatial processor model for object colour perception. *Journal of the Franklin Institute*, *310*(1), 1–26. https://doi.org/10.1016/0016-0032(80)90058-7

Burns, S. A., Cohen, J. B., & Kuznetsov, E. N. (1989). Multiple metamers: Preserving color matches under diverse illuminants. *Color Research & Application*, *14*(1), 16–22. https://doi.org/10.1002/col.5080140106

Chauhan, T., Perales, E., Xiao, K., Hird, E., Karatzas, D., & Wuerger, S. M. (2014). The achromatic locus: effect of navigation direction in color space. *Journal of Vision*, *14*(1), 1–11. https://doi.org/10.1167/14.1.25

Churma, M. E. (1994). Blue shadows: physical, physiological, and psychological causes. *Applied Optics*, *33*(21), 4719–4722. https://doi.org/10.1364/AO.33.004719

Cohen, J. B., & Kappauf, W. E. (1982). Metameric color stimuli, fundamental metamers, and Wyszecki's metameric blacks. *The American Journal of Psychology*, *95*(4), 537–564.

Colour & Vision Research Laboratory (CVRL). (1995). Retrieved March 5, 2017, from http://www.cvrl.org/

Craven, B. J., & Foster, D. H. (1992). An operational approach to colour constancy. *Vision Research*, *32*(7), 1359–1366. https://doi.org/10.1016/0042-6989(92)90228-B

Daugirdiene, A., Kulikowski, J. J., Murray, I. J., & Kelly, J. M. F. (2016). Test illuminant location with respect to the Planckian locus affects chromaticity shifts of real Munsell chips. *Journal of the Optical Society of America A*, *33*(3), A77–A84. https://doi.org/10.1364/JOSAA.33.000A77

de Almeida, V. M. N., Fiadeiro, P. T., & Nascimento, S. M. C. (2004). Color constancy by asymmetric color matching with real objects in three-dimensional scenes. *Visual Neuroscience*, *21*(3), 341–345. https://doi.org/10.1017/S0952523804213074

de Almeida, V. M. N., Fiadeiro, P. T., & Nascimento, S. M. C. (2010). Effect of scene dimensionality on colour constancy with real three-dimensional scenes and objects. *Perception*, *39*(6), 770–779. https://doi.org/10.1068/p6485

Delahunt, P. B., & Brainard, D. H. (2004). Does human color constancy incorporate the statistical regularity of natural daylight? *Journal of Vision*, *4*(2), 1. https://doi.org/10.1167/4.2.1

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

DiCarlo, J. M., & Wandell, B. A. (2000). Illuminant estimation: Beyond the bases. *IS&T/SID Eighth Color Imaging Conference*, 91–96.

Faul, F., Ekroll, V., & Wendt, G. (2008). Color appearance: The limited role of chromatic surround variance in the “gamut expansion effect.” *Journal of Vision*, *8*(3), 30. https://doi.org/10.1167/8.3.30

Fitzgibbon, A., Pilu, M., & Fisher, R. B. (1999). Direct least square fitting of ellipses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *21*(5), 476–480. https://doi.org/10.1109/34.765658

Foster, D. H. (2011). Color constancy. *Vision Research*, *51*(7), 674–700. https://doi.org/10.1016/j.visres.2010.09.006

Foster, D. H., Amano, K., Nascimento, S. M. C., & Foster, M. J. (2006). Frequency of metamerism in natural scenes. *Journal of the Optical Society of America A*, *23*(10), 2359–2372. https://doi.org/10.1364/JOSAA.23.002359

Foster, D. H., Amano, K., & Nascimento, Sé. M. C. (2003). Tritanopic Colour Constancy Under Daylight Changes? In J. D. Mollon, J. Pokorny, & K. Knoblauch (Eds.), *Normal and Defective Colour Vision* (pp. 218–224). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198525301.003.0022

Foster, D. H., & Nascimento, S. M. (1994). Relational colour constancy from invariant cone-excitation ratios. *Proceedings. Biological Sciences*, *257*(1349), 115–21. https://doi.org/10.1098/rspb.1994.0103

Foster, D. H., Nascimento, S. M. ., Craven, B. ., Linnell, K. J., Cornelissen, F. W., & Brenner, E. (1997). Four issues concerning colour constancy and relational colour constancy. *Vision Research*, *37*(10), 1341–1345. https://doi.org/10.1016/S0042-6989(96)00285-4

Franklin, A., & Davies, I. R. L. (2004). New evidence for infant colour categories. *British Journal of Developmental Psychology*, *22*(3), 349–377. https://doi.org/10.1348/0261510041552738

Fründ, I., Haenel, N. V., & Wichmann, F. A. (2011). Inference for psychometric functions in the presence of nonstationary behavior. *Journal of Vision*, *11*(6), 16–16. https://doi.org/10.1167/11.6.16

Funt, B. V., Drew, M. S., & Ho, J. (1991). Color constancy from mutual reflection.

International Journal of Computer Vision, *6*(1), 5–24. https://doi.org/10.1007/BF00127123

Gegenfurtner, K. R. (2003). Cortical mechanisms of colour vision. *Nature Rev. Neurosci.*, *4*(July), 563–572. https://doi.org/10.1038/nrn1138

Gegenfurtner, K. R., Bloj, M., & Toscani, M. (2015). The many colours of “the dress.” *Current Biology*, *25*(13), R543–R544. https://doi.org/10.1016/j.cub.2015.04.043

Gegenfurtner, K. R., & Rieger, J. (2000). Sensory and cognitive contributions of color to the recognition of natural scenes. *Current Biology*, *10*(13), 805–808. https://doi.org/10.1016/S0960-9822(00)00563-7

Gegenfurtner, K. R., Wichmann, F. A., & Sharpe, L. T. (1998). The contribution of color to visual memory in X-chromosome-linked dichromats. *Vision Research*, *38*(7), 1041–1045. https://doi.org/10.1016/S0042-6989(97)00200-9

Giesel, M., Hansen, T., & Gegenfurtner, K. R. (2009). The discrimination of chromatic textures. *Journal of Vision*, *9*(9), 11–28. https://doi.org/10.1167/9.9.11

Granzier, J. J. M., Brenner, E., & Smeets, J. B. J. (2009). Can illumination estimates provide the basis for color constancy? *Journal of Vision*, *9*(3), 18.1-11. https://doi.org/10.1167/9.3.18

Granzier, J. J. M., & Gegenfurtner, K. R. (2012). Effects of memory colour on colour constancy for unknown coloured objects. *I-Perception*, *3*(3), 190–215. https://doi.org/10.1068/i0461

Granzier, J. J. M., & Valsecchi, M. (2014). Variations in daylight as a contextual cue for estimating season, time of day, and weather conditions. *Journal of Vision*, *14*(1). https://doi.org/10.1167/14.1.22

Haanpalo, J. (n.d.). Paper spectra. Retrieved March 5, 2017, from http://www.uef.fi/en/web/spectral/paper-spectra

Hanawalt, N. G., & Post, B. E. (1942). Memory trace for color. *Journal of Experimental Psychology*, *30*(3), 216–227. https://doi.org/10.1037/h0055672

Hansen, T., Giesel, M., & Gegenfurtner, K. R. (2008). Chromatic discrimination of natural objects. *Journal of Vision*, *8*(1: 2), 1–19. https://doi.org/10.1167/8.1.2

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience*, *9*(11), 1367–1368. https://doi.org/10.1038/nn1794

Hansen, T., Walter, S., & Gegenfurtner, K. R. (2007). Effects of spatial and temporal context on color categories and color constancy. *Journal of Vision*, *7*(4), 1–15. https://doi.org/10.1167/7.4.2

Hedrich, M., Bloj, M., & Ruppertsberg, A. I. (2009). Color constancy improves for real 3D objects. *Journal of Vision*, *9*(4), 16–16. https://doi.org/10.1167/9.4.16

Heider, E. R. (1972). Universals in color naming and memory. *Journal of Experimental Psychology*, *93*(1), 10–20.

Hering, E. (1920). *Grundzüge der Lehre vom Lichtsinn*. Berlin: Springer-Verlag.

Hiltunen, J. (n.d.). Lumber spectra. Retrieved March 5, 2017, from http://www.uef.fi/en/web/spectral/lumber-spectra

Holway, A. H., & Boring, E. G. (1941). Determinants of Apparent Visual Size with Distance Variant. *The American Journal of Psychology*, *54*(1), 21–37. https://doi.org/10.2307/1417790

Hurlbert, A. C. (2007). Colour constancy. *Current Biology*, *17*(21), R906–R907. https://doi.org/10.1016/j.cub.2007.08.022

Ishihara, S. (2004). Ishihara's Tests for Colour Deficiency. Tokyo, Japan: Kanehara Trading Inc.

Jaaskelainen, T., Silvennoinen, R., Hiltunen, J., & Parkkinen, J. P. S. (1994). Classification of the reflectance spectra of pine, spruce, and birch. *Applied Optics*, *33*(12), 2356–2362. https://doi.org/10.1364/AO.33.002356

Jin, E. W., & Shevell, S. K. (1996). Color memory and color constancy. *Journal of the Optical Society of America A*, *13*(10), 1981–1991. https://doi.org/10.1364/JOSAA.13.001981

Judd, D. B. (1940). Hue Saturation and Lightness of Surface Colors with Chromatic

Illumination. *Journal of the Optical Society of America*, *30*(1), 2–32. https://doi.org/10.1364/JOSA.30.000002

Judd, D. B. (1951). *Report of U.S. Secretariat Committee on Colorimetry and Artificial Daylight*. *Proceedings of the Twelfth Session of the CIE* (Vol. 1). Paris: Bureau Central de la CIE.

Judd, D. B., MacAdam, D. L., Wyszecki, G., Budde, H. W., Condit, H. R., Henderson, S. T., & Simonds, J. L. (1964). Spectral Distribution of Typical Daylight as a Function of Correlated Color Temperature. *Journal of the Optical Society of America*, *54*(8), 1031–1040. https://doi.org/10.1364/JOSA.54.001031

Kanematsu, E., & Brainard, D. H. (2014). No measured effect of a familiar contextual object on color constancy. *Color Research and Application*, *39*(4), 347–359. https://doi.org/10.1002/col.21805

Kleiner, M., Brainard, D. H., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*, ECVP Abstract Supplement.

Koffka, K. (1935). *Principles of Gestalt Psychology*. New York: Harcourt Brace.

Kohonen, O., Parkkinen, J. P. S., & Jääskeläinen, T. (2006). Databases for spectral color science. *Color Research & Application*, *31*(5), 381–390. https://doi.org/10.1002/col.20244

Kraft, J. M., & 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), 307–312. https://doi.org/DOI 10.1073/pnas.96.1.307

Kraft, J. M., Maloney, S. I., & Brainard, D. H. (2002). Surface-illuminant ambiguity and color constancy: Effects of scene complexity and depth cues. *Perception*, *31*(2), 247–263. https://doi.org/10.1068/p08sp

Krauskopf, J., & Gegenfurtner, K. R. (1992). Color discrimination and adaptation. *Vision Research*, *32*(11), 2165–2175. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1304093

Kulikowski, J. J., & Vaitkevicius, H. (1997). Colour constancy as a function of hue. *Acta Psychologica*, *97*(1), 25–35. https://doi.org/10.1016/S0001-6918(97)00022-X

Lafer-Sousa, R., Hermann, K. L., & Conway, B. R. (2015). Striking individual differences in color perception uncovered by “the dress” photograph. *Current Biology*, *25*(13), R545–R546. https://doi.org/10.1016/j.cub.2015.04.053

Land, E. H., & McCann, J. J. (1971). Lightness and Retinex Theory. *Journal of the Optical Society of America*, *61*(1), 1–11. https://doi.org/10.1364/JOSA.61.000001

Langford, M., & Bilissi, E. (2011). *Langford’s advanced photography : the guide for aspiring photographers*. Taylor and Francis.

Lee, H.-C. (1986). Method for computing the scene-illuminant chromaticity from specular highlights. *Journal of the Optical Society of America A*, *3*(10), 1694–1699. https://doi.org/10.1364/JOSAA.3.001694

Lee, R. J., Mollon, J. D., Zaidi, Q., & Smithson, H. E. (2009). Latency characteristics of the short-wavelength-sensitive cones and their associated pathways. *Journal of Vision*, *9*(12), 5.1-17. https://doi.org/10.1167/9.12.5

Ling, Y., & Hurlbert, A. C. (2008). Role of color memory in successive color constancy. *Journal of the Optical Society of America A*, *25*(6), 1215–1226. https://doi.org/10.1364/JOSAA.26.001544

Logvinenko, A. D. (2009). An object-color space. *Journal of Vision*, *9*(11), 5: 1-23. https://doi.org/10.1167/9.11.5

Logvinenko, A. D., Funt, B., & Godau, C. (2014). Metamer Mismatching. *IEEE Transactions on Image Processing*, *23*(1), 34–43. https://doi.org/10.1109/TIP.2013.2283148

Logvinenko, A. D., Funt, B., Mirzaei, H., & Tokunaga, R. (2015). Rethinking Colour Constancy. *PLOS ONE*, *10*(9), e0135029. https://doi.org/10.1371/journal.pone.0135029

Lotto, R. B., & Purves, D. (2002). The empirical basis of color perception. *Consciousness and Cognition*, *11*(4), 609–629. https://doi.org/10.1016/S1053-8100(02)00014-4

Luo, M. R., Cui, G., & Rigg, B. (2001). The development of the CIE 2000 colour-difference formula: CIEDE2000. *Color Research & Application*, *26*(5), 340–350. https://doi.org/10.1002/col.1049

Malkoc, G., Kay, P., & Webster, M. A. (2005). Variations in normal color vision IV Binary hues and hue scaling. *Journal of the Optical Society of America A*, *22*(10), 2154–2168. https://doi.org/10.1364/JOSAA.22.002154

Marszalec, E. (n.d.). Agfa it8.7/2 set. Retrieved March 5, 2017, from https://www.uef.fi/web/spectral/agfa-it8.7/2-set

Matsumoto, Y., Hiramatsu, C., Matsushita, Y., Ozawa, N., Ashino, R., Nakata, M., … Kawamura, S. (2014). Evolutionary renovation of L/M opsin polymorphism confers a fruit discrimination advantage to ateline New World monkeys. *Molecular Ecology*, *23*(7), 1799–1812. https://doi.org/10.1111/mec.12703

McCamy, C. S., Marcus, H., & Davidson, J. G. (1976). A Color-Rendition Chart. *Journal of Applied Photographic Engineering*, *2*(3), 95–99.

McCann, J. J., McKee, S. P., & Taylor, T. H. (1976). Quantitative studies in retinex theory a comparison between theoretical predictions and observer responses to the “color mondrian” experiments. *Vision Research*, *16*(5), 445–458. https://doi.org/10.1016/0042-6989(76)90020-1

Mollon, J. D. (2006). Monge: The Verriest Lecture, Lyon, July 2005. *Visual Neuroscience*, *23*(3–4), 297–309. https://doi.org/10.1017/S0952523806233479

Monge, G. (1789). Memoire sur quelques phénomènes de la vision Annales de Chimie, 3, pp. 131–147 (Commentary and Translation by Kuehni, R. G.). *Color Research and Application*, *22*(3), 199–203. https://doi.org/10.1002/(SICI)1520-6378(199706)22:3<199::AID-COL8>3.0.CO;2-M

Munsell Color Services. (1990). The Munsell Book of Color. New York: Gretag Macbeth.

Munsell Color Services. (2007). The Munsell Book of Color - Matte Collection. Grandville, MI: x-rite.

Murray, I. J., Daugirdiene, A., Vaitkevicius, H., Kulikowski, J. J., & Stanikunas, R. (2006). Almost complete colour constancy achieved with full-field adaptation. *Vision Research*, *46*(19), 3067–3078. https://doi.org/10.1016/j.visres.2006.03.011

Nascimento, S. M. ., de Almeida, V. M. N., Fiadeiro, P. T., & Foster, D. H. (2004).

Minimum-variance cone-excitation ratios and the limits of relational color constancy. *Visual Neuroscience*, *21*(3), 337–340. https://doi.org/doi:10.1017/S095252380421327X

Nascimento, S. M. C., & Foster, D. H. (1997). Detecting natural changes of cone-excitation ratios in simple and complex coloured images. *Proceedings of the Royal Society B: Biological Sciences*, *264*(1386), 1395–1402. https://doi.org/10.1098/rspb.1997.0194

Nascimento, S. M. C., & Foster, D. H. (2000). Relational color constancy in achromatic and isoluminant images. *Journal of the Optical Society of America A*, *17*(2), 225. https://doi.org/10.1364/JOSAA.17.000225

Newhall, S. M., Burnham, R. W., & Clark, J. R. (1957). Comparison of Successive with Simultaneous Color Matching. *Journal of the Optical Society of America*, *47*(1), 43–56. https://doi.org/10.1364/JOSA.47.000043

Olkkonen, M., Hansen, T., & Gegenfurtner, K. R. (2008). Color appearance of familiar objects: effects of object shape, texture, and illumination changes. *Journal of Vision*, *8*(5), 13.1-16. https://doi.org/10.1167/8.5.13

Olkkonen, M., Hansen, T., & Gegenfurtner, K. R. (2009). Categorical color constancy for simulated surfaces. *Journal of Vision*, *9*(12), 1–18. https://doi.org/10.1167/9.12.6

Olkkonen, M., Witzel, C., Hansen, T., & Gegenfurtner, K. R. (2010). Categorical color constancy for real surfaces. *Journal of Vision*, *10*(9), 1–22. https://doi.org/10.1167/10.9.16

Orava, J. (2002). Munsell colors glossy (all) (spectrofotometer measured). Retrieved March 5, 2017, from http://www.uef.fi/en/web/spectral/munsell-colors-glossy-all-spectrofotometer-measured

Parkkinen, J. P. S., Hallikainen, J., & Jaaskelainen, T. (1989). Characteristic spectra of Munsell colors. *Journal of the Optical Society of America A*, *6*(2), 318–322. https://doi.org/10.1364/JOSAA.6.000318

Parkkinen, J. P. S., Jaaskelainen, T., & Kuittinen, M. (1988). Spectral representation of color images. In *[1988 Proceedings] 9th International Conference on Pattern*

Recognition (pp. 933–935). IEEE Comput. Soc. Press. https://doi.org/10.1109/ICPR.1988.28405

Pearce, B., Crichton, S., Mackiewicz, M., Finlayson, G. D., & Hurlbert, A. C. (2014). Chromatic illumination discrimination ability reveals that human colour constancy is optimised for blue daylight illuminations. *PLoS ONE*, *9*(2). https://doi.org/10.1371/journal.pone.0087989

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*(4), 437–42.

Pérez-Carpinell, J., Baldoví, R., de Fez, M. D., & Castro, J. (1998). Color memory matching: Time effect and other factors. *Color Research & Application*, *23*(4), 234–247. https://doi.org/10.1002/(SICI)1520-6378(199808)23:4<234::AID-COL8>3.0.CO;2-P

Philipona, D. L., & O'Regan, J. K. (2006). Color naming, unique hues, and hue cancellation predicted from singularities in reflection properties. *Visual Neuroscience*, *23*(3–4), 331–339. https://doi.org/10.1017/S0952523806233182

Radonjić, A., & Brainard, D. H. (2016). The nature of instructional effects in color constancy. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(6), 847–865. https://doi.org/10.1037/xhp0000184

Radonjić, A., Cottaris, N. P., & Brainard, D. H. (2015a). Color constancy in a naturalistic goal-directed task. *Journal of Vision*, *15*(13), 3.1-3.21. https://doi.org/10.1167/15.13.3.doi

Radonjić, A., Cottaris, N. P., & Brainard, D. H. (2015b). Color constancy supports cross-illumination color selection. *Journal of Vision*, *15*(6), 13. https://doi.org/10.1167/15.6.13

Radonjić, A., Pearce, B., Aston, S., Krieger, A., Dubin, H., Cottaris, N. P., … Hurlbert, A. C. (2016). Illumination discrimination in real and simulated scenes. *Journal of Vision*, *16*(11), 2. https://doi.org/10.1167/16.11.2

Regan, B. ., Julliot, C., Simmen, B., Viénot, F., Charles-Dominique, P., & Mollon, J. D. (1998). Frugivory and colour vision in Alouatta seniculus, a trichromatic platyrrhine monkey. *Vision Research*, *38*(21), 3321–3327. https://doi.org/10.1016/S0042-6989(97)00462-8

Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(4), 1436–1441. https://doi.org/10.1073/pnas.0610341104

Rüttiger, L., Mayser, H., Sérey, L., & Sharpe, L. T. (2001). The color constancy of the red-green color blind. *Color Research & Application*, *26*(S1), S209–S213. https://doi.org/10.1002/1520-6378(2001)26:1+<::AID-COL44>3.0.CO;2-D

Sachtler, W. L., & Zaidi, Q. (1992). Chromatic and luminance signals in visual memory. *Journal of the Optical Society of America A*, *9*(6), 877–894. https://doi.org/10.1364/JOSAA.9.000877

Sangwine, S. J., & Horne, R. E. N. (1998). *The Colour Image Processing Handbook*. New York: Springer.

Schultz, S., Doerschner, K., & Maloney, L. T. (2006). Color constancy and hue scaling. *Journal of Vision*, *6*(10), 1102–1116. https://doi.org/10.1167/6.10.10

Schütt, H. H., Harmeling, S., Macke, J. H., & Wichmann, F. A. (2016). Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. *Vision Research*, *122*, 105–123. https://doi.org/10.1016/j.visres.2016.02.002

Shapley, R., & Hawken, M. J. (2011). Color in the cortex: single- and double-opponent cells. *Vision Research*, *51*(7), 701–717. https://doi.org/10.1016/j.visres.2011.02.012

Shepard, R. N. (1992). The perceptual organization of colors: An adaptation to regularities of the terrestrial world? In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *Adapted Mind* (pp. 495–532). Oxford: Oxford University Press.

Siple, P., & Springer, R. M. (1983). Memory and preference for the colors of objects. *Perception & Psychophysics*, *34*(4), 363–370. https://doi.org/10.3758/BF03203049

Smet, K. A. G., Ryckaert, W. R., Pointer, M. R., Deconinck, G., & Hanselaer, P. (2012). A memory colour quality metric for white light sources. *Energy and Buildings*, *49*, 216–225. https://doi.org/10.1016/j.enbuild.2012.02.008

Smithson, H. E. (2005). Sensory, computational and cognitive components of human colour constancy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1458), 1329–1346. https://doi.org/10.1098/rstb.2005.1633

Smithson, H. E., & Zaidi, Q. (2004). Colour constancy in context: Roles for local adaptation and levels of reference. *Journal of Vision*, *4*(9), 693–710. https://doi.org/10.1167/4.9.3

Stockman, A., & Sharpe, L. T. (2000). The spectral sensitivities of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype. *Vision Research*, *40*(13), 1711–1737. https://doi.org/10.1016/S0042-6989(00)00021-3

Stromeyer, C. F., Eskew, R. T., Kronauer, R. E., & Spillmann, L. (1991). Temporal phase response of the short-wave cone signal for color and luminance. *Vision Research*, *31*(5), 787–803. https://doi.org/10.1016/0042-6989(91)90147-W

Tate, J. D., & Springer, R. M. (1971). Effects of memory time on successive judgments. *Psychological Bulletin*, *76*(6), 394–408. https://doi.org/10.1037/h0031847

Taylor, A. H., & Kerr, G. P. (1941). The Distribution of Energy in the Visible Spectrum of Daylight. *Journal of the Optical Society of America*, *31*(1), 3–8. https://doi.org/10.1364/JOSA.31.000003

The MathWorks Inc. (2007). Matlab - The language of Technical Computing (Version R2007a). Natick, Massachusetts: The MathWorks Inc.

Tiplitz Blackwell, K., & Buchsbaum, G. (1988). Quantitative studies of color constancy. *Journal of the Optical Society of America A*, *5*(10), 1772–1780. https://doi.org/10.1364/JOSAA.5.001772

Troost, J. M. (1998). Why things look as they do. In V. Walsh & J. Kulikowski (Eds.), *Perceptual Constancy* (pp. 262–282). Cambridge: Cambridge University Press.

Troost, J. M., & de Weert, C. M. (1991). Naming versus matching in color constancy. *Perception & Psychophysics*, *50*(6), 591–602. https://doi.org/10.3758/BF03207545

Troost, J. M., & De Weert, C. M. M. (1991). Naming versus matching in color constancy. *Perception & Psychophysics*, *50*(6), 591–602. https://doi.org/10.3758/BF03207545

Troscianko, T., Benton, C. P., Lovell, P. G., Tolhurst, D. J., & Pizlo, Z. (2009). Camouflage and visual perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1516), 449–461. https://doi.org/10.1098/rstb.2008.0218

Uchikawa, K., Kuriki, I., & Tone, Y. (1998). Measurement of color constancy by color memory matching. *Optical Review*, *5*(1), 59–63. https://doi.org/10.1007/s10043-998-0059-z

Vazquez-Corral, J., O'Regan, J. K., Vanrell, M., & Finlayson, G. D. (2012). A new spectrally sharpened sensor basis to predict color naming, unique hues, and hue cancellation. *Journal of Vision*, *12*(6), 7. https://doi.org/10.1167/12.6.7

von Helmholtz, H. (1867). No Title. In *handbuch der Physiologischen Optik* (pp. 408–409). Leipzig: Leopold Voss.

von Kries, J. (1902). Theoretische Studien über die Umstimmung des Sehorgans Freiburg. In *Festschrift der Albrecht-Ludwigs-Universität* (pp. 143–158). Freiburg.

Vos, J. J. (1978). Colorimetric and photometric properties of a 2-deg fundamental observer. *Color Research & Application*, *3*, 125–128.

Vurro, M., Ling, Y., & Hurlbert, A. C. (2013). Memory color of natural familiar objects: Effects of surface texture and 3-D shape. *Journal of Vision*, *13*(7), 20. https://doi.org/10.1167/13.7.20

Walraven, J., Benzschawel, T. L., Rogowitz, B. E., & Lucassen, M. P. (1991). Testing the Contrast Explanation of Color Constancy. In *From Pigments to Perception* (pp. 369–377). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4615-3718-2_42

Webster, M. A., Miyahara, E., Malkoc, G., & Raker, V. E. (2000). Variations in normal color vision II Unique hues. *Journal of the Optical Society of America A*, *17*(9), 1545–1555. https://doi.org/10.1364/JOSAA.17.001545

Weiss, D., Witzel, C., & Gegenfurtner, K. R. (under revision). Sensitivity to hue explains "blue bias" in perceived illumination.

Westland, S., Shaw, J., & Owens, H. (2000). Colour statistics of natural and man-made surfaces. *Sensor Review*, *20*(1), 50–55. https://doi.org/10.1108/02602280010311392

Winkler, A. D., Spillmann, L., Werner, J. S., & Webster, M. A. (2015). Asymmetries in blue–yellow color perception and in the color of "the dress." *Current Biology*, *25*(13), R547–R548. https://doi.org/10.1016/j.cub.2015.05.004

Witzel, C. (2016). New Insights Into the Evolution of Color Terms or an Effect of Saturation? *I-Perception*, *7*(5). https://doi.org/10.1177/2041669516662040

Witzel, C., Cinotti, F., & O'Regan, J. K. (2015). What determines the relationship between color naming, unique hues, and sensory singularities: Illuminations, surfaces, or photoreceptors? *Journal of Vision*, *15*(8), 19. https://doi.org/10.1167/15.8.19

Witzel, C., & Franklin, A. (2014). Do focal colors look particularly "colorful"? *Journal of the Optical Society of America A*, *31*(4), A365–A374. https://doi.org/10.1364/JOSAA.31.00A365

Witzel, C., & Gegenfurtner, K. R. (2013). Categorical sensitivity to color differences. *Journal of Vision*, *13*(7), 1. https://doi.org/10.1167/13.7.1

Witzel, C., & Gegenfurtner, K. R. (2015). Chromatic Contrast Sensitivity. In R. Luo (Ed.), *Encyclopedia of Color Science and Technology* (pp. 1–7). Berlin: Springer.

Witzel, C., Hansen, T., & Gegenfurtner, K. R. (2008). Wie sich Farben mit den Betrachtern und mit den Zeiten ändern. *Paper Presented at the Tagung Experimentell Arbeitender Psychologen (TeaP), Marburg*.

Witzel, C., Maule, J., & Franklin, A. (under revision). Are red, yellow, green and blue particularly "colorful"?

Witzel, C., Maule, J., & Franklin, A. (2013). Focal colors as perceptual anchors of color categories. *Journal of Vision*, *13*(9), 1164. https://doi.org/10.1167/13.9.1164

Witzel, C., Racey, C., & O'Regan, J. K. (2017). The most reasonable explanation of "the dress": Implicit assumptions about illumination. *Journal of Vision*, *17*(2), 1. https://doi.org/10.1167/17.2.1

Witzel, C., Valkova, H., Hansen, T., & Gegenfurtner, K. R. (2011). Object Knowledge Modulates Colour Appearance. *I-Perception*, *2*(1), 13–49. https://doi.org/10.1068/i0396

Witzel, C., van Alphen, C., Godau, C., O'Regan, J. K., Hough, C. A., J., K. C., & Simonds, J. L. (2016). Uncertainty of sensory signal explains variation of color constancy. *Journal of Vision*, *16*(15), 8. https://doi.org/10.1167/16.15.8

Wuerger, S. M., Hurlbert, A. C., & Witzel, C. (2015). Variation of subjective white-points along the daylight axis and the colour of the dress. *Perception*, *33*(S1), 153.

Wyszecki, G. (1958). Evaluation of Metameric Colors. *Journal of the Optical Society of America*, *48*(7), 451–452. https://doi.org/10.1364/JOSA.48.000451

Wyszecki, G., & Stiles, W. (1982). *Color Science: concepts and methods, quantitive data and formulae* (2nd ed.). New York: John Wiley and Sons Inc. https://doi.org/10.1002/col.5080080421

Xiao, B., Hurst, B., MacIntyre, L., & Brainard, D. H. (2012). The color constancy of three-dimensional objects. *Journal of Vision*, *12*(4), 6. https://doi.org/10.1167/12.4.6

Xue, S., Tan, M., McNamara, A., Dorsey, J., & Rushmeier, H. (2014). Exploring the use of memory colors for image enhancement. In B. E. Rogowitz, T. N. Pappas, & H. de Ridder (Eds.), *Proc. SPIE 9014 Human Vision and Electronic Imaging XIX* (p. 901411). International Society for Optics and Photonics. https://doi.org/10.1117/12.2036836

Yang, J. N., & Maloney, L. T. (2001). Illuminant cues in surface color perception: tests of three candidate cues. *Vision Research*, *41*(20), 2581–2600. https://doi.org/10.1016/S0042-6989(01)00143-2

Young, T. (1807). *A course of lectures on natural philosophy and the mechanical arts (Vol. I)*. London: Joseph Johnson.

Zaidi, Q., Ennis, R., Cao, D., & Lee, B. (2012). Neural Locus of Color Afterimages. *Current Biology*, *22*(3), 220–224. https://doi.org/10.1016/j.cub.2011.12.021

Appendix

Supplementary Material: Perfect colour constancy in real-world settings

Figure S 1.1. Photographs of the 17 objects brought by our participants. Each object is shown next to the Munsell chip most often selected as a match. Notice the variety of materials, size and colour that represent the diversity of objects for which we found almost perfect colour constancy.

Figure S 1.2. Left: Photograph of the 1325 chips from the Munsell Glossy collection displayed in 40 plastic bins, ordered by hue as they were seen by observers after completing the adaptation task and before making their memory match selection. Right: Experimental room and window with filter (C) attached. Notebook and objects are arranged to show the effect of the illuminant and were never included during the experimental procedure.

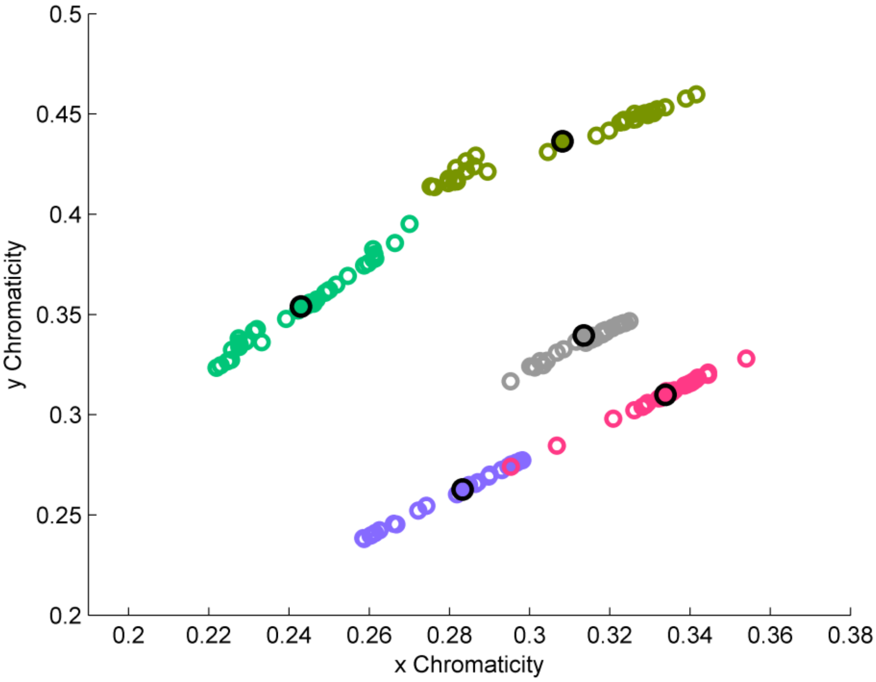


Figure S 1.3. Judd-Vos corrected CIE xy chromaticities of the five used illuminants measured in each session. Each colored circle represents the average of measurements before and after each session. Solid circles represent the chromaticity of one illuminant averaged over all measurements.

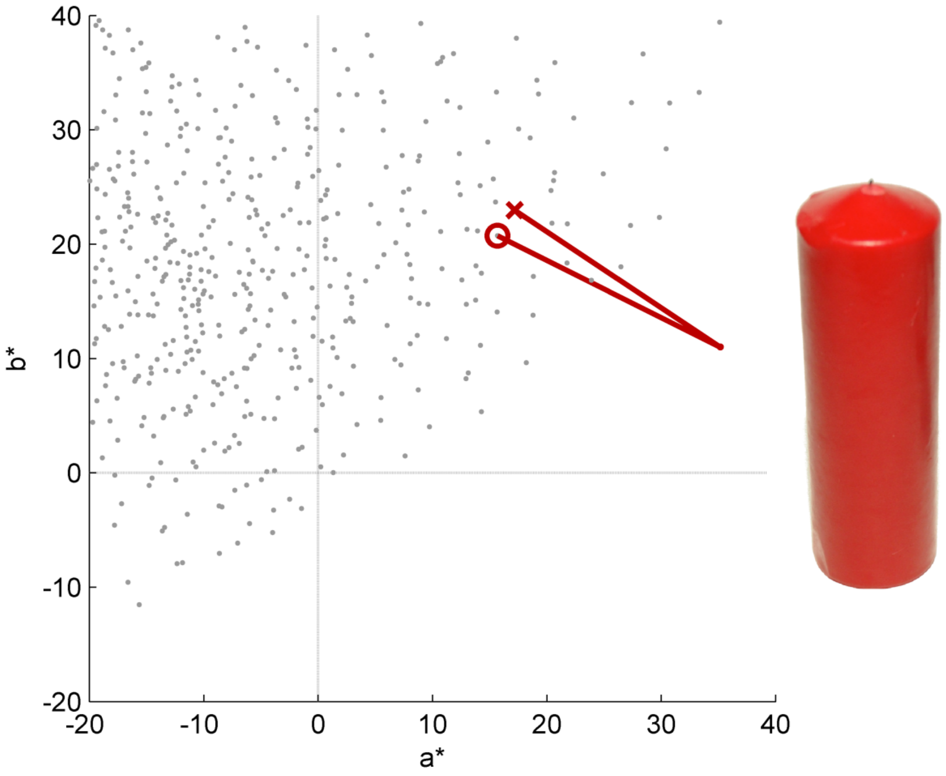


Figure S 1.4. Munsell collection rendered under illuminant B (yellow) and illuminant shift from daylight to illuminant B for chips selected for the red candle. The intersection of the Munsell collection and the line depicting the illuminant shift indicates the lower boundary of constancy for this object under illuminant B. The boundary was calculated by selecting the chip that leads to lowest constancy when taking all chips of the collection into account. This chip is closest to the coordinates of the long term memory match for the object under daylight and produces a constancy value as depicted for each object or illuminant condition by the shaded regions in figure 4.

Illuminant	Lee filter specification	x mean (SD)	y mean(SD)	Y(cd/m^2) mean (SD)
Daylight	-	0.3135 (0.0103)	0.3395 (0.0159)	186 (63)
Green (A)	242 LEE 4300K	0.2430 (0.0148)	0.3539 (0.0203)	80 (33)
Yellow (B)	138 Pale Green	0.3082 (0.0234)	0.4364 (0.0155)	175 (88)
Red (C)	035 Light Pink	0.3338 (0.0109)	0.3100 (0.0103)	67 (47)
Violet (D)	136 Pale Lavender	0.2832 (0.0143)	0.2627 (0.0143)	84 (31)

Table S 2.1. Illuminant specifications and averaged Judd-Vos corrected CIE xyY coordinates of the whitepoint measurements. The values are means (standard deviations) over all sessions for a specific illuminant.

	Size of Metameric mismatch volume			
	green	red	yellow	violet
'red book'	932.7	30.2	453.6	230.7
'light pink scarf'	1369.6	96.5	451.7	663.7
'tourquise dummy'	1840.7	138.4	718.4	1071.6
'blue spike ball'	768.8	48.1	825.7	461.2
'green t-shirt'	1928.0	208.6	600.6	1326.2
'tourquise sweater'	1079.4	111.2	475.0	776.6
'green pick'	1790.7	160.8	407.8	1075.2
'light yellow part of kitchen'	317.9	21.7	142.4	188.1
'wooden chess piece'	1977.1	154.6	398.8	724.9
'orange sweater'	1173.1	70.,4	307.4	433,.6
'brown sweater'	2724.8	197.4	940.3	1117.8
'green scarf'	2756.7	216.8	878.0	1379.8
'red candle'	1306.2	53.0	678.0	419.8
'pink rhino'	1266.7	84.9	432.7	588.0
'red metallic biketool'	998.6	34.4	534.7	255.1
'blue elephant'	823.8	50.2	873.7	476.0
'tourquise potato'	1740.4	117.6	739.5	974.1

Table S 2.2. Size of metameric mismatch volumes under each illuminant change from neutral to chromatic illumination in CIE-L*a*b.

Illuminant change	Metameric mismatch volume	
	r	p
Neutral - green	-.18	.51
Neutral - pink	-.32	.22
Neutral - yellow	-.01	.96
Neutral - violet	-.13	.83

Table S 2.3: Comparison of metameric mismatch volumes and colour constancy for each illuminant change.

	Number of metamers
'red book'	16
'light pink scarf'	6
'tourquise dummy'	0
'blue spike ball'	8
'green t-shirt'	0
'tourquise sweater'	0
'green pick'	0
'light yellow part of kitchen'	0
'wooden chess piece'	4
'orange sweater'	4
'brown sweater'	4
'green scarf'	42
'red candle'	0
'pink rhino'	6
'red metallic biketool'	0
'blue elephant'	8
'tourquise potato'	0

Table S 2.4. Number of metamers for the objects tested in our study. Frequencies are given for the five illuminants used in our experiment, out of 11,302 reflectance samples.

Supplementary Material: Sensitivity to hue explains "blue bias" in colour constancy

Experiment 1

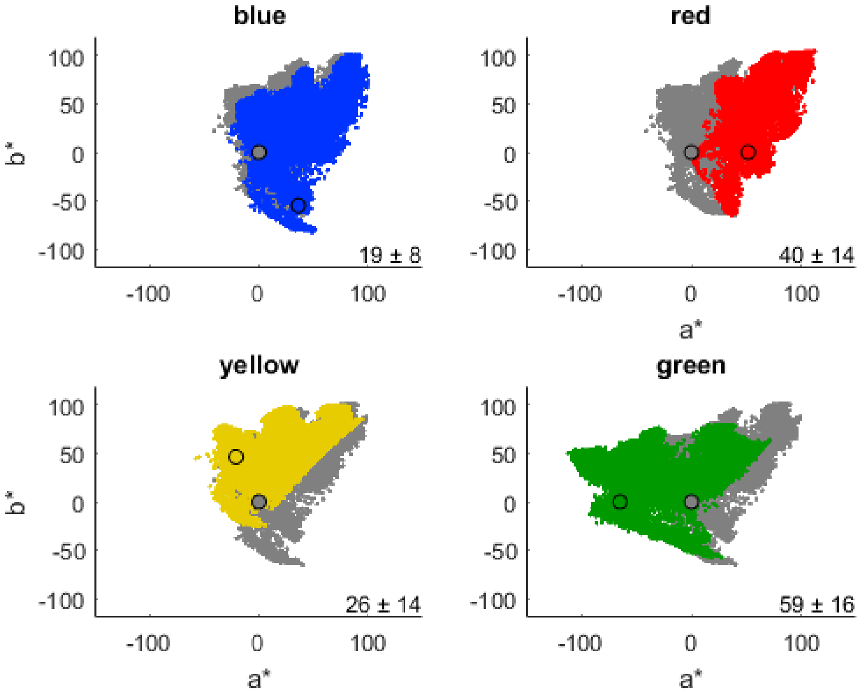


Figure S 3.1. Color shifts of the scene under different illuminations in CIELAB. Format is as in **Figure 3.4**.

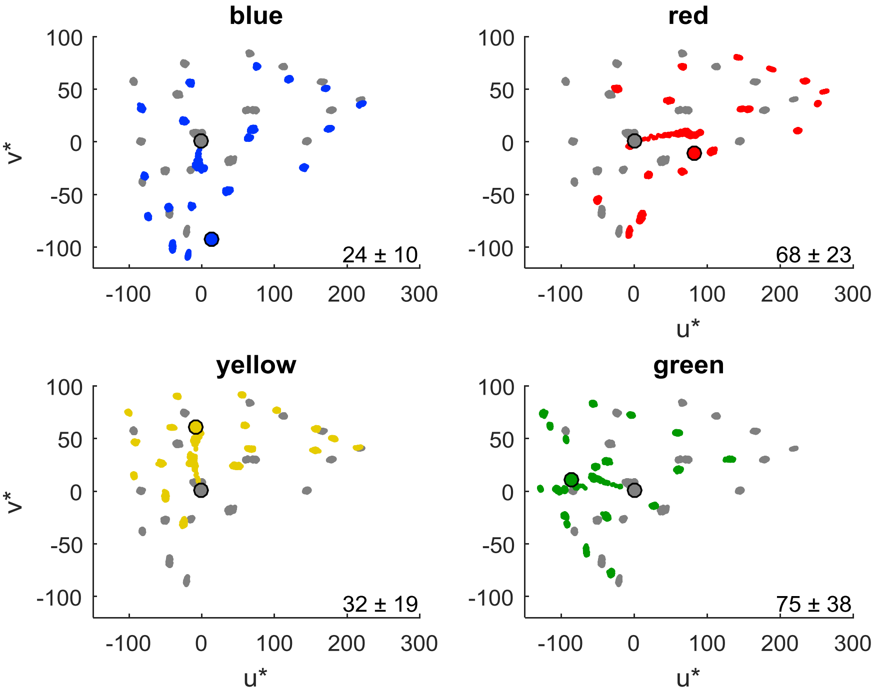


Figure S 3.2. Color shifts of the ColorChecker chart under different illuminations in CIELUV. Format is as in **Figure 3.4**.

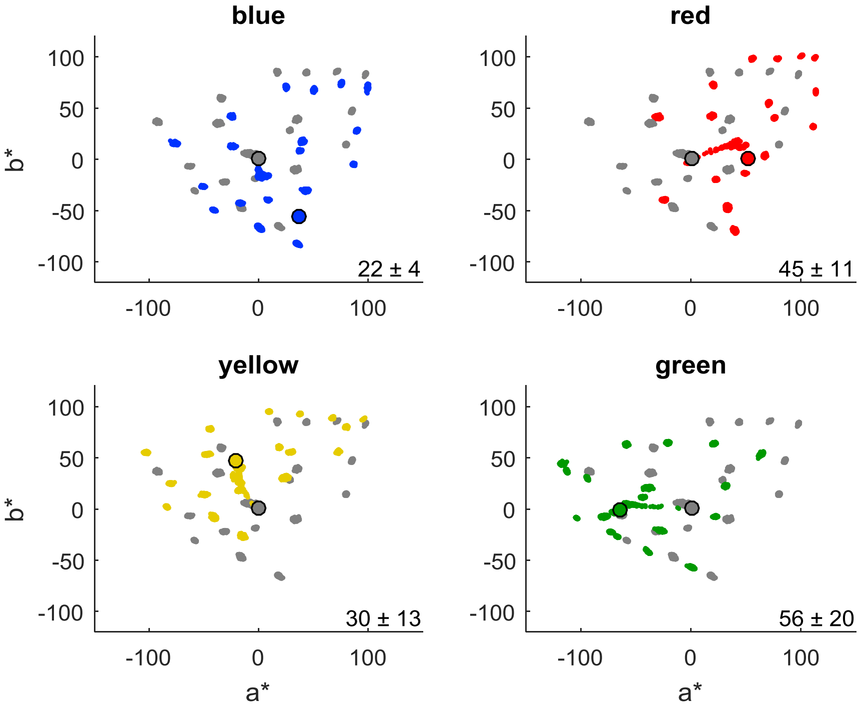


Figure S 3.3. Color shifts of the ColorChecker chart under different illuminations in CIELAB. Format is as in **Figure 3.4**.

Munsell notation	x	y	Y
2.5R7/4	0,287	0,338	46,61
5R7/4	0,281	0,339	50,22
7.5R7/4	0,281	0,333	49,21
10R7/4	0,284	0,326	47,75
2.5YR7/4	0,284	0,321	48,49
5YR7/4	0,286	0,315	48,32
7.5YR7/4	0,286	0,310	50,18
10YR7/4	0,289	0,306	49,29
2.5Y7/4	0,293	0,303	48,50
5YR7/4	0,295	0,302	47,77
7.5Y7/4	0,295	0,297	50,66
10Y7/4	0,299	0,298	48,67
2.5GY7/4	0,301	0,295	48,69
5GY7/4	0,303	0,291	49,75
7.5GY7/4	0,315	0,293	48,67
10GY7/4	0,325	0,294	49,77
2.5G7/4	0,344	0,298	52,83
5G7/4	0,358	0,306	52,94
7.5G7/4	0,367	0,312	52,89
10G7/4	0,372	0,317	53,13
2.5BG7/4	0,376	0,325	53,28

Munsell notation	x	y	Y
5BG7/4	0,380	0,333	53,59
7.5BG7/4	0,385	0,339	53,40
10BG7/4	0,382	0,345	52,46
2.5B7/4	0,378	0,350	51,62
5B7/4	0,375	0,355	51,37
7.5B7/4	0,373	0,358	51,83
10B7/4	0,368	0,363	50,38
2.5PB7/4	0,365	0,367	51,77
5PB7/4	0,360	0,370	49,98
7.5PB7/4	0,355	0,377	52,44
10PB7/4	0,349	0,380	49,85
2.5P7/4	0,343	0,383	48,77
5P7/4	0,335	0,388	49,56
7.5P7/4	0,322	0,395	51,11
10P7/4	0,310	0,395	51,28
2.5RP7/4	0,299	0,388	51,01
5RP7/4	0,290	0,375	51,81
7.5RP7/4	0,286	0,363	50,83
10RP7/4	0,283	0,352	51,95
NEUTRAL	0,327	0,342	48,70

Table S 3.1. Illuminant specifications

Supplementary Material: Determinants of colour constancy and the blue bias

Colour Naming Experiment

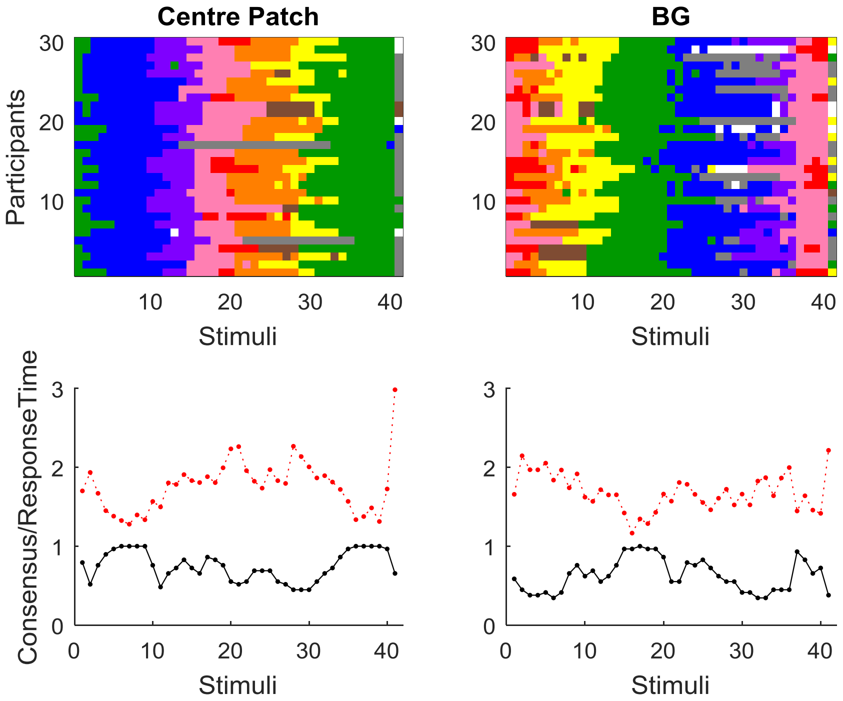


Figure S 4.4. Individual colour naming data. The black line in the lower graph indicates consistency, the dotted red line indicates response time.

DKL azimuth in degree	x	y	Y	Opponent Munsell Hue
213.8	0.287	0.338	46.61	2.5R
211.5	0.281	0.339	50.22	5R
218	0.281	0.333	49.21	7.5R
228.2	0.284	0.326	47.75	10R
234	0.284	0.321	48.49	2.5YR
240	0.286	0.315	48.32	5YR
243.6	0.286	0.310	50.18	7.5YR
250.2	0.289	0.306	49.29	10YR
254.4	0.293	0.303	48.50	2.5Y
257.2	0.295	0.302	47.77	5YR
260.3	0.295	0.297	50.66	7.5Y
263.7	0.299	0.298	48.67	10Y
266.3	0.301	0.295	48.69	2.5GY
269.2	0.303	0.291	49.75	5GY
281.3	0.315	0.293	48.67	7.5GY
292.2	0.325	0.294	49.77	10GY

315.2	0.344	0.298	52.83	2.5G
336.9	0.358	0.306	52.94	5G
351.4	0.367	0.312	52.89	7.5G
358.5	0.372	0.317	53.13	10G
8.400	0.376	0.325	53.28	2.5BG
18.60	0.380	0.333	53.59	5BG
24.70	0.385	0.339	53.40	7.5BG
29.70	0.382	0.345	52.46	10BG
36.50	0.378	0.350	51.62	2.5B
43.20	0.375	0.355	51.37	5B
45.60	0.373	0.358	51.83	7.5B
53	0.368	0.363	50.38	10B
59.90	0.365	0.367	51.77	2.5PB
65.10	0.360	0.370	49.98	5PB
76.80	0.355	0.377	52.44	7.5PB
83.70	0.349	0.380	49.85	10PB
91.30	0.343	0.383	48.77	2.5P
102.6	0.335	0.388	49.56	5P
119.1	0.322	0.395	51.11	7.5P
130.3	0.310	0.395	51.28	10P
145.2	0.299	0.388	51.01	2.5RP
163.9	0.290	0.375	51.81	5RP
179.8	0.286	0.363	50.83	7.5RP
193.3	0.283	0.352	51.95	10RP
NEUTRAL	0.327	0.342	48.70	N6.5

Table S 4.1. Illuminant specifications. Judd-corrected chromaticity and luminance of all 41

Achromatic Adjustments

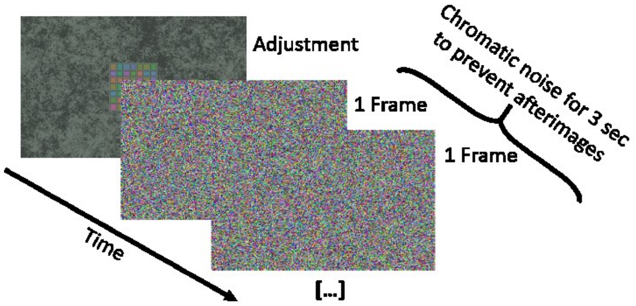


Figure S 4.5. A trial in the achromatic adjustment task. Note the chromatic noise patterns were presented after the adjustments to prevent after-images.

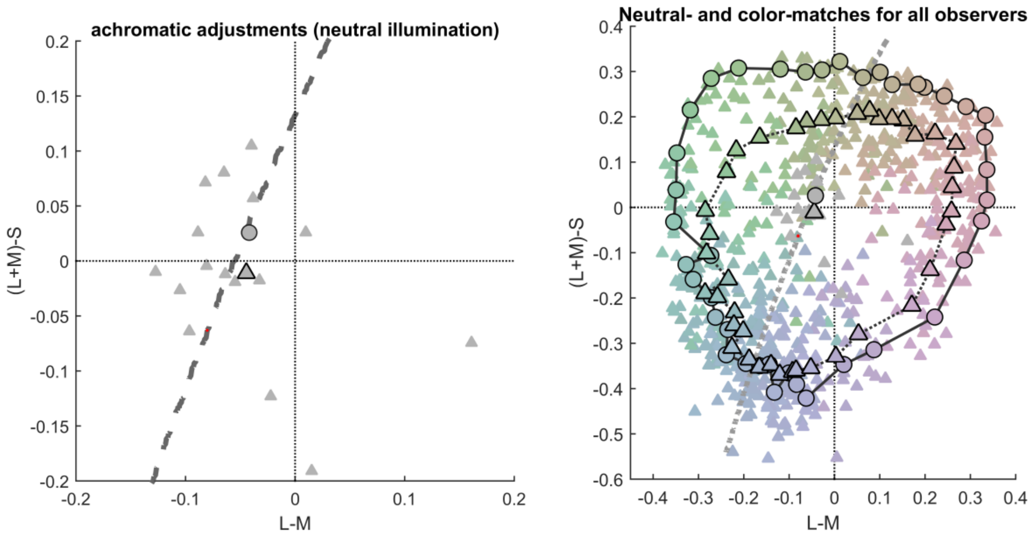


Figure S 4.6. Left: Individual adjustments under the neutral illumination. Triangles depict neutral matches for 16 observers, the neutral illuminant is indicated by the circle with the back contour. The dotted line depicts the daylight locus from 5000K (top) to 8500K (bottom). The right panel shows individual adjustments (triangles) for all 41 illuminations tested in the experiment. Black-framed triangles depict average adjustments, black-framed discs indicate the coordinates of the illumination as measured from the brighter background surface.

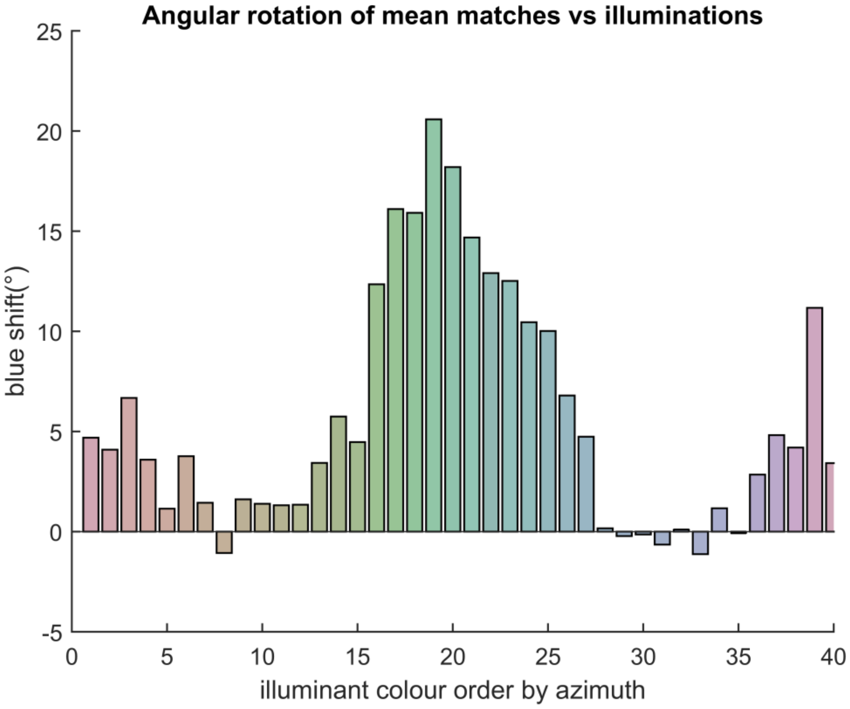


Figure S 4.7. Blue bias in angular rotations of mean matches relative to the according illuminations from yellow (78°) to daylight blue (238°). Positive values indicate that the average adjustment angle is shifted towards the angle of the bluish daylight illumination. The graphic demonstrates that the further illumination hues are away from the daylight blue–yellow axis, the more they are rotated towards blue.

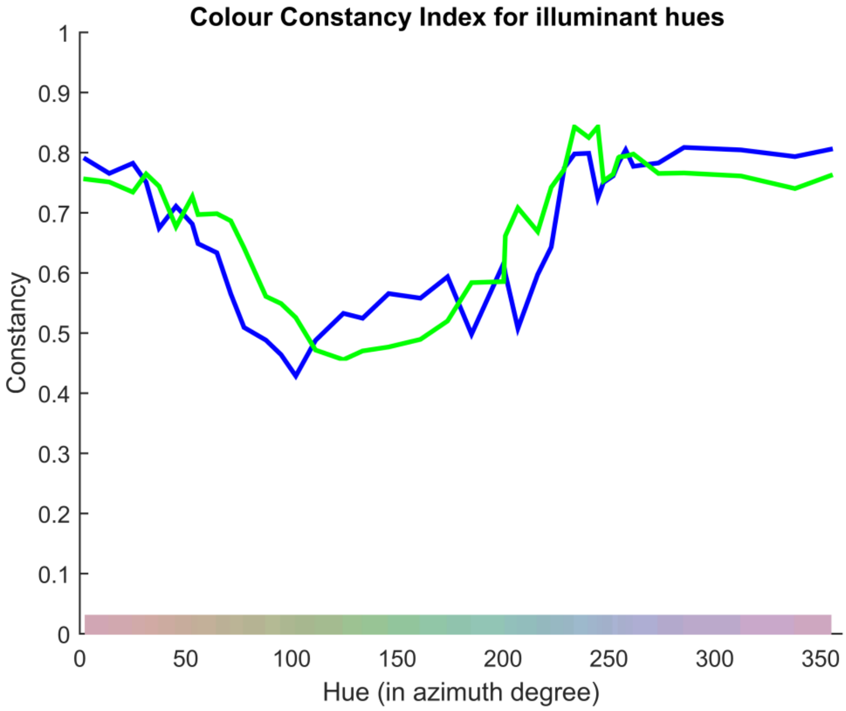


Figure S 4.8. Colour Constancy Index in DKL-space (green) and CIELAB (blue).

Comparison with illumination discrimination and chromatic detection

	Illumination discrimination		Chromatic Detection	
	r	p	r	p
CCI	-.003	.99	.08	.81
BR	.53	.08	.35	.26
Adj. shift (M)	.12	.71	-.04	.91
Adj. Shift (SD)	-.08	.80	.03	.92
Interindi. Var.	.25	.43	.36	.25

Table S 4.2. Correlations between illumination discrimination, chromatic detection, and colour constancy in DKL-space.

Candidate determinants

	df	Illumination shift		Blue bias		BG categories			
						Consensus		RT	
		r	p	r	p	r	p	r	p
CCI	39	.06	.71	**-.52**	**< .001**	**-.50**	**.001**	**.32**	**.043**
Adj. error	39	**.47**	**.002**	**.57**	**< .001**	**.68**	**<.001**	**-.49**	**.001**
Interindi. Var.	39	**.43**	**.005**	**.67**	**< .001**	**.68**	**<.001**	**-.51**	**<.001**

	Patch categories				Metamer mismatch area	
	Consensus		RT			
	r	p	r	p	R	p
CCI	.19	.24	-.23	.14	.16	.33
Adj. error	-.01	.95	.21	.19	.16	.33
Interindi. Var.	-.08	.62	.29	.07	.19	.22

	Sensory singularities				Cone ratio predictions			
	Correlations		Partial corr.		Correlation		Partial corr.	
	r	p	r	p	r	p	r	p
CCI	-.15	.35	-.09	.57	-.11	.48	-.17	.29
Adj. error	**.35**	**.02**	.12	.47	**.38**	**.015**	.17	.29
Interindi. Var.	**.41**	**.007**	.22	.17	**.41**	**.008**	.24	.14

Table S 4.3. Correlations between colour constancy and candidate determinants in CIELAB. Bold faced numbers indicate significant correlations. Partial corr. = Partial correlations controlling for illumination shift (i.e. chroma of surfaces under neutral illumination).

Illumination shift

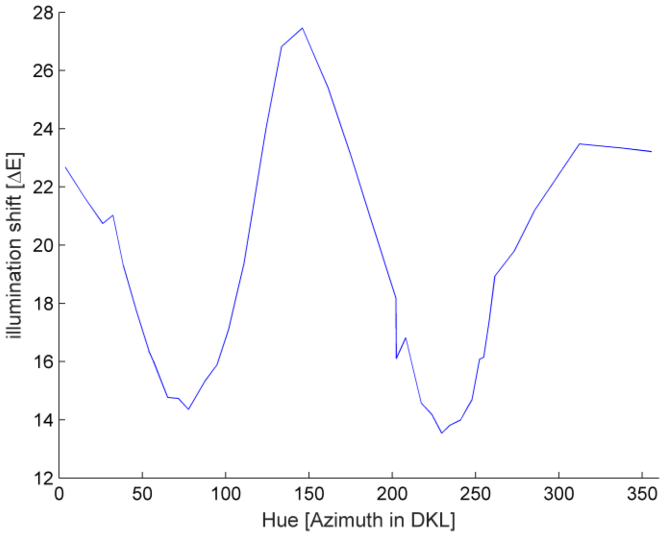


Figure S 4.9. Illumination shift in CIELAB.

Colour categories

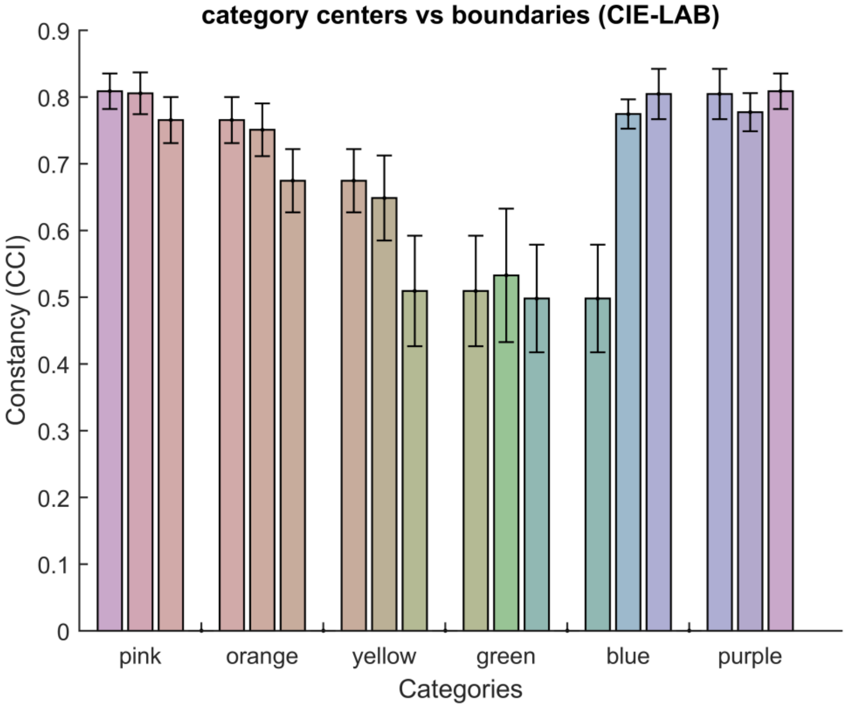


Figure S 4.10. Categorical colour constancy. The graphic compares the Colour Constancy Index between colour at the category centres and boundaries. Each group of three bars corresponds to one of the six categories. The bars in the centre of each group correspond to the category centres, the other two to the category boundaries. The y-axis represents the Colour Constancy Index. In case of categorical colour constancy, the centre colours (centre bars) should yield higher constancy (higher bars) than the boundary colours. This is not the case for any of the categories.

Erklärung

Ich erkläre: Ich habe die vorgelegte Dissertation selbstständig und ohne unerlaubte fremde Hilfe und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Ich stimme einer evtl. Überprüfung meiner Dissertation durch eine Antiplagiat-Software zu. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis“ niedergelegt sind, eingehalten.

Gießen, 04. April 2017,

David Weiß