UAB

Universitat Autònoma de Barcelona

Constancy and inconstancy

in

categorical colour perception

A dissertation submitted in fulfilment of the requirements

for the Degree of **Doctor en Informàtica**

at the Universitat Autònoma de Barcelona.

Bellaterra, November 2012

Jordi Roca i Vilà

Supervisors:

Dr. C. Alejandro Parraga
Universitat Autònoma de Barcelona
Dept. Ciències de la Computació and Computer Vision Center

Dr. Maria Vanrell
Universitat Autònoma de Barcelona
Dept. Ciències de la Computació and Computer Vision Center

The research described in this book was carried out at the Computer Vision Center, Universitat Autònoma de Barcelona, Bellaterra, Catalonia.

iv

a la Mireia,

Agraïments

Primer de tot, voldria agrair a la Mireia el seu suport i la seva generosa comprensió, els quals han estat constants en la quotidianitat dels anys en què aquest projecte s'ha estès. També voldria agrair als meus pares, a la família i als amics els seu suport i, en concret, agrair la paciència reiterada per mantenir-me connectat malgrat les meves intenses tribulacions.

Seguidament, voldria expressar el meu agraïment a la Maria i a l'Alejandro, els meus directors de tesi. Primer de tot, per donar-me la oportunitat d'emprendre un projecte tan definitori en el procés formatiu d'una persona com n'és la realització d'un doctorat. També la seva constant disponibilitat i paciència per tal de discutir de forma enraonada i reiterada tots i cadascun dels aspectes de la tesi. I reconèixer que el tracte d'afecte i entesa mútua en el qual s'ha desenvolupat aquest projecte ha resultat en una enriquidora amistat. Però per damunt de tot, el meu agraïment pel seu esforç constant per formar-me com a científic.

També voldria expressar el meu agraïment a l'Anya Hurlbert, la qual em va acollir en el seu laboratori i em va permetre complementar i estendre la meva formació científica, compromís el qual ha mantingut al llarg d'aquests darrers anys de col.laboració científica. També a l'Angela, a la qual li agraeixo, entre moltes altres coses, el seu entossudiment per fer-me parlar en anglès. A totes dues, el meu agraïment sincer per l'amistat consolidada i l'afecte que em vau demostrar durant l'estada a Newcastle.

Finalment, voldria agrair a la gent del grup de color, i en particular a en Javier, en Ramon, en Xavi, en Marc, l'Olivier, en Robert, en Joost i també a la resta, la seva disponibilitat per participar com a subjectes en els meus experiments, i també per la seva predisposició a discutir i proposar solucions als problemes concrets que m'han anat sorgint al llarg d'aquest projecte. Per acabar, voldria agrair als companys de despatx i en general al Centre de Visió per Computador el clima regnant de bona entesa i disponibilitat.

A tots, el meu agraïment i el reconeixement que la vostra participació ha contribuit decisivament a la consecució exitosa d'aquest projecte.

Acknowledgement

First, I would like to thank Mireia for her support and generous understanding, which have been constant throughout the period of time in which this project has been developed. I would like to thank my parents as well, my family and friends, for their support and specially for their endless patience to keep me connected to the world despite my worries.

Next, I would like to express my gratitude to Maria and Alejandro, my PhD supervisors. Firstly, for giving me the chance to undertake a project as defining in the learning process of a person such as carrying out a PhD. Also their constant availability and patience in order to discuss every and each of the thesis aspects in a rational and tireless way. And to acknowledge that the warmth and mutual understanding in which this project has been developed has resulted in an enriching friendship. But above all, my most sincere gratitude for their constant efforts to train me as a scientist.

I would also like to state my gratitude to Anya Hurlbert, who welcomed me in her lab and allowed me to complement and extend my scientific education, a commitment that has endured throughout the latest years of scientific collaboration. Also, I would like to thank Angela amongst lots of other things, for her stubbornness in making me speak English. To both of them, my most sincere gratitude for the strengthened friendship and affection they showed me during my stay in Newcastle.

Finally, I am grateful to the people in the colour group, specially to Javier, Ramon, Xavi, Marc Olivier, Robert and Joost and the rest, for their availability and for taking part in my experiments as subjects and also for being open to discuss and putting forward solutions to specific problems that arose throughout this project. Last, I want to express my gratitude to my workmates and to the "Centre de Visió per Computador" in general for the prevailing climate of understanding and availability.

To all of you, my most heartfelt gratitude and acknowledgment of your participation, which has contributed decisively in the successful completion of this project.

Abstract

To recognise objects is perhaps the most important task an autonomous system, either biological or artificial needs to perform. In the context of human vision, this is partly achieved by recognizing the colour of surfaces despite changes in the wavelength distribution of the illumination, a property called colour constancy. Correct surface colour recognition may be adequately accomplished by colour category matching without the need to match colours precisely, therefore categorical colour constancy is likely to play an important role for object identification to be successful.

The main aim of this work is to study the relationship between colour constancy and categorical colour perception. Previous studies of colour constancy have shown the influence of factors such the spatio-chromatic properties of the background, individual observer's performance, semantics, etc. However there is very little systematic study of these influences. To this end, we developed a new approach to colour constancy which includes both individual observers' categorical perception, the categorical structure of the background, and their interrelations resulting in a more comprehensive characterization of the phenomenon.

In our study, we first developed a new method to analyse the categorical structure of 3D colour space, which allowed us to characterize individual categorical colour perception as well as quantify inter-individual variations in terms of shape and centroid location of 3D categorical regions. Second, we developed a new colour constancy paradigm, termed *chromatic setting*, which allows measuring the precise location of nine categorically-relevant points in colour space under immersive illumination. Additionally, we derived from these measurements a new colour constancy index which takes into account the magnitude and orientation of the chromatic shift, memory effects and the interrelations among colours and a model of colour naming tuned to each observer/adaptation state.

Our results lead to the following conclusions: (1) There exists large inter-individual variations in the categorical structure of colour space, and thus colour naming ability varies significantly but this is not well predicted by low-level chromatic discrimination ability; (2) Analysis of the average colour naming space suggested the need for an additional three basic colour terms (turquoise, lilac and lime) for optimal colour communication; (3) Chromatic setting improved the precision of more complex linear colour constancy models and suggested that mechanisms other than cone gain might be best suited to explain colour constancy; (4) The categorical structure of colour space is broadly stable under illuminant changes for categorically balanced backgrounds; (5) Categorical inconstancy exists for categorically unbalanced backgrounds thus indicating that categorical information perceived in the initial stages of adaptation may constrain further categorical perception.

Resum

El reconeixement d'objectes és potser la tasca més important que un sistema autònom, ja sigui biològic o artificial, necessita realitzar. En el context de la visió humana, això s'aconsegueix parcialment a través del reconeixement del color de les superfícies malgrat els canvis en la distribució espectral de la llum, propietat anomenada constància de color. El reconeixement correcte del color de les superfícies és pot realitzar adequadament mitjançant la correspondència entre categories de color sense la necessitat d'ajustar exactament els mateixos colors, aleshores la constància de color categòrica juga probablement un paper important per tal d'aconseguir amb èxit el reconeixement d'objectes.

El principal objectiu d'aquest treball és estudiar la relació entre la constància de color i la percepció categòrica del color. Estudis anteriors de constància de color han mostrat la influència de factors tals com les propietats espai-cromàtiques de l'entorn, particularitats individuals dels observadors, semàntica, etc... Malgrat tot, aquestes influències només s'han estudiat breument de forma sistemàtica. Per solucionar-ho, hem desenvolupat una nova aproximació a la constància de color, la qual inclou la percepció categòrica dels individus, l'estructura categòrica de l'entorn, i les seves interrelacions, resultant en una caracterització més comprensiva del fenomen.

En el nostre estudi, primer hem desenvolupat un nou mètode per tal d'analitzar l'estructura categòrica 3D de l'espai de color, la qual ens ha permès caracteritzar la percepció categòrica de cada individu i també quantificar les variacions entre individus en termes de la forma i la localització dels centroides, de les regions 3D categòriques. Seguidament, hem desenvolupat un nou paradigma de constància de color, anomenat "chromatic setting", el qual permet mesurar de forma precisa la localització de nou punts categòricament rellevants en l'espai de color sota una ill.luminació envolvent. Adicionalment, hem derivat a partir d'aquestes mesures un nou índex de constància de color, el qual té en compte la magnitud i la orientació cromàtica de l'il.lumiant, influències de la memòria i les interrelacions entre colors, i també un model d'assignació de noms de color ajustat a l'estat d'adaptació de cada observador.

A partir dels nostres resultats concloem: (1) Existeixen àmplies variacions entre individus respecte l'estructura categòrica de l'espai de color, i pertant l'abilitat d'assignar noms de color varia significativament però aquesta no està ben predita per les abilitats discriminatives de baix nivell; (2) L'anàlisi de l'espai mitjà d'assignació de noms de color suggereix la necessitat d'afegir tres nous "basic colour terms" ("turquoise", "lilac" i "lime") per tal d'optimitzar la comunicació de color; (3) El "chromatic setting" ha millorat la precisió dels models lineals més complexos de constància de color, així suggerint que altres mecanismes que no pas els d'adaptació al guany dels cons poden ser més adequats per tal d'explicar el fenomen de la constància de color; (4) L'estructura categòrica de l'espai de color és en general estable sota canvis d'il.luminant quan s'usen entorns categòricament balancejats; (5) Existeix inconstància de color per entorns categòricament no balancejats i pertant indicant que la informació categòrica percebuda en les etapes inicials de l'adaptació pot condicionar la percepció categòrica posterior.

Related publications

Journals

- *Chromatic settings and the structural colour constancy index.* Roca-Vila J., Parraga C. A., and Vanrell M. Journal of Vision (2nd revision).

- *A compact description of colour naming ability: Quantifying individual variations*. Roca-Vila J., Owen A., Jordan G., Ling Y., Parraga C. A., and Hurlbert A. (In progress).

International conferences

- *Do basic colours influence chromatic adaptation?* Parraga C. A., Roca-Vila J, and Vanrell M. (2011). [Abstract] Journal of Vision, September 23, 2011 (10.1167/11.11.349).

- *Inter-individual variations in colour naming and the structure of 3D colour space.* Roca-Vila J., Owen A., Jordan G., Ling Y., Parraga C. A., and Hurlbert A. (2011). [Abstract] Journal of Vision, September 23, 2011 (10.1167/11.11.386).

- *Categorical focal colours are structurally invariant under illuminant changes.* Roca-Vila J., Vanrell M., Parraga C. A., (2011). [Abstract]Perception 40 ECVP Abstract Supplement, page 196.

- *Predicting categorical colour perception in successive colour constancy.* Roca-Vila J., Parraga C. A., and Vanrell M. (2012). [Abstract] Perception 41 ECVP Abstract Supplement, page 183.

- *What is constant in colour constancy?* Roca-Vila J., Vanrell M., Parraga C. A. (2012). 6th European Conference in Colour Graphics, Imaging and Vision.

Internal reports

- *A new paradigm for focal colour measurements under adaptation.* Jordi Roca, C. Alejandro Parraga, & Maria Vanrell. (2010). (M.Rusiñol, D.Ponsa & A.Fornes, Eds.). CVCRD. Bellaterra, Barcelona, Ediciones Gráficas Rey.

- *Human and Computational Colour Constancy.* Jordi Roca, C. Alejandro Parraga, & Maria Vanrell. (2009). (X.Baro, S.Escalera, & M.Ferrer, Eds.). CVCRD. Bellaterra, Barcelona, Ediciones Gráficas Rey.

Other publications

- *Variations within the Colour Naming ability of Normal Trichromats and Colour Anomalous Individuals.* Angela Owen, Jordi Roca-Vila, Gabriele Jordan, Yazhu Ling, Anya Hurlbert. (In progress).

Contents

List of Figures

xvii

List of Tables

xix

Chapter 1 Introduction

This chapter introduces the phenomenon of colour constancy and stresses its fundamental role in colour perception. Also, it introduces the concept of categorical colour perception and states some interesting questions on how categorical perception may vary under illumination changes. Next, it introduces a series of formal hypothesis which delineate how these questions will be addressed in this work.

1.1 Background

In our everyday life, changes of illumination occur frequently due to several factors, for instance the spectral content of sunlight changes from noon to dawn, or due to weather conditions or as we move from outdoor to indoor environments with artificial light. Despite all these illumination changes the perceived colours of objects remain fundamentally the same. This property is called colour constancy (Land 1964; Jameson and Hurvich 1989; Smithson 2005; Hurlbert 2007; Foster 2011). Colour constancy is a fundamental property of the human visual system (HVS), and in order to achieve colour constancy the HVS needs to disentangle the spectral properties of the illumination and object's surfaces, which are mixed into the information from the light reaching the photosensitive cells in the eye (Hubel 1988). Colour constancy is supported by a complex set of mechanisms located at several neural levels, which are mutually or exclusively activated from cues present in the visual scene (Hurlbert and Wolf 2004; Smithson 2005; Foster 2011).

Categorization is a fundamental human attribute and humans tend to classify colours into several categories. An important method of colour categorization is colour naming, which may assign the

same term to two coloured samples with different saturation or lightness properties. Names of some commonly used colour categories are red, green, blue and yellow, purple, brown, orange, pink, white, black and grey which are also called *basic colour terms* (Berlin and Kay 1969). The main functionality of colour naming seems to facilitate colour memory and colour communication, due to the large number of colours perceived by the HVS (Linhares, Pinto et al. 2008). Correct surface colour recognition may be adequately accomplished by colour category matching without the need to match colours precisely, therefore categorical colour constancy is likely to play an important role for object identification to be successful (Jameson and Hurvich 1989).

Figure 1.1 contains a schematic example of colour constancy under an illumination change. Suppose that you are immersed in a visual scene where there is only one diffuse *achromatic* illumination and a single object with several coloured patches, as represented in panel a. Next, the illuminant's chromaticity is changed to highly saturated *greenish*, as represented in panel b. Just after the illumination change, the ongoing processes of colour adaptation will change the initial colours perceived under *achromatic* illumination, until they become stable after several minutes under the *greenish* illumination.

(a) *Achromatic* illumination   (b) *Greenish* illumination   (c) *Adapted* illumination
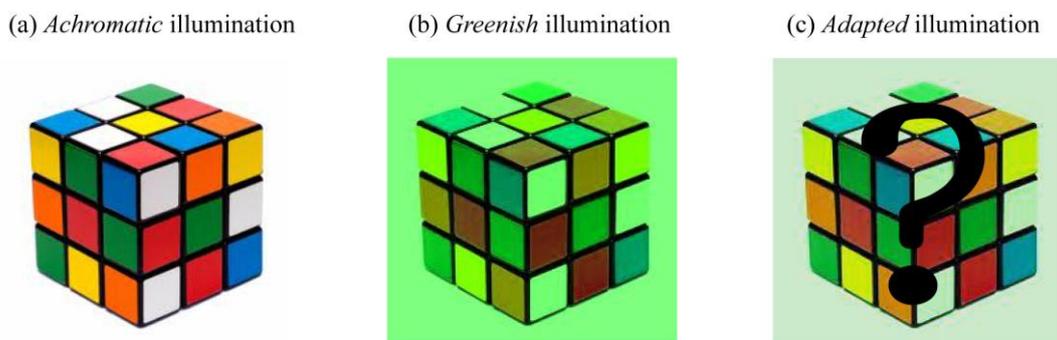
Figure 1.1 Schematic example of colour constancy. Panel a and b corresponds to the same object illuminated under two different illuminations, an *achromatic* and a *greenish* illumination respectively. Panel c shows an hypothetic colour appearance of cube colours after adaptation to the *greenish* illumination.

If colour constancy was complete in the previous example, the colours perceived after the adaptation period (hypothesised in panel c) would be the same as the ones under the *achromatic* illumination. Previous research studied the degree of colour constancy in natural or laboratory conditions (Smithson 2005; Foster 2011), and revealed a considerable range of degrees of colour constancy depending on the cues present in the visual scene (Kraft and Brainard 1999; Shevell and Kingdom 2008). However there is little systematic research on what happens to chromatic categories (*categorical colour constancy*), and most studies focus on using uniform achromatic backgrounds (Kulikowski and

Vaitkevicius 1997; Hansen, Walter et al. 2007; Olkkonen, Hansen et al. 2009; Olkkonen, Witzel et al. 2010), instead of the most realistic complex coloured backgrounds. In particular, some interesting questions arise from Figure 1.1:

- Are the newly perceived colour categories dependent on the illumination?
- Are the inter-relations among the perceived colours constant under illumination changes?
- Are the newly perceived colours equally categorized by different individuals?
- Do the initial categorical properties of the scene colours influence the outcome of the previous questions?

Consider now the previous example but using only achromatic patches with different lightness. After adaptation the lighter patches will be perceived as shifted towards the illumination chromaticity, the mid ones as greyish and the dimmer ones as shifted towards a chromaticity opposed to that of the illumination. This phenomenon is known as the Helson-Judd effect (Helson 1938; Judd 1940). This effect reveals how the simplest case (only achromatic patches) results in a remarkable failure of colour constancy. Also, it reveals how colour perception after adaptation may depend on the chromatic properties of the illuminant, i.e., adaptation under a highly saturated *greenish* illuminant will produce categorically different colours than a highly saturated *reddish* illuminant.

Colour constancy research makes use of psychophysical experiments which study the colour appearance after adaptation to a parameterized visual scene (Smithson 2005; Foster 2011). In these experiments there are traditionally two main constraints: (a) the information contained in the objects of the scene and (b) the selection of illuminants to be tested. Along these lines, both the colour constancy example represented by Figure 1.1 (where colour categories are determined by the patches on the cube) and the Helson-Judd effect (where colour categories are a consequence of the illumination) represent different ways of introducing categorical colour information to the visual scene.

1.2 Motivation

The aim of this thesis is the study of categorical colour perception and its relation with the colour constancy phenomenon. In particular, we hypothesize that categorical colour perception influences colour adaptation, constraining the outcome of colour constancy processes.

In the paragraphs below we list the minor hypotheses we need to address in order to tackle the previous issues.

There is considerable evidence of inter-individual variations in the perception of unique hues (Kuehni 2004) and basic colour categories (Boynton and Olson 1987; Sturges and Whitfield 1995) however, most of the work explored a limited set of colours centred around the focals instead of the whole 3D categorical structure of colour space. So, in order to include categorical or non-categorical information in our stimuli we need to know how these differences extend to the whole categorical regions in colour space:

> Hypothesis 1: There are significant inter-individual variations in the structure of 3D categorical regions in colour space.

The most common paradigms to measure colour constancy are based on measuring single or multiple points in colour space (McCann, Mckee et al. 1976; Brainard 1998), in many cases resorting to haploscopic (Lucassen and Walraven 1996) or alternating viewings under different illuminants (Kulikowski and Vaitkevicius 1997) instead of normal free viewing and immersive illumination. Given the complexity of the colour constancy phenomenon we wonder whether measuring more than one point improves our knowledge of colour constancy when adapted to long periods of immersive illumination:

> Hypothesis 2: Colour Constancy is better described by measuring multiple points in colour space.

Previous research demonstrated the categorical stability of colour space by means of the location of centroids and borders of categorical regions, which were fit from a large set of categorized samples (Hansen, Walter et al. 2007; Olkkonen, Hansen et al. 2009; Olkkonen, Witzel et al. 2010). However, coloured samples were only tested in simple achromatic backgrounds using a colour naming task, which needs large sets of measurements. This leaves scope to explore how the structure of colour space is transformed under immersive changes of illuminations in more complex coloured backgrounds.

> Hypothesis 3: The categorical structure of colour space is stable under illumination changes.

There have been extensive research on the influence of visual scene information on colour constancy where attributes of the scene e.g. number of surfaces (Linnell and Foster 2002), chromatic distribution of colours (Lucassen, Gijsenij et al. 2008), 3D shapes (Hedrich, Bloj et al. 2009), etc were systematically changed. However, there is a lack of understanding of the influence of categorical colour information on colour constancy.

Hypothesis 4: The categorical colour information of the visual scene under the neutral illumination influences colour constancy when the scene is viewed under a coloured illumination.

Although there is evidence that the chromatic properties of the illumination may influence colour constancy (Delahunt and Brainard 2004; Lucassen, Gijsenij et al. 2008), most of this research used colour constancy indices which only measure the overall degree of adaptation (Foster 2011). Also, there is scope for exploring whether categorical colour perception of variegated backgrounds is influenced by the chromatic properties of the illumination, as is the case in the Helson-Judd effect.

Hypothesis 5: The chromatic properties of the illumination influence categorical colour constancy.

1.3 Contributions

In this section we detail the contributions of this thesis as they are relevant to the set of hypothesis mentioned before.

We tested Hypothesis 1 by studying individual differences in the 3D categorical colour structure of colour space. In doing so, we used a set of colour naming tests where each observer categorized a large set of samples under an *achromatic* illumination (*Experiment I*). Next, we used *convex sets* (Gärdenfors 2000; Jäger 2010) to represent each region corresponding to the basic colour terms, and developed a new method to extract features among the interrelations of these regions. These features allowed to quantify inter-individual differences for a set of 23 normal trichromats. The new method provided the location and description of categorical colour regions in 3D colour space and also an observer-averaged colour naming space, which informed on colour communication features. Our results show (Roca-Vila, Owen et al. ; Roca-Vila, Owen et al. 2011) that:

- There exist remarkable inter-individual variation in the categorical structure of colour space.
- Individual colour naming ability is not well predicted by low level chromatic discrimination abilities.
- An analysis of the observer averaged colour naming space suggests the need for an additional three basic colour terms (turquoise, lilac and lime) for optimal colour communication when using only the basic colour terms.

Hypothesis 2 was tested by means of a newly developed colour constancy paradigm (called *chromatic setting*), which allows to measure the precise location of nine categorically relevant points in colour

space under immersive conditions of illumination. We designed a psychophysical experiment to test the ability of observers to perform in the new paradigm (*Experiment II*). Over the experiment also were tested several different stimuli, which were combinations between three illuminants and three background types. Backgrounds were a collage of patches with a limited number of colours. Additionally, we derived from these measurements a new structural colour constancy index (SCI) which takes into account the magnitude and orientation of the chromatic shift, memory effects and the interrelations among colours. Our results show (Roca-Vila, Parraga et al. ; Parraga, Roca-Vila et al. 2011; Roca-Vila, Vanrell et al. 2011) that:

- The new colour constancy paradigm is feasible and its measurements are consistent with previous colour constancy paradigms.
- Using multiple points improves the precision of more complex linear colour constancy models, suggesting that mechanisms other than cone gain might be necessary to explain colour constancy.

Hypothesis 3 was tested by means of a psychophysical experiment (*Experiment III*) where observers first adapted to illuminated backgrounds before performing a chromatic setting task. Since the paradigm provided a more precise location of nine categorically relevant points, we developed a method to quantify the degree of deformation in colour space and the interrelations of these chromatic settings when illumination was changed. Our results show (Roca-Vila, Vanrell et al. 2011; Roca-Vila, Parraga et al. 2012) that:

- The overall interrelations among chromatic settings are mostly stable under illumination changes, suggesting the stability of the categorical colour space.

Hypothesis 4 was tested by means of a psychophysical experiment (*Experiment II*) where observers performed chromatic settings once adapted to a particular stimulus. Since individual variations in the structure of categorical colour space exist, the colours present in the stimuli backgrounds were selected according to each subject's categorical properties. Two background types were designed to enclose the dichotomy between maximized and minimized categorical information under a reference (*achromatic*) illumination, and measurements were performed under several test illuminations. Our results show (Roca-Vila, Parraga et al. ; Parraga, Roca-Vila et al. 2011; Roca-Vila, Vanrell et al. 2011) that:

- According to our analysis from colour constancy indices and structural interrelations among chromatic settings, there are no colour constancy effects related to the backgrounds we tested.

6

Hypothesis 5 was tested by means of a psychophysical experiment (*Experiment IV*) where observers performed chromatic settings and a colour naming task. We tested six different illuminations with different chromatic properties, in terms of the magnitude and direction of the illuminants shift. Backgrounds were similar to those of Experiment II except that they had only three different colours. Furthermore, a model of categorical colour perception was developed to complete categorical information from untested regions of colour space. Our results show (Roca-Vila, Parraga et al. ; Roca-Vila, Parraga et al. 2012) that:

- The interrelations among chromatic settings and the structural colour constancy index are more disrupted under the *yellowish* than the *greenish* illuminants.
- There are asymmetries in the stimuli colours after adaptation; the perception of green was stable under all illuminations.
- The colours after adaptation were a mixture of the colours under the test and reference illuminations, and little but consistent adaptation differences (categorical inconstancies) existed according to the tested illuminants.

1.4 Outline

This thesis is conformed by seven chapters. In Chapter 1 we introduced the motivation of the thesis, formulated the main hypothesis, and delineated its main contributions.

Chapter 2 contains a bibliographical review of the two main topics of this research. First, we introduce fundamental concepts in the fields of visual perception and colour vision. Next, the topics of colour constancy and categorical colour constancy are introduced, and special attention is given to previous approaches which tackled both issues at the same time.

Chapter 3 studies the structure of the individual categorical colour space through an extensive colour naming experiment, *Experiment I*.

Chapter 4 introduces the new colour constancy paradigm (termed *chromatic setting*) and it also contains a new psychophysical experiment called *Experiment II*, which tests the feasibility of the paradigm.

Chapter 5 applies the chromatic setting paradigm to measure the stability of the categorical colour structure of colour space when illumination is changed. It introduces a new psychophysical experiment, called *Experiment III*, which involves testing subjects on several illuminated backgrounds.

Also, a new method to measure interrelations among chromatic settings under different illuminations is introduced.

Chapter 6 introduces the *Experiment IV* which used the chromatic setting and a constrained colour naming task. In this experiment there were six different illuminations and two different backgrounds which contained only three colours each. Results were interpreted in terms of categorical changes, and categorical information for each adaptation was expanded by means of a new model of categorical colour prediction, which uses the categorical information contained in chromatic settings as well as their precise location.

Finally, Chapter 7 contains a summary of the results from the previous chapters. A general discussion states the main contributions of this thesis, proposing future lines of research and linking its conclusions to others in the field of Computer Vision.

Chapter 2 Review of related research

This chapter begins with an introduction to the basic concepts in the fields of vision and colour vision, within the general framework where this work develops. It contains a bibliographical review of the two main topics of this work; categorical colour perception and colour constancy. First, we introduced categorical colour perception; its main definitions and features. Second, we introduced the colour constancy topic, common definitions and important issues such as how it is measured, quantified and modelled. After that, we introduced the problem of colour constancy in computer vision, linking it to human colour constancy, since another goal of this thesis is to contribute to solve this problem in the computer vision field, following the fruitful tradition of knowledge interchange between the biological and artificial vision fields.

2.1 Basic concepts

Vision

The first and primary component for human vision to take place is the existence of light since without light there is no vision. Light is first processed by several photosensitive cells in the back layer of the human eye. These cells are sensitive to the spectral properties of light, and only process it when its spectral wavelength falls approximately between 400 to 750 nm. Light with such properties is called *visible light* (Hubel 1988; Wandell 1995; Snowden, Thompson et al. 2006).

The capacity to interpret the surrounding environment from the information contained in visible light is called *vision* or *visual perception*. Such definition frames the process of vision as an information-processing task (Marr 1982). The physiological components involved in supporting human vision are

complex and extensive, ranging from the behaviour of single cells to that of extended regions of the brain, and they are collectively referred as the Human Visual System (HVS). Due to its high complexity, the process of Vision is necessarily studied from several scientific fields such as philosophy, psychology, cognitive science, neuroscience, molecular biology, neurophysiology, physics, computer science, and psychophysics.

Seeing starts when visible light is focused by the lens of the eye into the retinal surface, a photosensitive membrane at the back of the eye that converts patterns of light into neural signals. The only purpose of all eye components is to support the function of the retina, which is actually a part of the brain. The photoreceptive cells of the retina, the *cones* and *rods*, detect the photons of light and produce neural impulses which are processed in the retina itself by several layers of interconnected neurons (Hubel 1988; Wandell 1995; Snowden, Thompson et al. 2006). Rods are active in dim light and provide only achromatic vision while cones are mainly active in daylight and support the perception of colour. Several important features of visual perception can be traced back to this stage. Retinal signals leave the retina through the optic nerve to the Lateral Geniculate Nucleus (LGN) located in the thalamus prior to reaching the Visual Cortex (VC) located in the occipital brain lobe, in the back of the brain (see Figure 2.1). At the same time the LGN receives feedback connections from the VC. Also, signals from the retina travel directly to other parts of the brain and they are processed in a hierarchical fashion (Hubel 1988; Wandell 1995; Snowden, Thompson et al. 2006). The function of the LGN is mostly unknown but it is likely to help the HVS to focus its attention on the most important information (Hubel 1988; Wandell 1995; Snowden, Thompson et al. 2006). The VC is the primary location for processing visual signals in the brain and it contains a large number of functional units that are interconnected through both feed-forward and feedback connections. These cortical areas include: V1 which receives inputs from the LGN; V2 which received inputs from different V1 areas; V3 which integrates signals of different pathways and V4 which is thought to play an important role in processing colour information (Hubel 1988; Wandell 1995; Snowden, Thompson et al. 2006).

Colour Vision

The existence of three physiologically different cones in the retina makes the perception of colour possible. These cones are distinguished by the type of photo pigment molecules they contain and they are called L, M and S since they are selective for *long*, *middle* and *short* wavelengths respectively. This is called the *trichromacy* principle (Wandell 1995) and in practical terms implies that the perception of any visible light can be approximately described by three numbers. Once a photon is absorbed by a cone its reaction will be the same regarding the photon's wavelength, this fact is known as the principle of *univariance* (Rushton 1972). This means in practical terms that an increase in photon count can be due to an increase in light intensity, a change to a more favourable wavelength, or

both. The cones' spectral sensitivities are not limited to reduced intervals of wavelength but are broad and overlapping, i.e., what varies with wavelength is the probability that a photon is absorbed. These two principles allow the existence of the phenomenon of *metamerism*, i.e., any two different lights that produce the same triplet of cone responses under the same illuminant will be perceived as identical (Hurlbert 1997; Gegenfurtner and Kiper 2003; Shevell 2003).
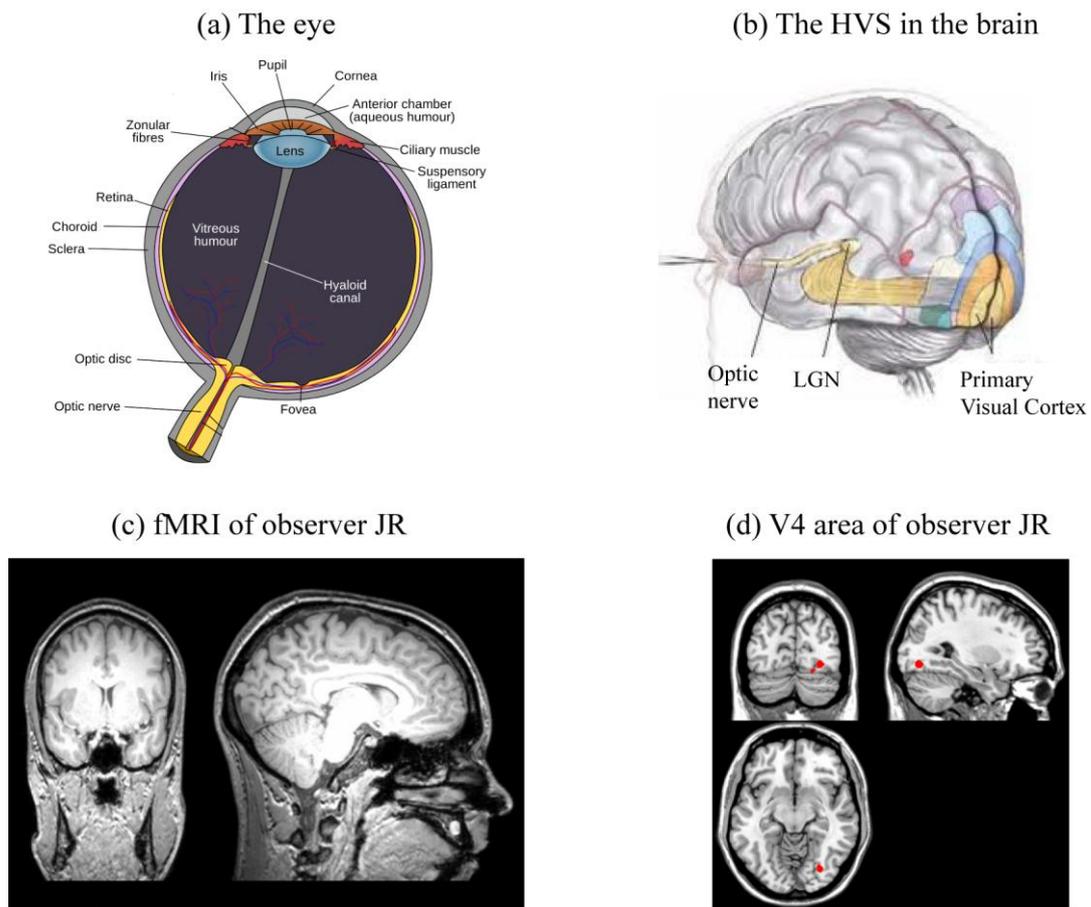
Figure 2.1 Stages of the Human Visual System (HVS). Panel a: schematic description of the main components of the human eye. Panel b: schematic location of the LGN and Visual Cortex in the human brain. Panel c: Sectional brain image from a fMRI scan. The subject was the author, while participating in a colour perception experiment at ION (Institute of Neuroscience, UK). Images courtesy of Prof Anya Hurlbert. Panel d: Normalized image of a brain scan of subject JR with area V4 highlighted in red. It has been hypothesized that area V4 is specialized in the perception of colour (Gegenfurtner and Kiper 2003).

After this first stage the output signals of each cone type are compared and transformed into action potentials by a complex network of cells in the retina itself. Next the information is sent to the visual cortex via the LGN in at least three separate colour-opponent channels, whose existence has been characterized psychophysically, physiologically and computationally (Gegenfurtner and Kiper 2003).

These three colour opponent channels are: a luminance opponent axis where the L and M cone signals are added; a red-green opponent axis where the opposing signals of L and M cones are computed; and a blue-yellow opponent axis where the S cone signal is subtracted from the sum of L and M cone signals (Hurlbert 1997; Gegenfurtner and Kiper 2003). From this, it follows the existence of wavelength distributions for which the opponent signals of one these channels are approximately balanced. These "hues" which do not contain perceptual components of other hues, e.g., a unique red hue is neither yellowish nor bluish, are called *unique hues* (Kuehni 2001). Notice that despite the colour term nomenclature used to define the colour opponent axes, their directions do not align with the axes predicted by traditional colour appearance theories of unique hues (Stockman 2010).
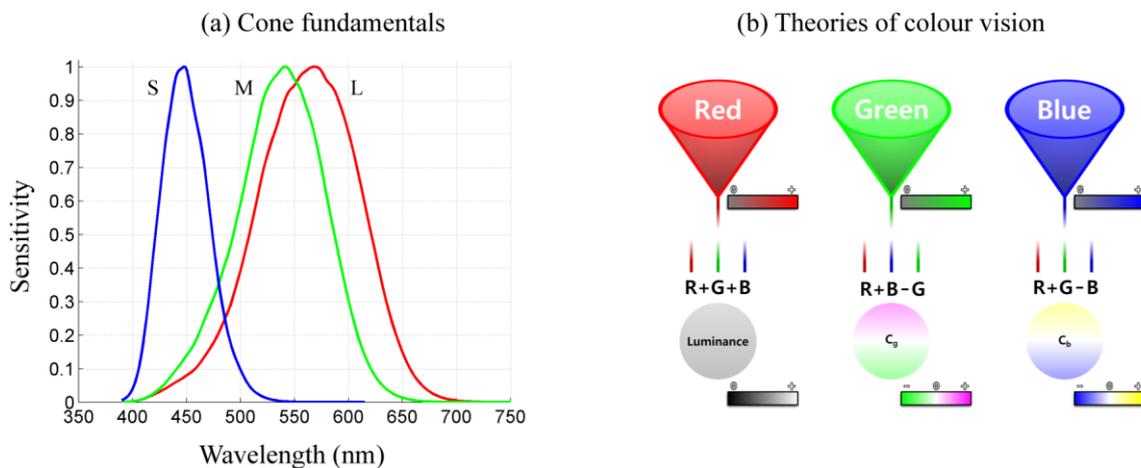
Figure 2.2 Fundamentals of colour vision. Panel a: Spectral sensitivity functions of the three different types of cone photoreceptors. Panel b: Schematics of the opponent theory of colour vision.

Despite this simple explanation of the physiological basis of colour vision (the existence of trichromatic sensitivity and of the three colour-opponent channels) the complete explanation of colour perception is far from solved and higher order chromatic mechanisms are required to account for it (Hurlbert 1997; Gegenfurtner and Kiper 2003; Shevell and Kingdom 2008). For instance, recent studies have shown that colour is not analysed separately from object motion and that neurons in certain visual areas simultaneously encode colour and form (Gegenfurtner and Kiper 2003).

The perception of colour, from the light reflected from the surface of objects in the visual scene, has several functionalities: coloured objects are more easily detected, more easily identified, more easily grouped, and more easily remembered than achromatic objects (Shevell and Kingdom 2008). Therefore, colour is deeply integral to vision, resulting from the combined activity of neurons in many different visual areas (Hurlbert 1996; Gegenfurtner 2003; Gegenfurtner and Kiper 2003).

Classical approaches to study colour vision are based in general on psychophysical experiments with isolated stimuli, but colour appearance of objects/surfaces depends strongly on the light reflected from other objects/surfaces in the visual scene (Shevell and Kingdom 2008). In natural viewing, variegation in the retinal image is abundant over space and time and such complex visual stimuli excite neural mechanisms that are not revealed by isolated stimuli. The simplest context used in laboratory conditions is a uniform background field, which has been used extensively over the last century. However, there has been demonstrated that this simple stimuli cannot account for appearance shifts caused specifically by chromatic variegation within the scene (Shevell and Kingdom 2008).

2.2 Categorical colour perception

The main functionality of colour perception seems to be colour communication, but also to facilitate colour memory due to the large number of colours perceived by the HVS. Due to the abundance of colour signals available humans resort to classify colours into several categories, often assigning the same colour term to coloured samples with different LMS responses. An important method for studying this categorization is *colour naming*, which aims at dividing the 3D colour space into several regions each linked to a different colour name.

There are two main theories proposed to explain our ability to categorize colour perception: (1) the Sapir-Whorf hypothesis (Roberson, Davies et al. 2000), which states that colours are perceived categorically only because they happen to be named categorically and these names vary across cultures and languages (Roberson, Davies et al. 2000); and (2) Berlin and Kay universalist theory (Berlin and Kay 1969) which sustains that most cultures and languages subdivide and name the colour space in ways that are innate to humans. The eleven colour categories proposed by Berlin and Kay, shared by most evolved languages, are: red, green, blue, yellow, purple, brown, orange, pink, white, black and grey (Berlin and Kay 1969). As a result of this, colour categorization is framed in terms of the old nature-nurture debate (Hardin 2005), i.e., whether it stems from childhood learned patterns or it is physiologically-based. Although there is solid evidence of the physiologically-based explanation (Kay and Regier 2003; Lindsey and Brown 2006), one of the main functionalities of colour categorization is colour communication and thus, language provides an essential feedback role in ensuring the universality of colour categories, especially when the only constraints in their formation are fundamental sensory discrimination abilities (Belpaeme and Bleys 2005; Baronchelli, Gong et al. 2010). Consequently, a combination of these two factors seems to be the most plausible answer (Loreto, Mukherjee et al. 2012).

2.2.1 Colour naming and a categorization of colour space

Colour naming is a common element of everyday social communication as well as vital to diverse behavioural tasks, including visual search (Dzmura 1991; Yokoi and Uchikawa 2005; Amano, Foster et al. 2012) and object identification (Zaidi and Bostic 2008), which in turn are critical for a number of jobs, for example, fire-fighting (Margrain, Birch et al. 1996), police (Birch and Chisholm 2008), transport, and medical diagnostics (Spalding 1999).

Previous studies, based on colour naming usage patterns, have focussed on the universality of some colour terms which allow a partitioning of colour experience into eleven basic colour categories (Berlin and Kay 1969; Heider 1972; Hardin 2005; Lindsey and Brown 2006). The structure of colour space has then typically been described using this reduced set of basic categories, supplemented by some much smaller "hard to name" regions (Boynton and Olson 1987; Sturges and Whitfield 1995; Chichilnisky and Wandell 1999; Guest and Van Laar 2000; Paggetti, Bartoli et al. 2011).

Two concepts typically have been used to characterize colour naming behaviour (Boynton and Olson 1987; Sturges and Whitfield 1995; Paggetti, Bartoli et al. 2011): *consistency,* which is defined as the giving of the same name to a particular sample on repeated presentations by one observer; and *consensus,* which is defined as the commonality of a particular colour categorisation across all observers. The categorical colour space structure is summarized by the *focal* locations, which are defined as the fastest to name samples in the consensus set for each category, and the *centroid* locations, defined as the average locations of classified samples for each category (Boynton and Olson 1987). For example, focal locations of the 11 basic colour categories have been defined in several colour spaces: the OSA (Sturges and Whitfield 1995; Paggetti, Bartoli et al. 2011), Munsell (Berlin and Kay 1969; Boynton and Olson 1987) and 1976 UCS (Guest and Van Laar 2000). Despite the demonstrated universality of basic colour terms, successful colour communication is weakened by large variations in colour perception across individuals. For example, Kuehni (Kuehni 2004) concludes from a study of unique hues: *"...Comparison of spectral light data indicates that one observer's unique blue can be another's unique green and vice versa, and the same for yellow and green,"* a finding supported by Webster et al (Webster, Miyahara et al. 2000; Xiao, Wuerger et al. 2011). Possible sources of inter-individual variability range from individual physiological and cognitive factors such as differences in the macular pigment density or cone spectral sensitivity peaks (Webster and Macleod 1988) to cultural factors such as language and professional activity (Webster, Webster et al. 2002). Yet evidence suggests that higher-level perceptual or cognitive factors may compensate for sensory variations; for example, although the physiological equipment of the human visual system changes throughout lifespan, the perception of unique hues remains essentially constant

(Raskin, Maital et al. 1983; Kuehni 2004). Furthermore, observers with reduced visual acuity and altered colour vision show consistency in naming colours and its characteristic shifts and confusions (Nolan, Riley et al. 2008; Uchikawa 2008). Recent results also demonstrate that, on the one hand, the sharing of a common language is not sufficient to prevent inter-individual variations in usage patterns of basic terms (Lindsey and Brown 2009), but, on the other, language provides an essential feedback role in ensuring the universality of colour categories when the only constraints on their formation are fundamental sensory discrimination abilities (although, note that the latter finding assumes that the measured discrimination ability itself has not been influenced by language) (Belpaeme and Bleys 2005; Baronchelli, Gong et al. 2010).

2.2.2 Basic Colour Terms

In their seminal work, Berlin and Kay (Berlin and Kay 1969) observed that colour terms translated easily between different languages and systematically studied their usage in several languages. Their approach was based on a simple unconstrained colour naming task where observers assigned a colour word to coloured surfaces (chips) extracted from a subset of the *Munsell colour space* (Fairchild 2005). Observers were also asked to specify the best example of each colour category, and the set of chips of the array to which they would assign the same colour name. Basic colour terms were defined as colour names with the following properties:

- Monolexemic, a single lexical term.

- They refer to the colour of objects.

- Are applicable to a wide range of objects.

- Are of frequent use.

The experiment used 20 subjects and was completed with information from published works including other 78 languages. Their main conclusions were: (i) there are only eleven basic colour terms and in the case of English they are: white, black, grey, red, yellow, green, blue, purple, orange, brown and pink; (ii) colour terms are acquired by languages in an evolutionary order and this means that although languages can have a different number of colour terms the way the incorporate new colour terms follow a fixed order (Berlin and Kay 1969).

Furthermore, in their research they also studied which basic colour terms describes best any particular colour. They obtained information about the boundary of colour terms by showing observers various colours and asking them to point out which colours lie in on the boundary of a specified colour term, or ask them to identify the colours that best describe a particular colour term. Participants gave best

and faster answers to the second task implying that colour categories are structured around focal colours, rather than around colours at the boundaries of colour terms. These results support that focal colours play a crucial role in our internal representation of colour categories.

2.3 Colour constancy

The colour constancy phenomenon has been acknowledged explicitly at least for the last two centuries but it has not been until the last decades that has been studied systematically under controlled stimuli conditions (Foster 2011). Since the literature related to colour constancy is extensive there are several comprehensive reviews on the topic which summarize the current and past approaches dealing with the phenomenon. Previous reviews start with colour constancy definitions, follow with measuring methods and quantifications about the extension of the phenomenon and detail methods on how to predict colour constancy behaviour and indicate possible explanations to the phenomenon (Jameson and Hurvich 1989; Smithson 2005; Hurlbert 2007; Foster 2011).

2.3.1 Definitions and main features

Colour constancy is not a property of objects; it is a perceptual phenomenon resulting from mechanisms located at different levels in the eye and the brain (Hurlbert 1996; Hurlbert and Wolf 2004). The simultaneous activation or inhibition of such mechanisms depends on the visual cues present in the visual environment (Kraft and Brainard 1999). Since humans are not perfectly colour constant, a natural goal for experimentation is to explore up to what extent humans are colour constant and from this, to find principles and to develop models that allow us to predict colour appearance in complex scenes (Brainard, Brunt et al. 1997).

Although colour constancy is a perceptual phenomenon, its roots are deep in the physics of the retinal image formation and in two features of the visual environment: (1) the illuminant's spectral power distribution, $E(\lambda)$, which specifies the amount of power in the illuminant at each wavelength; and (2) the surface spectral reflectance function, $S(\lambda)$, which specifies the fraction of incident power reflected at each wavelength. The colour signal, $C(\lambda)$, is defined as the spectral power distribution of the light reaching the observers eye, $C(\lambda)=E(\lambda)S(\lambda)$. Since this expression confounds the illuminant's spectral power distribution and the surface's reflectance function, a full recovery of the spectral properties of either the illumination or objects surface reflectance by the trichromatic eye does not have a unique solution (Foster 2011). Several possible strategies have been proposed to make colour constancy possible. These include restrictions on the number and dimensionality of the spectral reflectance's and

illuminants available (Maloney and Wandell 1986), normalizations with respect to the illumination (Brainard and Wandell 1986), assumptions about the brightest visible object (Land and McCann 1971) or the average colour of the world (Buchsbaum 1980), higher order statistical properties of the environment and other regularities (Golz and MacLeod 2002; Hordley 2006), or a combination of these. However, none of the explanations proposed so far provides a complete representation of how a visual scene is perceived under an illumination shift in naturalistic, complex, unconstrained conditions. For instance, the degree of colour constancy may depend on internal criteria derived from different judgments of the scene, as demonstrated by the hue-saturation vs. paper-matches of Arend and Reeves (Arend and Reeves 1986). Other confounds may depend on the ability of subjects to attribute changes in the scene to either changes in the spectral composition of the illuminant or the reflecting properties of objects in that scene (Foster and Nascimento 1994). High level visual memory may also play an important role in judgments of surface colour, as demonstrated by Hansen *et al* (Hansen, Olkkonen et al. 2006).

2.3.2 Measuring the phenomenon

The degree and quality of colour constancy experienced by observers is usually measured by a variety of psychophysical techniques. A typical experiment compares the colours an observer perceives under a given state of illuminant adaptation to the colours perceived under another state and the differences are then interpreted using models and indices (Brainard, Brunt et al. 1997; Foster 2011). Figure 2.3 shows a Mondrian image under several different illuminations. Mondrians are defined as a collage of 2D flat surfaces with several colours but without specularities and highlights, i.e., following a Lambertian model of illumination (Hurlbert 1998). This kind of stimuli allows easy manipulation of its geometrical and statistical chromatic features and has been used extensively in laboratory conditions (Land and McCann 1971; Arend, Reeves et al. 1991). The most popular (Smithson 2005; Foster 2011) colour constancy paradigms are described as follows:

Asymmetric matching. Asymmetric colour matching (Wyszecki and Stiles 1982; Arend and Reeves 1986) compares binocular or dichoptical stimuli under different illuminants, presented either simultaneously or successively. Subjects adjust a patch under one illumination to match another under a different illumination. This method requires that the state of adaptation follows closely the change of illumination, a strong assumption especially in the case of alternate viewing paradigms (Foster 2011). Simultaneous matching has the drawback that adaptation to the two scenes is determined by the pattern of eye movements across the two halves of the scene. Successive matching allows experimental control of adaptation to the two illuminants, but performance will ultimately depend on

the observer's ability to remember colours. Dichoptical matching allows separate adaptation states in the two eyes, but removes binocular cues to the scene geometry (Smithson 2005).

Achromatic setting. Under different contextual conditions of scene configuration and illumination the subject is asked to adjust a particular part of the scene to appear achromatic. Achromatic setting measures the perceptual stability of the achromatic locus under a change of adaptation by asking subjects to modify a stimulus until it appears "achromatic". It has been pointed out that this is a local measurement that may or may not be influenced by manipulations of other regions of the scene and also, that one measure may not be enough to estimate the stability of perceived colours away from the neutral point (Foster 2003; Delahunt and Brainard 2004; Schultz, Doerschner et al. 2006; Foster 2011).

Colour naming. The observer is asked to assign a colour term to a particular part of the screen. It can be done using a restricted set of colours, usually the 11 universal terms of Berlin and Kay or can be done without restrictions, a free-naming task. Colour naming paradigms rely on the subjects' internal colour categories by asking them to classify samples under different illuminants. It has been argued that colour naming provides a more direct method for measuring colour constancy (Foster 2011) on the grounds that it is less sensitive to the instructions given to subjects (Arend and Reeves 1986; Troost and de Weert 1991). The main setback of the this method is the large number of discernible colours (>2 million), much larger than the number of possible names (Pointer and Attridge 1998; Linhares, Pinto et al. 2008), resulting in limited accuracy (Foster 2011). Variants include determining unique hues and estimating the degree of colour constancy from the response categories of large numbers of samples and the position of colour boundaries (Chichilnisky and Wandell 1999; Smithson and Zaidi 2004) under different states of adaptation (Kulikowski and Vaitkevicius 1997; Hansen, Walter et al. 2007; Olkkonen, Hansen et al. 2009; Olkkonen, Witzel et al. 2010).
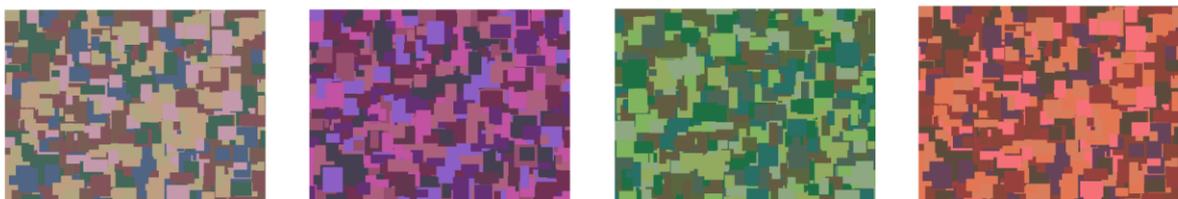
Figure 2.3 Examples of Mondrian stimuli images under different illuminants used in colour constancy experiments. The image on the left contains eight different colours under *D65* illumination and the rest correspond to three different simulated illuminations applied to the reflectance version of leftmost image.

2.3.3 Colour constancy mechanisms and visual cues

Colour constancy is composed of multiple mechanisms that are activated simultaneously or exclusively according to the cues present in the visual scene (Hurlbert and Wolf 2004; Smithson 2005; Foster 2011). According to Hurlbert *et al* (Hurlbert and Wolf 2004), these can be classified regarding to the type of computational and neural level they activate in *sensory*, *perceptual* and *cognitive*.

Sensory mechanisms are usually represented by linear transformations of the photoreceptor responses such as the traditional Von Kries adaptation (Von Kries 1905), which normalizes each photoreceptor signal by its average cone type over the image. However, recent physiological discoveries do not support Von Kries-type adaptations. Take for instance the presence of inter-connections among different cone types through the horizontal cells, as early as the outer retina (Vanleeuwen, Joselevitch et al. 2007). Perceptual mechanisms require a first step on scene processing in order to segment reflection and surface components. From a computational standpoint, mutual reflections and specular highlights are important visual cues that a perceptual mechanism may use to achieve successful colour constancy (Kraft and Brainard 1999; Maloney 2002; Hurlbert and Wolf 2004). Cognitive mechanisms require the complete segmentation of the scene into objects and their recognition, examples are familiar objects, memory colours, etc (Hansen, Olkkonen et al. 2006).

2.3.4 Quantifying colour constancy

The extent of colour constancy achieved by observers is traditionally measured by a normalised one-dimensional index (or *colour constancy index*) which compares measurements conducted under two illumination conditions: *reference* and *test* (Smithson 2005; Ling and Hurlbert 2008; Foster 2011). Values of colour constancy indices ideally range between 0 and 1; values lower than 1 indicate a colour constancy *failure* and values larger than 1 indicate an *overcompensation*. It is important to notice that colour constancy indices represent an over-simplification of the phenomenon and must be interpreted according to their particular experimental context.

Measurements obtained under the reference and test illuminations and the computations applied to these are usually interpreted in some colour space. Each colour space has its own characteristics and to make a good selection we must consider the particular features that need to be extracted from the experimental data. Two-dimensional colour spaces such as CIE1931 *xy* (Wyszecki and Stiles 1982) and CIE1976 *uv* (Wyszecki and Stiles 1982) are widely used, being one of their advantages a lack of reference white point specification, i.e. they do not need to incorporate any further chromatic transform to specify the adaptation of the subject. CIE1976 uv also has the advantage of being *perceptually uniform* meaning that Euclidean distances between pairs of colours are monotonically

related to their perceived dissimilarities (Brainard, Brunt et al. 1997; Fairchild 2005; Ling and Hurlbert 2008). Perceptually uniform colour spaces that include a *lightness dimension* (such as CIELab, CIELuv, and CIECAM02) are also popular in the colour constancy literature (Fairchild 2005; Foster 2011).
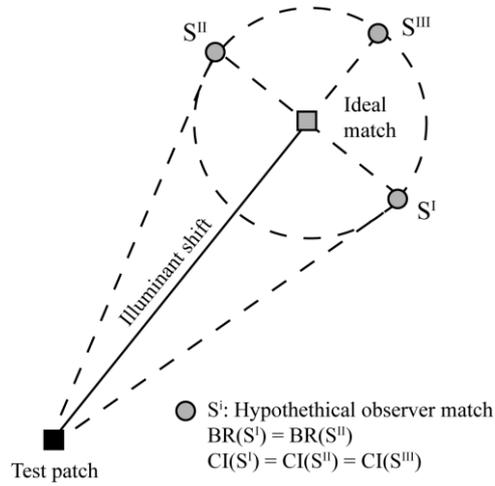
Figure 2.4 Failures of traditional colour constancy indices. The Figure illustrates how both CI and BR fail to differentiate between quite different observer matches. All points along the circle will produce the same CI index value and points S^I and S^{II} will produce the same BR index value.

Equations 2.1 and 2.2 detail the computation of the most common and simple colour constancy indices, the *Colour Constancy (CI)* index (Arend, Reeves et al. 1991) and the *Brunswich Ratio (BR)* index (Troost and de Weert 1991). These formulas are based on Euclidean distances among the colour coordinates of the *test patch* the *ideal match* and the *observer match*. The test patch is the surface patch selected for studying the phenomenon and illuminated by a reference (usually neutral) illuminant, the ideal match corresponds to the coordinates of the surface patch when illuminated by another illuminant and the observer match corresponds to the coordinates of the illuminated surface patch as perceived by the subject.

$$CI = 1 - \frac{\|Ideal\ match - Observer\ patch\|_2}{\|Test\ patch - Ideal\ match\|_2} \qquad (2.1)$$

$$BR = \frac{\|Test\ patch - Observer\ patch\|_2}{\|Test\ patch - Ideal\ match\|_2} \qquad (2.2)$$

20

The implicit assumption behind CI and BR indices is that observer adaptation lies somewhere in the joining line between the test and ideal match coordinates, which is not always the case as shown in the example of Figure 2.2 where both CI and BR have the same value for different observer matches. In the case of the achromatic settings paradigm (where the surface patch is neutrally coloured -see section 2.3.2) the chromatic coordinates of the test and ideal matches correspond to the chromatic coordinates of the respective test and reference illuminants. In subsection 4.5.3 we discuss these issues further.

2.3.5 Colour constancy in Computer Vision

Computer vision is the discipline concerned with the acquisition, processing, analyzing and ultimately understanding of images by computational means. A major line of research within this field has tried to replicate the abilities of human vision, which propitiated a flow of inspirational ideas from theories of visual perception. Conversely, methods and algorithms developed in the computer vision field have been a source of debate in visual perception. This exchange of knowledge between both fields, became apparent with Marr's influential book "Vision" (Marr 1982; Brainard 2003). This thesis is framed within this dialog.

Colour constancy is a particularly relevant example of this biological-artificial vision interaction (Brainard 2003). The understanding of images acquired with cameras by computer vision systems face similar challenges as the those captured by the retina and processed by the human visual system, i.e., in a variable environment where spectral information changes, both systems need to provide a stable interpretation of the world. Computer vision algorithms have a long tradition of dealing with the colour constancy phenomenon; however their aims are not necessarily the same as those of biological organisms. For instance, the output of a commercial camera needs to be highly consistent with human colour perception while the output of an industrial camera in a quality control process may only require stability under a very limited set of conditions.

According to Hurlbert (Hurlbert 1998) the computational colour constancy problem is: "*to recover the invariant spectral reflectance properties of object surfaces from the image irradiance, in which reflectance is entangled with surface illumination*". However, Jameson and Hurvich affirm that this approach could be an oversimplification of the phenomenon since the HVS may not need to estimate the spectral properties of the illuminant nor the objects' surface reflectance (Jameson and Hurvich 1989). A large body of research in computer vision is grounded on the theoretical framework of computational colour constancy and its algorithms strive to find an estimation of the chromatic coordinates of the illuminant in some colour space, prior to normalization over the image. However,

when the aim is to mimic the operation of the HVS and in particular colour adaptation processes, the performance of computer vision algorithms should be evaluated against perceptual and not physical "ground truths" (Vazquez-Corral, Parraga et al. 2009).

Finlayson *et al* present a correlation framework within which to consider illuminant estimation algorithms (Finlayson, Hordley et al. 2001). As in other studies (Forsyth 1990; Brainard and Freeman 1997), they state that the problem of illuminant estimation does not have a unique solution and that a practical solution is to select a set of candidate illuminants and to look for the best one. They recover a measure of the likelihood that each member of a set of possible illuminants was the scene illuminant, i.e., likelihood is assigned according to the correlation between the colours in an image and prior knowledge of the probability of different colours occurring under different lights. Finally, a threshold procedure returns the most likely illuminant or set of illuminants. A notable feature of Finlayson approach is that it incorporates several previous colour constancy algorithms in his theoretical framework. Following this approach a new study (Vazquez-Corral, Vanrell et al. 2011) has focused on the selection of illuminants according to the effects they produce on the final categorical properties of colours, i.e., it includes perceptual constraints that are computed on the corrected images. In other words, it weights a large set of candidate illuminants according to their ability to map the corrected image onto specific colours regions in RGB colour space. These colours regions are chosen as the basic colour categories which have been psychophysically measured.

Our research might be interpreted as a complement to the previous computational approach, since it aims to study whether the HVS may also use this categorical information to achieve colour constancy.

2.4 Colour constancy and categorical colour perception

In addition to using achromatic patches as a reference, some researchers have included multiple colour references to study colour constancy (Arend 1993; Kulikowski and Vaitkevicius 1997; Hansen, Walter et al. 2007; Olkkonen, Hansen et al. 2009; Olkkonen, Witzel et al. 2010) and to determine properties such as the boundaries between colour categories (Zaidi and Smithson 2004). Some of these studies have measured the colour appearance of several coloured patches under different illuminants (Arend 1993; Kulikowski and Vaitkevicius 1997; Speigle and Brainard 1997) while others have used colour naming to derive a conclusion about the categorical structure of colour space (Troost and de Weert 1991; Hansen, Walter et al. 2007; Olkkonen, Hansen et al. 2009; Olkkonen, Witzel et al. 2010). In the direct measures, colour constancy seems to hold best for hues corresponding to 'typical' colours as compared with the adjacent hues, however this effect may be residual (Kulikowski and Vaitkevicius

1997). Through the use of colour naming techniques and a large set of coloured samples, Hansen (Hansen, Walter et al. 2007) and Olkkonen (Olkkonen, Hansen et al. 2009; Olkkonen, Witzel et al. 2010) achieved different levels of colour constancy according to the degree of information provided. They modelled the transformations of the perceptual colour space under different illuminations by computing the boundaries of the colour categories (Hansen, Walter et al. 2007) and computing the colour constancy indices of the categorical prototypes (Olkkonen, Hansen et al. 2009; Olkkonen, Witzel et al. 2010). They concluded that the categorical structure of colour space has a high degree of robustness under changes of illumination which could be explained by linear models. However Hansen (Hansen, Walter et al. 2007) reported small rotations away from the illumination colour.

Chapter 3 Individual variations in colour categorization

Previous studies reported considerable variations in the categorical colour perception of different observers (Kuehni 2004; Xiao, Wuerger et al. 2011). For this reason, the first step in our analysis is to describe the inter-individual variability of categorical colour structure in an adequate colour space. This chapter quantifies this variability through the colour naming ability of individuals. To do so, we performed a psychophysical experiment (*Experiment I*) which tested normal trichromat observers in a series of colour naming tasks with an achromatic uniform background under *D65* illumination. This study of the colour naming abilities of our observers allowed us to formulate a direct link to the categorical colour structure of their internal colour spaces. We developed six indices to compactly measure their colour naming abilities, and from their values we derived our conclusions. Results indicate that the categorical structure of colour space is broadly uniform across observers: similar when close to the basic categorical centroid regions but different when close to their border regions. Also, using only the eleven basic terms may not be enough to achieve a successful colour communication in some regions of colour space, in other words, this eleven-term vocabulary needs to be expanded with three additional terms.

Part of the work presented in this chapter was done in collaboration with other co-authors (Roca-Vila, Owen et al. ; Roca-Vila, Owen et al. 2011). The original contribution of the author consisted on the development of new analytical methods, the analysis of the experimental data, the computational analysis and the discussion of results.

3.1 Introduction: Categorical colour perception from colour naming ability

Colour naming ability refers to an individual's capacity to communicate comprehensively the perception of surface and light colours using comprehensible colour terms.

The purpose of this chapter is to describe the colour naming abilities of individuals and to quantify their inter-individual variations. We did so by deriving a quantitative description of the normal colour space in terms of the centroid locations, volumes and shapes of distinct regions corresponding to the basic mono-lexemic colour categories as well as to significant multi-lexemic non-basic categories. Using a convex hull approach (Gärdenfors 2000; Berg 2008; Jäger 2010), we constructed a convex 3D region from punctuate naming data for each basic colour category. Then, using these 3D regions, we quantified the range of naming *inconsistency* from the overlap between category regions and the disagreement between surface-based and light-based categorisations. Naming *consensus* was derived from the intersection of all observers' same-category convex regions, which also informed calculations of the deviation in category shape from the region of maximal-agreement across normal observers. By defining the normal inter-individual variations in this quantitative description, we derived quantitative descriptors of naming ability, which may be used to assess individual performance on tasks where communication of colour terms is potentially critical, as well as differences in naming ability across age (Fu, Xiao et al. 2011), sex, culture and colour vision type. We compare these quantitative indices with the results of standardised tests of colour discrimination. According to previous studies of colour vision deficiencies, there is a lack of tests to measure colour perception beyond low-level sensory abilities (Cole and Maddocks 1998; Nolan, Riley et al. 2008; Uchikawa 2008). The current results provide indices which compactly describe colour naming abilities in terms of individual deformations of the categorical perceptual space. These results suggest that colour naming ability is not predicted solely by low-level chromatic discrimination ability.

Furthermore, we derive an average normal colour space which may be compared with models designed to predict the colour name of a given point in colour space (through assigning probabilities of belonging to a particular category (Lammens 1995; Benavente, Vanrell et al. 2008)). Our aim here, thus, is primarily to define variations in individual performance and individual colour spaces, and secondarily to characterise the categorical structure of average normal colour space.

3.2 Methods

Observers underwent a series of colour naming tasks using a total set of 439 distinct colour samples, either CRT stimuli (*light-based*) or Munsell chips (*surface-based*), with both forced- and free-choice colour naming paradigms. For each observer, we then defined his/her colour solid as the set of three-dimensional (3D) convex hulls computed for each basic colour category from the relevant collection of categorized points in CIELab colour space. From the parameters of the convex hulls, we then derived several indices to characterize the 3D structure of the colour solid and its inter-individual variations.

3.2.1 Observers

23 normal trichromatic observers (14 females and 9 males) participated, of mean age 27 years (range 17-50). All observers undertook and scored in the normal range on three standardized colour vision tests, the Neitz paper-based test (Neitz and Neitz 2001), Ishihara plates (Ishihara 1917) and Farnsworth-Munsell 100-Hue test (Farnsworth 1957) (mean score 13.22; 13.24 SD; range 4-64) performed under D65-metameric illumination (Verivide daylight simulation bulb) in a viewing cabinet. All observers were unpaid volunteers, naïve to the experiment's purpose. All had excellent English language skills; all but two were native English-speaking. All procedures were approved by the Newcastle University Psychology Ethics Committee (REF 060041).

3.2.2 Experimental setup and procedure

The colour naming tests (four in total) consisted of two types using different sample types and carried out in two different environments: the first used *surface*-based samples viewed under natural light (the *free naming* test), and the second used *light*-based samples displayed on a CRT monitor (the *forced-choice naming*, *pick-best* and *pick-all* tests).

Surface-based test: free naming

The *free-naming surface*-based test was performed using a standard colour chart (Gretag MacBeth Digital ColourChecker SG; (C. S. McCamy 1976)) consisting of 140 matte painted chips arrayed in a 10x14 grid, of which 99 chips are unique. At the viewing distance used here (80 cm), each chip subtended one square degree of visual angle, and the black border outlining each chip subtended 0.3 degrees. Observers viewed the chart illuminated by natural diffuse daylight through a large glass window, sitting at a bare grey table.

Observers were asked to name each chip using the most appropriate term, without any constraint of time or language, working row by row across the chart from top to bottom. Colour names were recorded by hand immediately and exactly as the observer said them, and subsequently coded for further analysis as described below.

Light-based tests: forced-choice naming, pick-best, pick-all

Each of the three light-based experiments began with a 60-second adaptation phase in which the observer viewed a uniform neutral colour (the "neutral background"; CIE Yxy coordinates [22.15 0.310 0.326]) filling the CRT display (41 x 35 degrees of visual angle).

Forced-choice naming. On each trial, the observer viewed a centrally placed single patch of uniform colour (5 degrees square) against the neutral background, appearing 2 degrees above the 11 basic colour terms (Berlin and Kay 1969) arranged in a 5 x 16 degree block of black text (BLACK WHITE RED GREEN / YELLOW BLUE BROWN ORANGE / PINK PURPLE GRAY). The observer's task was to move the mouse to select, as quickly as possible, the colour term that best named the displayed colour. Immediately after the selection, the colour patch was replaced with a multi-coloured mask of the same size, held for 200msec, after which the next trial began. The core test consisted of 340 trials, each presenting a different standard Munsell colour, as described below. (For each observer, an additional 4-13 "confusion" colour patches corresponding to particular colour vision deficiencies were included).

Forced-choice 'best exemplar' (pick-best). On each trial, the observer viewed an array of 141 uniformly-coloured patches, a subset of the forced-choice naming set, arranged in a 11 x 12 (+ 1 x 9) grid against the neutral background, each subtending 1.5 x 1 degrees with inter-patch spacing of approximately 0.5 degrees (total array size approximately 24 x 18 degrees). A block of black text appeared 1 degree below the array, containing the instruction: 'Select the best *[colour]*', where *'[colour]'* was one of the 11 basic colour names. The observer's task was to select the best example of the specified colour from the array, as quickly as possible. Immediately following the selection, the next trial began. There were five trials per colour name, the order of the 55 trials was randomised for each session, and the spatial configuration of patches in the array was re-randomised on each trial.

Forced-choice 'all exemplars' (pick-all). The stimuli and procedure for the pick-all test were exactly the same as for the pick-best test, except that the instructions read: 'Select all the [colour]' and the observer's task was to select all of the patches of the specified colour. There were two trials per colour name.

3.2.3 Stimuli

The 340 colours used in the *light*-based experiments (now called "*light*" samples) were selected by sampling 40 hues around the Munsell hue circle at each of three values (4,6,8), including all displayable chroma values from the set (4,6,10,12,14). CIE tristimulus coordinates were computed for each of the sampled Munsell colours, using the published Munsell paper spectral reflectance functions (Munsell Colour Company, 1929) and assuming D65 illumination, then converted to RGB coordinates for display using the calibration method described above. A subset of 140 colours from this set were selected to include at least one mid-chroma sample at each value, and one maximum-chroma sample at each of two values, per hue. The CIE tristimulus coordinates of the 140 patches in the *surface*-based test (now called "*surface*" samples), under diffuse daylight, were measured using a spectroradiometer (SpectraScan PR-650).

All tristimulus coordinates of the surface- and light-based samples were converted to the coordinates in the perceptually uniform CIELab space (Wyszecki and Stiles 1982), using the appropriate neutral point as anchor (CIE Yxy [22.15 0.310 0.326] for light samples, corresponding to the colour of the neutral grey background, and [57.70 0.314 0.343] for surface samples, corresponding to the mid-grey Digital ColourChecker chart patch under natural daylight). Colour space conversions and calibration routines relied partly on customised software (*kccv*; (Wolf 2011)) . Figure 3.1 illustrates the full set of samples in CIELab space; the values are also tabulated in Appendix A. It is evident that the two sets of samples are complementary in spanning colour space, with the surface-based samples tending to be more saturated, and the light-based samples providing denser coverage. Overall, the surface samples expand the volume of the light-based tested region in CIELab space by 50%.
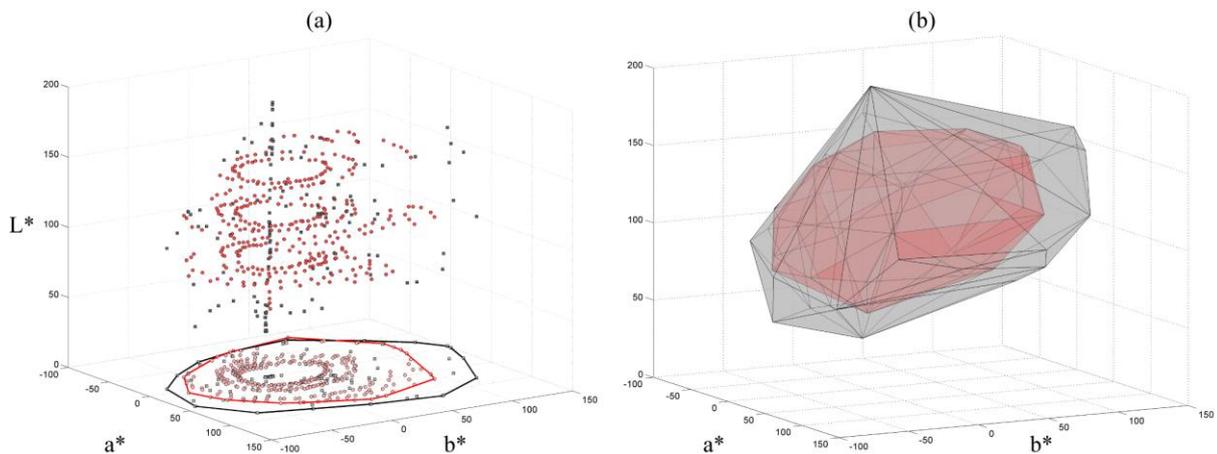
Figure 3.1 Tested coloured samples in CIELab colour coordinates.

3.3 Data analysis: the individual colour solid

3.3.1 Classification of colour names

For the forced-choice, pick-best and pick-all tasks, each sample was classified by a single colour name from the dictated set of 11 basic colour names. For the free-naming task, we used the following two methods to assign a colour name to a given sample based on the observer's response. (i) For the analysis of the number and frequency of colour terms used (see subsection 3.3.2), all colour names that were used as single names were counted individually, whether or not these were basic colour terms (e.g. *red*, *teal*), while for compound or qualified colour names (e.g. *greeny-white* or *pale pink*) only the base colour name was counted (*white* and *pink*, respectively). (ii) For the analyses of colour space, each colour name was assigned to one or more of the eleven basic colour categories following these rules: single basic colour terms were assigned to their corresponding category; compound or qualified colour names where the base name was a basic colour term (e.g. *creamy yellow*) were assigned to the base name category; and single colour terms that were not basic colour terms were assigned proportionately to the colour categories they straddled (e.g. *turquoise* was assigned 50% membership of the blue category and 50% membership of the green category).

Colour name	Codification		Colour name	Codification	
Turquoise	Green	Blue	Coffee	Brown	-
Terracotta	Brown	Orange	Coral	Pink	-
Sand	Yellow	Brown	Emerald	Green	-
Cream	Yellow	White	Flesh	Yellow	Pink
Fawn	Yellow	Brown	Fuchsia	Pink	-
Burgundy	Purple	Red	Lavender	Blue	Purple
Mint	White	Green	Lemon	Yellow	-
Lilac	Purple	White	Magenta	Red	-
Charcoal	Grey	-	Mushroom	Grey	Brown
Beige	Yellow	White	Mustard	Yellow	Brown
Violet	Purple	-	Ochre	Yellow	Brown
Salmon	Pink	Orange	Scarlet	Red	-
Cyan	Blue	Green	Stone	Grey	-
Mauve	Purple	-	Tan	Brown	-
Navy	Blue	-	Wine	Red	Purple
Light	White	-	Lime	Green	-
Peach	Orange	Pink	Khaki	Green	Brown
Maroon	Brown	Red	Slate	Grey	-
Magnolia	Yellow	White	Gold	Yellow	-
Skin	Pink	Brown	Aquamarine	Blue	-
Indigo	Blue	-	Praline	Yellow	Brown
Teal	Green	Blue	Taupe	Brown	Grey
Aqua	Blue	-	Amethyst	Purple	-
Aubergine	Purple	-	Tangerine	Orange	-
Olive	Green	-	Amber	Yellow	Orange
Sandstone	Yellow	Brown	Cerise	Red	-

Table 3.1 Codification of non basic colour terms in the free-naming experiment. Each non basic colour term is codified into basic colour categories for the analysis of colour space.

3.3.2 Definition

Our main analysis gathered all surface and light data together. Thus for each individual, for each colour category, we obtained a set of points in CIELab space that were classified by that category's name. Equation (3.1) defines this collection of points, C_i^j where $\bar{n}$ is the total number of observers, and M_i^j is the number of points $p \in \mathbb{R}^3$ in CIELab space classified by observer j as belonging to category i. We also define T_i^j as the corresponding set of response times, where each t_k is the time spent classifying p_k in seconds.

$$C_i^j = \left\{ p_{k_1}, \ldots, p_{k_{M_i^j}} \right\} \; ; \; p_{k_l} \in \mathbb{R}^3 \; l = 1, \ldots, M_i^j \; ; \; j = 1, \ldots, \bar{n} \; and \; i = 1, \ldots, 11 \tag{3.1}$$

$$T_i^j = \left\{ t_{k_1}, \ldots, t_{k_{M_i^j}} \right\} \; ; \; t_{k_l} \in \mathbb{R} \; l = 1, \ldots, M_i^j \; ; \; j = 1, \ldots, \bar{n} \; and \; i = 1, \ldots, 11 \tag{3.2}$$

The bounding region of the set of points C_i^j is a three-dimensional solid, which we modelled as its convex hull. The convex hull of a set of points X is defined as the intersection of all convex sets containing X. A set X in $\mathbb{R}^3$ is convex if, and only if, the line segment joining any pair of points of X lies entirely in X (Berg 2008). In our context, this means that if two samples are categorized by the same name then all points lying on their joining segment will also be assigned to the same colour category. Equation (3.3) gives its algebraic definition:

$$ConvexHull(X) = \left\{ \sum_{i=1}^{k} \alpha_i x_i \; ; \; x_i \in \mathbb{R}^3, \alpha_i \in \mathbb{R}, \alpha_i \geq 0, \sum_{i=1}^{k} \alpha_i = 1 \right\} \tag{3.3}$$

From now on, we refer to the convex hulls defined by C_i^j as

$$CH_i^j = ConvexHull\left(C_i^j \right) \tag{3.4}$$

We then define the *individual colour solid* for each observer i as the collection of the eleven convex hulls CH_i^j corresponding to the basic colour categories. Figure 3.2 illustrates the computed colour solids computation for three different observers. Notice the inter-individual differences in category volumes, shapes and intersections. For example, blue is relatively expanded for observer VV, while green is relatively expanded for observer IO. Also for observer IO, the green and blue categories overlap relatively little compared to the other two observers. The white category disappears for observer YL. Our following aim is to quantify the characteristics of the individual colour solids in a concise and precise way so that such inter-individual differences may also be quantitatively defined and used to describe individual colour naming abilities.

Previous studies demonstrated the rationale of using convex sets in CIELab colour space to represent colour categories corresponding to the basic colour terms (Gärdenfors 2000; Jäger 2010).
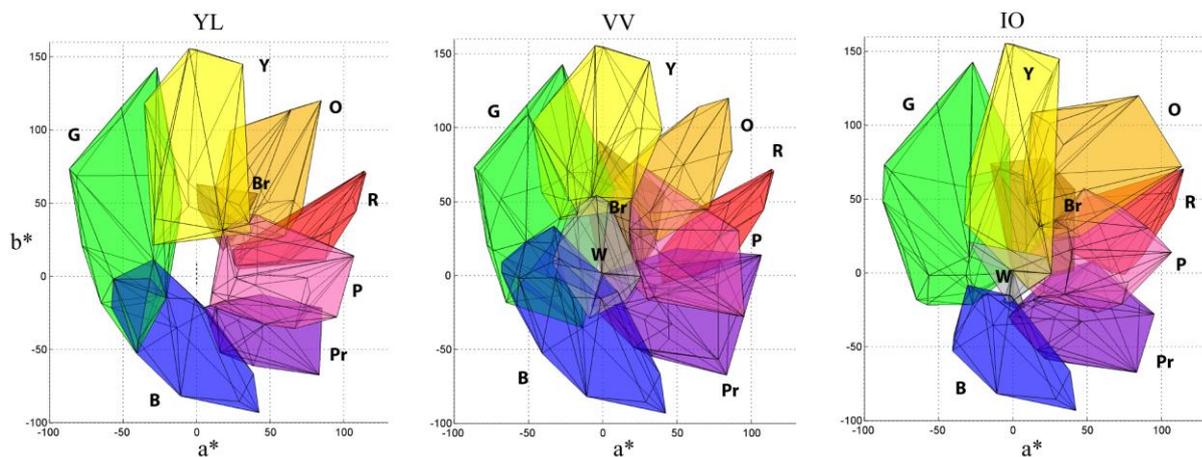
Figure 3.2 Computed colour solids for three normal trichromatic observers (YL, VV and IO). Each colour solid (set of 11 convex hulls corresponding to the basic colour categories), constructed from light and surface naming tasks as described in the text, is shown projected onto the chromaticity plane in CIELab colour space. Colour categories are labelled and colour-coded with their representative colours (R-red; G-green; B-blue; Y-yellow; N-grey; W-white; K-black; P-pink; O-orange; Pr-purple; Br-Brown).

3.3.3 Quantification

We developed a set of six indices that compactly describe the geometrical features of the three-dimensional colour solid and its inter-individual variations, and which may be directly related to features of naming behaviour described by other researchers, e.g. naming consistency and consensus (Boynton and Olson 1987; Sturges and Whitfield 1995). These indices are of two types: absolute indices that quantify the internal features of the colour solid (*volume*, *time*, *category inconsistency* and

surface-light inconsistency), and relative indices (*structure deviation* and *centroid deviation*) that quantify the deviations of the categories from a predefined average "normal" colour solid. Each index is defined for each observer *j* and each category *i*.

Volume (V). The *Volume* index is calculated straightforwardly from the convex hull and expressed in CIELab cubic units. The index indicates how large the colour category is and its range therefore goes from zero to the volume of the convex hull generated from all tested points.

$$V_i^j = Volume\left(CH_i^j\right) \tag{3.5}$$

Time (T). In the *forced-choice naming task* the response time for each trial was recorded, thus enabling us to calculate the average classification time (in seconds) for each colour category, the *Time* index. Its expected values range from milliseconds up to two or three seconds.

$$T_i^j = mean\left(\mathcal{T}_i^j\right) = \frac{1}{M_i^j}\sum_{l=1}^{M_i^j} t_l \tag{3.6}$$

Category Inconsistency (CI). As Figure 3.2 clearly illustrates, categorical regions often overlap in individual colour solids. This overlapping will necessarily arise when the observer uses different basic colour terms to classify the same colour sample on different occasions, a phenomenon that has been identified and quantified as naming "inconsistency" in previous studies ((Berlin and Kay 1969; Sturges and Whitfield 1995; Guest and Van Laar 2000)). In our analysis, the overlap may also arise when nearby samples are classified differently and thus fall into distinct convex hulls; in that sense, our analysis is able to include inconsistencies that are directly predicted by the convexity of categories even if not explicitly tested. We therefore define *Category Inconsistency* as an extension of direct naming inconsistency: the total volume of overlap between the given category's convex hull and all other categories, as a proportion of the given category's convex hull volume. The index therefore ranges from 0 (no intersection with any other regions) to 1 (all points in the category also fall in at least one other category).

$$CI_i^j = Volume\left(\bigcup_{\substack{k=1 \\ k \neq i}}^{11} CH_i^j \cap CH_k^j\right)\bigg/ V_i^j \tag{3.7}$$

Surface-Light Inconsistency (SLI). The *Surface-light Inconsistency* index was designed to measure categorization differences between data from the *surface* and *light* colour naming experiments. We therefore defined S_i^j, the *surface* set, in the same way as C_i^j but without the points from the *light* experiments, and $\overline{CH}_i^j$, the *light* convex hull, in the same way as CH_i^j but without the *surface* points. For each *surface* point we found the nearest convex hull in $\overline{CH}_i^j$. We counted the number of surface points whose surface categorisation was coincident with the category of the nearest light category. Where no points are coincident, the index is one; where all points are coincident, the index is 0, as below:

$$\Phi(p) = arg\left(\min_{i=1,\dots,11}\{dist(p,\overline{CH}_i^j)\}\right) \in \{1,\dots,11\} \; and \; p \in S_i^j \tag{3.8}$$

$$SLI_i^j = 1 - \left(\frac{\#\{p \in S_i^j, \Phi(p)=i\}}{\#\{S_i^j\}}\right), where \; \#\{\mathcal{C}\} = number \; of \; \mathcal{C} \; elements \tag{3.9}$$

Structure Deviation (SD). This index quantifies the regularity of category shape according to a predefined normal shape. The normal shape for a particular category is defined for a specified group of N normal observers as a *layered category* (Equation 3.10). Each successive layer includes votes from successively more observers. The Nth, innermost, layer for, say, the "green" category, is supported by all N observers; it is the intersection of all N convex hulls for green. The (N-1)th layer is supported by N-1 observers, and so on down to the 1st, outermost, layer. The *Structure Deviation* index for a given individual category is then calculated from a weighted sum of the intersections between the normal layers and the given category, normalised by the volume of the latter (Equation 3.11). For any given category, the maximum regularity possible is when it coincides perfectly with the innermost layer, agreeing with all N observers. The index value is then zero, indicating null shape deviation. If the intersection with all normal layers is empty, then the index value is one, indicating total shape deviation.

$$L_i(\alpha) = \left\{ p \in \mathbb{R}^3 \mid \exists r_1, \dots, r_\alpha \text{ and } p \in CH_i^{r_l} \; \forall l = 1, \dots, \alpha \; ; \; r_l \in \{1, \dots, N\} \right\} \text{ where } \alpha = 1, \dots, N \tag{3.10}$$

$$SD_i^j = 1 - \sum_{\alpha=1}^{N} \left(\frac{\alpha}{N}\right) \left(\frac{Volume\left(ConvexHull\left(L_i(\alpha)\right) \cap CH_i^j\right)}{V_i^j} \right) \tag{3.11}$$
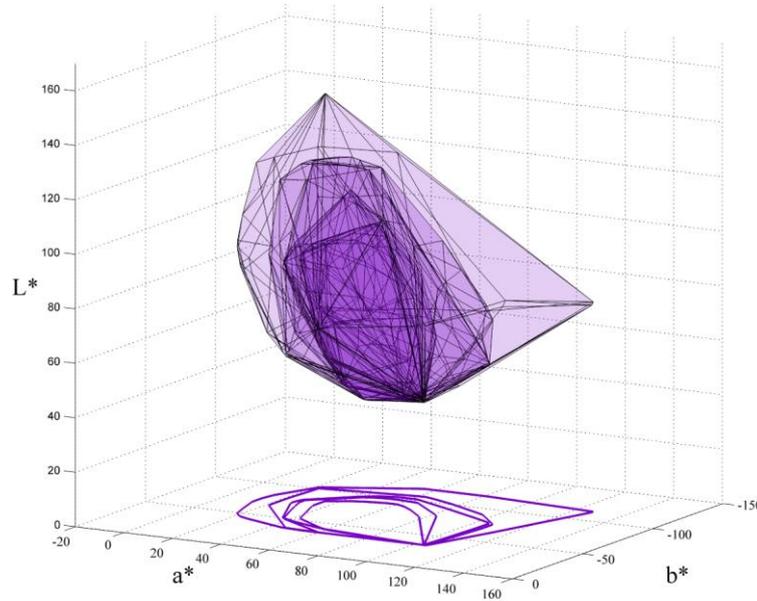
Figure 3.3 Purple layered category from 23 NT in CIELab. It was used in the computation of the Structure Deviation index.

Centroid Deviation (CD). This index quantifies the normality of the positions of the categories in colour space relative to each other. Following standard practice (Boynton and Olson 1987), we define the *centroid* of each category as the mean location of all its points, F_i^j (Equation 3.12). We then calculate the normal category focal difference between two categories as the mean difference in the centroids of the two categories, averaged over all N observers in the specified normal set: this is *T(i,c)* for colours *i* and *c* (Equation 3.13). For each observer *j*, and each colour *i*, the *Centroid Deviation* index is then the maximum of the absolute difference between one and the ratio of the observer's category focal distance to the normal category focal distance, taken over all other colours in the basic set. The index therefore indicates, for a given observer, which colour categories are most displaced

relative to all other categories; e.g., a high index for *green* indicates that there is at least one category from which *green* is displaced, relative to the normal distance between *green* and that category.

$$F_i^j = \frac{1}{M_i^j} \sum_{l=k_1}^{k_{M_i^j}} p_l \ \ where \ p_l \in C_i^j \ as \ defined \ in \ Equation \ (3.1)$$

$$\text{(3.12)}$$

$$T(i,c) = \frac{1}{N} \sum_{j=k_1}^{k_N} \left\| F_i^j - F_c^i \right\|_2 \ \ where \ i,c = 1,\dots,11$$

$$\text{(3.13)}$$

$$and \ k_l \ indicates \ subjects \ from \ the \ normal \ set$$

$$CD_i^j = \max_{c=1,\dots,11} \left\{ \left| 1 - \frac{\left\| F_i^j - F_c^j \right\|_2}{T(i,c)} \right| \right\}$$

$$\text{(3.14)}$$

	Notation		Computational description	Colour naming framework	Range of values/units
General Descriptors	V	Volume	3D volume of category convex hull	Number of usages of a given colour name	0 to total volume/ CIELab cubic units
	T	Time	Average time to name a sample (forced-choice naming)	Naming time	Unlimited/seconds
Naming behaviour	CI	Category Inconsistency	Overlap between given category and all other categories, as proportion of total volume	Inconsistency in classification of given sample	0 to 1
	SLI	Surface-Light Inconsistency	Proportion of surface sample names in disagreement with light categories	Inconsistency in classification of given sample when presented as light vs surface	0 to 1
Category Geometry	SD	Structure Deviation	Deviation in category shape from region of maximal-agreement across normal observers	Deviation from typicality of given colour name usage	0 to 1
	CD	Centroid Deviation	Deviation from normal distance between given category centroid and other category centroids	Deviation of selected focal colour relative to normal	0 to $T < \infty$ (positive number)

Table 3.2 The introduced indices, their notation, computational description and corresponding feature in colour naming studies.

3.4 Results

3.4.1 Overall colour category locations

Each observer performed 760 (on average) classifications in total (340 in forced-choice naming, 55 in pick-best, 225 in pick-all, on average, and 140 in free naming). Of the samples classified, 439 were unique (340 *light* samples; 99 *surface* samples). To obtain an overall picture of the classification data, we pooled data from all observers across all tasks into the eleven basic colour categories (using the second method to classify each response in the free-naming task, as described in subsection 3.2.1).
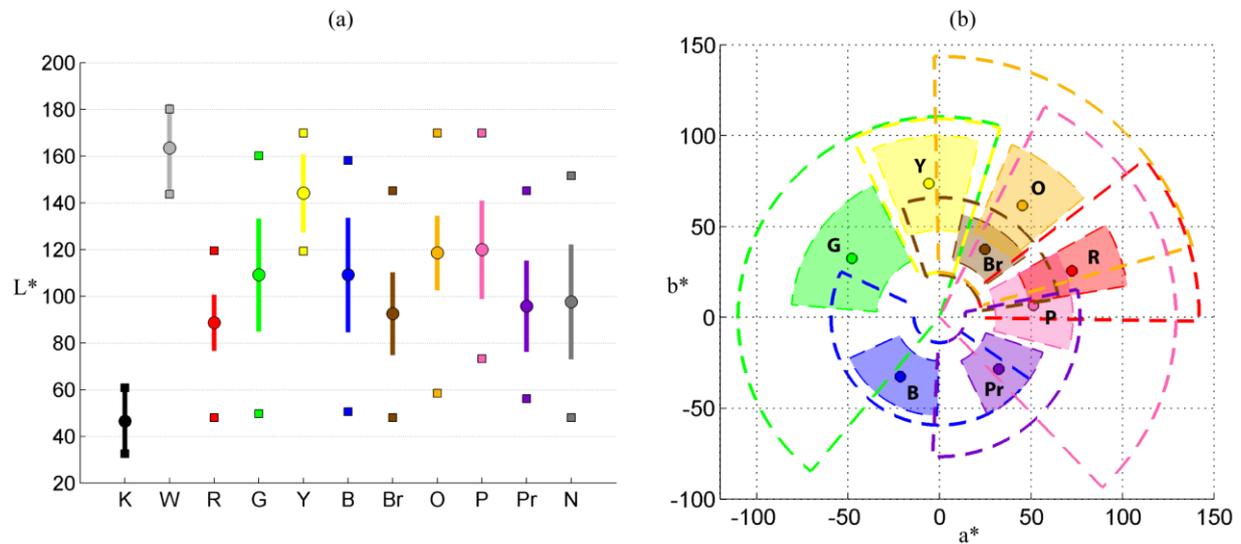
Figure 3.4 CIELab coordinate locations of mean and range of categorized samples in each basic colour category, calculated from all naming data (light and surface sets) across all observers. Panel a: Mean (open circle), standard deviation (coloured bar), and maximum and minimum values of L* (open squares) of all colours classified with the indicated colour term. Panel b: Mean chromaticity coordinates in the a*b* plane of chromatic colours classified with the indicated colour term. Shaded areas indicate the standard deviation in hue and saturation of all colours classified with the indicated colour term, calculated in LHS space. Dotted lines indicate the maximum and minimum hue and saturation range. Achromatic categories are omitted for clarity.

Figure 3.4 illustrates the resulting mean and range of the lightness, hue and saturation values for each basic colour category. Figure 3.4 reveals some notable features. First, the overlapping between shaded regions (indicating one standard deviation) is nearly nonexistent; combining the CIE L* information from panel *a* with the hue and saturation information from panel *b*, we see that there is minimal overlap only between brown and orange, and red and purple regions, suggesting stable classification naming behaviour for samples around the mean position in terms of inter-individual variability. Second, for three categories (yellow, red and purple), the full-range sectors (indicated by dotted lines in panel *b*, and by filled squares in panel *a*) are relatively little expanded compared to the shaded

regions, whereas for others (green, blue, orange and pink) they are greatly expanded. Third, the overlapping between the full-range sectors is much greater on the right-hand side of the chromaticity plane (regions corresponding to yellow, orange, red, brown and pink) than the other (regions corresponding to green, blue and purple). These observations indicate that, despite the core stability, there will be large variability in naming in some regions of colour space.

The indices we report below quantify these observations and the underlying naming patterns in terms of intra- and inter-individual variations in naming ability. Basic and non basic colour terms usage

3.4.2 Basic and non-basic colour terms usage

The free naming test provided a total of 3220 named samples (140 samples for each of 23 observers). A total of 46 distinct colour terms were used. The 11 basic terms were used for 90.5% of distinct samples, and non-basic terms for the remaining 9.5%. Only the basic terms were used by all observers. Amongst the 35 non-basic terms used, the highest-usage terms were *turquoise*, used by 15 observers, *peach* (12), *lilac* (9) and *beige* (8).
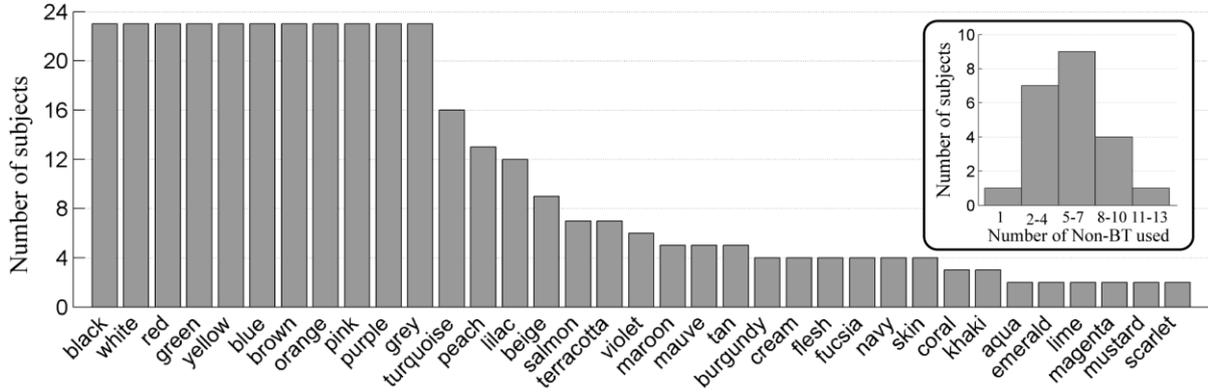
Figure 3.5 *Free naming* test results. The main histogram indicates the number of observers (total of 23) that used each particular term given on the horizontal axis. The inset histogram indicates the number of non-basic terms used (omitting one observer who used 21 different terms).

Results are summarized in Figure 3.5; its horizontal axis lists all terms that were used at least twice and bar height indicates the number of observers using the term. 74% of observers used 7 or fewer non-basic terms and only two observers used more than 11 non-basic terms, as shown in the inset histogram. The overall set of non-basic terms (*turquoise, peach, lilac, beige, salmon, terracotta, violet, maroon, mauve, tan ...*), and their usage frequency are similar to previous reports (Boynton and Olson 1987; Sturges and Whitfield 1995) despite differences in the number of tested samples; *turquoise*, in

particular, is in the top three most-used non-basic terms for all studies. There is a trend for females to use more non-basic terms than males (15.8 vs 9.4) and to use a larger number of distinct non-basic terms (7.1 vs 5.4).

3.4.3 Naming indices results

Unless otherwise stated, each index is computed for each category from the results of all tasks, and averaged over all observers.

Volume *(Figure 3.6a).* We normalize the volume index by the sum of all category volumes, for each observer separately, before averaging across observers, in order to allow comparison with naming frequencies reported in previous studies (Boynton and Olson 1987; Sturges and Whitfield 1995; Paggetti, Bartoli et al. 2011) (illustrated with squares [Boynton and Olson 1987] and circles [Sturges and Whitfield 1995] in Figure 3.6a) and with the naming frequencies calculated from our forced-choice naming task (triangles in Figure 3.6a). (Reported rates from earlier studies combine consistent and inconsistent results for a fair comparison with our data, which at this stage does not differentiate between overlapping and non-overlapping regions). Green is the largest category, occupying nearly a third of the total volume, followed by blue and several other midsized categories (pink, orange, yellow and purple). The smallest categories are brown, red and white. The forced choice naming frequencies are significantly correlated with the normalised volume index values across categories ($r=0.9312$; $p<0.01$), and follow the same trend as the naming frequencies from previous studies.

(Because the black and grey volume indices are close to zero, due to possible under-sampling of the achromatic region of colour space and/or augmented chromatic sensitivity near the achromatic axis (Sturges and Whitfield 1995), we are forced to discard their corresponding *category inconsistency* and *structure deviation* indices, due to their usage of the convex hull element.)

Time. The *Time* index was calculated for data from forced choice naming tests. The mean T was 1.85 seconds (0.45 SD) per classification, with only small differences between colour categories. On average, samples farthest from the category centroid required 30% more time per classification than samples close to the centroid. These results are in broad agreement with earlier studies, which reported values between 1.46s and 2.55s (Boynton and Olson 1987), and 1.31s and 2.34s (Sturges and Whitfield 1995) depending on whether samples were *consistent/ consensus* or *inconsistent* types, respectively.
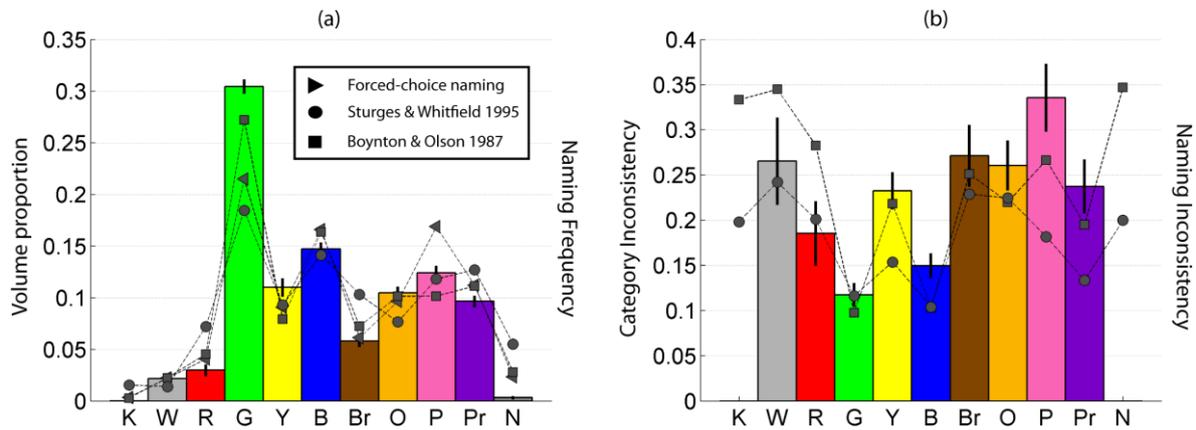
Figure 3.6 Volume and Category Inconsistency indices (coloured bars) compared to previous literature results (triangles correspond to the forced-choice naming test, squares to B&O1987 and circles to S&W1995). Black lines indicate the standard error. Panel a: Bars indicate the Volume index values normalized by overall volume and markers indicate the naming frequencies of the corresponding basic terms (see text for details). Panel b: Bars indicate the Category Inconsistency index values and markers correspond to the ratio of inconsistent to (consistent+inconsistent) frequencies reported by B&O and S&W. Values corresponding to K and N are not reported/analysed due to their small or non-existent volume.

Category Inconsistency (Figure 3.6b). The mean CI value varies significantly across colour ($F(10,220)=12.528$, $p<.001$) from 0.1 to 0.35, with a mean of 0.23 (0.13 SD), equating to roughly one-quarter of the category volume. Green and blue have the lowest CI values (mean=0.13; 0.05 SD), significantly lower than brown and pink (mean=0.27; 0.16 SD) ($t(22)=-4.371$, $p<0.001$; two-tailed). Note also that CI values for green and brown are significantly correlated ($r=.624$, $p=0.001$). The CI values follow the same pattern across categories as the normalised naming inconsistency values from previous studies (calculated as the ratio of reported inconsistent to the sum of consistent and inconsistent classifications; shown as squares (Boynton and Olson 1987) and circles (Sturges and Whitfield 1995)).

Surface-Light Inconsistency (Figure 3.7a). The mean SLI index value for chromatic categories is 0.39 (0.19 SD), indicating roughly 60% coincidence between names in the surface and light presentations. For the achromatic categories, the mean SLI value is much lower at 0.08 (0.07 SD). Of the chromatic colours, blue has the lowest SLI, in keeping with the fact that its category volume spans almost the whole of the lightness dimension. The difference between the achromatic and chromatic categories may be partly explained by the achromatic categories having relatively smaller volumes and larger numbers of surface samples. In general, these results are confounded by the negative correlation

between the average number of surface samples classified and SLI index per category *(r=-0.64; p=0.0319)*.

Structure Deviation (Figure 3.7b). The mean SD index value over all colour categories (excluding black) is 0.25 (0.09 SD), indicating significant deviations in category shapes between observers. The index varies significantly across colours (F(9,135)=71.088, p<0.001). Green and blue categories have the lowest values (close to 0.1); yellow, brown, purple, orange and pink have mid values (close to 0.25); and white and red have the largest values (0.6 and 0.37, respectively).

Centroid Deviation (Figure 3.7c). The CD index varies significantly across colours (F(10, 220)=5.231, p<0.001), being lowest for green (0.08) and highest for red (0.16) with a mean value of 0.13 (0.05 SD). The low values overall (around 10% maximum deviation from normal focal distances) suggests relative stability in centroid locations across observers.
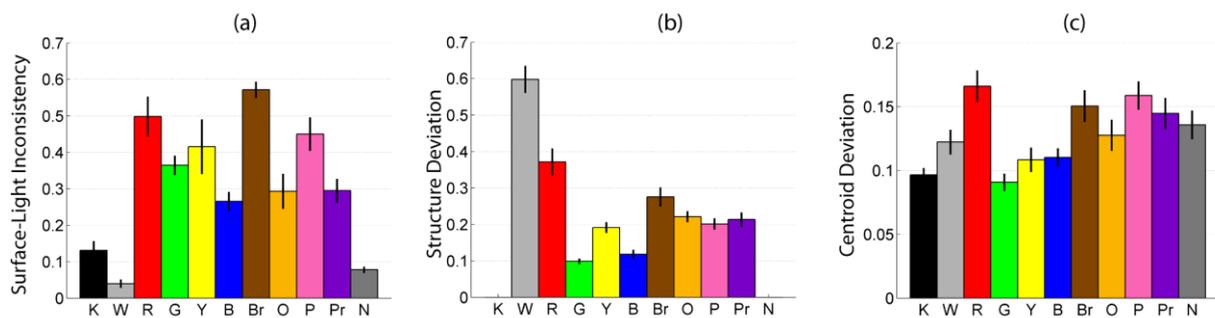
Figure 3.7 The Surface-Light Inconsistency, Structure Deviation and Centroid Deviation index values (panel a, b and c respectively). For each panel coloured bars and black lines indicate observer averaged values and the standard error, respectively. As before, the K and N structure deviation indices are not reported because of their near-zero volume indices. Note that panel c uses a different scale.

3.4.4 Inter-individual variations: the coefficient of variation

Comparison between the three colour solids displayed in Figure 3.2 suggests notable differences in their inner configurations. To capture this inter-individual variation, we introduce the *coefficient of variation (CV)* for each index, defined as the quotient between the standard deviation and the mean of the given index (i.e., a normalized measure of dispersion). Table 3.3 summarises the coefficients of variation, per index and per colour category. The average value of 0.42 indicates substantial inter-individual differences in colour naming abilities and categorical colour spaces. From now on, in extracting features of colour naming behaviour, we will classify *CV* values as *low* (less than 0.25), *medium* (between 0.25 and 0.5) and *high* (greater than 0.5).

41

In detail, for a particular colour category, the *volume index CV* indicates the variability between observers in the number of times that the particular colour term is used. Volume CV values are high for red and white, low for green and medium for the rest. For example, the green CV value is 0.07, indicating that the green convex hull volume, and therefore the frequency of usage of the term *green*, was nearly the same over all observers. The *time index CV*, on the other hand, varies little across colour categories, with a low mean value of 0.24, indicating that observers spend similar amounts of time in classifying colours on the forced-choice naming task. Inter-individual variability in naming inconsistency is assessed by the *category inconsistency index* CV, which is high for white and red, and medium for the rest. Therefore, despite the relatively low absolute value of CI for red, its high CV suggests that some observers highly confound red terms with others (see *Geography of the average colour naming space* subsection for details on which sets of basic terms are confounded). In general, though, colour categories with lower *CI* values tend to have lower CV values, especially for green and blue which have the lowest *CI* values. The *surface-light inconsistency* CV index (*SLI*) varies significantly across colours $(F(10,220)=18.279, p<0.001)$, with high CV values for black, white, yellow and orange, low CV values for brown and medium CV values for the remaining categories. Lastly, the CV values for the *structure deviation* and *centroid deviation* indices follow similar patterns, both varying little across categories, with medium average values of 0.39 and 0.37, respectively. The two together suggest that observers vary moderately regarding where in colour space they assign a particular colour term.

	V	T	CI	SLI
Black	-	0.31	-	0.93
White	1.07	0.33	0.87	1.37
Red	0.61	0.24	0.92	0.52
Green	0.07	0.18	0.43	0.34
Yellow	0.22	0.23	0.37	0.86
Blue	0.16	0.19	0.38	0.46
Brown	0.45	0.28	0.40	0.18
Orange	0.22	0.23	0.46	0.78
Pink	0.22	0.25	0.48	0.49
Purple	0.28	0.17	0.50	0.52
Grey	-	0.28	-	0.53
Mean	0.37	0.24	0.53	0.63

Table 3.3 *Coefficient of variability* according to colour categories and indices. Note that values for black and grey categories on the V and CI indices have been disabled due to their convex hull existence restrictions.
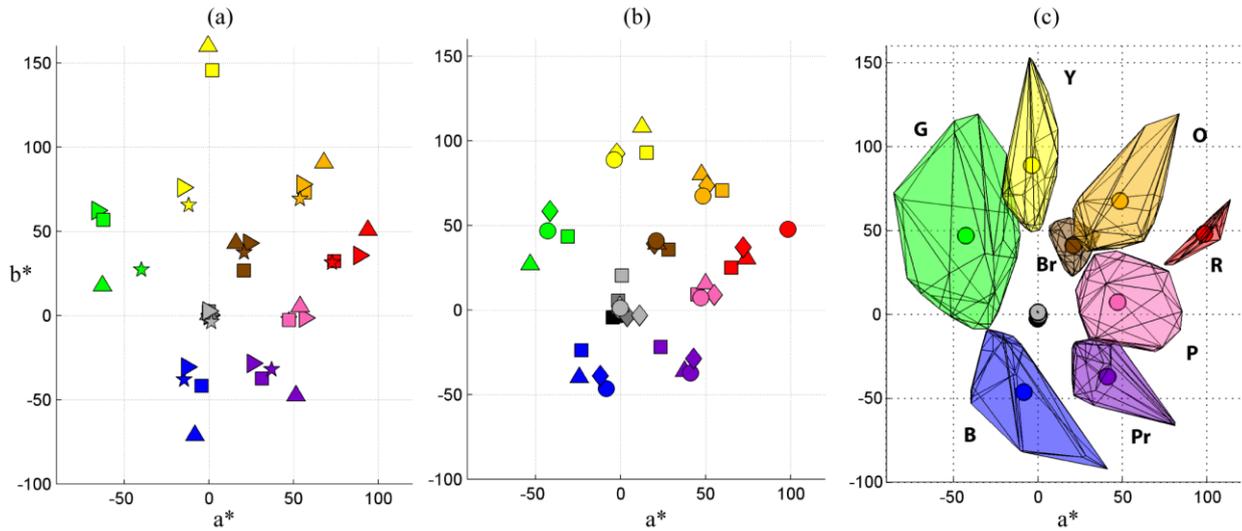
Figure 3.8 Colour category regions location in the a*b* chromaticity plane. Panel a: focal colour locations for forced-choice naming data (pentagon), pick-best data (rotated triangle), B&O (triangles) and S&W (squares). Panel b: centroid locations for all our categorized samples (rhomboids), B&O (triangles), S&W (squares) and the inner layer of each layered category (circles). Panel c: Each coloured region represents the inner layer of each corresponding layered category (see Structure Deviation index definition in Methods section).

3.4.5 Geography of the average colour naming space

From the full set of data over all tasks (approximately 17480 classifications from 23 observers and 760 samples each) we constructed an *average colour naming space (ACNS)* which preserves information about both the commonality and variability of the individual colour solids. This construction may be compared with previous attempts to build universal colour spaces (Lammens 1995; Guest and Van Laar 2000; Menegaz, Troter et al. 2007; Benavente, Vanrell et al. 2008) .

To describe the space systematically and comprehensively, we gathered all data into a single multidimensional variable. First, we sampled the three-dimensional region spanned by the union of all colour solids in CIELab space, in steps of one unit. Second, each point in the resulting three-dimensional matrix was expanded into an eleven-dimensional *nameability vector*, in which each coefficient corresponded to one particular colour category. Each coefficient was computed from counting the number of times that the given CIELab point fell inside one of the coefficient's category's convex hulls. Because there are as many convex hulls as observers per category, the *nameabilty vector* coefficient values range between zero and the total number of observers (23). Last, the matrix was smoothed by applying a low-pass Gaussian filter and normalizing each *nameablity vector* by the sum of its coefficients. This final step abstracts the number of observers used in the experiment, gives the same information consistency to all CIELab points and allows further comparisons with fuzzy sets approaches (Lammens 1995; Benavente, Vanrell et al. 2008).
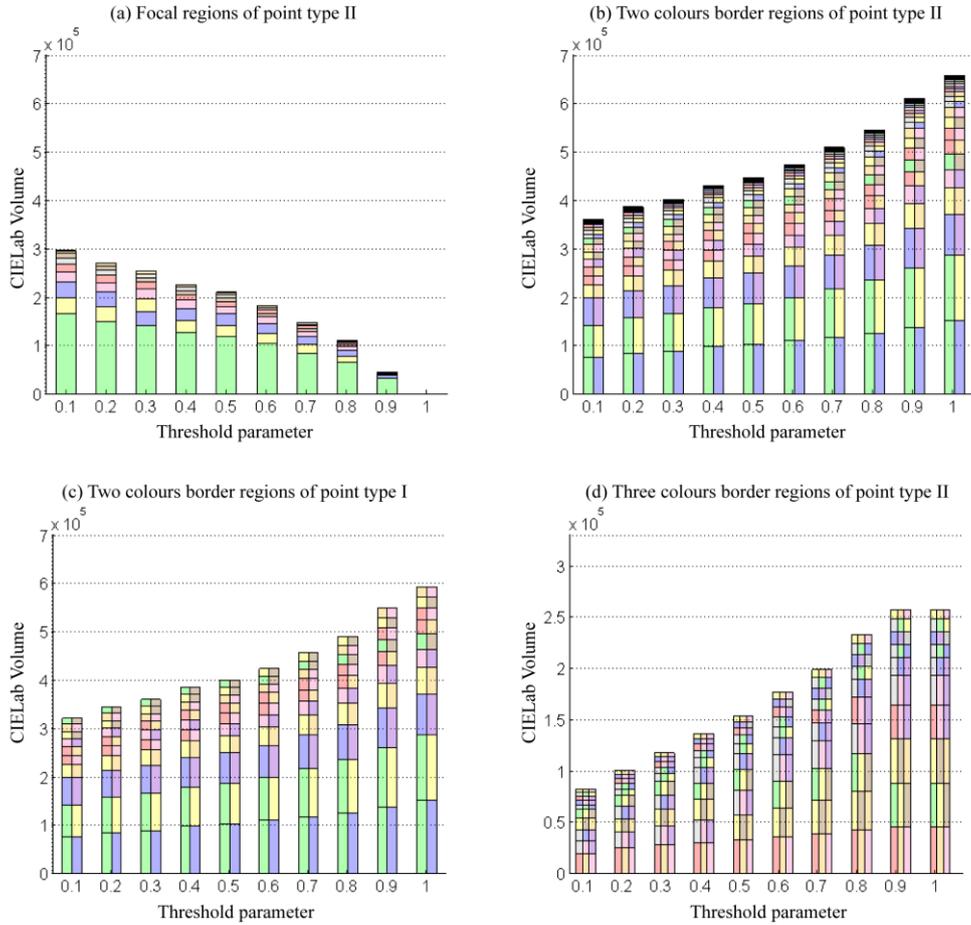
Figure 3.9 Influence of the threshold parameter in the volume of ACNS regions.

Using the information in this multidimensional matrix, we then classified the category belongingness of each point in the CIELab region using two descriptors*: Point Type*, and *basic/border*. The point's *Point Type* is the number of non-zero coefficient values of its *nameablity vector*, i.e. the number of different categories to which it belongs. The point is further labelled as *basic* to category i if the absolute distances between the *ith* coefficient and other coefficients in the nameability vector are all larger than a specified threshold parameter, or as *border to* categories j if all of its related non-zero coefficients are below this threshold. For example, a Type II point will have only two non-zero coefficients in its nameability vector, say at positions k and l. If 0.5 is the threshold, and its values are 0.4 and 0.6, then their absolute difference is 0.2 and being lower than 0.5 dictates that the point will be labelled as *border* to both categories. If the values are 0.2 and 0.8, respectively, their absolute distance is 0.6 and the point will be labelled as basic to category l. Note that the *basic* labelling is exclusive; a given point may be *basic* to only one category and cannot simultaneously be *border* to others. Varying

the threshold parameter from 0.3 to 0.7 (see Figure 3.9) made no qualitative difference to ACNS, so for this analysis we set it to 0.5.

Our single variable multidimensional approach allowed us to easily produce 3D reconstructions and visualizations of all regions. This can be achieved following the same approach as Menegaz et al (Menegaz, Troter et al. 2007) where each region surface was effectively rendered by means of the *marching cubes* algorithm (Lorensen and Cline 1987).

As Figure 3.10a illustrates, the proportion (and volume) of points decreases exponentially as Point Type rank increases, with point types I to III taking up nearly 92% of the ACNS volume. *Basic* points (light grey) constitute overall 67% of the volume, and *border* points (dark grey) the remaining 33%. Table 3 and Figure 3.10b report the CIELab centroid coordinates, and proportional and cumulative volumes for the 35 largest regions (*basic* or *border*), respectively. The proportional volumes for basic categories follow roughly the same order as their mean volume indices, with green and blue the largest and red, brown and achromatic categories the smallest. The largest 11 regions, though, include 3 border regions of Point Type II: green-blue, green-yellow, and blue-purple. The largest Type III border region is the red-orange-pink region; the largest Type IV border region is white-blue-pink-purple; and the largest Type V border region is white-yellow-brown-orange-pink.
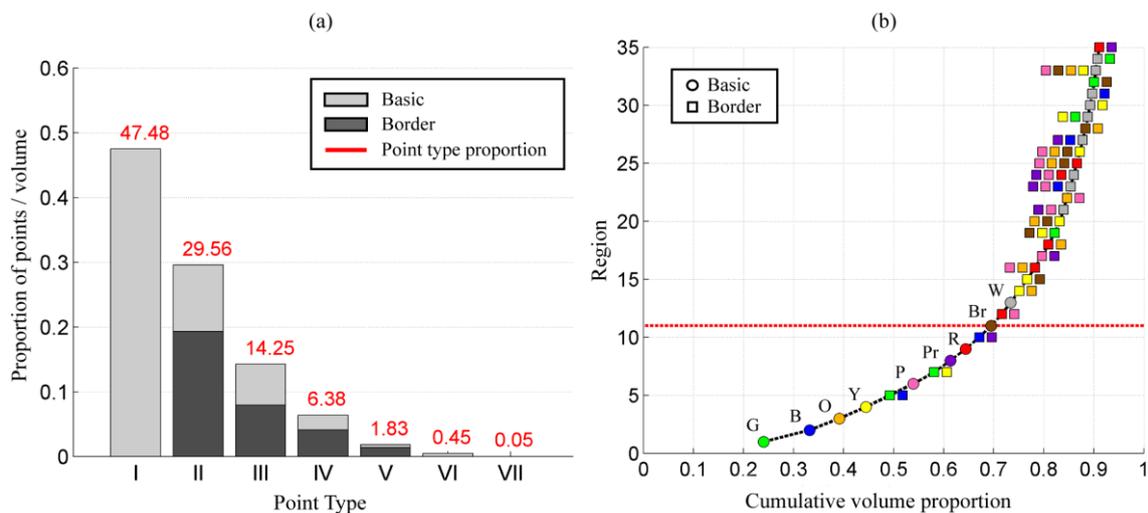
Figure 3.10 Composition of the Average Colour Naming Space. Panel a: Point type histogram. Light and dark grey bars indicate basic and border regions respectively as described in the text. Panel b: Cumulative proportion of total volume of the average colour naming space spanned by basic and border category regions, sorted, from left to right, by size of contributing region. Each region is represented by one or multiple markers according to whether it corresponds to a basic region (circle) or a border region (multiple inline squares) from 2-,3-,4- or 5- colours. The red dotted line delimits the first 11 regions, which contribute 69.5% of total volume.

45

The total cumulative volume reaches approximately 90%. The remaining 10% consists of regions smaller than those reported in Table 3 as well as non-analysed parts of point type IV and higher, whose inclusion in the analysis does not alter the overall picture.

Region					Centroid coordinates			Volume proportion
1	**2**	**3**	**4**	**5**	**a***	**b***	**L***	
Green					-36.4	57.8	112.1	23.99
Blue					-9.58	-38.66	103.3	9.21
Orange					63.5	86.8	119.6	5.91
Yellow					-7.1	80.9	134.6	5.35
Green	Blue				-34.1	-14.7	114.7	4.81
Pink					71.5	5.9	120.1	4.67
Green	Yellow				-15.5	75.06	129	4.18
Purple					52.5	-27.8	90.7	3.25
Red					67	38.2	86.8	3.02
Blue	Purple				23	-34.2	101.6	2.74
Brown					21.5	29.2	77.5	2.33
Red	Pink				70.6	18.08	104.6	2.18
White					5.4	4	156	1.75
Yellow	Orange				27.4	86.2	136.4	1.71
Yellow	Brown				7.1	55.4	109.8	1.59
Red	Orange	Pink			66.5	29.8	115.3	1.55
Pink	Purple				51.4	-13.6	108.5	1.38
Red	Orange				73.3	42.4	101.4	1.31
Green	Yellow	Brown			2.5	46.5	106.9	1.27
Yellow	Brown	Orange			26.5	80.4	126.81	0.98
White	Pink	Purple			43.8	-14.5	124.3	0.77
Orange	Pink				56.9	27.5	121.8	0.72
White	Blue	Pink	Purple		24.8	-20.9	117.6	0.67
White	Red	Pink	Purple		43.5	-6.6	107	0.66
Red	Brown	Orange	Pink		43	28.8	103.6	0.62
Yellow	Brown	Orange	Pink		26.3	45.6	122.4	0.60
White	Blue	Purple			20.5	-36.6	115.46	0.59
Brown	Orange				34	70.2	145.7	0.51
White	Green	Yellow			-11.5	27.6	145.7	0.48
White	Yellow				-0.5	29.8	149.4	0.45
White	Blue				5.3	-17.11	150.43	0.44
Green	Brown				3.5	32.4	77.7	0.38
White	Yellow	Brown	Orange	Pink	21.5	31.2	135.2	0.34
White	Green				-16	21.8	146.8	0.33
Red	Purple				54.9	-4.6	73.8	0.33

Table 3.4 The 35 largest regions of the *average colour naming space* (*basic* in bold) and their centroid CIELab coordinates, sorted by volume proportion (rightmost column). First row indicates the number of categories in each

Figure 3.11a shows the proportional volume occupied by each basic and Type II border colour category as a function of hue angle in CIELab space; the integral of the green line corresponds to the reported volume for the green category in Table 3.4. Note that each curve has the same fundamental shape, i.e., continuous with only one clear maximum. The same analysis for luminance and saturation

revealed similarly shaped curves (see Appendix B), thus indicating that each region of the ACNS is fully connected as confirmed by luminance plane sections (see Figure 3.11b for the L*=93 section). The centroid locations of the 35 largest regions (both basic and border) are shown in Figure 3.11c. Here it is clear that the higher the point type rank (the more categories contribute to the region), the closer the region's centroid is to the origin. This is expected as the centroid values signal the central tendency of the samples that they encapsulate, but also suggests that the richest part of colour space is the innermost one, in terms of colour naming singularities.

We studied systematically the locations of centroids of border regions of two relative to their constituent basic categories. For example, the green-blue region centroid was projected onto the line that joins the blue and green region centroids and proximity proportions were computed (see Figure 3.11). Over all such regions, only three border regions fall exactly in the mid-position: purple-blue, orange-yellow and brown-yellow. Two additional significant features emerged: (1) when the border region involves the green category then the centroid is always displaced towards the opposing category and; (2) conversely, when the border region involves grey, its centroid is always displaced towards grey.
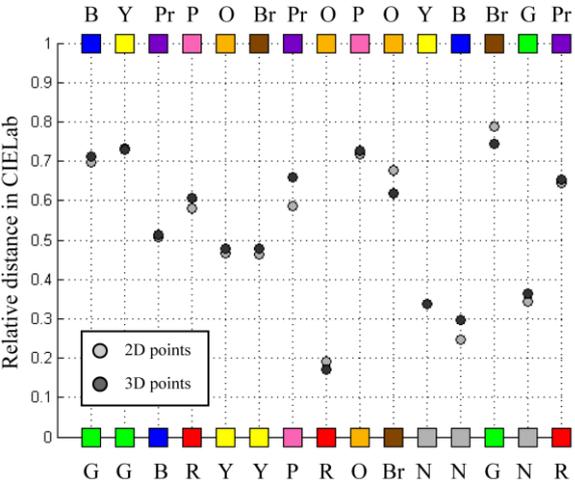
Figure 3.11 Relative distance location between ACNS region centroids of two colours and their constituent basic colour categories. Light grey markers indicate computations using only a* and b*, and dark grey markers indicate using all three dimension of CIELab space. See text for further computation details.

Additionally, we compared the locations of the border regions with the locations of the centroids of the most-used non-BCT categories (turquoise, peach, lilac and beige). To do so, we removed the non-BCT naming data from the results and re-computed the ACNS (the BCT ACNS), then computed the mean locations of all free-naming terms used more than four times. Of all the identified regions, the

centroid location of turquoise is closest to the centroid of the blue-green border region (distance 17.5 ΔE in Lab units in the BCT ACNS; 13.6 ΔE in the full ACNS), which suggests that this non-BCT arose from the need to resolve uncertainty at the borders of the two categories. Similarly, the centroid location of the non-BCT lilac corresponds well to the 3rd largest border region: blue-purple. Its centroid is closest to the white-blue-purple centroid in the full ACNS (18.2 ΔE*) and second closest to the blue-purple centroid (20.7 ΔE*). The centroids of peach and beige are themselves nearly overlapping (12.9 ΔE*), and also very close to the centroids of the less-commonly used terms skin and fawn. The centroid location of peach is closest to the centroid of the five-border region of white-yellow-brown-orange-pink in the full ACNS. We suggest that these very similar terms arose not out of uncertainty over naming in a border region but instead out of the need for finer delineation of the sub-region of colour space corresponding to skin colours.

Conversely, the green-yellow border region, the second largest border regions, does not appear to have given rise to a common-use non-BCT. The term lime is the nearest non-BCT (distance 25.4 ΔE*) but in our population was used fewer than 4 times.

3.4.6 Influence of coding method for non basic colour terms

We examined the effects of the way we coded single non-basic colour terms on our results by performing the same analyses on the data set excluding all non-basic terms. We find minimal effects of the coding method. For example, the volume index ranks for the basic categories from the two methods (see Figure 3.12a) are highly correlated (r=0.99, p<0.01). Also, the intersections between region pairs are nearly the same for the two methods (see Figure 3.12b), yielding similar values for the category inconsistency index across categories. Finally, the ACNS volume ranks for the 20 largest regions (see Figure 3.12c) are highly correlated in both methods (r=0.98, p<0.01), with the order of the largest three and smallest three preserved in both.
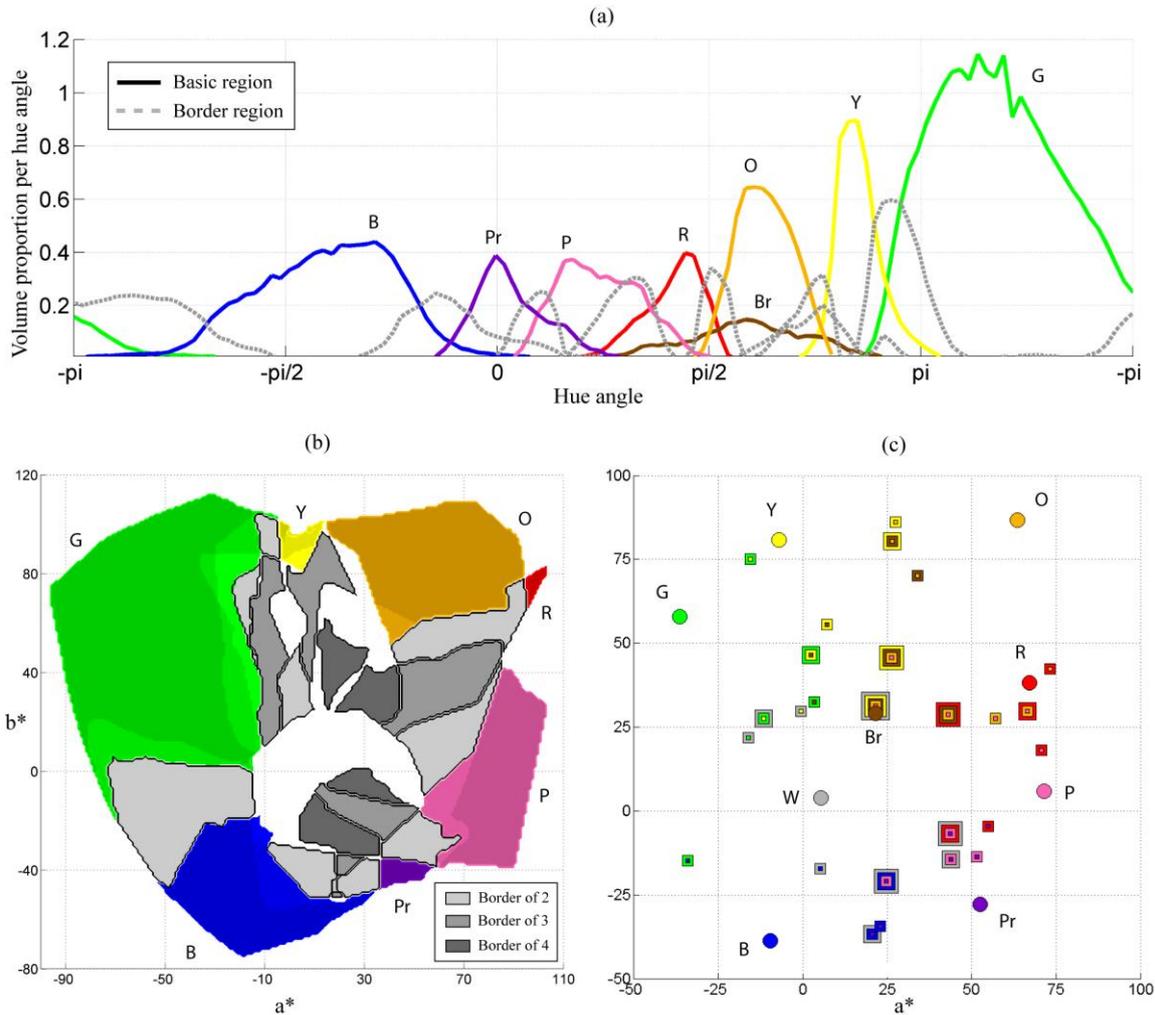
Figure 3.12 Structure of the average colour naming space (ACNS). Panel a: Volume proportion histogram of *basic* (coloured lines) and *borders of two* (grey dotted lines) regions according to their hue. The same analysis was performed for Saturation and Luminance and similar shapes were obtained. Panel b: Luminance section of the ACNS at $L^* = 93$. Darker greys indicate higher-order border regions (see text for details). Panel c: Centroid locations in the a*b* chromaticity plane for the 35 largest regions listed in Table 3. Circles indicate basic regions; squares indicate border regions, nested according to number of categories.

3.5 Discussion

3.5.1 Features of our model

Previous colour naming studies used naming frequencies to assess the reliability and the commonality of individual responses using surface or light samples (Boynton and Olson 1987; Sturges and Whitfield 1995; Paggetti, Bartoli et al. 2011). Our approach has similar aims but differs in its methods and samples and thereby enables a richer examination of both universal features and individual variations in colour naming. We used a combined set of light and surface samples. Instead of

analysing naming frequencies we modelled individual responses as a collection of 3D regions and quantified their interrelations. The model assumes that regions are convex and connected in CIELab colour space, a feature that follows from its construction as an opponent colour space (Fairchild 2005). Other studies have implicitly made the same assumption without explicit use of convex hulls (e.g. when computing centroids by averaging named locations); the assumption is also supported by extensive empirical categorizations (Boynton and Olson 1987; Olkkonen, Hansen et al. 2009). A recurrent conclusion of previous studies is that the number of tested samples and their distribution in colour space influence the resulting categorisations (Boynton and Olson 1987; Sturges and Whitfield 1995; Paggetti, Bartoli et al. 2011). In our study, the tested samples were not equally distributed but biased to the positive-positive quadrant in the CIELab chromaticity plane (see Table 3.5), enabling the densest sampling where the greatest number of category centroids are located (see Table 3.6 and Figure 3.8).

Quadrant	++	-+	--	+-
Light	0.33	0.30	0.15	0.22
Surface	0.45	0.24	0.13	0.18
Light+Surface	0.35	0.28	0.15	0.21

Table 3.5 Fraction proportion of test samples in each quadrant of the a*b* plane, for the light-based samples (top row), surface-based samples (middle row) and all samples (bottom row). Each quadrant is labelled by the signs of the a* and b* axes.

Despite differences in number and distribution of samples, our results are consistent with previous studies, as indicated by the high correlation between our *Volume* index and others' naming frequencies (Figure 3.6a), by the agreement between our *Category Inconsistency* index and others' inconsistent naming frequencies (Figure 3.6b), and by the agreement in focal and centroid locations (Figure 3.8a; the only significant difference lies in the yellow focal location, because our tested samples were not as saturated as those used for other studies). The convex hull approach thus enables a consistent expansion of information from a small set of categorized samples to large regions of non-tested samples, i.e., all points contained in the convex hulls. In turn, the indices report information not only on the tested samples but the whole of colour space (up to the sampling borders).

Quadrant/Category	Red	Green	Yellow	Blue	Brown	Orange	Pink	Purple
++	**0.99**	0.01	0.26	0	**0.97**	**1.00**	**0.65**	0.02
-+	0	**0.92**	**0.74**	0.04	0.03	0	0	0
--	0	0.07	0	**0.86**	0	0	0	0
+-	0.01	0	0	0.10	0	0	0.35	**0.98**

Table 3.6 Fraction proportion of tested samples according to basic colour categories and a*b* plane quadrants. In bold the quadrant with highest proportion for each colour category.

3.5.2 Colour naming ability

We derived quantitative indicators of colour naming ability from two main analyses: the compact descriptions of individual colour solids in terms of indices; and the region-based analysis of the ACNS. The usage patterns of basic and non-basic colour terms in the free-naming test also provided quantitative information about naming behaviour, on both the individual and group level.

Results from the free-naming test revealed an extensive usage of the 11 basic colour terms (BCTs) by our observers – 90.5% of samples were named with BCTs, and all observers used all of the 11 BCTs – thus vitiating the use of only the 11 BCTs in the other forced-choice naming tests. Nonetheless, nearly 10% of samples elicited non-basic colour terms (non-BCTs), and all observers used at least one non-BCT, despite not being restricted to use monolexemic terms and therefore allowed to qualify BCTs. This pattern supports the notion of "hard-to-name" regions (Boynton and Olson 1987; Sturges and Whitfield 1995) and, given that BCTs should span the whole of colour space, might suggest a need for more BCTs, a point to which we return when considering the ACNS analysis.

Analyses of the index values revealed significant differences in naming behaviour between colour categories. *Volume* and *Category Inconsistency* indices confirmed green and blue categories to be the largest (Figure 5a) and the most stable (Figure 5b) of all categories, in accordance with previous studies (Boynton and Olson 1987; Sturges and Whitfield 1995; Paggetti, Bartoli et al. 2011). This result implies that the usage of the terms *green* and *blue* in colour communication is both more extensive and more reliable than other terms. Furthermore, differences in the *Structure Deviation* index between colour categories indicate that observers are more likely to concur when naming green or blue samples and disagree when naming white and red. The low value of the *Centroid Deviation* index across all colour categories indicates that the categorical perceptual structure of colour space is similar for all our observers. The close concordance of our calculated centroids with those previously reported (see Figure 7b) also suggests similarity across tested populations. Communication of colours close to the centroids should therefore be highly successful in comparison to those further away, a conclusion reinforced by the fact that samples close to the centroid are categorized 30% faster. Lastly, our approach allows us to make comprehensive characterisations of colour naming ability by

combining several indices. For example, we are able to conclude that the green category is the most stable when it has to be named (*CI*) and the most-often named (*V*), and that the most agreement is reached (*SD* and *CD*) in where to name it.

Analysis of the ACNS, which summarized all the colour naming information provided by our observers, provided the description of colour space in terms of *basic* and *border* regions, of which we reported the largest 35. As expected from the forced-choice naming tasks and previous studies (Menegaz, Troter et al. 2007), most of the space is spanned by regions corresponding to the basic terms. Yet there are also large regions corresponding to intermediate border areas, in particular green-blue, green-yellow and blue-purple border regions, which together constitute nearly 12% of the total volume. This fact, together with the moderate usage of some non-BCTs in the free-naming test (e.g. turquoise), suggests that additional BCTs in the experimental vocabulary are necessary to reduce the size of the border regions. We also suggest that it is critical to successful colour communication to maximise the volume covered by basic terms and minimise the size of intermediate regions because these are where observers vary most in their naming behaviour. Our results indicate that adding just three basic terms, corresponding to the three largest border regions, would make a significant difference in their coverage of colour space (from 59.5% to 71.2% coverage) and a corresponding improvement in colour communication. Analysis of the locations of monolexemic non-BCT centroids suggests that the two commonly-used terms *turquoise* and *lilac* may already act to cover the first and third largest of these border regions. These three border regions also behave differently from other border regions, in that their constituent terms do not form the stems for higher-order concatenations. The latter, formed by incrementally adding colour terms, as in *yellow-brown*, *yellow-brown-orange*, *yellow-brown-orange-pink*, and *yellow-brown-orange-pink-white*, generally support the existence of "hard to name" regions described by Boynton and Olson 1987. The latter five-colour region may also correspond to the commonly-used non-BCT *peach*.

Analysis of the ACNS also provides centroid locations for the border categories, an addition to previous reports of basic category centroids (Boynton and Olson 1987; Sturges and Whitfield 1995; Paggetti, Bartoli et al. 2011). Most locations are biased towards one category, rather than located at mid-point between the constituent BCT categories. The dominance of green is also reflected in the displacement of border regions to green *away* from the *green* centroid. The heavy density of border regions around grey, and the displacement of border centroids towards grey, fits with the observation that discrimination is increased around neutral points (von der Twer and MacLeod 2001) and also suggests that the inner regions of colour space will be most susceptible to deviations from colour constancy.

In general, the complexity of the 3D category shapes revealed by the ACNS analysis, especially of the border categories, indicates that it is not appropriate to generalise from linear borders in the 2D plane (as in Hansen et al. (Hansen, Walter et al. 2007)) to planar borders in 3D space. Modelling of colour categories as fuzzy sets in which smooth functions mediate between basic regions (Lammens 1995; Benavente, Vanrell et al. 2008) may also be an oversimplification.

3.5.3 The relationship of naming indices to FMHT performance

Standardised tests of colour vision examine chromatic discrimination at the sensory level. Although the ability to distinguish between two colours must logically be a prerequisite to giving them different names, chromatic differences are clearly not sufficient to ensure categorical differences. Many discriminable *greens* still fall in the same green category, while two barely discriminable colours may fall either side of a categorical divide. Perceptual categorisation and linguistic labelling, and any interaction between these (Gilbert, Regier et al. 2006; Brown, Lindsey et al. 2011), are thus likely to involve perceptual and/or cognitive processes well beyond the sensory level, unexamined by standardised discrimination tests. Thus, we expect that the naming indices introduced here will provide a complementary tool to sensory discrimination tests in assessing colour perception.

To evaluate the complementary information conveyed by the indices, we first examined potential redundancies between the indices themselves, and then examined their relationship with the Farnsworth-Munsell 100-hue test (FMHT) scores. For the former, we performed Pearson correlation on all possible pairs of the mean indices, averaged over all colour categories (excluding white and black) for each observer, across the 23 observers (see Table 3.7). Because the total volume overall is limited by the maximal volume of the tested set (Figure 3.1b), increases in the mean category volume tend to correlate with increasing overlap of individual categories. We therefore expect, and find, a significant correlation between the mean *Volume* (non-normalised) and *Category Inconsistency* (*CI*) indices *(r = .858; p < 0.01)*. But this correlation does not hold for the individual colours green, brown, orange, pink, and purple, indicating that the two indices provide independent information on the individual category level. In the case of green, it is clear from the coefficients of variation and indices value that the green category is large and relatively stable in size across all observers, and that therefore its variation in naming inconsistency is due to other factors. Similarly, although the correlation between the mean *Surface Light Inconsistency (SLI)* and *V* is significant (r = 0.478, p < 0.05), the correlation is insignificant at the individual category level, except for orange (r = 0.591; p < 0.01) and pink (r = -0.542; p < 0.01). The lack of significant correlation of *SLI* and *CI* for 5 of the 9 colour categories (excluding white and black) confirms that the two indices measure different types of inconsistency.

The lack of correlation between *Centroid Deviation (CD)* and *Structure Deviation (SD)* both at the mean level and for all individual colour categories (except red; $r = 0.506$, $p < 0.05$) indicates that the two measure different structural features. Mean *SD* correlates with mean *V* ($r = 0.465$, $p < 0.05$), whereas *CD* does not, on either level. The correlation between mean *V* and *SD* arises because the larger volume categories are comprised of a greater number of outer regions receiving a fewer number of votes, i.e. a larger number of layers that do not perfectly intersect with individual convex hulls. The correlation between *V* and *SD* is significant within each colour category alone, indicating that *SD* is largely determined by the presence of outer layers.

	Score/Index	FMHT	V	CI	SLI	CD	SD
General Descriptors	Farnsworth-Munsell 100-Hue Test (FMHT)	1	-	-	-	-	-
	Volume (V)	-.104	1	-	-	-	-
Naming Behaviour	Category Inconsistency (CI)	-.207	.858**	1	-	-	-
	Surface-Light Inconsistency (SLI)	-.191	.478*	.436*	1	-	-
Category Geometry	Centroid Deviation (CD)	-.217	.339	.423*	.174	1	-
	Structure Deviation (SD)	-.116	.465*	.360	.321	.191	1

Table 3.7 Correlations between indices values and Farnsworth-Munsell 100-hue test scores. Each value in the table indicates the Pearson Correlation coefficient for the corresponding row and column indices or test scores. **. Correlation is significant at the 0.01 level (1-tailed). *. Correlation is significant at the 0.05 level (1-tailed).

Second, we examined the correlations between the FMHT scores and mean indices across observers. There are no significant correlations (Table 3.7). The lack of correlation between FMHH and indices is further illustrated by factor analysis on the correlation matrix of the mean index and test scores. Principal component analysis (PCA) with varimax rotation reveals two components that together explain 61.4% of the variance across observers. The V, CI, SLI and SD indices load highly and positively on the first component; the FMHH score loads negatively and the CD index positively on the second component. This result supports the hypothesis that low-level chromatic discrimination ability does not perfectly predict naming behaviour, in accordance with previous studies which suggest the contribution of higher-level factors (Webster, Webster et al. 2002). It is important to note, though, that these results hold only for the population of normal trichromats tested here. We expect the descriptors to vary differently for populations of colour-anomalous observers.

3.5.4 Choice of colour space

The analysis colour space used in the previous methods has been the CIELab colour space, which is a tri-dimensional colour space with distances perceptually uniform (Fairchild 2005). Distances in CIELab colour space are distorted when the reference white point is selected, however according to Fairchild the CIELab colour space is very good when achromatic colours are used as a reference white point, which is our case (Fairchild 2005). Also, most of our analysis is based on relative and not absolute values, and since the formulae to compute CIELab values is smooth (continuous and differentiable) we could expect that our relative values are mostly resistant to this wrapping.

3.6 Conclusions

Colour naming ability may be compactly described by a reduced set of indices which characterize the 3D structure of the individual categorical colour space and allow quantification of its inter-individual differences. Although the population of normal trichromat observers tested here had broadly similar categorical structures, there was significant variation amongst observers in the centroid locations of categories, 3D shapes of the outer layers, and overlap between neighbouring categories. These variations were not explained by the concomitant inter-individual variation in sensory discrimination ability, as tested by the Farnsworth-Munsell 100-hue test. The indices may be further used to extract descriptors of naming behaviour for other populations, differing in age, sex and colour vision deficiency type. It will also be of interest to examine the variation in index values under different states of adaptation due to different scene configurations or illumination contexts, to supplement measures of colour appearance or colour constancy. The tested samples used here were embedded in a uniform grey background, and we expect that introducing more complex backgrounds may alter the indices; i.e. naming behaviour will vary with context as well as the individual.

Chapter 4   The chromatic setting paradigm

In Chapter 3 we described a considerable inter-individual variation in the categorical colour structure of colour space. We also described how most points in colour space were categorized using only eleven basic colour terms. Our results indicate that the categorical structure of centroid regions was mostly stable, but in-between regions had a higher potential to fail in terms of colour communication. Therefore, when studying categorical colour constancy we could restrict our analysis to basic colour categories, and also tune the categorical colour information present in the stimuli to each particular observer. This chapter introduces a newly developed colour constancy paradigm and a psychophysical experiment (*Experiment II*), to measure colour constancy under more extensive periods of immersive adaptation to the illumination. We used this paradigm to compare adaptations under two different illuminations, as is common in colour constancy studies, and since our experiment relies strongly on colour memory ability, we studied its validity by testing the stability of the internal references used by our subjects. To check whether our paradigm provides a more comprehensive measure of the colour constancy phenomenon, we applied linear models to characterize our observer's behaviour and to quantify to what extent these models are capable of absorbing the growing data complexity that results from measuring multiple points. Finally, we developed a new colour constancy index that arguably captures the intrinsic complexity of the phenomenon in a single value, while still in agreement with the previous colour constancy literature. Chapters 5 and 6 apply the new colour constancy paradigm to obtain a new set of measures and show its usefulness for studying categorical colour constancy under different illuminations.

4.1 A new psychophysical paradigm to measure colour constancy

We developed a colour constancy paradigm that tries to minimize the weaknesses while keeping the strongest points of previous paradigms (Foster 2011) and can be seen as an extension of the achromatic setting paradigm which, instead of using only the internal "grey" reference, uses several categorical colours, exploiting the ability of subjects to consistently replicate focal colours over time. Focal colours (Berlin and Kay 1969; Boynton and Olson 1987) are by definition the most representative colours of each naming category and there is strong evidence of they are influenced by language (Heider 1972; Kay, Siok et al. 2009) and memory (Hansen, Olkkonen et al. 2006; Ling and Hurlbert 2008; Nemes, McKeefry et al. 2010). Although the ability of subjects to match a memorized colour decreases in general with increasing inter-stimulus intervals (Nemes, McKeefry et al. 2010), there is some evidence that focal colours can be remembered more accurately than other colours (Heider 1972). In our method, the measurements are conducted under a permanent state of adaptation to the illumination, thus avoiding potential illuminant-switching issues that occur in other methods. The stimuli were combinations of three different two-dimensional Mondrians and three different illuminants. The subject's task was to select and then to reproduce a particular colour from memory. In order to rule out memory failings the experimental procedure included a series of repeatability tests.

4.1.1 Workings of the new paradigm

Our paradigm consists of two steps as illustrated by Figure 4.1. In the first step, subjects were asked to select colours that best represented basic colour terms within a limited region of the colour space (*Bounding Cylinder*, represented by a red circle). These were grey, green, blue, purple, pink, red, brown, orange and yellow (Berlin and Kay 1969). The squares within the red circle in Figure 4.1 symbolize the colours selected during this first step which we called *reference session*. We termed these colours *Selected Representatives (SRs)*. In the second step which we called *regular session,* the same subjects were asked to reproduce these SRs under different conditions of background and illumination. The squares outside the red circle in Figure 4.1 correspond to these colours, and the arrow represents the change in adaptation state. Since the new paradigm can be seen as an extension of the achromatic setting paradigm to multiple colours, we named it *Chromatic Setting*.

The red circle in Figure 4.1 corresponds to the projection of a cylinder in the a*b* plane. This cylinder was introduced to limit subjects choices, thus avoiding highly saturated colours that fall outside the CRT gamut when "illuminated". Details on the Bounding Cylinder implementation can be found below.
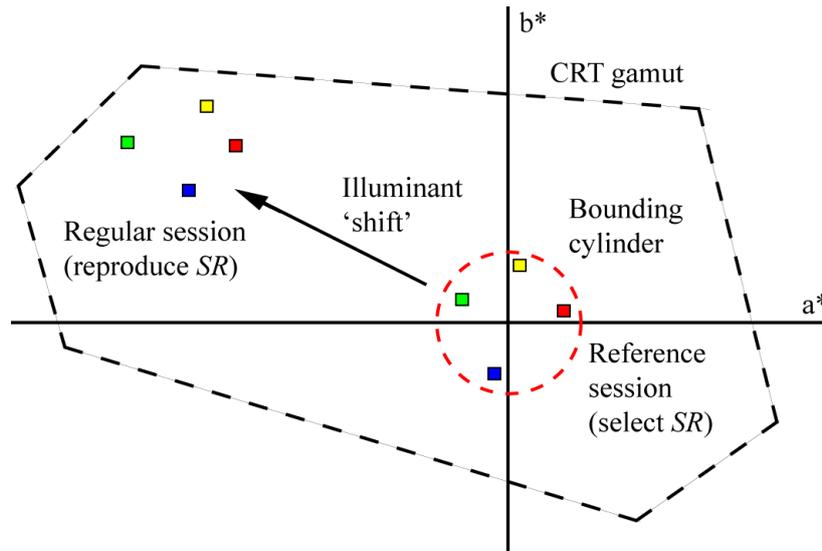
Figure 4.1 Schematics of the chromatic setting paradigmin the a*b* plane of CIELab colour space. The black broken lines represent the boundary of the CRT gamut. The squares inside the red circle represent the colours selected in the reference session. The squares outside the red circle represent the colours reproduced once adapted to the new illuminant in a regular session. The arrow represents the chromatic shift induced by the illumination.

4.1.2 The Bounding Cylinder

In the reference sessions, the palette of possible colours was limited in saturation and lightness by a cylinder whose main axis was the lightness dimension of CIELab (L* between 30 and 70 and radius equal to 22 ΔE*). The purpose of the cylinder was strictly technical as illustrated in Figure 4.1: we wanted subjects to find reasonably representative samples while still allowing these colours to be "illuminated" later without exceeding the CRT-monitor gamut. This limitation and the shape of the monitor's gamut in CIELab also determined our choice of illuminants. The value of 22 ΔE* for the radius was chosen after our own (unpublished) measurements indicated that colours closer than 12 ΔE* to the achromatic locus were usually categorized as "grey". Subjects naturally tended towards choosing saturated colours, and to stop them from using the borders of the cylinder as a reference, i.e. to increase saturation until hitting the cylinder limit, the experimental program "bounced back" the stimulus inside the cylinder by a small random amount once the boundary was reached. The Bounding Cylinder was not present in regular sessions.

4.2 Methods

4.2.1 Observers

Ten subjects, six male and four female took part in our experiments. They were between 20 and 44 years old and their colour vision was normal as tested by the Ishihara colour vision test (Ishihara 1972) and the Farnsworth-Munsell D15 hue test (Farnsworth 1957). All had self-reported normal or corrected to normal visual acuity. Three of the subjects were the authors. The rest were naïve to the purposes of the experiment and of these, three were paid.

4.2.2 Experimental setup

All sessions were conducted inside a dark room, with all walls lined in black. The experiment was programmed in Matlab and the stimuli were displayed on a CRT Mitsubishi Diamond Pro 2045SU monitor at 100Hz, driven by a ViSaGe graphics card from Cambridge Research Systems Ltd. (CRS - www.crsldt.com) with 12 bits colour resolution per channel. The CRT screen measured 389 mm in height by 292 mm in width subtending approximately 22x17 deg and was the only light source in the room. Its resolution was 1024x768 pixels. Viewing was binocular and unrestrained. The monitor was calibrated regularly using a Minolta *ColourCal* colorimeter and CRS software. We used the COLOURLAB (Malo and Luque 2002) toolbox to get the colour space conversions needed. Subjects modified the test stimuli by navigating the CIELab colour space using six different buttons, two for each colour space dimension on a commercial gamepad. The reference white point was D65, Lum = 100 Cd/m^2.

4.2.3 Stimuli

Our basic stimulus consisted of a Mondrian background pattern, i.e. a set of randomly overlaid coloured rectangles, distributed across the screen. The average rectangle size was 50x50 pixels. There were three types of backgrounds:

Type 0. It was built from 7 intensity levels of the same D65 chromaticity. They were equally spaced between 40 and 70 Lab lightness units and their luminances in Cd/m^2 were: 11.25, 14.54, 18.42, 22.93, 28.12, 34.05 and 40.75. Its mean was 22.66 Cd/m^2.

Type I. It was built from the SRs chosen by each subject in reference sessions (see details below). There were 8 colours in total: green, blue, purple, pink, red, brown, orange and yellow. Their averaged luminance range was between 12.77 and 39.29 Cd/m^2, mean = 25.11 Cd/m^2.

60

Type II. It was built from 8 hues halfway between those of type I, with similar saturation and lightness: blue-purple, purple-pink, purple-red, red-orange, orange-yellow, orange-brown, yellow-green and green-blue. Their averaged luminance range was between 16.87 and 35.54 Cd/m^2, mean = 24.35 Cd/m^2.

The number and sizes of rectangles were manipulated so that the pixel average chromaticity of all background types prior to illumination was that of D65. Backgrounds Type I and II did not contain achromatic D65 rectangles to avoid giving the observer cues about the illuminant (Foster 2011). Unique randomized Mondrians were created for each experimental trial: no observer saw the same Mondrian twice. To illuminate the Mondrian pattern, we first assigned to each rectangle a spectral reflectance function, interpolated from the set of Munsell chips assuming a Lambertian reflectance model -see COLOURLAB (Malo and Luque 2002) for implementation details. Illumination was simulated by performing the spectral product of each rectangle's reflectance by one of three illuminants (*D65*, *greenish* and *yellowish*), whose CIE *xy* chromaticities are shown in Table 4.1. The luminance range in Cd/m^2 for the illuminated stimuli was between 11.25 and 40.74 for the D65 illuminant; between 11.24 and 40.73 for the greenish illuminant and between 11.20 and 40.56 for the yellowish illuminant. The mean values in Cd/m^2 were 24.04, 23.7 and 24.37 respectively.

Illuminant	x	y
D65	0.312	0.329
Greenish	0.296	0.453
Yellowish	0.453	0.434

Table 4.1 CIE xy chromaticity of the illuminants used in Experiment II.

4.2.4 Procedure

The experiment consisted of sixteen sessions divided in three groups: *reference*, *regular* and *repeatability tests*. Figure 4.2 shows the time sequence of the experiment. First there was a *training* period followed by the *reference session*, after which the main body of the experiment started. It consisted of nine *regular sessions* and three interleaved repeatability tests (occurring at the beginning, halfway and at the end of the regular sessions) whose aim was to track variations in subject's responses. Subjects completed all experiments in less than three weeks and no more than two sessions per day were allowed. Details of the different sessions were as follows:

Reference session. It consisted of a single session with Type 0 background and D65 illumination and it started just after the training was completed. Subjects were instructed to select the most representative colours for each of the eight basic chromatic categories. The choice of available colours was constrained by the Bounding Cylinder (see squares within the red circle in Figure 4.1).

Regular sessions. They consisted of nine sessions combining the three illuminants and three background Types described before. Each regular session followed a similar protocol as the reference session, except that subjects were instructed to reproduce the same SRs they had selected in the reference session without any constraints (no Bounding Cylinder).

Repeatability test. It consisted of three groups of two sessions each. In the first session, subjects were asked to reproduce the *SR* chosen before, this time under Type 0 background, D65 illuminant and within the Bounding Cylinder. This is equivalent to a Reference session where subjects reproduce instead of selecting the colours. The second session was a regular session with Type II background and *greenish* illumination.

Training. It was done at the very beginning and consisted of repeating two consecutive sessions: a reference session followed by a regular session both with Type 0 background and D65 illuminant (i.e. in the second session there was no Bounding Cylinder). The objective of this was for subjects to understand the different instructions in both cases. Pilot sessions with the authors as subjects, showed that in regular sessions it was possible to reach a precision of 5 ΔE^* at reproducing the same colours after about two sessions and this did not improve significantly afterwards. We used this value as a criterion to determine the end of training.

Panel a of Figure 4.2 shows the common schematics of the reference and regular experimental sessions. Each session started with 120 seconds adaptation to a uniform D65 screen (luminance equal to 30 Cd/m^2) followed by 180 seconds of adaptation to a Mondrian under the same simulated illumination to be used later in session. After that, subjects were prompted auditorily and visually (by a word written in black at the bottom of the screen) to the colour category requested and manipulated the gamepad to either select or reproduce the colours according to their instructions. Each trial ended by pressing a "next trial" button on the gamepad which followed re-adaptation to a geometrically randomized version of the original Mondrian and illuminant for 10 seconds before proceeding to the next trial. There were 44 trials: in the first four subjects were asked to produce "grey" and in the following, they were asked to produce the other eight colours 5 times each in random order. Test patches occurred simultaneously at multiple random locations in the Mondrian and were adjusted by the observer with no time constraints. They were spatially distributed in a random manner in every trial with the aim of forcing subjects to average test locations thus reducing local chromatic induction

effects (Shevell and Wei 2000; Otazu, Parraga et al. 2010). The average number of test patches was randomly determined following a normal distribution around 25 (2.4 SD) for the Type I backgrounds and 4.1 (0.75 SD) for the Type I and II backgrounds and occupied between 4 and 7 percent of the total display area respectively. In the cases where "grey" was requested, we randomized the chromaticity of the initial test patches around the expected value to avoid influencing the subject's response -see Brainard's *basic starting rule* (Brainard 1998). In all other cases, the starting value of the test patches was randomly distributed around each subject's selected "grey". To obtain a single measure of a *SR* colour we averaged its individual trials adjustments. Each trial lasted approximately 30 seconds and each session approximately 25 minutes.
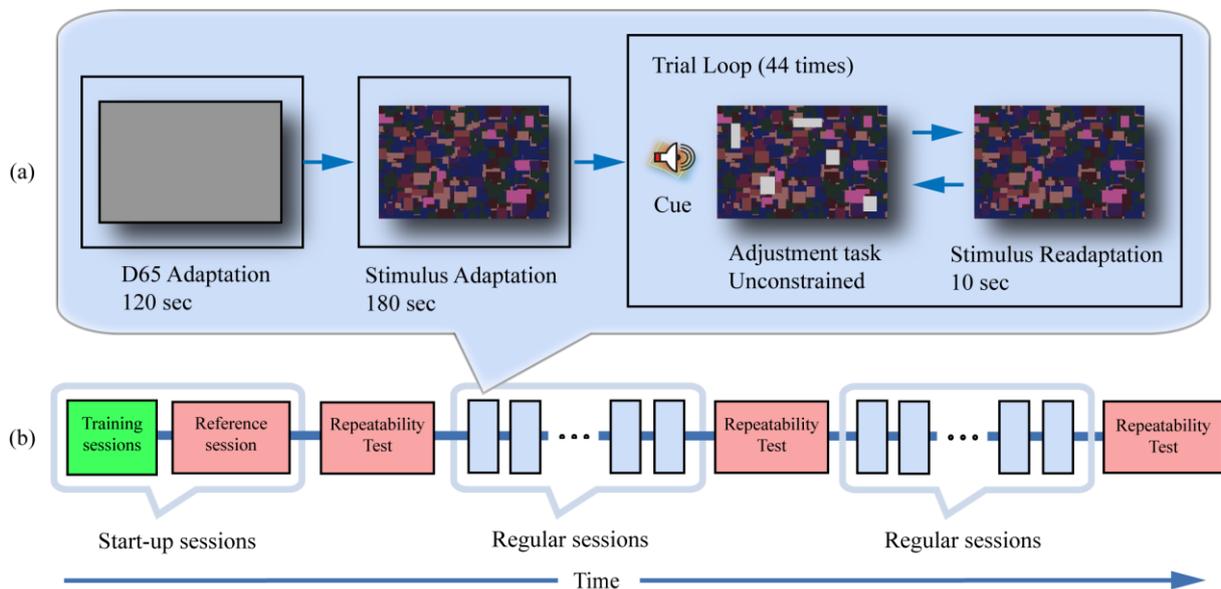
Figure 4.2 Temporal sequence of the chromatic setting paradigm .Panel a shows the common schematics for a reference or regular session. Panel b illustrates the setup of the whole experiment. Start-up sessions consisted in both training and reference sessions. In a reference session, subjects selected their most representative colour for each category. Regular sessions were similar, except that subjects had to reproduce the same colours they had chosen in the reference session. Repeatability Tests were designed to assess subject ability to reproduce the colours selected in the reference session.

4.3 Results

4.3.1 Selected representatives and their repeatability

Figure 4.3 shows the CIELab location of selected representatives chosen by all subjects (D65 was used as a reference white point). Panel a shows the data projection into the lateral surface of the Bounding

Cylinder and panel b shows their projection into the a*b* plane. The limits of the Bounding Cylinder are shown as a red circle in panel b. The coloured areas highlight the inter-subject variability, which is largest in the lightness dimension (Webster and Kay 2007; Foster 2011), particularly for green, blue and purple. From the two panels it can be inferred that there is no volumetric overlap among the different coloured areas, i.e. subjects were consistent in selecting colours within categories. The figure also shows good agreement between the hue locations of our categories and the hues of Boynton and Olson's focals (Boynton and Olson 1987), plotted beyond the cylinder boundaries in panel b.
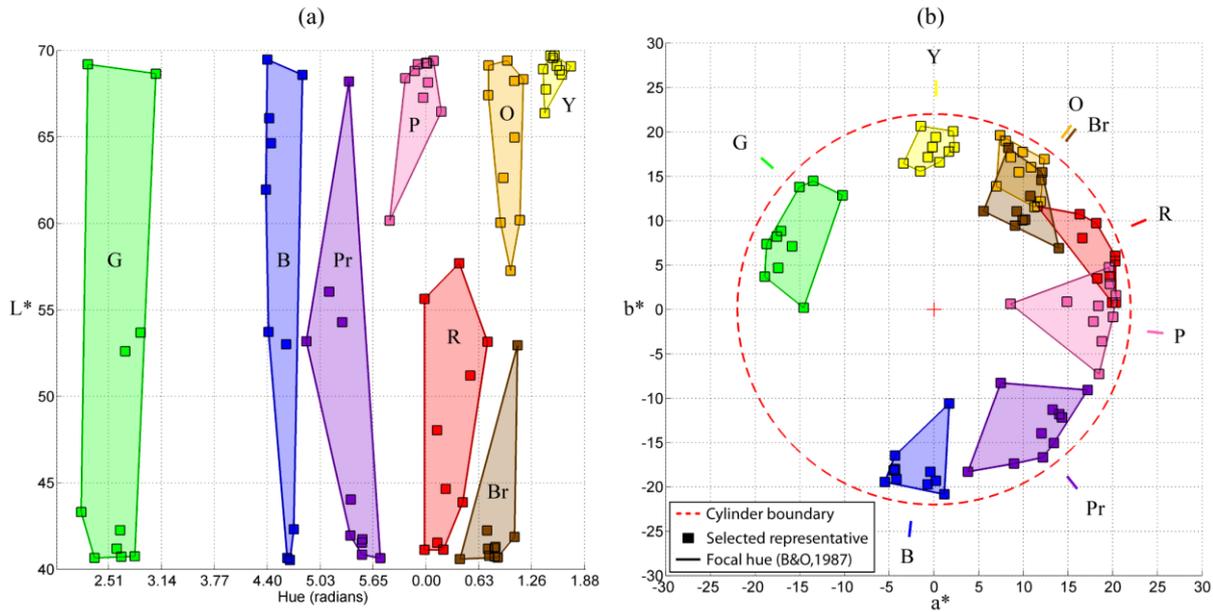
Figure 4.3 CIELab locations of the selected representatives adjusted in the reference sessions by all 10 subjects. Square markers in both panels indicate the average location (5 trials) of each colour category and subject. Colour categories are labeled and colour-coded with their representative colours (R-red; G-green; B-blue; Y-yellow; N-grey; W-white; K-black; P-pink; O-orange; Pr-purple; Br-Brown). Panel a shows the projection of the data in hue and lightness. Panel b shows the same data projected on the a*b* plane. The red circle shows the boundary constraints imposed by the method in the reference sessions.
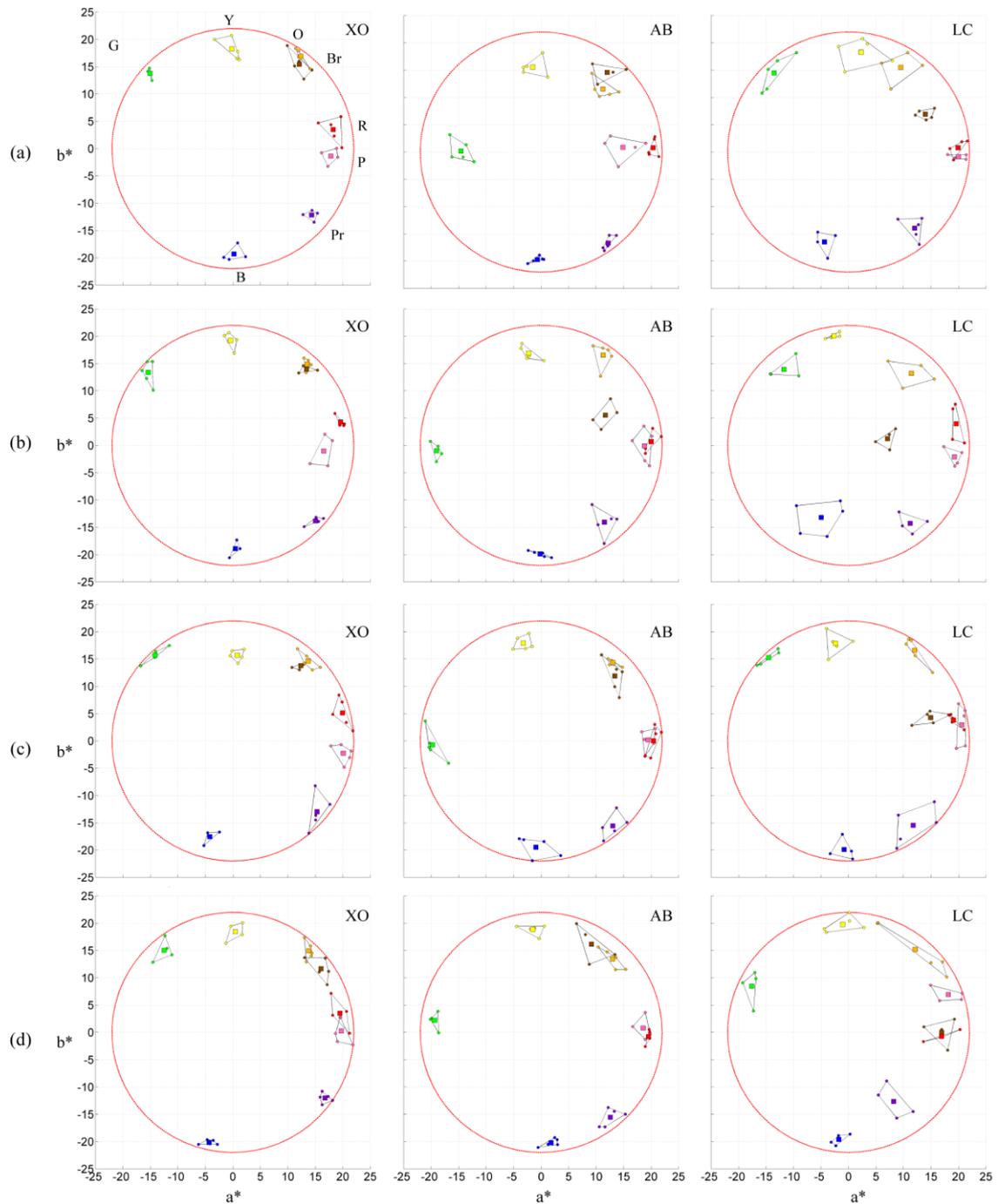
Figure 4.4 Chromatic settings from the reference session and the repeatability sessions. Row a shows the selected representatives chosen by three subjects in the reference session. Rows b, c and d show the corresponding settings for the three subsequent repeatability tests. Square markers represent the average of individual trials (small dots joined by lines) and the large red circle corresponds to the Bounding Cylinder in a*b* chromaticity plane.

Along the experiment we kept track of the accuracy of responses over time by means of the repeatability tests as detailed in the Methods section and Figure 4.2b. These were conducted regularly at approximately three days' intervals and included a reference session where observers were asked to reproduce the original SR colours. Plots in Figure 4.4 were arranged in rows and columns. Columns correspond to two typical subjects (XO and AB) and the most inconsistent subject (LC) over time. Rows correspond to measurements taken over three days' intervals. The first row corresponds to the chromatic settings of the reference sessions and the others (rows b, c and d) correspond to the repeatability tests in temporal sequence. We looked for inconsistencies in the repeatability data by applying a Student's t-test (p<0.05) to the same categories across different rows in each CIELab dimension. Our results show that the means of the results populations considering all subjects were equal in 95% of the cases. Some subjects complained that red and/or orange selections were not saturated enough to be called "representatives". Crucially, this did not seem to impair their capacity to remember the same colour throughout the rest of the experiment even for close categories such as brown and pink.

Repeatability tests also contained a regular session with Type II background and *greenish* illuminant. Figure 4.5 shows a summary of these results. Each panel corresponds to the same observer as before (XO, AB and LC) and each square marker corresponds to a measurement taken over three days' intervals. Notice the data shift corresponding to the change of illuminant. We applied the same approach as before and found that the means of the results populations were equal in 73% of the cases (t-test, p<0.05). This difference is likely to be due to the absence of the Boundary Cylinder which increased uncertainty in the saturation dimension.

Although the repeatability tests show that subjects can reproduce the same SR colours, we conducted another experiment to test longer term colour memory. These results which are consistent with Figure 4.5 are detailed in the Appendix C.

4.3.2 Chromatic settings under different illuminants

Figure 4.6 shows the averaged chromatic settings in CIELab obtained during regular sessions for all subjects, discriminated by backgrounds and separated in panels according to the illuminant. Over the regular sessions, our 10 subjects adjusted 5 times (4 for grey) each of the 9 basic colours for each of the 9 different stimuli, totaling 3960 adjustments. Only 1.4% of these adjustments were closer than 5 CIELab ΔE^* units from the CRT monitor gamut boundary, thus indicating that subjects did not use this boundary as a cue to find their SR colours.
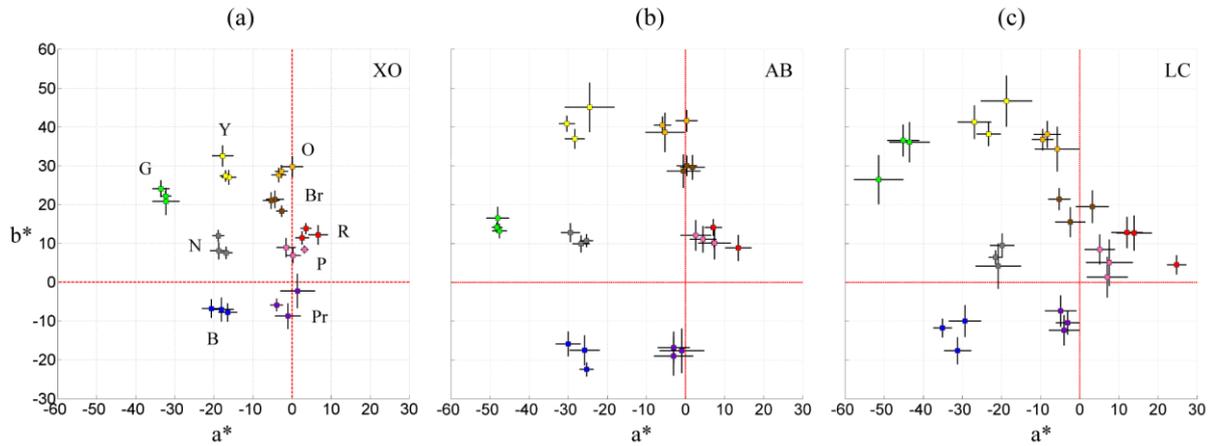
Figure 4.5 Chromatic settings for repeatability sessions. Results include settings for three subjects, Type II background, *greenish* illuminant and no Bounding Cylinder. Each point represents the average of 5 trials (4 for grey) and it was produced in different days over the experiment lifespan. Error bars show the SD. Panels a and b correspond to typical subjects and c shows the subject with the largest variability. Notice the shift of all points towards green, corresponding to the *greenish* illuminant. We chose D65 as a reference white point to highlight the effects of the illuminant for illustrative purposes. Again, for clarity's sake lightness information is not shown.

As before, we plotted these results from different illuminations under the same D65 reference white point in order to highlight the amount of illumination shift, hence the displacement of the data in the plots. Figure 4.6 shows a tendency for subjects to choose more saturated colours in the presence of coloured backgrounds than in the presence of achromatic backgrounds, i.e. squares are closer to the achromatic locus. This is true for all colours studied except for green, yellow and orange. A similar outcome was reported by Brown and McLeod (Brown and MacLeod 1997) in their comparison between the effects of low-contrast and high-contrast multicoloured surrounds. From the same figure we conclude that the type of background did not have a strong influence in the chromatic settings. However, since Type I and II backgrounds were customized for each subject according to their SR, the generalization may be masking individual effects.
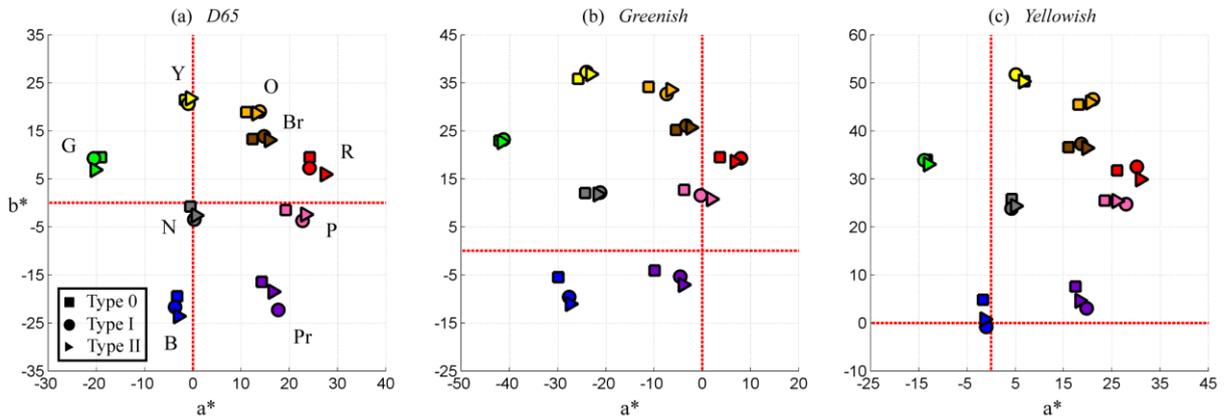
Figure 4.6 Chromatic settings of the selected representatives in regular sessions. The symbols show the chromatic settings for each background type: squares for Type 0, circles for Type I and triangles for Type II. Points were computed by averaging the corresponding SR for all subjects, for each particular background and illuminant. Panel a corresponds to D65, b to *greenish* and c to *yellowish* illumination.

Figure 4.7 shows a set of typical result plots, arranged in columns and rows. Each of the columns corresponds to a different illuminant and the rows to four exemplary subjects, all measured using Type II backgrounds. Inside the plots, each coloured square correspond to the average of 5 trials (4 for grey), which are shown as smaller points joined by lines. To quantify the amount of *variability* (δ) within each group of five trials we computed the average CIELab ΔE^* distance between each SR trial and the mean SR. As a white point for our calculations we used the corresponding chromaticity of each illuminant (see Table 1) at 100 Cd/m^2. Since there were differences in the dispersion of data around the mean depending on each subject and colour category, we summarized δ in Table 4.2 where each value corresponds to the average variability over illuminant-background combinations. The average δ value was 2.09 ΔE^* (1 SD) for the reference sessions and 4.60 ΔE^* (2.06 SD) for regular sessions. The difference between these values is likely to result from the Bounding Cylinder. According to our estimations, the precision of our method is consistent with that of achromatic setting studies (Brainard 1998), where accuracies between 4 and 5 ΔE^* are common.
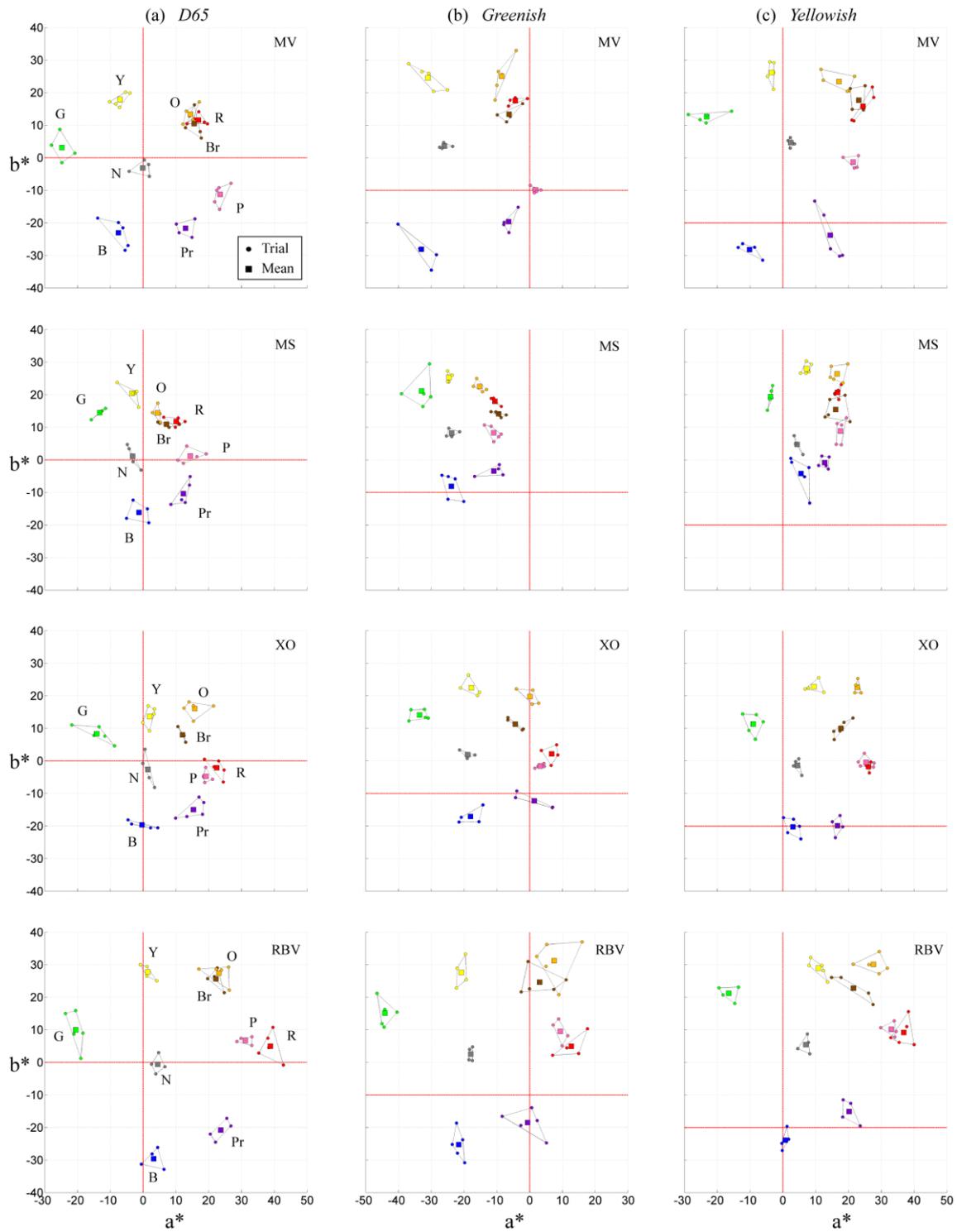
Figure 4.7 Typical chromatic settings from four different subjects for regular sessions. Column a: under *D65* illuminant; column b: under *greenish* illuminant; and column c: under *yellowish* illuminant. The background was Type II in all cases. Individual trials are represented by small dots joined by lines and their average is represented by a colour-coded square.

The results show that mean variability for colour categories is different: red (mean=4.03, 1.25 SD) and grey (mean=4.05, 1.49 SD) have lower variability than purple (mean=5.29, 1.15 SD) and orange (mean=5.02, 1.22 SD). When grouped according to background types no differences were found. The different illuminants affected the variability of our measures: *D65* illuminant has the lowest variability (mean=3.83, 1.52 SD), followed by *greenish* (mean=4.81, 2.08 SD), and *yellowish* (mean=5.16, 2.27 SD) illuminants.

	R	G	B	Y	N	Pr	P	O	Br	Mean
JR	2.14	3.82	3.44	3.42	2.35	3.87	3.26	3.42	2.60	3.18
CAP	4.40	3.56	3.53	3.45	3.81	5.96	4.66	5.75	4.96	4.45
MV	3.51	4.31	5.60	5.10	2.91	6.62	4.66	5.25	4.34	4.70
MS	2.44	4.05	5.82	4.65	3.39	3.78	5.43	3.37	2.78	3.97
XO	2.98	3.87	3.33	3.28	3.00	4.38	3.27	4.02	3.22	3.48
RB	3.65	3.51	3.61	3.42	7.39	4.44	3.62	4.71	6.70	4.56
LC	5.94	6.55	5.06	7.61	4.71	5.76	5.87	4.85	7.11	5.94
AB	4.29	5.66	5.23	4.34	5.09	6.55	5.55	5.21	4.47	5.15
RBV	5.17	5.20	4.93	5.06	3.08	4.87	4.94	6.71	5.14	5.01
JC	5.54	4.45	5.67	4.94	4.81	6.63	6.19	6.92	4.79	5.55
Mean	4.03	4.50	4.62	4.53	4.05	5.29	4.74	5.02	4.61	**4.60**

Table 4.2 Variability (δ) of mean chromatic settings in ΔE* units, averaged over illuminants and backgrounds. The columns show values according to colour category and the rows according to subject. The last column/row shows the means of the rows/columns. The value in bold corresponds to the overall mean.

We recorded the time subjects took to complete each trial. The average was 19.5 (5.7 SD) sec for the reference sessions and 20.7 (6.2 SD) sec for the regular sessions. Also, there were no remarkable time differences in the regular sessions according to illuminants and backgrounds, but there were differences according to colour categories: grey was the longest to adjust (mean=25.1, 7.6 SD), followed by brown (mean=22.1, 5.6 SD) which took longer than blue (mean=18.6, 5.7 SD), purple (mean=18.7, 4.5 SD) and pink (mean=17.9, 4.7 SD). Red (mean=21.1, 7.1 SD) and yellow (mean=20.7, 6.2 SD) took longer time than pink which was the fastest to adjust.

4.3.3 Colour constancy indices

We quantified the extent of colour constancy achieved by our subjects through three colour constancy indices: the *Constancy Index CI* (Arend, Reeves et al. 1991), the *Colour Constancy Index CCI* (Ling and Hurlbert 2008) and the *Brunswick ratio BR* (Yang and Shevell 2002; Smithson and Zaidi 2004;

Hansen, Walter et al. 2007), which takes into account the adaptation under the reference illumination. Equation 4.1 shows an example of how this was implemented for the case of BR.

When considering a particular subject's data, we noted a_c^i as the chromaticity coordinates of his/her selected representative c under illumination i (1 corresponds to *D65*; 2 to *greenish* and 3 to *yellowish*). Also, b_c^i are the chromaticity coordinates of the corresponding a_c^1 when the illuminant i was applied. The numerator computes the perceptual shift, i.e. the difference between SRs chosen under D65 illuminant and *greenish/yellowish* illuminants. The denominator computes physical shift, i.e. the difference between SRs chosen under *D65* and their chromatic coordinates when illuminated by *greenish/yellowish* illuminants. Following this arrangement, a value of 1 indicates perfect colour constancy and 0 no colour constancy.

$$\overline{BR}_c^i = \frac{\left\| a_c^1 - a_c^i \right\|_2}{\left\| b_c^1 - b_c^i \right\|_2} \; where \; i = 2,3 \; and \; c = 1, \dots, 9 \tag{4.1}$$

Although there is no assumption of any specific colour space in the index formulae, we choose CIE1976 uv, a perceptually uniform space which does not incorporate any white point normalization as CIELab does (Wyszecki and Stiles 1982; Brainard 1998). The values in Table 4.3 are the subject-averaged indices considering each colour category and illumination.

Remarkable differences were found in the mean values of the indices when our data was grouped according to colour categories. Table 4.3 shows the differences in the indices according to colour categories and illuminants. We highlighted these differences by showing the maximum and minimum values within each column in bold. We looked for correlations between the different indices when applied to data under the same illuminant and did not find any. In other words, the results obtained by colour categories are heavily dependent on the selected colour constancy index.

Cat/Index	CI		$\overline{BR}$		CCI		Mean
Cat/Illum	G	Y	G	Y	G	Y	
Red	**0.37**	0.63	0.69	0.65	0.82	0.76	0.65
Green	**0.73**	0.68	0.61	0.58	0.89	**0.88**	0.73
Blue	0.53	**0.55**	0.64	0.65	0.68	0.68	0.62
Yellow	0.71	**0.76**	**0.51**	**0.49**	0.72	0.75	0.66
Grey	0.55	0.56	0.62	0.63	**0.61**	**0.62**	0.60
Purple	0.49	0.58	0.72	**0.78**	0.77	0.79	0.69
Pink	0.55	0.64	0.54	0.58	0.64	0.68	0.60
Orange	0.62	0.75	0.53	0.51	0.73	0.76	0.65
Brown	0.50	0.70	**0.75**	0.57	**0.96**	0.82	0.72
Mean	0.56	0.65	0.62	0.60	0.76	0.75	0.66

Table 4.3 Three colour constancy indices applied to our measures and split by colour categories and illuminant type (G-*greenish*; Y-*yellowish*). All indices were computed in the CIE1976 UCS uv uniform colour space and averaged for all subjects and backgrounds. We highlighted in bold the maximum and minimum values in each column, which reveal considerable differences within colour categories.

4.3.4 Linear colour constancy models

As Table 4.3 indicates, the chromatic settings of our subjects were different for different illuminants. We modelled the effects of the illuminant change using linear models of colour constancy, i.e. a linear transformation matrix that relates two chromatic settings of the same colour under different illuminants. To be able to relate the parameters of our models to properties of the human visual system, we chose to operate in LMS cone excitation coordinates (Burnham, Evans et al. 1957; Jameson and Hurvich 1964; Brainard, Brunt et al. 1997), calculated from the Smith and Pokorny cone sensitivity functions (Smith and Pokorny 1975). Equation 4.2 formalizes the previous approach where **x** and **y** are the LMS cone excitations produced by the light reaching the observer from the CRT monitor: **x** corresponds to the reference illuminant (*D65*) and **y** corresponds to the test illuminant (*greenish* or *yellowish*).

$$\mathbf{y} = \mathbf{M}[\ \mathbf{x}\ 1\ ]^{\mathrm{T}}\ \ where\ \mathbf{M} = \begin{pmatrix} m_{1,1} & m_{1,2} & m_{1,3} & m_{1,4} \\ m_{2,1} & m_{2,2} & m_{2,3} & m_{2,4} \\ m_{3,1} & m_{3,2} & m_{3,3} & m_{3,4} \end{pmatrix} \in \mathbb{R}^{3x4} \tag{4.2}$$

The model is represented by the matrix **M** which can take one of several possible forms according to its non zero coefficients. These can also be understood in terms of models of visual mechanisms:

Diagonal (D). The diagonal model ($m_{i,j}=0$ if $i\neq j$) has only 3 free parameters. This model only allows for multiplicative gain changes that are specific to each one of the three cone classes. It is often referred as Von Kries adaptation (Von Kries 1905; Brainard and Wandell 1992).

Linear (L). The linear model ($m_{i,j}=0$ if $j=4$) has 9 free parameters. This model allows signals from each cone type to be modulated independently and can describe multiplicative gain changes both at the receptor level and after an opponent transformation (Brainard and Wandell 1992).

Affine (A). The affine model does not set any initial coefficient to zero and it has 12 free parameters. It contains nested versions of the previous two models. The first three columns of **M** include the linear model and the fourth column represents an additive process. This model can be thought as an instance of the two-process model proposed by Jameson and Hurvich (Jameson and Hurvich 1964; Brainard and Wandell 1992).

Diagonal plus Translation (DT). The diagonal plus translation model ($m_{i,j}=0$ if $i\neq j$ and $j<4$), has 6 free parameters and can be seen as a simplification of the affine model. The first three columns allow only for multiplicative gains for each cone class and the last column allows a further additive process.

We studied the predictive power of each model when multiple chromatic settings were used as data points. Equation 4.3 generalizes Equation 4.2 into a single system of linear equations when using more than one data point. This formulation allows using standard multiple linear regression methods to fit the model parameters, i.e., to minimize the mean-square difference between the measured and the predicted points. In Equation 4.3, the matrix **X** contains the LMS coordinates of *n* colours $\mathbf{x}_i$ under reference illuminant and matrix **Y** contains the settings of those same colours, $\mathbf{y}_i$, under test illuminant.

$$\mathbf{Y} = \mathbf{MX} \ \ where \ \mathbf{Y} = (\mathbf{y}_1| \cdots |\mathbf{y}_n) \in \mathbb{R}^{3xn} \ \ and \ \ \mathbf{X} = ([\mathbf{x}_1 \ 1]' | \cdots | [\mathbf{x}_n \ 1]') \in \mathbb{R}^{4xn} \tag{4.3}$$

Equation 4.4 describes $\mathcal{H}$, which contains all possible subsets of nine colours and their combinations according to their indices (1 for green, 2 for blue, 3 for yellow, etc.). Once a particular element of $\mathcal{H}$ was selected we could fit the model parameters to this element as described in Equation 4.5, substitute their LMS coordinates and solve the linear system using least squares. However, since LMS is not perceptually uniform, we chose to follow the approach described by Brainard et al (Brainard and Wandell 1992; Brainard, Brunt et al. 1997). They solved the linear system through a minimization process which determined the model parameters according to the mean CIELab ΔE^* colour difference between the N predictions and the data points. The function to minimize is described by Equation 4.6,

73

where φ is an operator that translates from LMS to CIELab coordinates. F_N was minimized using the Matlab Optimization Toolbox. Model precision was evaluated by computing the average ΔE^* difference between the whole set of nine chromatic settings and their predictions computed from the matrix **M**.

$$\mathcal{H} = \left\{ (k_1, \ldots, k_N) \, ; \, k_i \in \{1, \ldots, 9\}, k_i \neq k_j, N = 1, \ldots, 9 \ and \ i, j = 1, \ldots, N \right\} \tag{4.4}$$

$$\left(\mathbf{y}_{k_1} | \cdots | \mathbf{y}_{k_N} \right) = \mathbf{M}\left([\mathbf{x}_{k_1} \ 1]' \ | \ \cdots | \ [\mathbf{x}_{k_N} \ 1]' \ \right) \ where \ (k_1, \ldots, k_N) \in \mathcal{H} \tag{4.5}$$

$$F_N\left(\mathbf{M}, (k_1, \ldots, k_N) \right) = \frac{1}{N} \sum_{i=1}^{N} \left\| \varphi\left(\mathbf{M}[\mathbf{x}_{k_i} \ 1]' \right) - \varphi\left(\mathbf{y}_{k_i} \right) \right\|_2 \ \ where \ \varphi(LMS) \in CIELab \tag{4.6}$$

We considered all possible combinations of SRs, within the limits imposed by each model. For example, when fitting the linear system in Equation 4.5, the minimum number of points that the model can fit is determined by the number of free parameters contained in the model. This terminology is equivalent to a system of linear equations where there are larger, fewer or equal number of equations than unknowns. The underdetermined case occurs when the number of unknowns is larger than the number of the equations (the system is underconstrained). From this follows that the diagonal model admits any number of data points $N \geq 1$, diagonal plus translation admits $N \geq 2$ data points, linear admits $N \geq 3$ data points, and affine $N \geq 4$ data points. This is also valid for Equation 4.6.

Figure 4.8 summarizes our modeling results as described above. Panel a corresponds to *greenish* illuminant and panel b to *yellowish*. The y-axis shows the prediction error (in ΔE^* units) associated with each model as a function of the number of chromatic settings used to fit it. Following the approach of Brainard (Brainard, Brunt et al. 1997) we used the chromaticity coordinates of the corresponding illuminant as a reference white point in each case. The function specified in Equation 4.6 was minimized to fit chromatic settings **x** (corresponding to D65) and **y** (corresponding to *greenish* or *yellowish* illuminants) keeping the same background type. Take for instance panel a in Figure 4.8, where each point is the average model prediction error from all possible combinations of elements of $\mathcal{H}$ that contain the number of colours specified in the x-axis, across backgrounds and subjects. Consider the case when the nine SRs were measured both under D65 and *greenish* illumination using the same background type. We fitted the diagonal model to only one correspondence pair from the nine chromatic settings available, and used the same parameters to predict the positions of all nine

corresponding pairs. We repeated this for all the other pairs and calculated the average CIELab ΔE* distance between predicted and measured points for the nine chromatic settings pairs. We extended this to all subjects and backgrounds. The result of these calculations (average from 270 model predictions) is shown in panel a as the leftmost filled circle in the plot. To calculate the second leftmost circle in the plot, we fitted the diagonal model to two correspondence pairs from the nine chromatic settings available and predicted the positions of all nine pairs (36 possible combinations). This point represents the average across subjects and backgrounds (1080 model predictions). The other circles were calculated similarly, by fitting the diagonal model to increasingly more data points. The same reasoning was applied to the other models, shown as triangles and squares in Figure 4.8. Since the results of the minimization process in Equation 4.6 depend on the initial seed, we used 100 random seeds (for larger values results tend to stabilize) and the solution to the linear system specified by Equation 4.5 (Brainard and Wandell 1992) as a complementary seed. We selected the minimum optimization value all seeds.
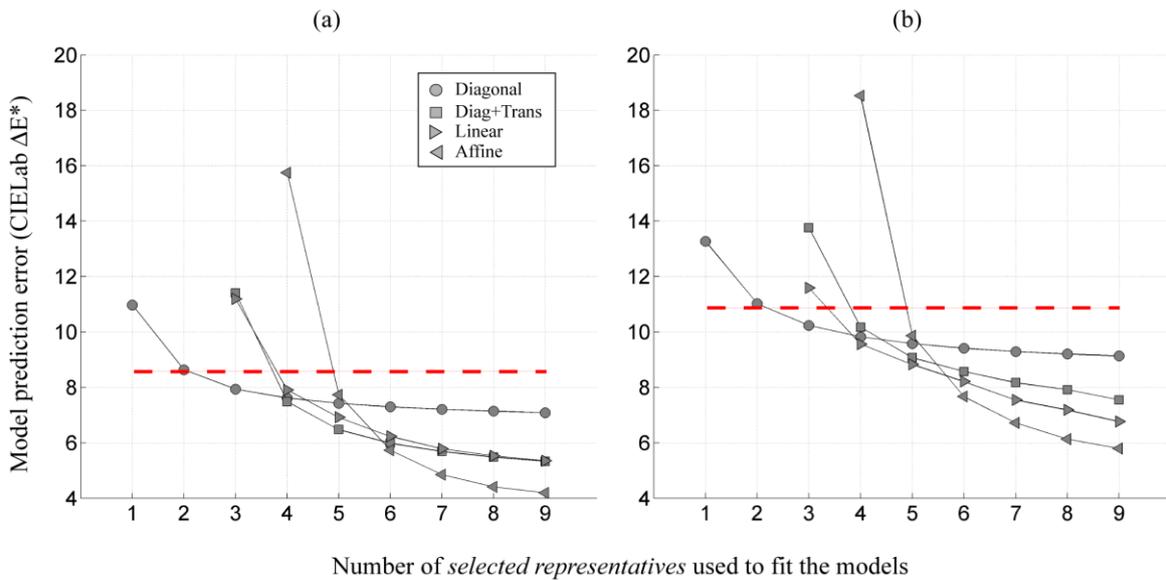
Figure 4.8 Model prediction error according to the number of colours used to estimate their parameters. Panel a corresponds to *greenish* test illuminant and panel b to *yellowish*. Each point corresponds to a particular model (circles for the Diagonal, squares for the Diagonal plus Translation, right-pointing triangles for the Linear and left-pointing triangles for the Affine), computed from all background types and subjects. For comparison we show the prediction error of Von Kries transformation applied to the achromatic locus as a horizontal red broken line. The values were calculated using the corresponding reference white point for each illuminant (*greenish* and *yellowish* -see Table 1).

Predictably, Figure 4.8 shows that adding more data points and increasing the number of free parameters lowers the model prediction error exponentially: the more free parameters a model has, the

more accentuated the decay is. For instance, the Diagonal model (circle symbols) improves less, from 10.9 to 7.1 ΔE* for the *greenish* and from 13.3 to 9.1 ΔE* for the *yellowish* as we add more fitting points. When the maximum number of fitting points (9) are used, the errors in ΔE* are: 7.09 (D), 5.33 (DT), 5.35 (L) and 4.19 (A) for the *greenish* illuminant and 9.13 (D), 7.55 (DT), 6.76 (L) and 5.79 (A) for the *yellowish* illuminant (see Figure 4.9). In general, model errors under *greenish* illuminant are lower than model errors under *yellowish* illuminant. Simpler models tend to perform better with a small number of fitting points whereas more complex models tend to perform better with larger numbers of fitting points. For instance the Linear and Affine models start to perform better than the simpler Diagonal when more than 5 points are considered. There are also quantitative differences regarding the illuminant: for up to five fitting points, error values are between 4 and 7.5 ΔE* for *greenish* and between 6 and 9.6 ΔE* for the *yellowish*.
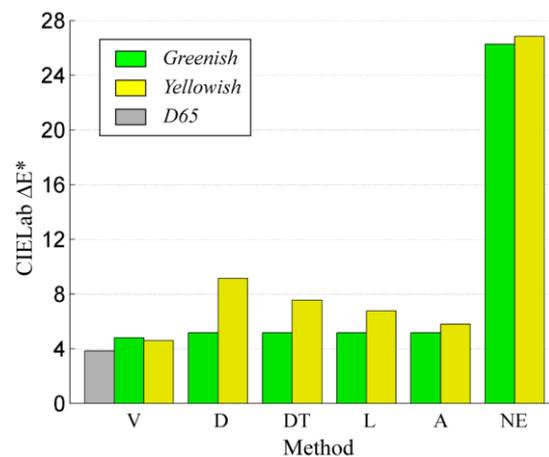
Figure 4.9 Models' prediction error when all nine SRs points were included. The first column (V) contains the subject average variability in the trials (see Table 2) and the last column (No Effect – NE) is a quantitative measure of the illuminant shift computed without any predictive model. The groups of bars labeled as D, DT, L and A correspond each to the Diagonal, Diagonal plus Translation, Linear and Affine models respectively.

4.4 Discussion

Our previous results show the feasibility of using several colours rather than a single colour as a metric for assessing the stability of colour appearance under a change of illumination. In the following session we discuss the usefulness of this new metric, showing that linear colour constancy models satisfactorily explain the transformations with a larger number of colours. At the end of the section we introduce a new colour constancy index that takes into account several aspects of colour constancy not considered before.

4.4.1 Does include more colours increase the precision of models?

Both graphs in Figure 4.8 illustrate clearly how the predictive power of all models is increased by adding more fitting points, something that is in agreement with previous studies (Hansen, Walter et al. 2007; Olkkonen, Hansen et al. 2009; Olkkonen, Witzel et al. 2010). However, the error curves tend to a constant value after eight SRs and this suggests that measuring more points would lead to minimal improvements. In this context, it is worth noticing that our current fitting points were not determined randomly but had a particularly even distribution over the colour space, thus our conclusions become more relevant when all nine fitting points are used. This highlights the advantage of measuring several colours instead of just grey and although it disagrees with previous results (Speigle and Brainard 1999), we believe it is unlikely to be the product of experimental artefacts. Figure 4.9 shows the portion of the phenomenon that is captured by the models. The large differences in height between the bar labelled as "No-Effect" (which summarizes the effects of the illumination) and the other bars suggest that all linear models succeed in modelling the phenomenon (Brainard, Brunt et al. 1997; Brainard 1998). However there is still a small part which is not captured by the models.

We tested the parsimony of the models to see whether they include more parameters than it is necessary by applying the *Akaike Information Criterion* (Burnham and Anderson 2002). This criterion measures the relative goodness of fit of a model in terms of the information lost when it is used to describe data (see Appendix D). The results show that the best models in Figure 4.8 are the simplest: Diagonal and Diagonal plus Translation, implying that the Linear and the Affine models are possibly over-fitting the data. The results also show a clear tendency for the Diagonal plus Translation to become the best in terms of number of free parameters and prediction error as we add more data points.

If we ignore the Linear and Affine models, in Figure 4.8 there are some common qualitative features for both illuminants that are worth mentioning:

Stability point at 5 SRs. All models approximately have the same precision when five SRs are used for the fit, i.e. three free parameters achieve similar results as twelve. This might reflect the fact that considering less than five points in our calculations allows for distributions of colours that are not symmetric with respect to the centre, something that is less likely when more colours are considered. Furthermore, models with more free parameters are more sensitive to these asymmetries.

Diagonal outperforms the Diagonal plus Translation before the stability point. This suggests a link between the number of colours available and the complexity of the colour constancy mechanism needed: in a simpler environment, a cone gain-based transformation outperforms the others.

Diagonal plus Translation outperforms the Diagonal after the stability point. This represents an improvement from the Diagonal model, and suggests the involvement of the additive process in a two-stage mechanism as proposed by Jameson and Hurvich (Jameson and Hurvich 1955).

Interestingly, the modelling of the chromatic settings performed under the *greenish* illuminant is better than under the *yellowish* one, and this effect is general to all models and fitting point numbers. This fact suggests a higher degree of dispersion in the chromatic settings which may result from the split of the resulting colours into several categories when illuminated by the *yellowish* illuminant, something that did not occur under the *greenish* illuminant (see further discussion below).

4.4.2 Further insights into the role of colour categories

The overall pattern of results shown in the previous sections is broadly uniform across colour categories, but some particularities exist. For example, we expected the behaviour of grey (the colour measured in achromatic settings) to be outstanding in terms of variability (δ), adjustment time and constancy index values and to summarize the behaviour of the whole set of chromatic settings. Interestingly, we have found that subject's ability to adjust grey and red are similar, closely followed by many other categories. Also, grey is the colour that takes longer time to adjust, may be because subjects can discriminate more finely near the achromatic locus (Boynton and Olson 1987). Furthermore, we expected colour constancy indices values for grey to be near the average and Table 4.3 shows that they are generally low and in the case of the CCI index, the lowest. Previous work found higher colour constancy for grey than for chromatic stimuli (Speigle and Brainard 1999; Olkkonen, Witzel et al. 2010) which is perhaps due to the fact that we used simulated surfaces and illuminants instead of real surfaces. We also found high colour constancy index values (0.66 in average), which is in accordance to similar studies (Murray, Daugirdiene et al. 2006; Hansen, Walter et al. 2007; Ling and Hurlbert 2008; Olkkonen, Hansen et al. 2009; Olkkonen, Witzel et al. 2010; Foster 2011), a fact that is supported by visual inspection of the plots in Figure 4.6, where inter-distances among measured colours are largely preserved. This supports the finding that the categorical structure of colour space is largely preserved under illuminant changes (Hansen, Walter et al. 2007; Olkkonen, Hansen et al. 2009; Olkkonen, Witzel et al. 2010).

The differences in colour constancy values found for different categories suggest different properties for different categorical colours. These properties could be determined by experimenting with other stimulus configurations or subjects' tasks. However, no significant differences were found for background types, a result which is similar to others (Brainard 1998).

4.4.3 SCI: a new Structural Colour Constancy Index

Colour constancy indices attempt to capture the extent of the phenomenon's effect in a single number. They relate perceptual data measured under a state of adaptation to the corresponding data predicted for "perfect" adaptation (i.e. physical colour shift). The simplest indices quantify Euclidean distances (*magnitude*) among the colours of the test surface, the ideal match and the observer match. Examples of these are the *Constancy Index (CI)* (Arend, Reeves et al. 1991), the *Brunswik Ratio (BR)* (Troost and de Weert 1991) and the *BRϕ* which incorporates the direction (*orientation*) between the perceptual and physical colour shifts (Foster 2011). Several improvements have been suggested. For instance, Ling and Hurlbert (Ling and Hurlbert 2008) proposed a new index *CCI* that incorporates the matching error in the absence of illumination change (*memory* shift) and Brainard (Brainard 1998) proposed to use the *Equivalent Illuminant* (*EI*) instead of the measured adaptation point, which is calculated from different measured points and thus captures the inter-distances among the colours considered under a given adaptation state (*structural*).

Property/Index	CI,BR,$\overline{BR}$	EI	CCI	SCI
Magnitude	Yes	Yes	Yes	Yes
Orientation	No	No	Yes	Yes
Memory	No	Yes	Yes	Yes
Structure	No	Yes	No	Yes

Table 4.4 Summary of some properties of colour constancy incorporated into each index.

Following the previous discussion, we introduced a new colour constancy index, termed S*tructural Constancy Index (SCI)* which captures all the features stated in Table 4.4. The new index is defined in terms of matrix norms, which are extensions of the notion of vector norms applied to matrices. As Equation 4.7 shows, the norm of a matrix **A** is obtained from the norm of vectors **x** and **Ax** and describes the maximum relative vector magnitude change under the linear transformation **A**.

$$\|\mathbf{A}\|_2 = \sup_{\mathbf{x} \neq 0} \frac{\|\mathbf{Ax}\|_2}{\|\mathbf{x}\|_2} = \max_{\|\mathbf{x}\|_2 = 1} \|\mathbf{Ax}\|_2 \tag{4.7}$$

In our context, the matrix **A** models the effects of the illuminant change, i.e., given the coordinates **x** of a colour sample under the reference illuminant, it returns the coordinates **Ax** of the same sample under the test illuminant in a given colour space. We define SCI as:

$$SCI(\mathbf{A}_{\text{percep}}, \mathbf{A}_{\text{phys}}) = \frac{\|\mathbf{A}_{\text{percep}}\|_2}{\|\mathbf{A}_{\text{phys}}\|_2} \cdot cos\big(angle(\mathbf{r},\mathbf{s})\big) \frac{\|\mathbf{A}_{\text{percep}}\|_2}{\|\mathbf{A}_{\text{phys}}\|_2} \cdot \frac{\mathbf{r}\mathbf{s}}{\|\mathbf{r}\|_2\|\mathbf{s}\|_2} \tag{4.8}$$

In Equation 4.8 SCI is defined as the product of two factors. The first factor is the quotient of two matrix norms, and computes the relative magnitudes of the perceptual and physical effects of the illuminant, as is commonly the case with constancy indices (Arend, Reeves et al. 1991; Yang and Shevell 2002; Smithson and Zaidi 2004; Hansen, Walter et al. 2007; Ling and Hurlbert 2008; Foster 2011). The second factor estimates how much the direction of the adaptation coincides with the direction of the actual illuminant change in the colour space considered. To compute this we need $\mathbf{A}_{\text{percep}}$ and $\mathbf{A}_{\text{phys}}$ to be affine matrices, i.e. to include vectors $\mathbf{r}$ and $\mathbf{s}$ in the last columns specifying a translation each.

The coefficients of $\mathbf{A}_{\text{percep}}$ are determined from pairs of corresponding chromatic settings under reference and test illuminants and can be obtained following the approach described in the modeling subsection above (Equation 4.3). Likewise, the coefficients of $\mathbf{A}_{\text{phys}}$ are determined from correspondences between the chromatic settings made under the reference illuminant and simulations of the same colours under a test illuminant. In this formulation, if matrices $\mathbf{A}_{\text{percep}}$ and $\mathbf{A}_{\text{phys}}$ are equal, then colour constancy is perfect. Finally, memory effects like those discussed by Ling and Hurlbert (Ling and Hurlbert 2008) are neutralized since our measurements were obtained from direct comparisons under reference and test illuminants.

Figure 4.10 illustrates the behavior of Equation 4.8 for several hypothetical cases. Panel a describes how the magnitude size of each transformation contributes to the value of the SCI. This contribution is always positive and can be smaller or larger than one according to the ratio between the norms of the $\mathbf{A}_{\text{percep}}$ and $\mathbf{A}_{\text{phys}}$. The latter case happens when observers correct for the illuminant more than they should. Panel b describes the contribution of the second term of Equation 4.8, i.e. a weighting factor to penalize for angular deviations from the direction of the simulated illuminant shift. As $\mathbf{r}$ and $\mathbf{s}$ become more perpendicular, their product $\mathbf{r}\mathbf{s}$ becomes closer to zero. Although negative values are possible in theory, in practice this weighting factor should be positive assuming that $\mathbf{r}$ and $\mathbf{s}$ are far from perpendicular.
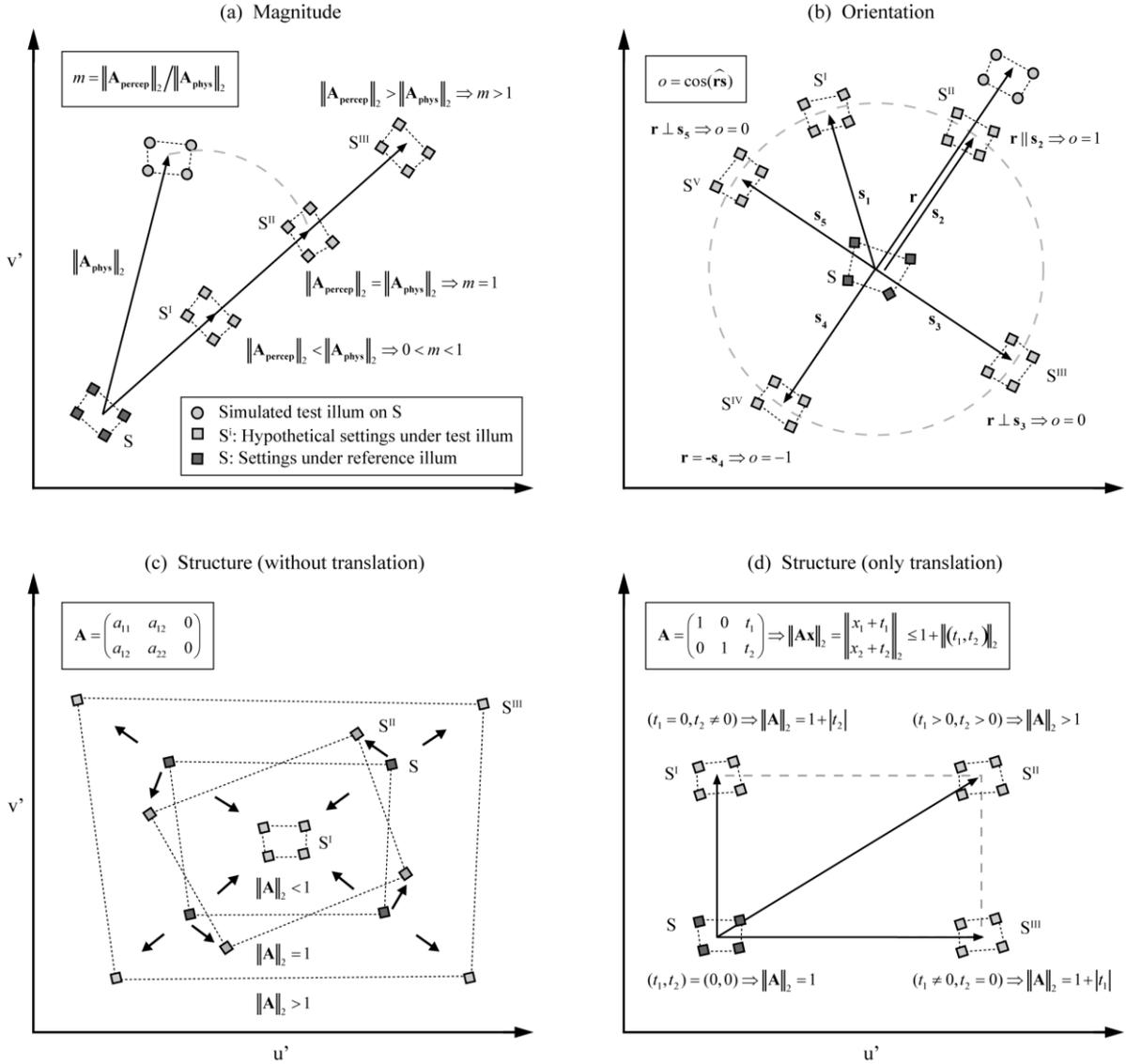
Figure 4.10 Hypothetical cases of chromatic settings and their contribution to SCI values in the CIE1976 uv colour space. Each panel illustrates the contribution of a particular feature of our index. Dark squares correspond to chromatic settings made under the reference illuminant, light squares correspond to hypothetical chromatic settings made under test illuminant and circles correspond to a simulated illumination of the chromatic settings made under the reference illuminant (dark squares). Panel a: effects of a shift in magnitude only with respect of a simulated illumination. Panel b effects of a change in the orientation from the simulated illuminant shift. Panels c and d effects of an expansion/contraction and a translation are captured and converted into a single number by the affine matrix norm.

Structural information of the colour constancy phenomenon is implicitly embedded in the affine matrix. Panels c and d illustrate how this information is summarized into a single positive number. Panel c illustrates the case when there is no translation (i.e. the last column of the affine matrix is null) and the matrix can be interpreted in terms of expansion ($\|\mathbf{A}\|_2 > 1$), retraction ($\|\mathbf{A}\|_2 < 1$) or rotation ($\|\mathbf{A}\|_2 = 1$). Panel d illustrates the case when only the translation part is operative and the value of the

norm reflects this translation. In theory, the SCI can assume values that are larger than 1 or negative, representing overcompensation or failures of colour constancy that may happen under certain illumination conditions such as multiple illuminants, non Lambertian surfaces, self-luminous or fluorescent materials, etc., that imply a violation of the initial conditions of this analysis.

Table 4.5 shows the average values obtained from applying four colour constancy indices ($\overline{BR}$, EI, CCI and SCI) to all subjects and background types, discriminated according to illumination. All indices were computed in the CIE1976 uv colour space. There was no effect of background types in the calculations. Interestingly not all indices gave the same values; EI and $\overline{BR}$ were generally lower and SCI was the highest. The differences between popular indices such as $\overline{BR}$ and CCI were reported by Ling and Hurlbert and attributed to the incorporation of memory shift into the index formula (Ling and Hurlbert 2008). SCI values are slightly higher than CCI values, and in the case of *greenish* illuminant larger than one. This fact is due to the incorporation of "structural" components, i.e. measures of the inter-distances among data points into the index calculation (see panel c in Figure 10), which can increase the total index value in some cases. We calculated the contribution of the different components in Figure 10 to the SCI values in Table 5 and found that, for *greenish* illuminant, the norm of $\mathbf{A_{percep}}$ is slightly larger than the norm of $\mathbf{A_{phys}}$ making the first term of Equation 8 slightly larger than one. The previous analysis implies that perfect colour constancy is achieved when SCI is equal to one and different values indicate either lack of constancy (SCI < 1) or overcompensation (SCI > 1). In our case, we expected indices values close to one due to the large adaptation period of immersive illumination.

Index/Illum	*Greenish*	*Yellowish*
$\overline{BR}$	0.62	0.61
EI	0.58	0.59
CCI	0.76	0.75
SCI	1.03	0.85

Table 4.5 The *Structural Constancy Index (SCI)* and other typical colour constancy indices computed in the CIE1976 uv. Each value corresponds to the average over subjects and background types, also for the CCI and EI averaged over colour categories.

We tested whether the high indices values we found in Table 4.5 were due to the fact that observers had the chance to see the Type I background colours (i.e., the colours to be adjusted) often; and hence subjects performed matches to the displayed colours instead of reproducing them from their memory. This was done by repeating the experiment with two new subjects using only Type II background, i.e.

they had not seen the Type I backgrounds colours before. Their results were in agreement with those of the rest of the subjects and indeed their colour constancy indices were not lower than those of Table 4.5.

Table 4.5 reveals that only SCI differentiates between the *greenish* and the *yellowish* illuminants. Further inspection of the *magnitude* and *orientation* contributions revealed that these differences originated in the norm of the perceptual matrix. In the previous modeling subsection, we found lower prediction errors for the *greenish* illuminant (see Figure 4.8), indicating that such data is better captured by the fitting of linear models, a process similar to the computation of SCI values. This explains why chromatic settings under *yellowish* illuminant have a higher degree of dispersion when compared to chromatic settings under D65 than in the *greenish* case. These differences manifest in Figure 4.6 as subtle variations in the location of the yellow, orange, brown, red and pink data points, which may account for the 18% difference between both illuminants in Table 4.5. We could hypothesize about the origin of this dispersion and say that *greenish*-illuminated colours fall inside the broad green category, whereas *yellowish*-illuminated colours fall into several categories and this initial (first milliseconds) categorical perception may influence the subject's adaptation and subsequent chromatic settings. However, this need to be settled by doing more experiments in the future.

4.4.4 Comparison to previous paradigms

Our contribution is complementary to the work of others who also studied successive colour constancy (Foster 2011) under large periods of immersive illumination and using simulated (Hansen, Walter et al. 2007; Olkkonen, Hansen et al. 2009) or real (Olkkonen, Witzel et al. 2010) surfaces. These studies categorized a large number of coloured samples with higher results variance, while we measured only nine relevant points with relatively higher precision. Hansen *et al* (Hansen, Walter et al. 2007) measured changes in the categorical boundaries of the colour space while Olkkonen et al (Olkkonen, Hansen et al. 2009; Olkkonen, Witzel et al. 2010) used a conventional constancy index (including shift magnitude and orientation) applied to the categorical prototypes. Our results qualitatively agree with their findings regarding the stability of the categorical structure of colour space under illuminant changes (Hansen, Walter et al. 2007; Olkkonen, Hansen et al. 2009; Olkkonen, Witzel et al. 2010).

The chromatic setting paradigm was primarily designed to deal with two main issues: (i) the state of adaptation closely following the change of illuminant (Foster 2011) and (ii) the effects of instructions regarding the nature of the stimuli (surface-match or colour-match criteria) (Troost and de Weert 1991). For this reason, it makes use of subject's colour naming abilities, asking them to select their own colours instead of reproducing arbitrary ones, thus improving on the chromatic resolution limits

of standard colour naming techniques (Foster 2011). The main disadvantage of the method is arguably the saturation restriction to the colours that subjects can initially select imposed by the CRT monitor gamut limitations. However low-saturation SRs were not particularly difficult to reproduce in regular sessions.

Possible chromatic induction (Shevell and Wei 2000) effects resulting from the local influence of neighbouring patches were avoided by embedding the multiple test patch within the Mondrian, randomizing its spatial and chromatic structure from trial to trial (while keeping its global statistics constant prior to illumination). In this manner, subjects have to look at several places and average the test patch colour before making a decision. Also, general memory effects (Ling and Hurlbert 2008) were isolated from constancy effects by analyzing memory matches with and without the illuminant change.

4.4.5 Are some subsets of colours more informative?

The results shown in Figure 4.8 pose another interesting question: is there a subset of *selected representatives* (red, pink, purple, blue, green, yellow, orange, brown and grey) that conveys more information about the colour constancy phenomena than the rest? Figure 4.11 shows the average error (over all subjects, illuminants and backgrounds) derived from adjusting the simplest model (diagonal) to each of the measured colours and testing with the rest. To construct the Figure 4.12 we analyzed all possible combinations among the SR. This approach discarded quantitative information in favour of qualitative information, i.e., we ranked the colours according to their contribution to the model's error. Panel a of Figure 4.12 shows this performance ranking (larger contribution to error to the right) for the diagonal model under *greenish* illumination. The bottom row illustrates the ranking of the colour points shown in Figure 4.11 when sorted according to the model's error. To obtain the row corresponding to 2 colours (second from the bottom) we considered all possible combinations pairs and their corresponding errors, and calculated the average error from all pairs in which a given colour was present. Performance order was obtained from ranking all errors. This was replicated in all other rows to obtain all other combinations (three colours, four colours and so on). The bottom row in Panel b of Figure 4.12 summarizes the information contained in Panel a. Subsequent rows summarize similar analyses for other combinations of model and illuminant. Notice that as we increase the number of data points (and the number of free parameters) vertically upwards, quantitative differences become less significant.
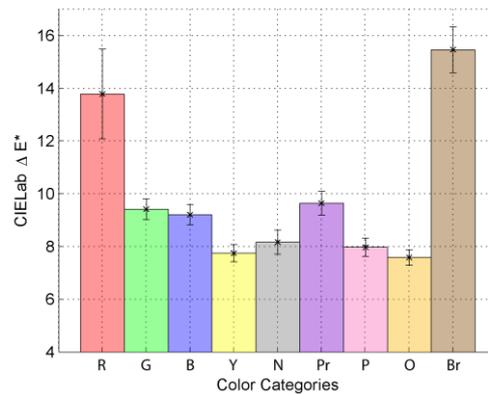
Figure 4.11 The Diagonal model's error for each *selected representative* used to set model's parameters. Results are clustered around two groups, the red-brown with higher error values and the rest.

From the previous approach we can extract some qualitative conclusions in regard of the number of points, model or illumination studied. From Panel a of Figure 4.12 it is possible to see that the performance rank of some colours remains stable when the number of data points grows. This is true for brown, red, purple and green. And the best colours (in the sense of introducing less error to the models) in most cases are yellow, orange and grey. Panel b shows some general trends such as the improvement or decline of performance of a given colour when more free parameters are added to the model. For example, green jumps from middle to first position and red from the last position to mid-rank as we increase the number of parameters. Purple, blue, pink, orange and yellow keep in roughly the same position (regardless of the model used) and grey clearly declines as we climb from the diagonal towards the affine model, possibly due to geometrical rotation and scaling. Brown is a special case, as it moves from the worst position in the diagonal and linear models to the best position in the diagonal plus translation and affine models. When we look at differences between illumination conditions, changes in ranking position become more frequent as we increase the model complexity.

From this we can conclude that grey is not necessary the most informative colour (for example, yellow could be used equally well as a predictor given that its ranking is quite stable across all conditions). For the same reasons purple should be avoided. A second conclusion is that performance rankings vary according to the model but keeps roughly stable across illuminant conditions. Regarding the inter-distance of the data points in colour space, model precision increases for colours that are further apart among themselves in CIELab (as expected).
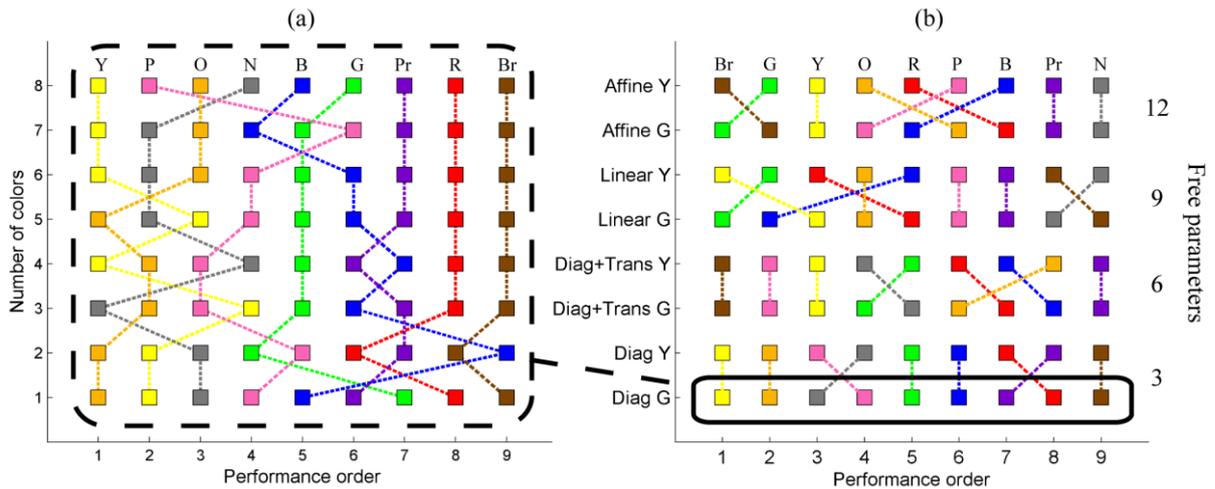
Figure 4.12 Selected representative ranking. The horizontal axis shows a ranking of the colours according to their contribution to the total error (largest contribution to the right). Panel a shows results under *greenish* illumination (all other variables were averaged). Panel b shows the results for all models combined discriminated by the type of illumination (*greenish* (G) or *yellowish* (Y)) and averaged in terms of subject and background. The bottom line in panel b corresponds to the model detailed in panel a (Diagonal model).

4.5 Conclusions

Colour constancy is usually measured by achromatic setting, asymmetric matching or colour naming paradigms, whose results are interpreted in terms of indexes and models which arguably do not capture the full complexity of the phenomenon. Here we propose a new paradigm, *Chromatic Setting*, which allows a more comprehensive characterization of colour constancy through the measurement of multiple points in colour space under immersive adaptation. We demonstrated its feasibility by assessing the consistency of subject's responses over time. The paradigm was applied to 2D Mondrian stimuli under three different illuminants, and the results were used to fit a set of linear colour constancy models. The use of multiple colours improved the precision of more complex linear models compared to the popular diagonal model computed from grey. Our results show that a diagonal plus translation matrix which models mechanisms other than cone gain might be best suited to explain the phenomenon. Additionally, we calculated a number of colour constancy indices for several points in colour space and our results suggest that interrelations among colours are not as uniform as previously believed. To account for this variability, we developed a new structural colour constancy index which takes into account the magnitude and orientation of the chromatic shift in addition to the interrelations among colours and memory effects. Our results do not show any quantitative difference regarding the types of coloured background tested.

Chapter 5 Constancy in categorical colour perception

The previous chapter introduced a new psychophysical paradigm to measure successive colour constancy. The paradigm, called chromatic setting, measures the location of nine categorically relevant points in colour space once the observer is adapted to particular conditions of illumination and background, therefore allowing us to make use of more sophisticated techniques for describing the adaptation. This chapter introduces a new technique to model the structure of chromatic settings which provides a more precise measure of the interrelations among these settings when illumination is changed, and thus to draw conclusions about the stability of categorical colour perception. This approach is based on defining a *graph* and a *graph distance* (Gross and Yellen 2004) to capture the structure of chromatic settings and its variations over different stimulus conditions. Also, we performed a new experiment which tested two additional illuminants to the ones tested in Experiment II. The new experiment used the chromatic setting paradigm to measure the observers' adaptation for stimuli combinations of background types II and III under three different illuminants (*D65*, *purplish* and *orangish*). The measured chromatic settings were interpreted in terms of graphs and their interrelations quantified using graph distances according to stimuli conditions, background types and illuminants. These results indicated a high degree of interrelations stability among chromatic settings when illumination was changed, suggesting that categorical colour features could be used to predict the overall behaviour of the colour constancy phenomenon.

5.1 Introduction: Structural colour constancy

Few studies of colour constancy studies have measured more than one point in colour space under immersive illuminant adaptation (McCann, Mckee et al. 1976; Brainard, Brunt et al. 1997; Kulikowski and Vaitkevicius 1997; Amano and Foster 2008) and fewer have measured enough points to address whether the subject's categorical perceptual *structure* (i.e. the interrelations among perceived colours) is kept constant under illuminant changes. The exceptions to this are colour naming paradigms, where subjects are asked to name several colours under different adaptation conditions. In this case, one of the main limitations is the restriction of choices presented to the observer, e.g. the limited number of colour samples used. For example, in two recent experiments (Hansen, Walter et al. 2007; Olkkonen, Hansen et al. 2009), one restricted its measurements to an equiluminant plane with 417 testing samples and the other to a set of 469 Munsell samples of the whole 3D space, which may constrain the method's precision. Also colour naming approaches rely in the categorical structure defined by the measurements, i.e., on the boundaries of the categorical regions or on the centroid locations of tested samples according to each colour category. Here we used the chromatic setting paradigm, which instead of measuring category borders or category centroids as before (Hansen, Walter et al. 2007; Olkkonen, Hansen et al. 2009), allows subjects to select their own *memorable* colours from a set close enough to the focal colours so as to make them easy to memorize and reproduce. The memorable colours were selected from a choice of approximately 45.000 different (volume of the Bounding Cylinder expressed in JND spaced units) CIELab samples which look continuous to the subjects.

5.2 Methods

Here we introduce a new psychophysical experiment, termed *Experiment III*, were observers performed chromatic settings adapted to a combination of two different backgrounds and three different illuminations. After five minutes of adaptation to this stimulus, observers were presented on the screen with the written name of a basic colour term and asked to match it to their own internal representations by manipulating the colour of patches by means of a gamepad. After that, they were required to reproduce the very same colours on different days under different conditions of background and illumination.

5.2.1 Observers

Four subjects took part in our experiment. They were between 31 and 44 years old, their acuity was normal (or corrected to normal) and their colour vision was tested by the Ishihara coloured plates and

the Farnsworth-Munsell D15 Hue Test. Of these, two were naïve to the experiment's purpose, one was not naïve, and the other was the author.

5.2.2 Experimental setup and procedure

The chromatic setting paradigm was used to measure the adaptation state of our observers and thus we used the same experimental procedure as described in subsection 4.3.4. Also the experimental setup was the same as described in subsection 4.3.2.

5.2.3 Stimuli

The stimuli were similar to the ones used in Experiment II. In this particular experiment we only used background configurations Type I and II (see subsection 4.3.3), however the illuminations tested were *purplish* and *orangish* instead of *greenish* and *yellowish* (see Table 5.1). As before, the spatial structure of our stimuli was a Mondrian pattern consisting of a set of overlaid coloured rectangles randomly distributed across the image (i.e. flat, without highlights or mutual reflections) similar to others (Arend and Reeves 1986; Arend 1993). The rectangle size frequency distribution was similar for all stimuli (mean square size was 50x50 pixels) and its geometrical distribution was uniform across the digital image. Backgrounds *Type I* and *Type II* were customised for each subject and did not contain grey patches to avoid giving the observer clues about the illuminant (Foster 2011). All Mondrians were in turn "illuminated" by performing the spectral product of each patch's reflectance times one of three simulated illuminations assuming a Lambertian reflectance model (Walsh and Kulikowski 1998). The illuminants were chosen so that the final product (the illuminated Mondrian) was as saturated as possible while still inside the CRT monitor's gamut. There was no "central patch" to look at, but a set of randomly distributed patches that were simultaneously adjusted in colour and lightness by manipulating the gamepad. These constituted up to 10% of the all patches and their positions were randomly selected in each trial. The object of this was to force the subject to average among patches that had different local surroundings, thus avoiding local chromatic induction effects. Each Mondrian was unique.

Illuminant	x	y
D65	0.312	0.329
Purplish	0.316	0.228
Orangish	0.437	0.343

Table 5.1 CIExy chromaticity for illuminants used in Experiment III.

5.3 Data analysis: *Graph distance* from chromatic settings

For each adaptation state the chromatic setting paradigm produced nine colour measurements which can be interpreted as points in colour space. Notice how these measures were not random but had a particular distribution, especially in terms of hue and lightness. To formally describe the overall interrelations among these sets composed of nine point-measurements we modelled them as a *graph*, (Gross and Yellen 2004) i.e., each *node* (Equation 5.1) corresponds to the coordinates of one chromatic setting and *edges* (Equation 5.2) are defined for all possible node combinations. Additionally, an *edge weight* can be associated to each edge and defined as the Euclidean distance between nodes (Equation 5.3). This approach is schematically shown in Figure 5.1.

$$\boldsymbol{node}: n_i = Coordinates\ of\ the\ i\ chromatic\ setting, where\ i = 1, \dots, 9 \tag{5.1}$$

$$\boldsymbol{edge}: e(i,j) = \overline{e_i e_j}\ , where\ i,j = 1, \dots, 9\ and\ i \neq j \tag{5.2}$$

$$\boldsymbol{edge\ weight}: w_{i,j} = \left\| n_i - n_j \right\|_2 \tag{5.3}$$

When all edge weights are considered together they conform a matrix called the *distance matrix*, we divided its coefficients by the mean distance from all nodes to the node corresponding to the "grey" chromatic setting (Equations 5.4 and 5.5). This normalization produces a distance matrix which contains the proportional distances between nodes relative to the distance to the central node, and thus allowing fair comparisons between different adaptations and observers.

$$\overline{w} = \frac{1}{8} \sum_{\substack{i=1 \\ i \neq j\ and\ j = "grey"}}^{9} w_{i,j} \tag{5.4}$$

$$B = \begin{pmatrix} b_{1,1} & \cdots & b_{1,N} \\ \vdots & \ddots & \vdots \\ b_{N,1} & \cdots & b_{N,N} \end{pmatrix} = \begin{pmatrix} \dfrac{w_{1,1}}{\overline{w}} & \cdots & \dfrac{w_{1,N}}{\overline{w}} \\ \vdots & \ddots & \vdots \\ \dfrac{w_{N,1}}{\overline{w}} & \cdots & \dfrac{w_{N,N}}{\overline{w}} \end{pmatrix} \tag{5.5}$$

Finally, we defined a *graph distance* which quantified the overall difference between two graphs, computed as the mean absolute difference between corresponding distance matrices (Equation 5.6). In this way, we can compare the categorical colour structure under two different adaptations.

$$dist(Graph_1, Graph_2) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} |b_{i,j} - \bar{b}_{i,j}| \tag{5.6}$$

The previous computations need to be performed in some colour space, ideally three-dimensional and perceptually uniform. Possible candidates are the CIELab colour space and the more perceptually uniform CIECAM02 (Fairchild 2005).
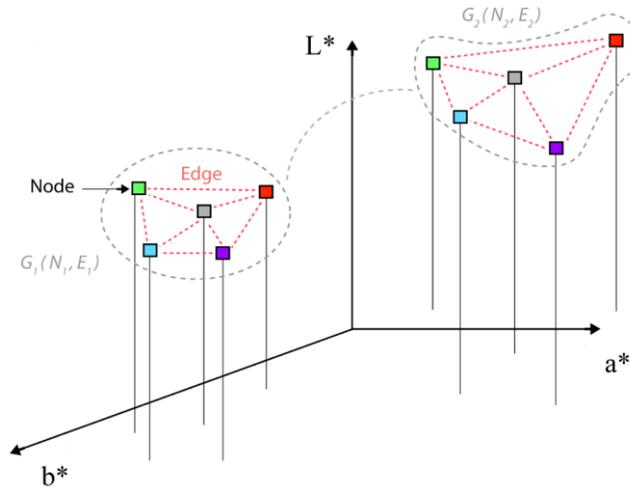
Figure 5.1 Schema of a graph and graph distance constructed from two sets of chromatic settings. A graph (G_1) is represented as a set of nodes (chromatic settings represented as coloured squares) and a set of edges (dotted red lines) for each possible pair of node combinations. The graph and its relative node distances arguably captures the structure of chromatic settings. Two graphs, G_1 and G_2, can be compared by evaluating the relative location of its conforming nodes. See main text for details.

5.4 Results

This section presents data from Experiments II and III. There was a lapse of six months between both experiments. Experiment II used 10 subjects, three illuminations (*D65*, *yellowish* and *greenish*) and three backgrounds (*Type* 0, I and II). Experiment III used 4 subjects, two backgrounds (*Type* I and II) and two illuminations (*purplish* and *orangish*). Our results reveal a high degree of invariance among the interrelations of chromatic settings under different illuminations.

5.4.1 Chromatic settings under *purplish* and *orangish* illuminations

Figure 5.2 illustrates the *selected representatives* chosen by three different subjects on the *reference session* (see section 4.4.1) of Experiment III. Each coloured circle shows the selected representatives (SR) for the corresponding category and the joining lines (Euclidean distances) help to illustrate their geometrical interrelations. Notice how these interrelations are different for each subject, for instance, subject JR's "blue" SR has high lightness while the other two subjects selected colours with low lightness. Also, when comparing subject CAP and XO, notice how their selection of orange and red also differs in lightness level: subject CAP selected higher lightness than XO. Finally, notice how the "red" and "pink" SRs of CAP were different in hue and lightness from those of XO. Following this, we conclude that each subject had his/her particular choice of *selected representatives*, which is expected from previous studies (Berlin and Kay 1969; Boynton and Olson 1987; Sturges and Whitfield 1995) and Figure 4.3. Also, notice from Figures 5.2 and 4.3 how the inter-observer pattern remains approximately the same. In the following sections we studied whether this pattern was also invariant under the different illuminant adaptations.
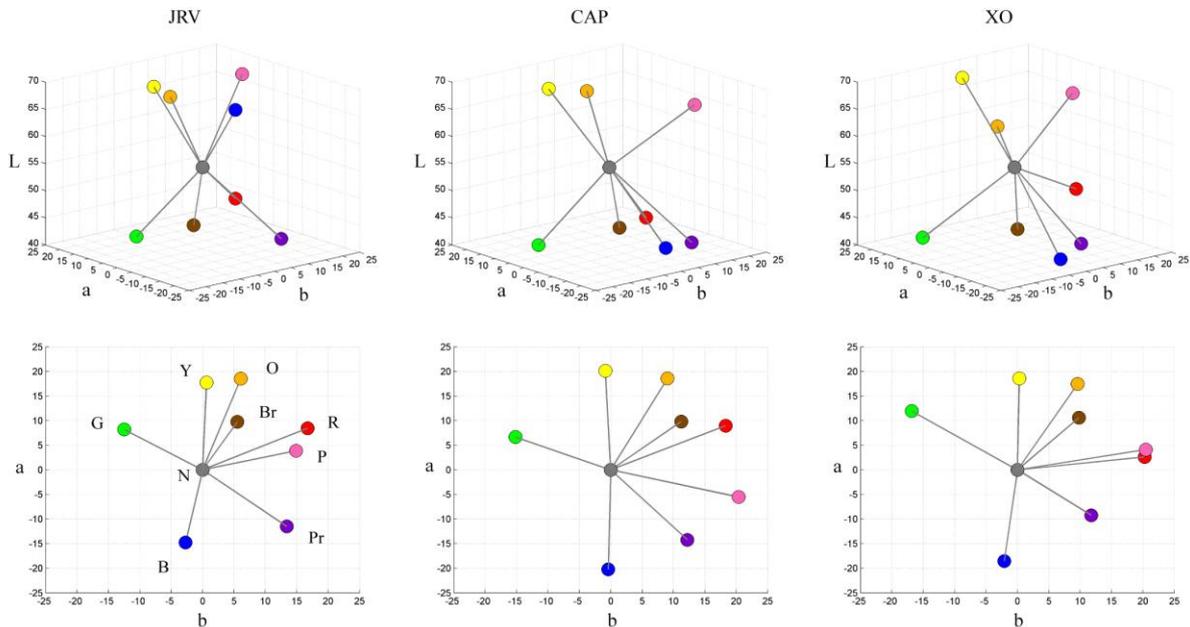
Figure 5.2 *Selected representatives* from three observers in Experiment III. Each column of plots corresponds to a subject and it contains two complementary views of the selected representatives obtained in the reference session. Results are plotted in CIELab colour space: the top plots contain an isometric view and bottom plots show the projection of the same data on the a*b* chromaticity plane. Key: G=green, B=blue, Pr=purple, P=pink, R=red, Br= brown, O=orange, Y=yellow, N= neutral (grey).

Figure 5.3 shows the location for chromatic settings of Experiment III in the $a_c b_c$ projection plane of CIECAM02 colour space (Fairchild 2005). These results are discriminated by illuminations (panels) and background types (markers) and averaged over observers. CIELab is a very good colour space when illumination is daylight or close to achromatic (Fairchild 2005). In our case, the *purplish* and *orangish* illuminations are highly saturated and so the computed CIELab coordinates for chromatic settings may include some distortions. In order to avoid these artefacts we used the CIECAM02 colour space which is more perceptually uniform than CIELab. It also allows disabling its chromatic transform, therefore allowing us to analyse our data without discounting any adaptation to the illumination. The *viewing condition* parameters of CIECAM02 were set to the mode *dark surround* which is indicated for viewing in a dark room (Fairchild 2005), but since we do not want CIECAM02 to impose its chromatic transform to our data, the degree of adaptation (F) was set to 0.

The experimental error was computed as the standard deviation from the mean for each colour category, averaged over subjects and sessions (the same procedure as in subsection 4.4.2). The overall standard deviation in Experiment III was 4ΔE* units.
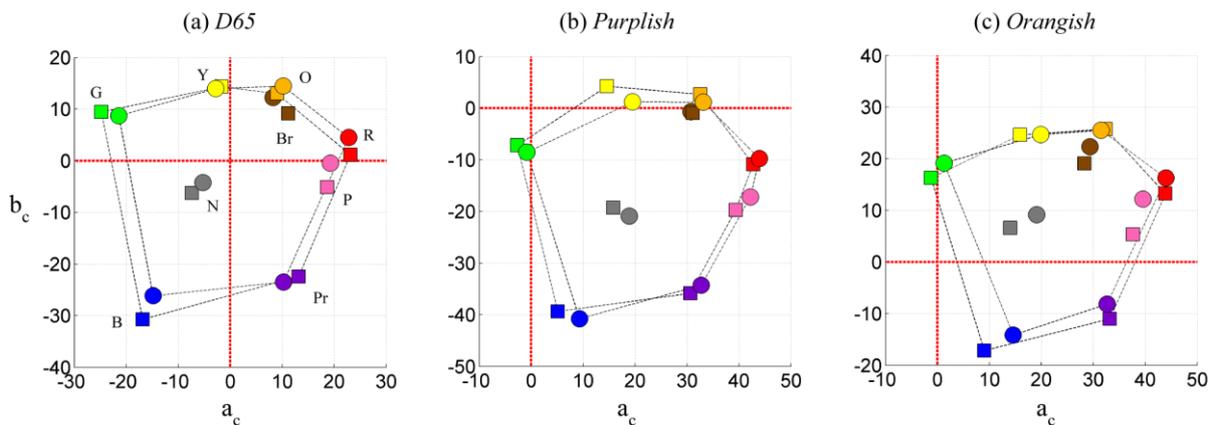
Figure 5.3 Average location in CIECAM02 of the chromatic settings adjusted in Experiment III. Coordinates are represented in the $a_c b_c$ plane, parameters set for *dark mode* and its chromatic transformation disabled. Each panel corresponds to the chromatic settings measured under each one of the three tested illuminations. Square and circle markers indicate adaptation to background type II and III, respectively.

5.4.2 Interrelations among chromatic

The chromatic setting paradigm produced a set of nine measured points for each adaptation state. Panel a of Figure 5.4 shows the chromatic settings of observer JR under the five illuminations tested in Experiments II and III, using the same background type II. Notice that measures under the same illuminant are linked by a coloured line, helping to visualize the stability of the chromatic settings

93

structure. Panel b shows the same tendency for observer XO. We described formally this observation through the concepts of *graph* and *graph distance* introduced previously (section 5.3).

When computing a colour constancy index is common practice to compare perceptual measurements with their physical expected values (Foster 2011). Following this idea we simulated the effect of the four chromatic illuminants (Tables 4.1 and 5.1) on the selected representatives adjusted under *D65* illumination (i.e. the *physical* colours) for each observer. Notice that these *physical* colours coincide with the colours present in background type I. Panels c and d of Figure 5.4 show the CIELab coordinates of such colours for observers JR and XO, and if colour constancy was to be complete, coloured markers from the top and bottom plots should overlap.

Visual inspection of the graphs in Figure 5.4 reveals two trends: (i) the perceptual representations in the top plots seem to have maintained the same proportions showing higher stability in terms of their interrelations, while their counterparts at the bottom plots have been slightly warped by the illumination; and (ii) the region spanned by the perceptual measures is more compact than the region spanned by the physical counterparts. The latter observation is wholly captured by standard colour constancy indexes, which measure distances between the perceptual and the physical grey (Foster 2011) while the former observation needs a more comprehensive approach to be grasped, for instance using the *graph* representation.

Although the visualization in Figure 5.4 shows the coordinates of perceptual and physical colours in the CIELab colour space (D65 as a reference white point), we computed the same data in CIECAM02 for the reasons outlined in section 5.4.1. The same CIECAM02 parameters (*dark surround* and F=0) were used in all the following computations.

Our primary aim was to quantify structural differences over illumination changes, then for each observer and adaptation state (illumination plus background type) we computed the corresponding graph from its chromatic settings. Next, while keeping the same observer and background type, we computed the graph distance between the graph under the reference illuminant (*D65*) and the graphs under the test illuminants (*greenish*, *yellowish*, *purplish* and *orangish*). For instance, looking back at panel a of Figure 5.4, each of its five sets of chromatic settings were modelled as graphs, and graph distances were computed from the graph centred at the achromatic locus.
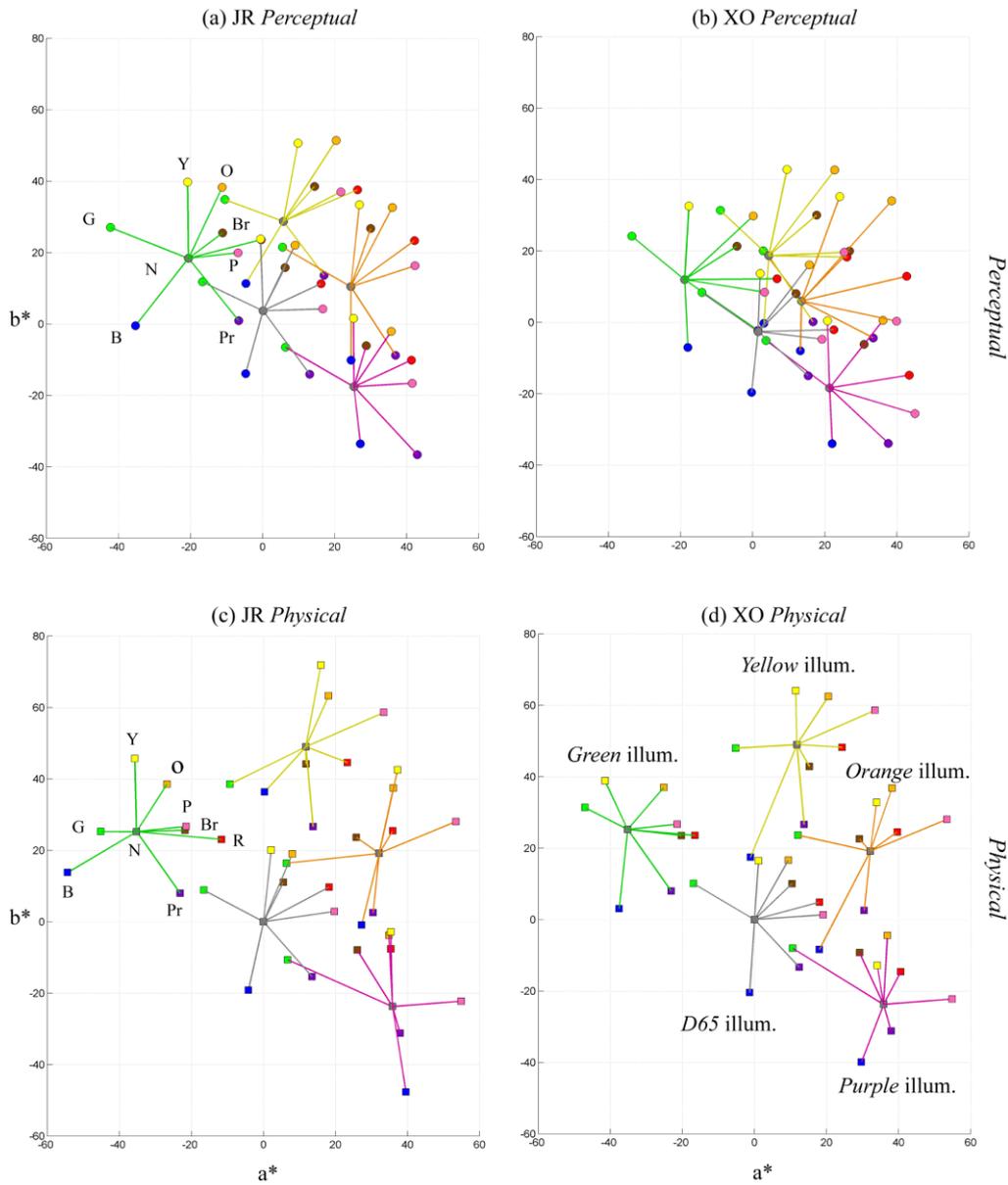
Figure 5.4 Chromatic settings obtained for two subjects and five illuminants in CIELab. .Each plot shows the measures obtained for 5 illuminants (each coloured line corresponds to one illuminant). The top plots show the chromatic settings adjusted by two observers and the bottom plots show the corresponding projections of the physical colours after illumination. Left plots correspond to subject JR with type I background and right plots to subject XO with the type II background. All measures are shown as projections on the a*b* chromaticity plane.

Panels of Figure 5.5 contain the results of this approach averaged over observers. Each panel corresponds to proportions of structural deformation (y-axis) under one test illumination discriminated according to background types (x-axis). The same graph distances were computed for the *physical* colours and their results are also shown in each panel. The overall average deformation proportion for perceptual colours is 0.13 (0.03 SD) and for physical colours is 0.24 (0.03 SD). These results reveal

that interrelations among chromatic settings were mostly constant (87%) under illumination changes, and 11% more stable than the physical ones.

All panels of Figure 5.5 reveal no structural differences for chromatic settings between background types I and II. However, when averaged over background types and discriminated according to illuminations, structural differences were: 0.12 (0.03 SD) for *greenish*; 0.16 (0.04) for *yellowish*; 0.11 (0.02) for *purplish*; and 0.11 (0.01) for *orangish*. Interestingly, chromatic settings under the *yellowish* illumination had a higher degree of distortion, in accordance with Chapter 4 results.
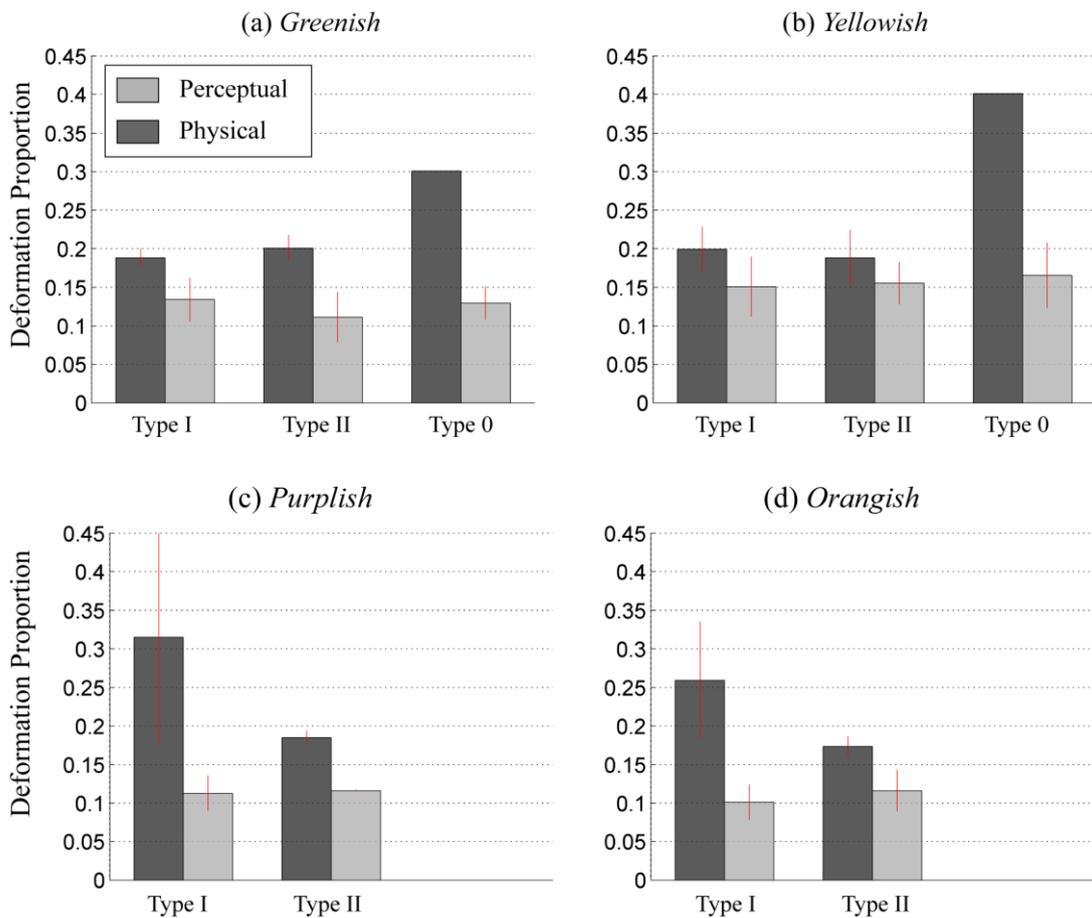
Figure 5.5 Proportion of structural deformation for chromatic settings under different illuminations. Each panel corresponds to one tested illumination. Results are discriminated according to background types and red lines indicate SD. Background type 0 was not used in Experiment III, hence panels c and d do not have their corresponding column bars. See text for further details on how deformation proportion was computed.

5.5 Discussion

5.5.1 Chromatic settings were not influenced by the CRT gamut boundary

In Chapter 4 we have shown that subjects were able to remember their selected representatives over an experimental period of several weeks (section 4.4.1). Another factor that could influence the outcome of our experimental procedure is the constraining effect of the CRT gamut boundary: while adjusting the chromatic settings, observers may be limited by the boundary in their choice. For this reason, we analysed our results looking for effects of this possible "artefact".

Panel a of Figure 5.6 shows all chromatic settings resulting from the Experiments II and III: 1026 colour measurements obtained from adjusting 5130 settings over 114 experimental sessions. Each data point in Figure 5.2 corresponds to a particular chromatic setting which is the average of 5 different trials. In order to ensure that subjects did not use the CRT gamut boundary as a reference when doing the adjustments requested, we computed for each of the 1026 colour measurements their CIELab distance to the CRT gamut boundary. This information is summarized by the histogram contained in panel b of Figure 5.6, i.e., the height of each bar indicates the number of chromatic settings that were at a particular distance interval from the CRT gamut boundary. Our results show that only 1.4% of the 1026 points were closer than $5\Delta E^*$ units from the boundary, and only 16.3% were between 5 and $10\Delta E^*$ units.

The boundary of the CRT gamut was computed by considering a (dense enough) sampling of the device dependent RGB colour space. In particular, we computed the XYZ coordinates of the RGB cube boundary through the CRS (www.crsldt.com) calibration software. Next, the conversion to CIELab was straightforward and D65 was used as a reference white point.

5.5.1 Methodological issues

Here we studied whether the previous results have been influenced by two factors: the particular choice of colour space; and the precision of the chromatic setting method.

Choice of colour space

Each colour space has its own particularities which makes it suitable for each specific application. In our case, the main requirements for a suitable colour space were three: (1) it should be a 3D colour space in order to grasp the complexity of colours such as brown, yellow, orange, red and pink; (2) it should be perceptually uniform in order to obtain meaningful distances and (3) it should be

sufficiently popular to allow us to compare our results with the literature. After careful consideration, we decided that CIELab and CIECAM02 satisfy these requirements (Fairchild 2005). Computing CIELab coordinates is straightforward from tristimulus coordinates (XYZ) once the reference white point is chosen, however it incorporates a *wrong* Von Kries transformation (computed in XYZ instead of LMS) and the choice of the white point distorts the space (Fairchild 2005), then the meaning of distances is diminished as we go away from the origin (Wyszecki and Stiles 1982). So, in our case with highly chromatic illuminants we could expect some degree of deformation for points away from CIELab's origin, using D65 as a reference white point. Because of these drawbacks we computed the previous graph distance in both CIELab and CIECAM02 and obtained similar results. This may be because of three reasons: (1) our graph distances are computed from relative distances, (2) our chromatic settings tend to be in small clusters, close to each other and (3) the wrapping of the CIELab space is smooth (continuous and differentiable, as a consequence of its formulation), thus modifying groups of points such as our chromatic settings in a uniform manner.
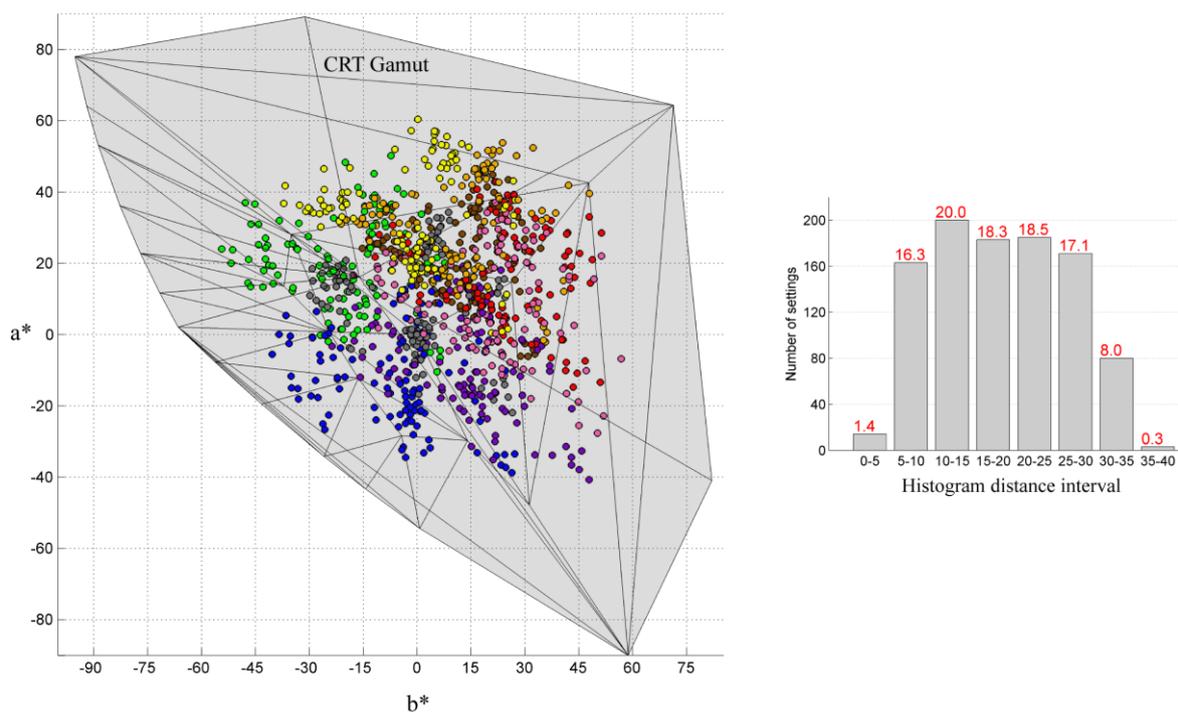
Figure 5.6 Chromatic settings and the CRT gamut in Experiments II and III. Panel a shows the CIELab chromaticity plane where the convex hull of the CRT gamut is projected and 1026 coloured dots corresponding to chromatic settings under different conditions. Each dot was computed as the average of 5 settings and its colour indicates the colour perceived by the subject during the adjustment. Panel b shows a histogram of the number of measured colours according to their distance to the boundary of the CRT gamut.

98

Because of the above, we are confident that our method is robust enough to small differences in the structure of CIELab colour space. However since CIECAM02 is more perceptually uniform and also allows to conveniently disable the chromatic transform (Fairchild 2005), we selected it as our main colour space for computations.

Method's precision

As previously reported, the chromatic setting experimental error was estimated to be about 4 ΔE*. We studied whether the magnitude of this error could account for the structural deformation results reported in Figure 5.5. To do so, we propagated the chromatic setting error to our structural deformation index and obtained an average value of about 5.2%, which is clearly lower than the average 13% of structural deformation. Thus, we discarded the chromatic setting error as the main explanation for the structural deformation shown in Figure 5.5.

5.5.2 Comparison to previous work

Previous work in the literature has also focused on the colour appearance of multiple points under illumination changes. Some researchers used real surfaces and a matching technique (McCann, Mckee et al. 1976; Brainard, Brunt et al. 1997; Kulikowski and Vaitkevicius 1997; Amano and Foster 2008) or CRT-simulated scenes (Brainard, Brunt et al. 1997; Amano and Foster 2008). Others, (Hansen, Walter et al. 2007; Olkkonen, Hansen et al. 2009) measured the categorical colour appearance of multiple points and studied them through the deformation of categorical boundaries or displacements of centroid locations in colour space, concluding that the categorical structure of colour space remains roughly stable.

In order to further compare our results to the literature, we have considered the pioneering work of McCann *et al* (McCann, Mckee et al. 1976). They reported the Munsell coordinates of 17 matches under 5 different illuminants. Figure 5.8 shows the CIELab colour space plots of MacCann *et al* results. The colours in the plot indicate the illuminant under which the matching was done and reference dots (in grey) are linked to their corresponding matching under a coloured illumination by black lines. We applied our structural approach, comparing their 'grey' set of points to the other four coloured sets and obtained a structural stability of 85% (2% SD) which is similar to our results and suggests a high degree of (but not perfect) colour constancy.

Since our results also indicate that the inter-distances among chromatic settings are mostly stable under illumination changes, we conclude that they are in agreement with those of McCann *et al* (McCann, Mckee et al. 1976).
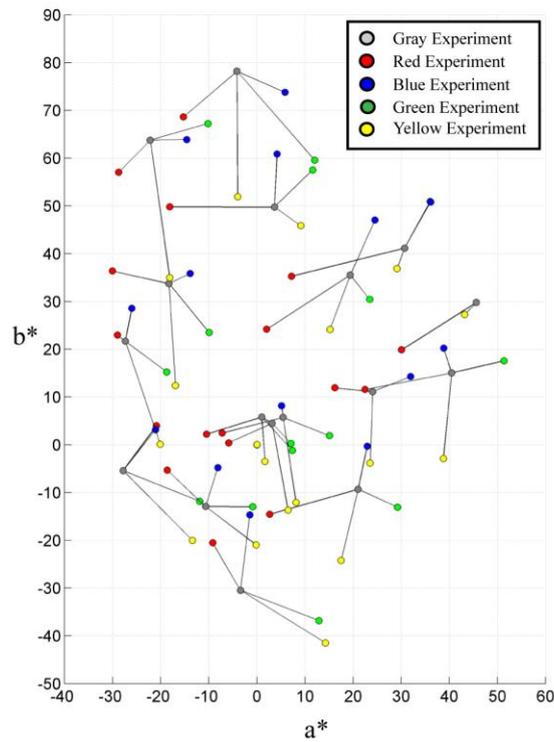
Figure 5.7 Projections in the a*b* plane of the results reported by MacCann *et al* (McCann, Mckee et al. 1976). Each coloured circle corresponds to the a*b* coordinates of surface matches under a five different illuminants, which are linked to their reference grey by black lines. Colour matches are colour coded according to test illumination. D65 was used as a reference white point to compute CIELab coordinates.

5.6 Conclusions

We collected information on the perceptual interrelations of coloured surfaces under illuminant changes and modelled these measurements using graphs. Our results show (see Figure 5.5) that these interrelations remained 87% constant under an illumination change, in contrast with the structural deformation undergone by the physical colours (76%). This is in accordance with previous studies that reported categorical stability using colour naming techniques (Hansen, Walter et al. 2007; Olkkonen, Hansen et al. 2009) and suggests that categorical colour perception may be used to guide colour constancy adaptation. Despite the reported structural stability, there was still a remaining 13% of "inconstancy" which may allow some changes of category under different illumination, in particular for those colour samples near the borders of basic colour categories.

Chapter 6 Inconstancy in categorical colour perception

In previous chapters we have shown that the structure of the categorical colour space has its own particularities for each observer and to study it, we introduced a new colour constancy paradigm which measures the location of nine points that are relevant to this categorical structure under changes of illumination. In addition, we demonstrated that the interrelations among these points were mostly stable when tested with several illuminated backgrounds. From our study two interesting conclusions arose: first, we observed small but consistent differences in adaption for the different illuminations; second, the backgrounds we studied had no effect on the final adaptation state regardless their differences in terms of the categorical information they contained. This might be a consequence of having tested categorically balanced backgrounds (where all basic categories are represented). This chapter furthers our examination of the possible influences of categorical colour perception in colour constancy, in particular whether having a less variegated background would change the adaptation process. To do so, we performed a new psychophysical experiment (*Experiment IV*) which tested adaptation under six different illuminations and two different Mondrian backgrounds with only three different colours each. In each experimental session observers performed both a chromatic setting *and a colour naming task*, providing categorical colour information for a small set of points in colour space. In order to expand this categorical information to other, unmeasured points, we developed a categorical colour prediction model which was tuned using the chromatic settings. Experimental and modelled data were interpreted in terms of categorical colour changes. This new model complements the previous paradigm by including adaptation and categorisation in the same framework.

6.1 Introduction

Most studies in the literature attempt to quantify the extent of colour constancy through colour constancy indices (Smithson 2005; Foster 2011), which only indicate the overall degree of adaptation without providing insights on the categorical appearance of stimuli after adaptation (see section 2.3.4). Also, the phenomenon has been modelled in order to predict the appearance of samples, but these models generally provide point coordinates with little information regarding their categorical colour appearance (Brainard, Brunt et al. 1997; Uchikawa, Emori et al. 2002). Of course, colour appearance models could be used to predict the appearance of samples, but the considerable inter-individual variability existing suggests that models based on *average* colour appearance may be too coarse to predict colour appearance for particular individuals.

According to Jameson and Hurvich surface object recognition under illumination changes may be adequately accomplished by category matching, and does not require the precise matching of hue, brightness and saturation (Jameson and Hurvich 1989). Also, Boynton and Olson (Boynton and Olson 1987) state that: "*Measurable colour changes that do not produce categorical shifts may not matter very much if memory is of basic sensations and their names, rather than colours per se*". Consequently, we may hypothesize that categorical colour perception is a key factor in successful colour constancy.

Consider now the following example: let us assume that we are adapted under an *achromatic* illumination and we see several unsaturated coloured objects, then the illumination chromaticity is changed to *greenish*. At first, the colour appearance of these objects will be mostly greenish but after a few seconds/minutes (while adaptation to the average scene chromaticity takes place and our reference white is moving closer to the average scene chromaticity) most of these objects will be perceived as having different colours (Webster 1996; Shevell and Kingdom 2008). Since in general adaptation under highly saturated illuminants is never 100% complete, after it reaches stability the objects may not have the same colour appearance as before under the *achromatic* illumination. Here we speculate that after an illumination change and while the ineludibly global contrast adaptation, which is driven by low level mechanisms (Webster 1996; Hurlbert and Wolf 2004) takes place, the HVS tries to keep the ongoing categorical information stable. This implies that the chromatic properties of the illumination (magnitude and orientation of the chromatic shift) may influence the resulting categorical colour perception during and after the adaptation.

Since our chromatic setting paradigm allows us to test these influences, we performed a new psychophysical experiment which tested two Mondrian backgrounds with only three colours each

under six different illuminations. In addition to this, observers performed a colour naming task. Our results were studied in terms of the categorical changes between the reference and test illuminations. Also, we expanded our analysis by developing a new model of categorical colour prediction (termed the *Customised Colour Category Predictor* or *CCCP*) based on the chromatic settings measurements, and thus on the particular colour perception of each individual adaptation state.

6.2 Methods

This section introduces a new experiment, called Experiment IV, where observers performed a chromatic setting task (see subsection 4.3.2) followed by a customized colour naming task in each experimental session. We tested two Mondrian backgrounds with only three different colours each under six different illuminations (12 different stimuli in total), all presented on a CRT screen in the same conditions as before.

6.2.1 Observers

Three observers, one female and two males took part in this experiment. They were between 26 and 32 years old and their colour vision was normal as tested by the Ishihara colour vision test (Ishihara 1972) and the Farnsworth-Munsell D15 Hue test (Farnsworth 1957). All had self-reported normal or corrected to normal visual acuity. Two of the observers were unpaid volunteers naïve to the experiment's purpose and the other was the author. All were non-native English speakers but had excellent English language skills.

6.2.2 Experimental setup and procedure

The main difference between Experiment IV and Experiments II and III was that observers here performed two different tasks in each session: (1) a chromatic setting task immediately followed by (2) a customized colour naming task, while adapted to the same viewing conditions.

The colour naming task consisted of assigning a single or compound colour name to each of the three colours of the Mondrian stimuli. These names are referred here as colour *descriptors*. Observers were instructed to choose their colour descriptors from combinations of basic colour terms (BCT) including their adjectival forms (BCT-ish, e.g. *greenish, yellowish, bluish, reddish, greyish, orangish, pinkish, purplish* and *brownish*) and encouraged to do this as quick as possible. Descriptors were recorded by hand by the observer itself on a white sheet of paper which was out of sight during the chromatic setting task.

As before, all sessions were conducted inside a dark room, with all walls lined in black. The chromatic setting was programmed in Matlab and the stimuli were displayed on a CRT monitor which was the only light source. Viewing was binocular and unrestrained. Subjects modified the test stimuli by navigating the CIELab colour space using six different buttons, two for each colour space dimension on a commercial gamepad. The reference white point was D65 with luminance equal to 100 Cd/m^2. This experimental setup was similar to the one used in Experiment II as described in section 4.3.

The whole experiment consisted of 16 sessions: 12 sessions corresponded to all possible combinations between 2 different Mondrian backgrounds and 6 different illuminations; a repeatability test conformed of 4 sessions (two of them with the bounding cylinder) with background Type 0 and D65 illumination. As before, the aim of these tests was to track changes in the observer's ability to reproduce the same colours throughout the experimental period (see section 4.3.4). The experiment lasted for two weeks with repeatability test sessions evenly distributed over such period. Each experimental session (chromatic setting and colour naming) took approximately 25 minutes and each chromatic setting trial took 30 seconds.

6.2.3 Stimuli

Since our general aim is to test whether the chromatic properties of the background/illumination influence categorical colour perception, we selected the chromaticities of the background most likely to propitiate categorical changes. For instance, we reduced the number of background colours in order to weaken colour constancy mechanisms (Linnell and Foster 2002) and selected colours other than SRs to facilitate categorical changes. These colours were also chosen to maximise changes in their corresponding descriptors as observers' categorical perception change. We also tested the influence of the average chromaticity of these colours by considering two background types with different mean spatio-chormatic information.

Our basic stimulus consisted of a Mondrian background pattern, i.e. a set of randomly overlaid coloured rectangles, distributed across the screen. The average rectangle size was 70x70 pixels. There were two types of backgrounds:

Type III. It was built from colours other than the SRs chosen by each observer in the reference session. There were three colours in total and they were selected to lie in the lines joining SRs in CIELab a*b* plane. These colours were labelled as C_1, C_2 and C_3 (see panel a of Figure 6.1). C_1 was located between yellow and green (25% of the joining distance away from green), C_2 between blue and purple (25% of the joining distance away from blue), and C_3 in between purple and pink (50% of the joining distance from purple).

Type IV. It was also built from colours other than the SRs chosen by each subject in the reference session. There were three colours in total: the first two were selected to lie halfway along the line joining blue and green, and along the line joining purple and pink. The third colour was selected randomly within the circle defined by the blue SR and radius 10 ΔE^*. These colours were labelled C_4, C_6 and C_5 respectively. As shown in panel b of Figure 6.1, this background type is clearly biased towards the blue side of colour space.
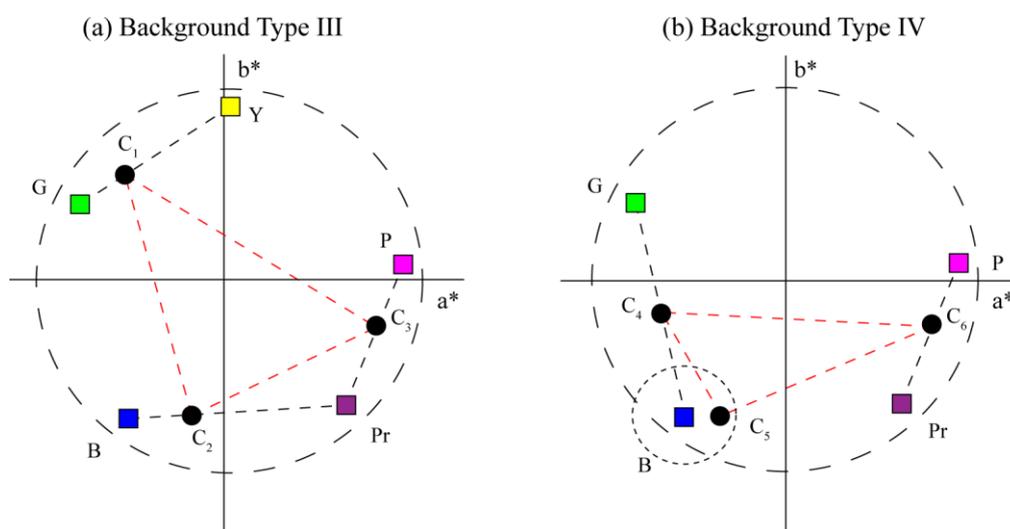
Figure 6.1 Schematics of the colours included in backgrounds type III and IV in CIELab a*b* plane.. Panel a and b corresponds to colours selected for background types III and IV respectively. See details of how they were chosen in the main text. Square colour-coded markers represent the selected representatives and filled black circles represent the colours chosen for the stimuli. The dotted circle represents the Bounding Cylinder (see section 4.2.2) limiting the region from where all SR were chosen and the black dotted lines highlight the relationships specified in main text.

Stimuli colour computations were performed in CIELab space using the same reference white point as before, i.e., D65 with luminance 100 cd/m^3. Figure 6.1 shows a scheme of the approximate location of the stimuli colours (circles) in relation to hypothetical selected representatives (squares). Notice that each observer had a different set of stimuli colours while keeping similar categorical information, thus in our nomenclature, observers were differentiated by means of a super-index. Table 6.1 details this notation and the CIELab coordinates of the colours C_i^j chosen under D65 illumination. The effects of illumination over these colours are shown in Figure 6.2.

Observer	Notation	Type III			Notation	Type IV		
		a*	b*	L*		a*	b*	L*
JR	C_1^1	-8.57	8.01	48.78	C_4^1	-8.85	-4.05	57.29
	C_2^1	-0.37	-13.18	66.88	C_5^1	-6.85	-14.26	76.15
	C_3^1	12.87	-6.20	56.96	C_6^1	12.87	-6.20	56.96
IR	C_1^2	-8.75	8.47	58.03	C_4^2	-8.85	-4.05	57.29
	C_2^2	0.82	-21.86	57.33	C_5^2	-2.44	-22.78	57.17
	C_3^2	12.55	-8.46	67.40	C_6^2	12.55	-8.46	67.40
JV	C_1^3	-6.61	10.82	49.40	C_4^3	-8.98	-2.00	48.52
	C_2^3	-1.49	-14.70	55.82	C_5^3	-9.18	-20.50	67.04
	C_3^3	14.94	-0.98	57.41	C_6^3	14.94	-0.98	57.41

Table 6.1 CIELab coordinates of the stimuli colours used in Experiment IV Mondrian colours were selected by taking into account each observer's SRs (see Figure 6.1). Their colour coordinates are grouped according to observers (rows) and dimensions and background types (columns). The reference white point was D65 with L=100 cd/m³. The column labelled as *Notation* specifies the notation used for each colour where the super-index corresponds to the observer (1 for JR; 2 for IR; 3 for JV) and the sub-index identifies each of the three Mondrian colours. See main text for explanation on how they were chosen.

Unique randomized Mondrians were created for each experimental trial: no observer saw the same Mondrian twice. To illuminate the Mondrian pattern, we employed the same techniques as before (see section 4.3.2). Table 6.2 shows the CIE *xy* and Munsell coordinates of the six illuminants considered. Given the asymmetric 3D shape of the CRT gamut we reduced the intensity of the bluish illumination by 40% in order to ensure that all illuminated colours fall inside the gamut.

Illuminant	Notation	x	y	Munsell coord.	Saturation (ΔE^*)
D65	I_1	0.317	0.329	-	-
Greenish	I_2	0.296	0.453	2.5G 7/10	43.5
Yellowish	I_3	0.453	0.434	2.5Y 7/8	48.0
Bluish	I_4	0.242	0.285	7.5B 7/8	20.5
Unsaturated greenish	I_5	0.316	0.405	2.5G 7/6	24.4
Unsaturated yellowish	I_6	0.401	0.393	2.5Y 7/4	26.6

Table 6.2 Illuminants used in Experiment IV. The first two columns show the notation assigned to the illuminants used in Experiment IV. The following columns show the illuminant's CIE 1931 *xy* coordinates and Munsell coordinates and the last column shows a measure of their saturation computed as distances between the reference white point (D65) and each illuminant in CIELab ΔE^* units.

The number and sizes of rectangles were manipulated so that the pixel average chromaticity of all background types under D65 illumination was as close as possible to D65. Background Type III was effectively centred on the origin but background Type IV was slightly shifted (6.06 ΔE^* in average) towards the negative part of the b* axis, for all three observers. This is due to the particular choice of colours for background Type IV which made it impossible to average them on the centre of the

CIELab space. Panel b of Figure 6.2 reflects this issue: D65 chromaticity falls outside the triangles formed by background type IV colours. Table 6.3 reports the average chromaticities of the Mondrian backgrounds in CIELab coordinates.

Obs/Back	Type III			Type IV		
Obs/Coord	a*	b*	L*	a*	b*	L*
JR	0.00	0.00	54.26	-0.61	-6.21	59.71
IR	-0.75	-0.94	60.89	-1.62	-7.88	60.28
JV	0.00	0.00	53.29	0.17	-4.10	54.42

Table 6.3 Average pixel chromaticity of Type III and IV Mondrian backgrounds under *D65* illumination in CIELab coordinates.

As specified in subsection 4.3.4 we used multiple test patches to perform the chromatic setting task. The average number of patches was 11 (3 SD) and their total average area corresponded approximately to 7% of the displayed image.
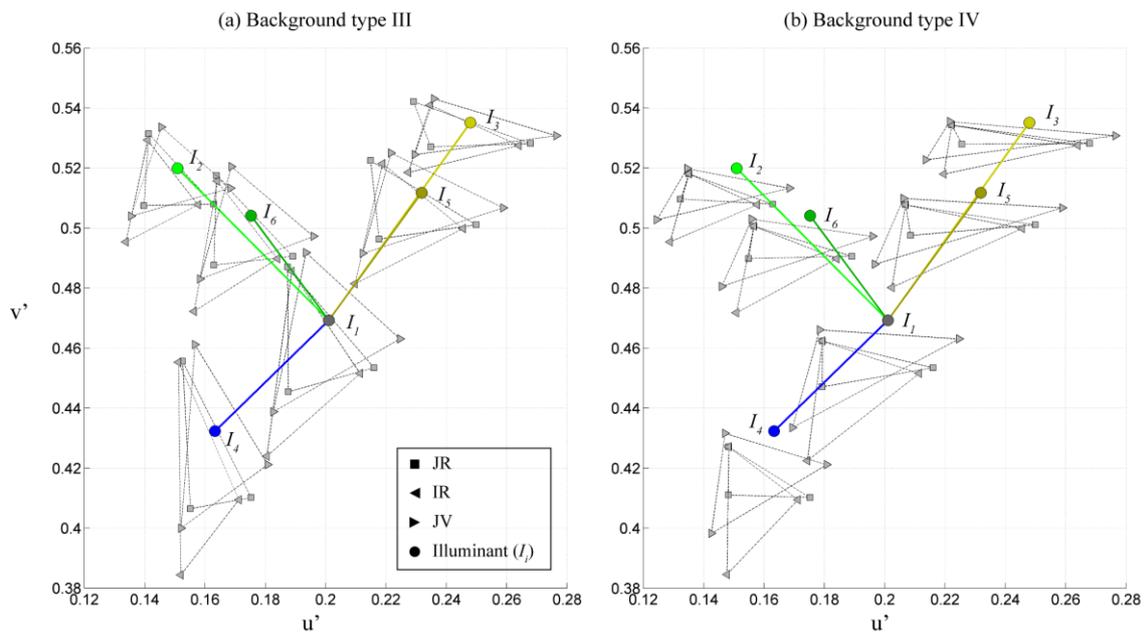
Figure 6.2 CIE 1976 uv coordinates of stimuli colours used in Experiment IV. Triangle and square markers joined by black dotted lines correspond to the three-coloured Mondrians under all six illuminations for all observers. Panels a and b correspond to background type III and IV respectively. Circle markers correspond to the illuminants coordinates and are colour-coded according to their representative colour. Coloured lines link D65 to the rest of the illuminants' coordinates, using the same colour-coding.

6.3 CCCP: A new model for categorical colour prediction

In this section we introduce the Customised Colour Category Predictor (CCCP): a new method to predict the categorical colour appearance of coloured samples after adaptation to a test illumination. This method is based on the measurements provided by the chromatic settings under the corresponding test and reference illumination which provide not only the location of nine points in colour space but also their categorical information. CCCP inherits the properties of the chromatic setting paradigm, i.e., it is tuned to the adaptation of each particular observer under specific background and illumination conditions.

The method has three stages. In stage 1 it uses chromatic settings to predict the coordinates of any colour sample under the reference illumination. In stage 2 it assigns a *categorical probability* in terms of only two BCT. In stage 3 it assigns a colour term of the form BCT or BCT BCT or BCT BCT adjective.

In stage 1, the parameters of our model are the chromatic settings under the reference and test illuminations, measured for each observer and background conditions. From these correspondences we compute a linear model of colour constancy based on a diagonal plus translation matrix similar to that of subsection 4.4.4. This allows us to predict the LMS coordinates of any colour sample $\mathbf{x}$ under the reference illuminant from the LMS coordinates of the same colour sample under the test illuminant. Since we want to work in a 3D perceptually uniform colour space, we compute the CIELab coordinates of the samples, using the reference illumination as a white point.

Stage 2 assigns to $\mathbf{x}$ a categorical probability in terms of only two basic colour categories, a procedure performed in CIELab space and summarized in Figure 6.3. In order to achieve this we generated a *Hue-Lightness map* by defining a cylinder of radius 22 ΔE*, centred on the "grey" chromatic setting ($\mathbf{x}_N$) and with its main axis parallel to the L* dimension. The dimensions of this cylinder were chosen to be the same as those of the bounding cylinder of section 4.2.2. The Hue-Lightness map in Figure 6.3 is made of several categorical *key points* ($\mathbf{q}_i$ -squares in panel b) with their hue defined from the hue of the chromatic settings (indicated by black arrows in panel a) and their lightness defined from previous studies (Benavente, Vanrell et al. 2008) and Chapter III results. Three colour categories (green, blue and purple) span the whole height of the cylinder (L*), while others (pink, orange and yellow) are only present in high lightness regions or in low lightness regions (brown and red). The Hue-Lightness map also contains *links* between these key points which will determine the BCT composition in terms of categorical probability. These links were defined following the rationale introduced by Boynton and Olson (Boynton and Olson 1987). To quantify the position of points

relative to the achromatic locus, we defined a threshold value that limits the region around the cylinder's axis, called the *achromatic threshold* $T_{achr} \in \mathbb{R}$. For points **x** further away from the axis than this achromatic threshold, we performed the following steps:

1. Project **x** to the cylinder surface (we note this projection $\mathbf{x}_p$), i.e., to the Hue-Lightness map.

2. Find the *link* with the smallest distance to $\mathbf{x}_p$. Categorical colour appearance is determined from the categorical belonging of $\mathbf{x}_p$ to either end of the link ($\mathbf{q}_i$ or $\mathbf{q}_j$). Find the categorical belonging (cat_i, cat_j) where cat_k is one of the BCTs.

3. Compute the point belonging to the link with the smallest distance to $\mathbf{x}_p$ (we note this point as $\mathbf{x'}_p$). Compute the *probability* of belonging to each candidate category (cat_i, cat_j) from the relative distances between the link extremes and $\mathbf{x'}_p$: $(p_1, p_2) = (d(\mathbf{q}_1, \mathbf{x'}_p)/d(\mathbf{q}_1, \mathbf{q}_2)$, $d(\mathbf{x'}_p, \mathbf{q}_2)/d(\mathbf{q}_1, \mathbf{q}_2))$. Notice that distances are computed on the cylinder surface and in the case where $\mathbf{q}_i$ and $\mathbf{q}_j$ have the same lightness values; the calculations will be in terms of angular distances (hue).
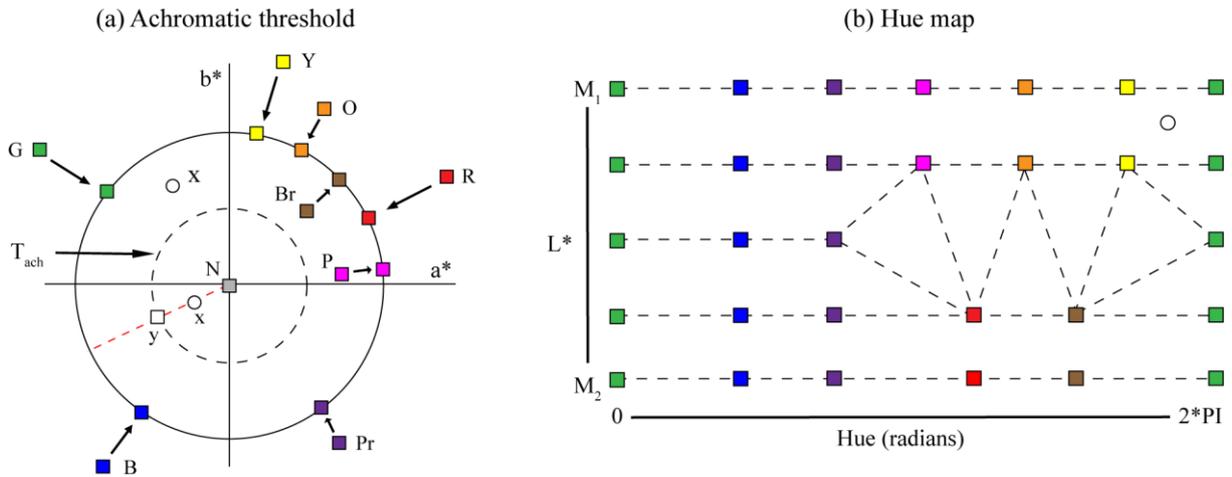
Figure 6.3 Schematics of the Hue-Lightness map used in the CCCP. Black arrows in panel a illustrate how the Hue-Lightness map was produced from the chromatic settings' hue dimension. The inner circle represents the achromatic region, delimited by the achromatic threshold ($T_{achr} \in \mathbb{R}$ see main text). Panel b shows the Hue-Lightness map with the location of *key points* ($\mathbf{q}_i$) and the *links* between them represented by broken lines. See text for further details.

For points **x** within the region delimited by the achromatic threshold ($T_{achr} \in \mathbb{R}$ -inner circle in panel a of Figure 6.3), we performed the following steps:

1'. Perform steps (1) to (3) and select the category with the highest probability (cat_i).

2'. Find the probability of belonging to both "grey" and cat_i by computing the relative distances in the a*b* plane between $\mathbf{x}$ and $\mathbf{x}_N$ and between $\mathbf{x}$ and $\mathbf{y}$: $(p_{gray}, p_{cati}) = (d(\mathbf{x},\mathbf{x}_N)/T_{achr}$, $d(\mathbf{y},\mathbf{x})/T_{achr})$, where $\mathbf{y}$ is the projection of $\mathbf{x}$ on the cylinder defined by the achromatic threshold (T_{achr} –the inner circle of panel a in Figure 6.3).

3'. Modify the previous probabilities by doing: $(p_{grey},p_{cat})=(0.5*(1+p_{grey}),0.5*p_{cat})$. This transformation ensures that all points inside the region delimited by the achromatic threshold will not be classified as belonging to the category p_{cat} alone.

Finally, in stage 3, the method assigns a descriptor to the computed pair of probabilities. Equation 6.1 shows how colour terms of the form BCT or BCT BCT or BCT BCT-ish are assigned from the computed (p_1,p_2) values.

$$Descriptor(p_1,p_2) = \begin{cases} cat_1cat_2 & |p_1 - p_2| < T_1 \\ cat_1\ cat_2\text{-ish} & T_1 < |p_1 - p_2| < T_2\ and\ p_1 > p_2 \\ cat_1\text{-ish}\ cat_2 & T_1 < |p_1 - p_2| < T_2\ and\ p_1 < p_2 \\ cat_1 & |p_1 - p_2| > T_2\ and\ p_1 > p_2 \\ cat_2 & |p_1 - p_2| > T_2\ and\ p_1 < p_2 \end{cases} \tag{6.1}$$

In Equation 6.1, descriptors are obtained according to the absolute distance between the p_1 and p_2 probabilities and a pair of thresholds (T_1 and T_2): if the distance is lower than T_1, we label the coloured sample with both colour categories; if the distance is larger than T_2, the coloured sample is only labelled with one colour category; and if the distance is between both thresholds, the coloured sample is labelled as belonging to both categories but an accent (-ish) is added to the one with smaller probability. Threshold parameters in Equation 6.1 satisfy $0 < T_1 < T_2 \leq 1$ and cat_i represents one of the BCT. See subsection 6.5.1 for a discussion on adequate threshold values.

Take for example the case where $(p_1,p_2) = (0.3,0.7)$ and between "yellow" and "orange". Their absolute distance is 0.4 and this is what determines which is the categorical relationship between both probabilities. The threshold T_2 determines the minimum absolute distance to classify the point as yellow or orange; if $T_2=0.5$ then the point of the example will be classified as belonging to *orange*. The threshold T_1 determines the maximum absolute distance to classify the point as belonging to both categories at the same time; if $(p_1,p_2) = (0.45,0.55)$ and $T_1=0.2$, then the point will be classified as *yellow-orange*. When the absolute value is between both threshold values the point will be classified as belonging to one category but complemented by the other category plus adjective. If $(p_1,p_2) = (0.35,0.65)$ and $T_1=0.2$ and $T_2=0.4$, then the point will be classified as *orange-yellowish*.

Notice that the structure defined by the Hue-Lightness map is isomorphic to variations of the cylinder radius (values larger than T_{achr}), which allows us to extend the categorization smoothly to the rest of the colour space (see subsection 6.6.1 below).

6.4 Data analysis: similarities between descriptors

In the colour naming part Experiment IV, of for each Mondrian colour each subject produced a set of three descriptors. We used the algorithm described in Appendix G, to match the answers obtained for the six different illuminations. Results from the colour naming tests indicate that observers never used more than two terms for each coloured sample, leading us to constrain all possible descriptors within the three following possibilities:

Descriptor: BCT | BCT BCT | BCT BCT-ish

BCT: 'Red' |'Green' | 'Blue |'Yellow' | 'Grey' |'Pink' | 'Purple' |'Orange' | 'Brown'

BCTish: 'Reddish' |'Greenish' | 'Bluish |'Yellowish' | 'Greyish' |'Pinkish' | 'Purplish' |'Orangish' | 'Brownish'

where *Descriptor* represents one answer for a given colour, the symbol '|' denotes different possibilities for this answer, and *BCT* and *BCT-ish* are representing the basic colour terms and their adjectival forms respectively.

In order to evaluate systematically how similar were the answers from two different illuminations, we developed a similarity index between a pair of descriptors, denoted as:

Similarity index = ⟨Descriptor$_1$, Descriptor$_2$⟩

This index has been defined for six possible combinations of descriptor pairs, namely, *archetypes*, which are described in Table 6.4. The index values were assigned following a two-fold rationale, first they had to be consistent with the rest of the cases in the same archetype and also to keep consistency with similar cases from different archetypes. Index values range between 0 and 1 which indicates null or complete similarity respectively.

Archetype		Rationale of cases				Index	Example
I. $\langle n_1, n_2 \rangle$	$n_1=n_2$	-	-	-		1	*<Green, Green>*
	$n_1 \neq n_2$	-	-	-		0	*<Green, Blue>*
II. $\langle n_1 m_1, n_2 m_2 \rangle$	$n_1=n_2$	$m_1=m_2$	-	-		1	*<Blue-Green, Blue-Green>*
		$m_1 \neq m_2$	-	-		0.5	*<Green-Blue, Green-Yellow>*
	$n_1 \neq n_2$	$m_1=m_2$	-	-		0.5	*<Blue-Green, Yellow-Green>*
		$m_1 \neq m_2$	-	-		0	*<Blue-Purple, Green-Yellow>*
	$n_1=m_2$	$m_1=n_2$	-	-		1	*<Blue-Green, Green-Blue>*
		$m_1 \neq n_2$	-	-		0.5	*<Blue-Green, Purple-Blue>*
	$n_1 \neq m_2$	$m_1=n_2$	-	-		0.5	*<Blue-Green, Green-Yellow>*
		$m_1 \neq n_2$	-	-		0	*<Blue-Green, Yellow-Orange>*
III. $\langle n_1, n_2 m_2 \rangle$	$n_1=n_2$	-	-	-		0.75	*<Green, Green-Yellow>*
	$n_1 \neq n_2$	$n_1=m_2$	-	-		0.75	*<Green, Yellow-Green>*
		$n_1 \neq m_2$	-	-		0	*<Green, Yellow-Orange>*
IV. $\langle n_1, n_2 a_2 \rangle$	$n_1=n_2$		-	-		0.875	*<Green, Green-Yellowish>*
	$n_1 \neq n_2$	$a_2 \simeq n_1$	-	-		0.125	*<Blue, Green-Bluish>*
		$a_2 \not\simeq n_1$	-	-		0	*<Blue, Green-Bluish>*
V. $\langle n_1 m_1, n_2 a_2 \rangle$	$n_1=n_2$	$a_2 \simeq m_1$	-	-		0.875	*<Blue-Green, Blue-Greenish>*
		$a_2 \not\simeq m_1$	-	-		0.5	*<Blue-Green, Blue-Purplish>*
	$n_1 \neq n_2$	$m_1=n_2$	$a_2 \simeq n_1$	-		0.875	*<Green-Blue, Blue-Greenish>*
			$a_2 \not\simeq n_1$	-		0.5	*<Blue-Green, Blue-Purplish>*
		$m_1 \neq n_2$	$a_2 \simeq m_1$	-		0.125	*<Green-Blue, Purple-Bluish>*
			$a_2 \simeq n_1$	-		0.125	*<Green-Blue, Purple-Greenish>*
			$a_2 \not\simeq n_1$	$a_2 \not\simeq m_1$		0	*<Green-Blue, Purple-Pinkish>*
VI. $\langle n_1 a_1, n_2 a_2 \rangle$	$n_1=n_2$	$a_1=a_2$	-	-		1	*<Green-Bluish, Green-Bluish>*
		$a_1 \neq a_2$	-	-		0.75	*<Green-Bluish, Green-Yellowish>*
		$a_1=a_2$	-	-		0.25	*<Green-Bluish, Purple-Bluish>*
		$a_1 \neq a_2$	-	-		0	*<Green-Bluish, Purple-Pinkish>*
	$n_1 \neq n_2$	$n_1 \simeq a_2$	$n_2 \simeq a_1$	-		0.75	*<Green-Bluish, Blue-Greenish>*
			$n_2 \not\simeq a_1$	-		0.125	*<Green-Bluish, Blue-Pinkish>*
		$n_1 \not\simeq a_2$	$n_2 \simeq a_1$	-		0.125	*<Green-Purplish, Blue-Purplish>*
			$n_2 \not\simeq a_1$	-		0	*<Green-Yellowish, Blue-Purplish>*

Table 6.4 A similarity index between two compound colour terms. The first column contains the archetype of the descriptors (n_1, n_2, m_1 and m_2 represent BCT; a_1 and a_2 represent BCT-ish and the $\not\simeq$ symbol indicates coincidence of BCT category), the following four columns show all possible cases and the sixth column contains the index value (see main text for details on how index values were assigned). Finally, the last column contains an example for the corresponding row case.

Archetypes were created using the following criteria:

Archetype I corresponds to the similarity between two single BCT, noted in Table 6.4 as $\langle n_1, n_2 \rangle$. Only two cases for Archetype I are possible: coincidence or not, and thus its values are 1 or 0 respectively.

Archetype II corresponds to the similarity between two compound colour terms which used only BCT, noted in Table 6.4 as $\langle n_1 m_1, n_2 m_2 \rangle$. Possible cases are: coincidence for both basic colour terms;

coincidence for one basic colour term and no coincidence for the other, with values of 1, 0.5 and 0 respectively.

Archetype III corresponds to the similarity between one BCT and a compound colour term conformed by two BCT, noted in Table 6.4 as $\langle n_1, n_2 m_2 \rangle$. This case only allows coincidence for one BCT or no coincidence at all, and thus the index values assigned are 0.75 and 0 respectively. A value of 0.75 may seem too high in terms of Archetype III values, but we selected it to keep inter-archetype consistency. Take for instance the case *<Green, Green-Yellowish>* of Archetype III and the previous one *<Green-Blue, Green-Yellow>* of Archetype II. If we think in terms of categorical/geometrical distances in colour space, the green centroid is closer to the green-yellow centroid than the green-blue and green-yellow are. From here it follows the need for a higher value for the *<Green, Green-Yellowish>* case.

Archetype IV is similar to Archetype III but instead of having two BCT in the compound term it has a BCT and a BCT-ish, noted in Table 6.4 as $\langle n_1, n_2 a_2 \rangle$. Here there are only three possible cases: categorical coincidence between both BCT, between the first BCT and the BCT-ish and no coincidence at all, with values of 0.875, 0.125 and 0 respectively. Notice here another case of inter-archetype consistency: take for instance the *<Green, Green-Yellowish>* archetype, following the same geometrical rationale as before we assigned it a higher index value than to *<Green, Green-Yellow>*.

Index values assigned to Archetype V and VI follow the same rationale described in the previous archetypes and cases (see *Rationale of cases* column in Table 6.4). Notice that for Archetype VI there exists a singular case where none of the BCTs coincide (e.g. *<Green-Bluish, Blue-Greenish>*). Here the previous rationale leads to a low index value (0.25), but the categorical quality of both colour terms is so close that we decided to assign a higher index value (0.75).

6.5 Results

6.5.1 Selected representatives and their repeatability

The top graph of each panel in Figure 6.4 shows the location in the a*b* plane of selected representatives chosen by each of the three observers (Panel a for JR; b for IR; and c for JV). Notice how these results are consistent with the selected representatives of Experiment II which are reported in Figure 4.3.
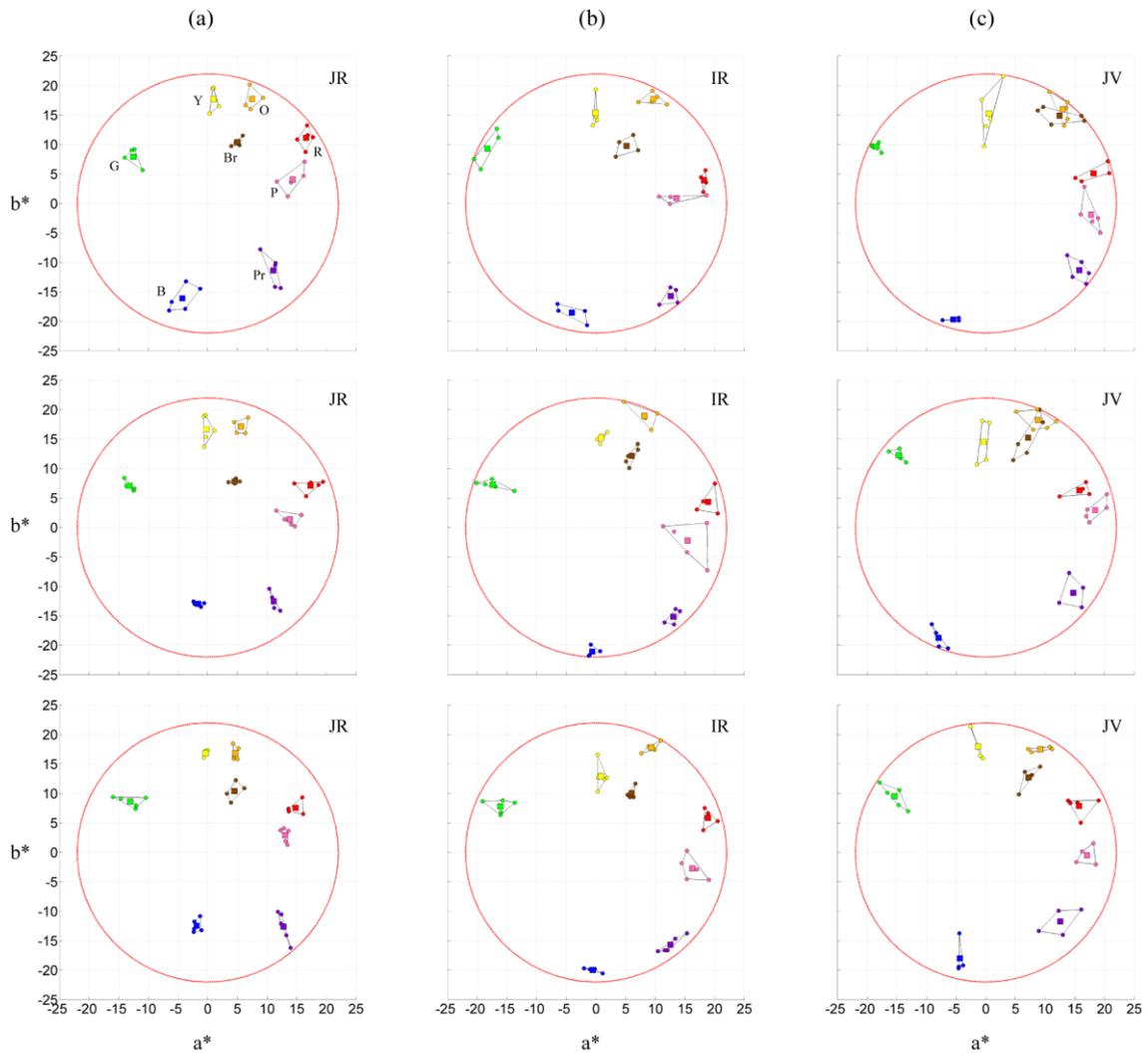
Figure 6.4 Chromatic settings and their repeatability in Experiment IV. The first row of each panel shows the selected representatives chosen by the three subjects in the reference session. The following rows show the corresponding settings for the three subsequent repeatability tests. Square markers represent the average of individual trials (small dots joined by lines) and the large red circle corresponds to the Bounding Cylinder in a*b* chromaticity plane.

As detailed in the methods section, observers participated on a series of sessions where their ability to reproduce the same colours was assessed over certain period. Middle and bottom rows in each panel of Figure 6.4 show the results of the repeatability sessions with the Bounding Cylinder in temporal sequence. We applied an ANOVA test to the hypothesis that population samples were drawn from the same mean (null hypothesis). In average, 79% (6% SD) of the tests failed to reject this null hypothesis, with significance p=0.05. Also, the average distance among same-category chromatic settings (squares in Figure 6.4) was very small: 2.06 ΔE* (1.07 SD). Appendix F contains the results for the repeatability tests without the Bounding Cylinder and, in that case statistical tests failed up to 65% (8% SD) to reject the null hypothesis. The average distance among chromatic settings was of slightly

higher than before: 2.22 ΔE* (1.38 SD). These results lead us to the conclusion that our three observers were able to reproduce approximately the same colours over the experimental period.

6.5.2 Chromatic settings under six different illuminants

Figure 6.5 shows the chromatic settings in CIELab colour space averaged over all three observers, discriminated by background type (squares and circles), and separated in panels according to the six illuminations. Although CIELab is defined in relationship to the illuminant at each adaptation state, the CIELab coordinates of these results were computed using the same reference white point (D65) in order to highlight the effect of illuminant shift (direction and magnitude), hence the displacement of the data in plots.

The precision of the chromatic settings was computed as the average ΔE* distance to the trials' mean for each colour category, as described in subsection 4.4.2. The average experimental error of chromatic settings was 3.2 ΔE* and no remarkable differences were found according to illuminants, colour categories or background types. Average time spend in each trial was 24.3 seconds (5.5 SD).

All six panels of Figure 6.4 show a consistent shift in the location of chromatic settings (linked by a black line) according to background type. This tendency may be summarized as a broadly uniform shift, in magnitude and direction (represented by the same colour-coded markers in each panel of Figure 6.4). This shift was modelled for each colour category as a vector (**v**) defined from the chromatic settings under background type III and IV (same colour-coded squares and circles). Its average magnitude was 7.2 ΔE* (2.1 SD) and its direction was computed as the angular distance between **v** defined under the reference illuminant and **v** defined under the test illuminants for each colour category. The average value of this angular distance was 30 degrees (20 SD). The extent and direction of this shift is consequence of the biased chromaticity of background type IV.
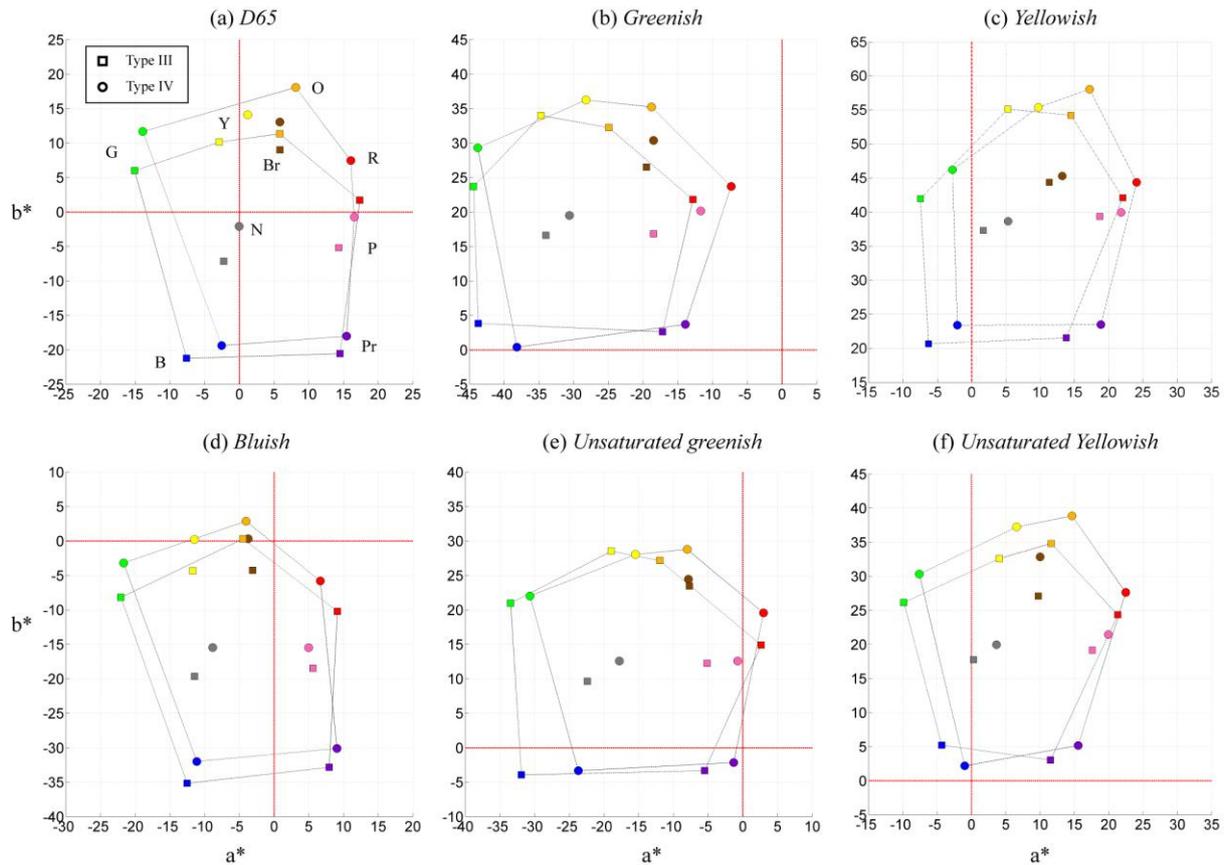
Figure 6.5 Chromatic settings averaged over all three observers for all six illuminants. Each panel corresponds to one particular illuminant (a to D65, b to greenish, c to yellowish, d to bluish, e to unsaturated greenish and f to unsaturated yellowish), and squares and circles correspond to background type III and IV respectively. Notice the consistent shift according to background type over all panels; this is due to biased overall chromaticity of background type IV. See main text for details.

Figure 6.6 shows the CIE 1976 uv coordinates of chromatic settings corresponding to the "grey" category for all observers and background/illumination conditions. As expected from the results of previous studies in successive colour constancy (Foster 2011) the location of this particular point lies close to the joining line between reference and test illuminants coordinates. Notice that its proximity to the coordinates of the test illuminants suggests a high degree of colour constancy. Also, the small shift for measurements done under D65 confirms global contrast effects (Webster 1996), a consequence of the biased average chromaticity of background type IV.

We quantified the extent of colour constancy for each chromatic setting computing the *Brunswick Ratio* between measurements done under both the reference (D65) and test illuminants. Our computations followed the same approach as described in subsection 4.4.3 and Equation 4.1. The overall average Brunswick ratio value was 0.83 (0.17 SD), confirming the existence of a high degree of colour constancy. The Brunswick ratio values over colour categories were broadly but not

116

completely uniform, it's variation was of 15% since their averaged standard deviation over categories within each illuminant was of 0.124 (coefficient of variation: 0.124/0.83=0.15). Also, in order to get a more comprehensive measure we computed the SCI and obtained an average value of 1.02 (0.19 SD) which suggest nearly complete colour constancy. Both index measures are in accordance with results shown in Figure 6.5.
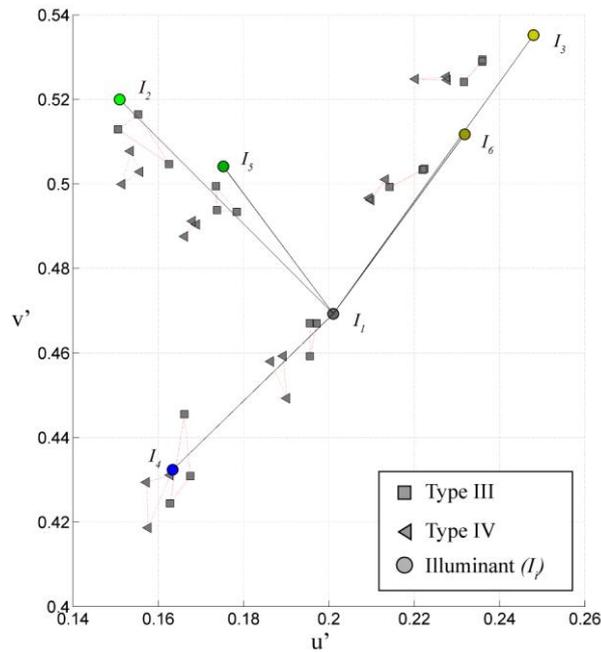
Figure 6.6 Coordinates of chromatic settings corresponding to *grey* in CIE 1976 uv colour space for all observers and background/illuminant conditions. Squares and circles correspond to background type III and IV respectively. Coloured circles indicate the illuminants' location and are colour-coded according to the colour names assigned to the illuminants in Table 6.2.

6.5.3 Interrelations among chromatic settings under different illuminants

Each set of chromatic settings in Figure 6.5 seems to somehow preserve its inter-distances, something that was also observed and quantified in Chapter 5 using *graphs*. We applied the same technique (see subsection 5.3) in order to find overall structural differences in the interrelations of chromatic settings under illumination changes. Notice that for our *graph* computations we did not use the CIELab coordinates shown in Figure 6.5 but the equivalent CIECAM02 coordinates and these computations were done on individual points instead of averages.

Initially, we computed graph distances for measurements under different background types and the same illumination. The average distance was only 0.04 (0.03 SD), suggesting that background types

had a small influence on the adaptation. Table 6.5 contains graph distances among all possible illuminant combinations; each of its values was computed from averaging over observers and background types. The average proportion of structural deformation when comparing to *D65* illumination (bold values in Table 6.5) is 0.15 (0.04 SD). The first column of Table 6.5 contains graph distances between the reference (*D65*) and test illuminants: 0.14 for *greenish*, 0.16 for *yellowish*, 0.14 for *bluish*, 0.12 for *unsaturated greenish* and 0.18 for *unsaturated yellowish*. Results for the *greenish* and *yellowish* illuminants are in agreement with SCI values reported in Chapter 4 and structural values in Chapter 5, which suggest a higher degree of structural disruption for chromatic settings under the *yellowish* illuminant. Interestingly, this is even more so for unsaturated illuminants.

Illuminant/Illuminant	I_1	I_2	I_3	I_4	I_5	I_6
I_1: *D65*	0	-	-	-	-	-
I_2: *Greenish*	**0.140**	0	-	-	-	-
I_3: *Yellowish*	**0.164**	0.166	0	-	-	-
I_4: *Bluish*	**0.136**	0.167	0.168	0	-	-
I_5: *Unsaturated greenish*	**0.120**	0.145	0.172	0.150	0	-
I_6: *Unsaturated yellowish*	**0.176**	0.175	0.134	0.178	0.165	0

Table 6.5 Graph distances among chromatic settings adjusted under different illuminants. Results are averaged over observers and background types. The proportions of structural deformation when compared to chromatic settings under *D65* illumination are highlighted in bold. Graph and graph distances were computed as in Chapter 5.

Despite that Table 6.5 suggests a high degree of stability (85% in average) in the structure of chromatic settings, the graph computations could hide cancellations among relatively small chromatic settings movements. Also, results from our colour naming task revealed the existence of categorical changes (see subsection 6.4.4).

We further analysed our data in CIECAM02 colour space by removing the saturation dimension from the analysis and focusing only in hue and luminance. The hue dimension was computed for each chromatic setting as the contraclockwise angle between the horizontal and the vector defined between "grey" and the chromatic setting considered. The abscissas in Figure 6.7 correspond to these hue values, and the ordinates correspond to the lightness. Each panel corresponds to a different background type; panel a to type III, panel b to type IV, and panel c to the averaged chromatic settings of types 0, I and II (obtained from Experiment II).
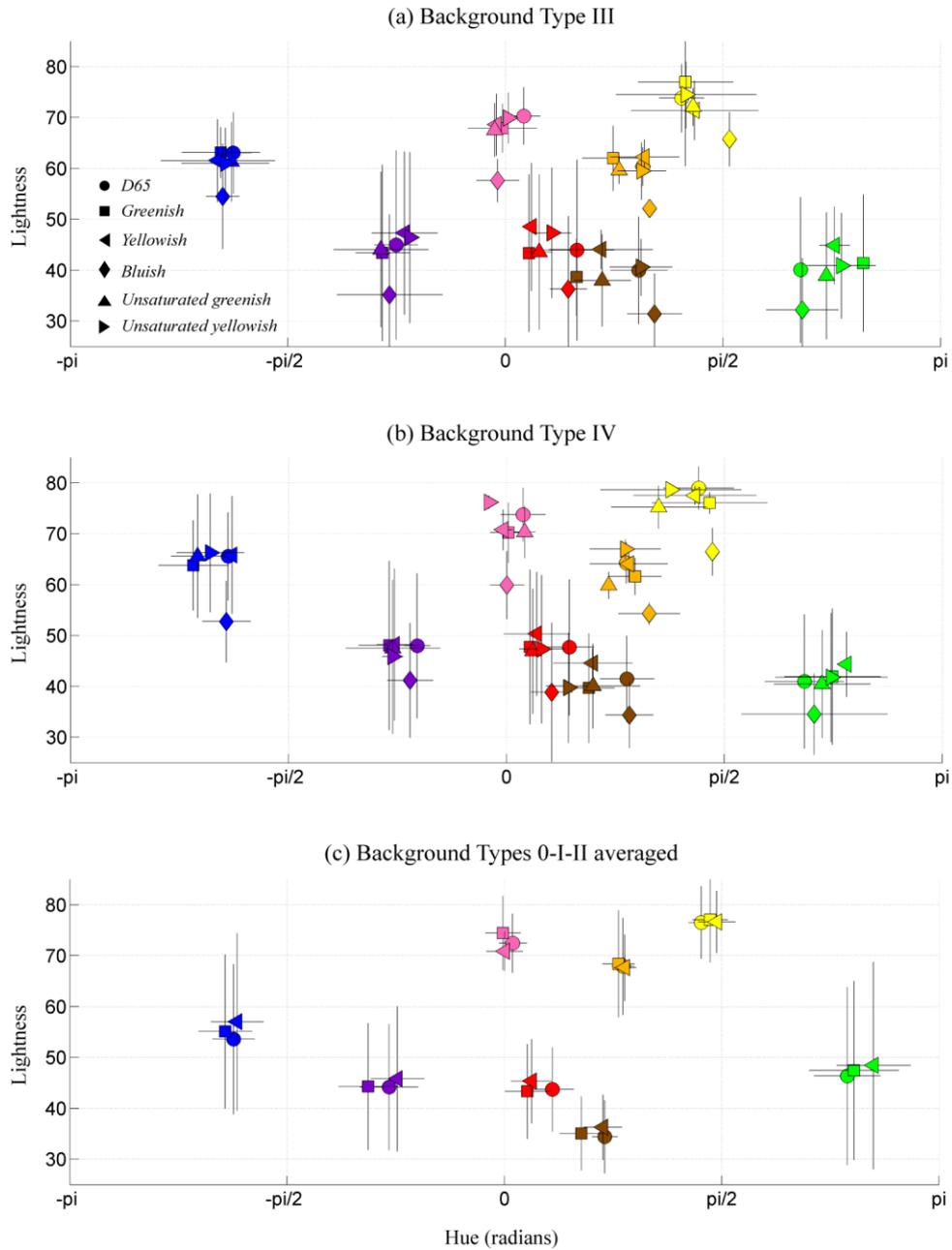
Figure 6.7 Hue and lightness differences in chromatic settings due to illuminant changes computed in CIECAM02. Markers are colour-coded according to the chromatic setting colour category and their shape indicate the adaptation illuminant. Each panel corresponds to results obtained under different backgrounds (see main text for details). Crosses indicate standard deviation.

Three observations emerge from Figure 6.7. First, lightness information is not a distinctive factor despite that chromatic settings under the bluish illumination (diamonds in panels a and b) have lower lightness values, probably due to the lower intensity of the bluish illuminant. Second, panels a and b show remarkable hue differences according to the illuminants for some colour categories. Third, the hue range of some chromatic settings under the *greenish* and the *yellowish* illuminants is larger (panels

119

a and b than in panel c), confirming that colour constancy is *weakened* when the number of colours in the stimuli is reduced (Linnell and Foster 2002).

Variations in Hue in Figure 6.7 were summarized by computing the difference (in degrees) between the corresponding chromatic setting hue under the reference and test illuminants (i.e. between the circle marker and the others in Figure 6.7). Thus, when averaged over observers and backgrounds and colour categories, the angular variations were: 11.8º (7.5 SD) for *greenish*, 8.9º (6.2 SD) for *yellowish*, 5.6º (4.8 SD) for *bluish*, 9.9º (5 SD) for *unsaturated greenish*, and 8.5º (6.6 SD) for *unsaturated yellowish*.

6.5.4 Usage of colour terms in the colour naming test

Table 6.6 contains the results of the colour naming test; a total of 108 samples (*descriptors*), 54 for each background type (each of the three observers named 3 colours under 6 illuminants). Rows in bold correspond to the descriptors assigned under the reference illumination (*D65*). As reported previously, correspondences between descriptors were confounded over sessions in the colour naming test. For this reason, the descriptors in Table 6.6 were previously sorted by the algorithm described in Appendix G (descriptors were reordered in a way that maximized their categorical coincidence with the colours seen under the reference illuminant).

Obs		JR			IR			JV	
Illum	C_1^1	C_2^1	C_3^1	C_1^2	C_2^2	C_3^2	C_1^3	C_2^3	C_3^3
I_1	**B**	**G**	**P,Pr-ish**	**B**	**G**	**Pr,P-ish**	**B,R-ish**	**G,N-ish**	**P,Pr-ish**
I_2	B,G	G,Y-ish	N	B,G	G	N,O-ish	B,O-ish	G	N,O-ish
I_3	N	G,Y-ish	P,R-ish	B,N-ish	G,Y-ish	P,O-ish	N	G	R-P
I_4	B	G	Pr,N-ish	B	G	Pr,N-ish	B	G	Pr,N-ish
I_5	B,G	G	Pr,P-ish	B	G	N,Pr-ish	B,R-ish	G	N,R-ish
I_6	N,B-ish	G,Y-ish	P,R-ish	B	G	P	N,B-ish	G	P,R-ish

Illum	C_4^1	C_5^1	C_6^1	C_4^2	C_5^2	C_6^2	C_4^3	C_5^3	C_6^3
I_1	**B, N**	**G,N-ish**	**P,R-ish**	**B**	**G**	**P,N-ish**	**B**	**G,N-ish**	**R,P-ish**
I_2	B,G	G	Pr,R-ish	B,G	G ,Y-ish	G	B,O-ish	G	R,N-ish
I_3	N,Y-ish	G,Y-ish	R	B,N-ish	G	O,P-ish	B,N	N,B-ish	R,P-ish
I_4	B	G,N-ish	Pr,R-ish	B	G	Pr,N-ish	B	G,B-ish	R,Pr-ish
I_5	B,G	G,N	Pr,P-ish	B	G	Y,P-ish	B,O-ish	G	R,Br-ish
I_6	N	G,N-ish	P,R-ish	B	G	P,O-ish	B,N-ish	G,N-ish	R,P-ish

Table 6.6 Results of the colour naming test for all three observers. Rows correspond to illuminants and background types. Columns correspond to observers and Mondrian colours. In bold are the results obtained under *D65* illumination. Notice that colours reported were previously sorted according to the method introduced in Appendix G.

Notice that observers used at most two colour terms to describe the Mondrian colours and their usage pattern was limited to the following cases: [BCT] or [BCT, BCT] or [BCT, BCT-ish]. Figure 6.8

contains histograms which describe the usage pattern of these descriptors. Observers used in average 8.8 times the [BCT] descriptor, 7.5 times the [BCT, BCT-ish] descriptor, and only 1.7 times the [BCT, BCT] descriptor. Panels b and c show the same information as in panel a but this time sorted according to illuminants and background types. When compared across background types, histograms in panels b and c are similar but there are some differences regarding illuminants; the [BCT] descriptor was mostly used under *greenish* and *bluish* illuminations while under *yellowish*, *unsaturated greenish* and *unsaturated yellowish* subjects preferred to use the [BCT, BCT-ish] descriptor.
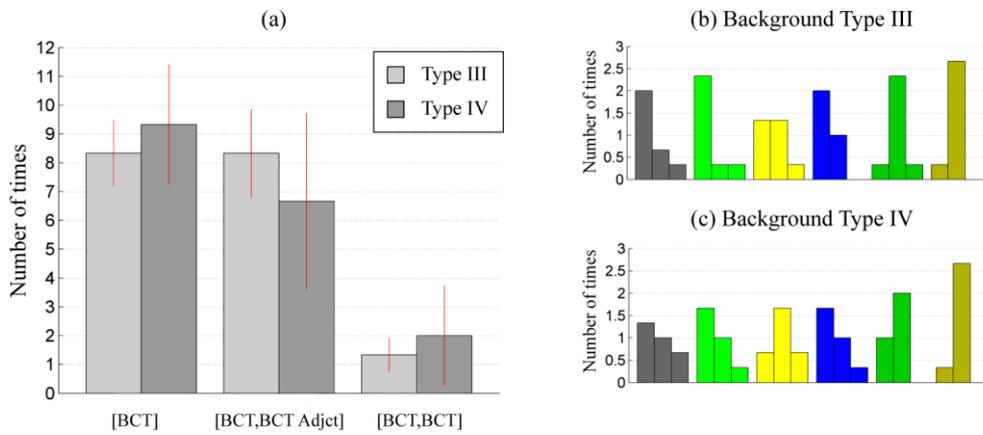
Figure 6.8 Usage pattern of colour terms in the colour naming test. Panel a: the histogram shows the number of times that each descriptor was used, averaged over observers and illuminants. Panels b and c show the same information but sorted according background types and colour-coded according to illuminants.

If colour constancy was complete then descriptors in Table 6.6 would be the same in each column, but this is not the case. We quantified these categorical variations using the similarity index introduced in subsection 6.3.2, i.e., we computed the index between the colour term assigned under D65 illumination and the colour term assigned under the test illumination. For instance, the colour assigned by observer IR to C_1 colour under I_1 was *green* and under I_2 was also *green*, then the similarity index value was 1. Table 6.7 contains the values of the similarity index for all colours. The rows in bold contain the averaged values according to each background type.

The average of all values in Table 6.7 is 0.66, indicating that categorical colour perception was mostly stable for the three-coloured backgrounds tested. However, a close analysis reveals that this categorical stability only hold for colours C_1 (0.60), C_2 (0.92), C_4 (0.91) and C_5 (0.81), and not for colours C_3 (0.32) and C_6 (0.39) which have a much lower index of categorical similarity. In other words, green and blue were the most stable colours when illumination changed and combinations of pink, red and purple under D65 illumination were the less stable.

Obs	JR			IR			JV		Mean
Illum	c_1^1	c_2^1	c_3^1	c_1^2	c_2^2	c_3^2	c_1^3	c_2^3	c_3^3	
I_2	0.75	0.875	0	0.75	1	0	0.875	0.875	0	**0.57**
I_3	0	0.875	0.75	0.875	0.875	0.125	0	0.875	0.5	**0.54**
I_4	1	1	0.125	1	1	0.75	0.875	0.875	0.125	**0.75**
I_5	0.75	1	0.75	1	1	0.125	0	0.875	0	**0.61**
I_6	0.125	0.875	0.75	1	1	0.25	0.125	0.875	0.75	**0.62**
Mean	**0.52**	**0.92**	**0.47**	**0.92**	**0.97**	**0.22**	**0.37**	**0.87**	**0.27**	**0.62**

(Back III)

Illum	c_4^1	c_5^1	c_6^1	c_4^2	c_5^2	c_6^2	c_4^3	c_5^3	c_6^3	
I_2	0.875	0.5	0	0.875	0.75	0	0.75	1	0.875	**0.62**
I_3	0.75	0.5	0.125	1	0.875	0.125	1	0.75	0.125	**0.58**
I_4	1	0.75	0	1	1	0	0.75	1	0.75	**0.69**
I_5	0.875	0.5	0.125	1	1	0.125	0.75	1	0.875	**0.69**
I_6	1	0.75	1	1	1	0.75	1	0.875	1	**0.93**
Mean	**0.90**	**0.60**	**0.25**	**0.97**	**0.92**	**0.20**	**0.85**	**0.92**	**0.72**	**0.71**

(Back IV)

Table 6.7 Similarity index computed between the colour naming results of *D65* and test illuminations. This table quantifies the degree of categorical change from Table 6.6 (see main text for details). Averaged values according to background colours (rows) or illuminants (last column) are in bold.

6.5.5 Asymmetries in categorical colour constancy

In this section we further studied the extent and quality of categorical colour constancy; in particular we focused on the relationship between the perception of background colours under the reference illumination and the test illumination once adaptation was discounted. To do so, we calculated the LMS coordinates of the colours perceived after adaptation using a linear model of colour constancy, i.e., chromatic settings under the test and reference illuminants were used to fit a *diagonal plus translation* matrix which predicted the LMS coordinates of background Mondrian colours as if they were under the reference illumination, (see subsection 4.4.4). This procedure was applied to each observer and adaptation state.

Figure 6.9 shows the results of such approach for each observer (rows) and background type (columns). Each plot shows the location in the a*b* plane of the three Mondrian colours as predicted by the *diagonal plus translation* model, coded by markers which discriminate by illuminant, and colour-coded according to the descriptors assigned in the colour naming test (see figure caption for details). Colours in Figure 6.9 are linked to their reference colour (as seen under D65) by a black line.
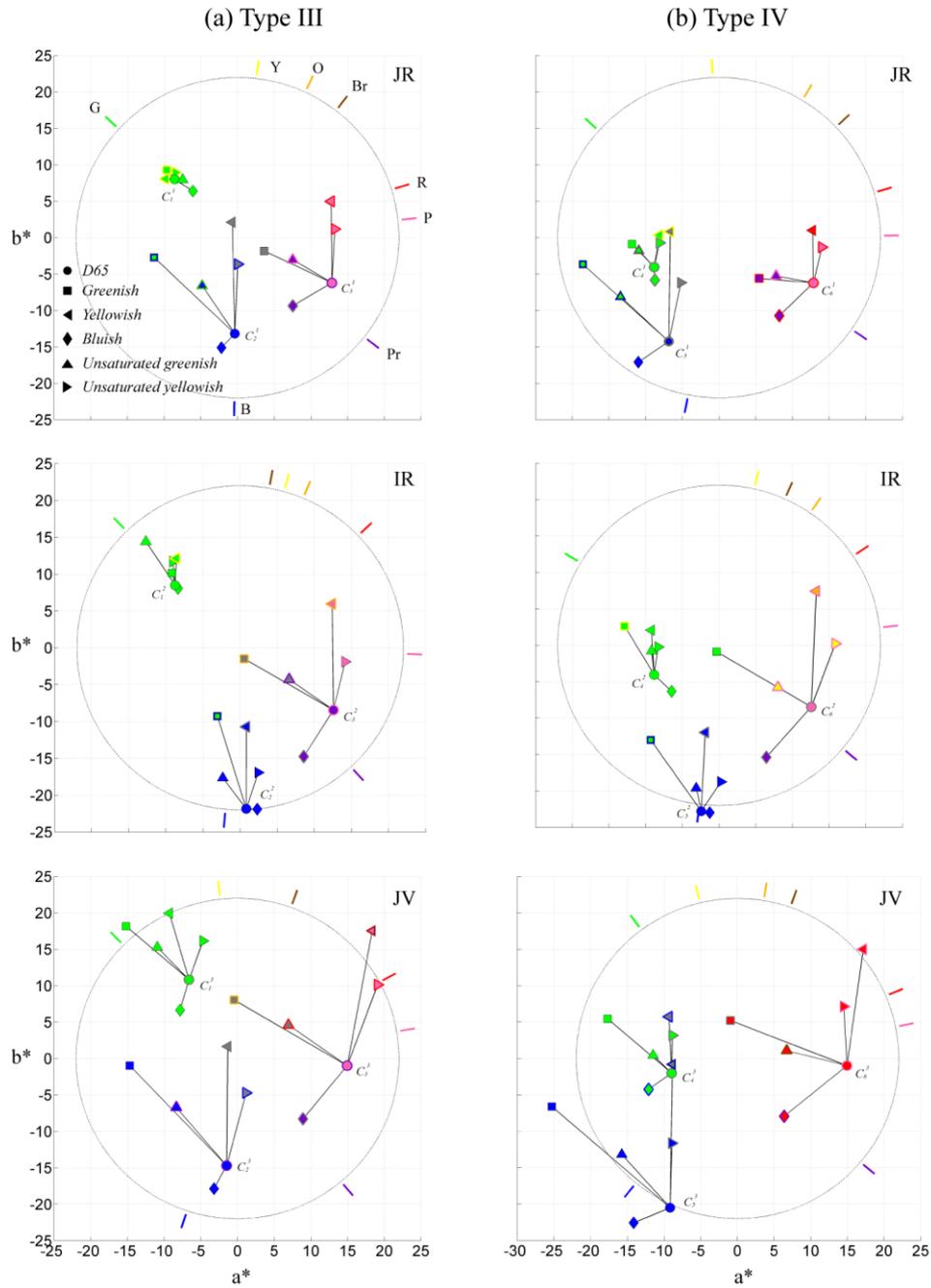
Figure 6.9 Background colours once adaptation was discounted, as predicted by the *diagonal plus translation* model. Plots on the left correspond to Type III and plots on the right to Type IV backgrounds. The three observers are represented by the different rows. In each graph different markers indicate the colours perceived under different adaptation and their colour-coding is related to the colour naming results. We represented [BCT] descriptors with one colour; [BCT, BCT] descriptors with two colours; and [BCT, BCT-ish] descriptors with one colour plus a coloured border. The short lines located outside the circle represent the hue of the chromatic settings under D65. Black lines join the colours perceived after adaptation to a coloured illuminant to the same colours as seen under the reference D65 illumination (circles). See Table 6.1 for numerical values.

From Figure 6.9, notice how the model-predicted colours (non-circle markers) are distributed around the *expected* "perfect constancy" location (circle). If adaptation was complete, all markers would coincide with the circles. We quantified this failure of constancy by computing ΔE^* distances between the model-predicted colours and their expected location. Table 6.8 contains these distances summarized according to illuminants and background types.

Illum / Colour	Type III			Type IV			Mean
	C_1	C_2	C_3	C_4	C_5	C_6	
Greenish	5.16	15.85	14.15	8.02	16.54	13.31	**12.17**
Yellowish	5.24	14.49	14.86	6.38	15.45	13.16	**11.60**
Bluish	3.98	4.31	9.14	3.92	4.44	10.03	**5.97**
Greenish unsaturated	5.08	8.01	7.84	3.61	7.62	6.47	**6.44**
Yellowish unsaturated	3.43	8.47	8.72	4.19	7.54	7.58	**6.65**
Mean	**4.58**	**10.23**	**10.94**	**5.22**	**10.32**	**10.11**	**8.57**

Table 6.8 Distance (ΔE^*) between the reference and predicted background colours. This table contains the observer-averaged values of the distances highlighted with back lines in Figure 6.9. As usual ΔE^* distances were computed using D65 as a reference white point.

Table 6.8 shows two interesting patterns: (1) the last column reveals that distances (failures of constancy) increase with illuminant saturation and (2) the last row reveals that Mondrian colours C_1 and C_4 kept approximately the same distance to their reference throughout the experiment.

The coloured lines located outside the circle in Figure 6.9 indicate the hue of the chromatic settings obtained under *D65* illumination. A visual inspection reveals the agreement between the position of these hues and the markers' colours, i.e. between the results of the chromatic setting and the colour naming task in the hue dimension.

6.5.6 Simulating categorical colour perception with CCCP

The colour naming task described above only provided categorical information for three points in colour space under each adaptation state, and its results suggest that categorical colour constancy was not complete. The CCCP (introduced in subsection 6.3.3), allows us to expand this categorical information to any point in colour space. Since the parameters of CCCP are chromatic settings, its predictions are more reliable near the central part of colour space. Due to the similarity between adaptation under background type III and IV (see Figure 6.9), we restricted our analysis to background type III.

Achromatic locus

We considered the colours belonging to the achromatic locus and predicted their categorical colour appearance under different adaptations. To do this, the "achromatic" axis was sampled by seven points which were equally spaced between 40 and 70 L* units. Tristimulus XYZ coordinates of these points were produced from the spectral distributions of the illuminants, a flat spectral reflectance function and the CIE colour matching functions (see subsection 4.3.3).

Figure 6.10 shows the categorical colour predictions obtained by CCCP for all observers (panels). Columns labelled as I_i correspond to CCCP predictions for samples a_i obtained from the XYZ values described above. Columns labelled as A_i correspond to CCCP predictions for the samples a_i described above and transformed via the diagonal plus translation model into perceived colours. All CCCP predictions in Figure 6.10 are colour-coded using circles following the same convention as in Figure 6.9: [BCT] descriptors were represented with one colour; [BCT, BCT] descriptors with two colours; and [BCT, BCT-ish] descriptors with one colour plus a coloured border. The first column of each panel contains the categorical colour prediction for the "achromatic" points a_i sorted according to decreasing lightness from top to bottom. The next five pairs of columns contain side by side predictions for I_i and A_i colours. CCCP parameters (see section 6.5) were: T_1=0.25, T_2=0.5 and T_{achr}=10. For details on how they were chosen see subsection 6.6.1.
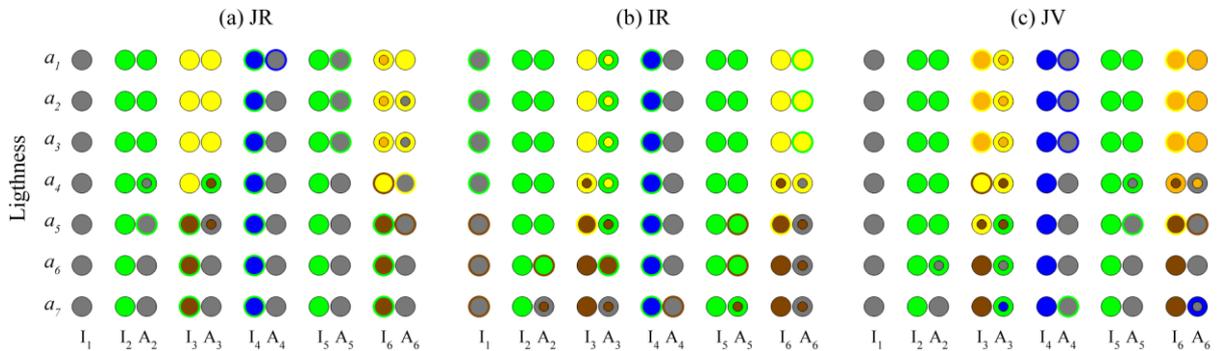
Figure 6.10 Categorical colour predictions for the achromatic axis. Each panel corresponds to one observer and coloured circles are coded as in Figure 6.9. Each column corresponds to CCCP predictions for this achromatic axis under a different illumination (I_i), or after adaptation (A_i). See text for further details.

We used the categorical similarity index from section 6.4 to quantify the categorical relations in Figure 6.10. In particular we studied whether descriptors A_i were more similar to their corresponding I_1 (D65) descriptor than to their corresponding test illuminant I_i (i= 2, 3, 4, 5) descriptor or vice versa. Categorical comparisons were computed for each pair of points with the same lightness. Table 6.9 shows these results ($<I_1, A_i>$) averaged over all three observers and it reveals two tendencies: (1)

illuminants with higher saturation produce a lower similarity index, and (2) points with lower lightness produce a higher similarity index.

Colours /Illum	Greenish	Yellowish	Bluish	Greenish unsaturated	Yellowish unsaturated	Mean
a_1	0.04	0.04	0.87	0.33	0.00	**0.26**
a_2	0.04	0.04	0.92	0.33	0.25	**0.32**
a_3	0.04	0.04	0.92	0.33	0.25	**0.32**
a_4	0.29	0.04	0.96	0.62	0.71	**0.52**
a_5	0.29	0.29	0.96	0.62	0.87	**0.61**
a_6	0.58	0.62	0.96	0.67	0.96	**0.76**
a_7	0.96	0.62	0.96	0.71	0.96	**0.82**
Mean	**0.32**	**0.24**	**0.93**	**0.52**	**0.56**	**0.51**

Table 6.9 Categorical similarity index between descriptors A_i and their corresponding I_1 (D65). $<I_1, A_i>$ index values were computed from the predictions shown in Figure 6.10 (see main text). Rows correspond to lightness and are sorted from top (higher) to bottom (lower). Values were observer-averaged. Notice that this similarity index may act a categorical colour constancy index. Repeated coefficients are due to the discontinuous nature of the similarity index coefficients in Table 6.4.

The analysis performed in Table 6.9 was reproduced in Table 6.10, but instead of comparing A_i descriptors to the reference illumination descriptors I_1 we compared them to their corresponding test illuminant I_i (i= 2, 3, 4, 5) descriptors, obtaining $<I_i, A_i>$. Notice how the illuminant-averaged results in the last column of both tables are, in broad terms, inversely correlated.

Colours /Illum	Greenish	Yellowish	Bluish	Greenish unsaturated	Yellowish unsaturated	Mean
a_1	1.00	0.87	0.08	0.71	0.83	**0.70**
a_2	1.00	0.87	0.04	0.71	0.75	**0.67**
a_3	1.00	0.87	0.04	0.71	0.75	**0.67**
a_4	0.92	0.46	0.00	0.58	0.37	**0.47**
a_5	0.71	0.50	0.00	0.33	0.25	**0.36**
a_6	0.54	0.29	0.00	0.29	0.25	**0.27**
a_7	0.00	0.25	0.00	0.25	0.25	**0.15**
Mean	**0.74**	**0.59**	**0.02**	**0.51**	**0.49**	**0.47**

Table 6.10 Categorical similarity index between descriptors A_i and their corresponding I_i (i= 2, 3, 4, 5) descriptors. $<I_i, A_i>$ index values were computed from the predictions shown in Figure 6.10 (see main text). Rows correspond to the lightness and are sorted according to bottom(lower) to top(higher). Values were observer-averaged.
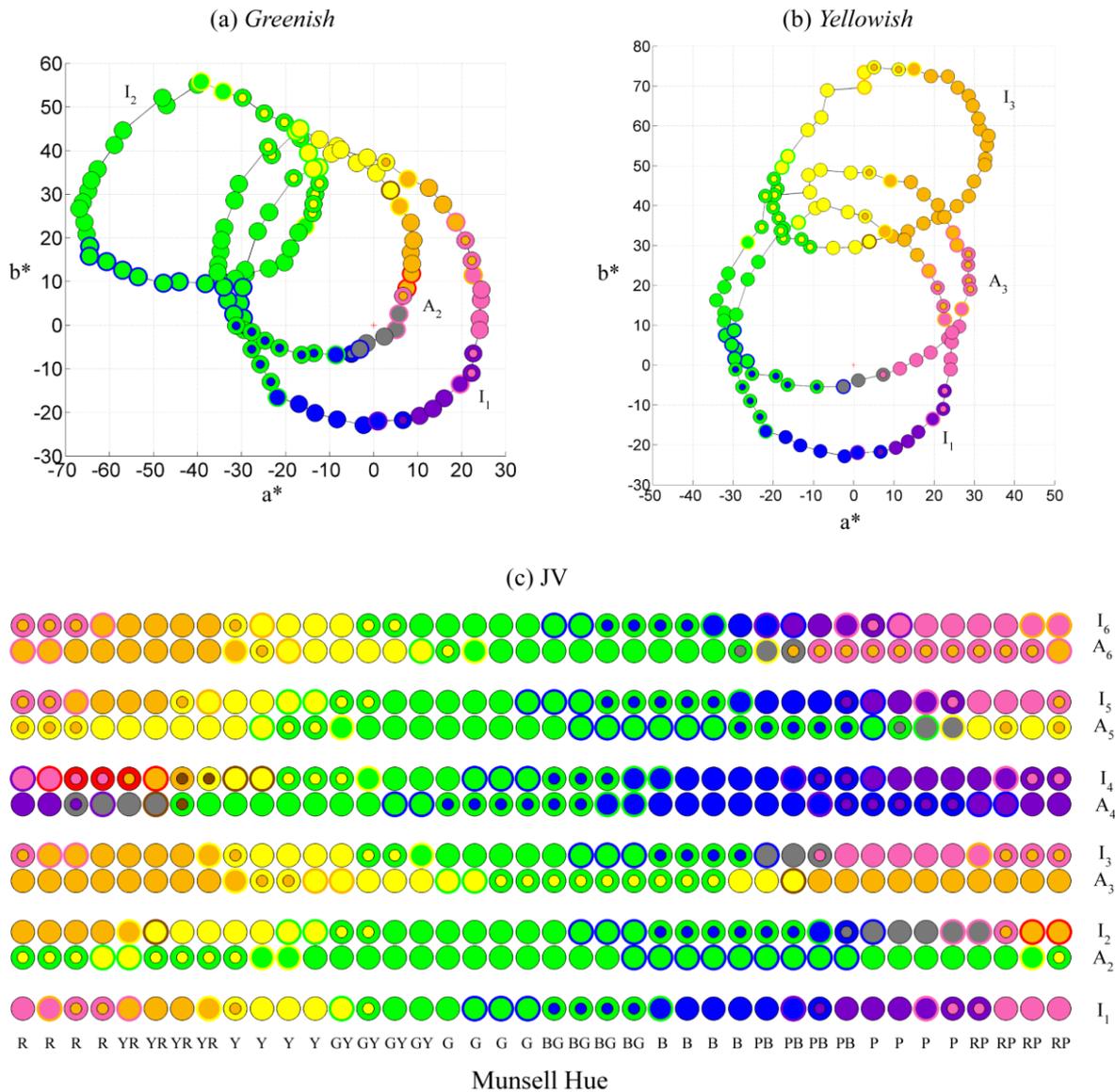
Figure 6.11 Categorical colour predictions for the Munsell samples of Value and Chroma equal to 6, for observer JV and background type III. Round markers follow the same colour-coding as those of Figure 6.10. Panels a: The *circle* labelled as I_1 contains CCCP predictions for descriptors under D65, the circle labelled as I_2 contains the predictions for descriptors under greenish illumination, and circle labelled as A_2 contains the predictions for descriptors of perceived colours after adaptation. Panel b shows the same information as panel a but for yellowish illuminant. Panel c summarises the categorical predictions for all illuminations. Horizontal labels in panel c contain the Munsell coordinates for hues corresponding to 2.5, 5, 7.5, and 10.

Munsell Circle

We further investigated the categorical adaptation of our observers by reproducing the previous analysis on a representative set of Munsell (Fairchild 2005) samples. We considered the Munsell samples with *Value* and *Chroma* equal to 6 which occupy a central part of the Munsell colour space, and consequently have a rich categorical representation. We reproduced the previous analysis on these

127

colours for observer JV. The results are shown in Figure 6.11 following the same colour-coding and organization as Figure 6.10. Panel a shows CCCP predictions for I_1 and I_2 descriptors for 40 Munsell samples under *D65* and *greenish* illumination, and predictions for A_2 descriptors under the *greenish* illumination. Panel b has the same organization as panel a, but shows categorical predictions for the *yellowish* illumination. The information contained in Panels a and b is also displayed with a different arrangement in the five bottom rows of Panel c. These results are summarized and further discussed in subsection 6.6.2.

6.6 Discussion

6.6.1 The parameters of the CCCP

The Customised Categorical Colour Predictor depends on two sets of parameters: an achromatic threshold (T_{achr}) which limits the region considered as achromatic, and two categorical thresholds (T_1 and T_2) which discriminate between pairs of descriptors. We studied the influence of these parameters on the model performance by: (1) evaluating how well data from the colour naming task (see section 6.2.2) is predicted and; (2) comparing CCCP predictions of the central part of colour space with those of a standard colour naming model (Benavente, Vanrell et al. 2008).

Test 1:

The first test has a twofold purpose: it measures the model's consistency with colour naming data and helps to select the best parameters for the model. The approach considered a partition of the parameters' space (T_{achr} ranging from 0 to 20 in steps of 1 and T_i ranging from 0 to 1 in steps of 0.01, $T_1<T_2$) and for each of its points we computed an average similarity index between data from the colour naming test (Table 6.6) and their CCCP predictions. This similarity index was computed by averaging the results of all similarity index cases between colour naming data and their categorical colour predictions once adaptation was discounted (Figure 6.11 can be interpreted as showing the location of Munsell colours after adaptation was discounted, in terms of colour naming results).

Figure 6.12 shows two visualizations of these results: panel a shows the influence of the achromatic threshold parameter T_{achr} on the similarity index (error bars indicate the variability induced by the other two threshold parameters); and panel b shows the variability of the similarity index according to the value of T_1 and T_2 once T_{achr} was fixed and equal to 10 ΔE^*. Panel a shows how the prediction's accuracy decreases as we increase the value of the achromatic threshold, i.e., increasing the size of the region corresponding to grey. Consequently, we selected a value of 10 as our achromatic threshold

T_{achr} since this is the last value before the slope of the curve in panel a decreases markedly. It also implies that values within 5 ΔE* will be classified as "grey" by CCCP, in agreement with previous studies (Benavente, Vanrell et al. 2008). Panel b shows the influence of threshold parameters from the surface resulting from assigning the similarity index value to all possible threshold values ($0 < T_1 < T_2 < 1$) keeping T_{achr} fixed and equal to 10. Surface values range between 0.65 and 0.78 with a mean value of 0.72 (0.03 SD). These high values demonstrate that CCCP succeeded, at least for the majority of the samples we tested. Also, notice the smooth properties of the surface, i.e., it is continuous and differentiable and with only one local maximum, thus supporting the consistency of our approach (the maximum of the surface shown in panel b corresponds to (0.45, 0.5)). Although a value of $T_2=0.5$ seems reasonable, a value of the $T_1=0.45$ seems slightly high, at least from the results of panel a in Figure 6.8 which indicate a more extended usage of BCT adjectives. These results might be an artefact of the small number and limited location of the samples we tested, and thus we tried the more canonical value of 0.25 for T_2.
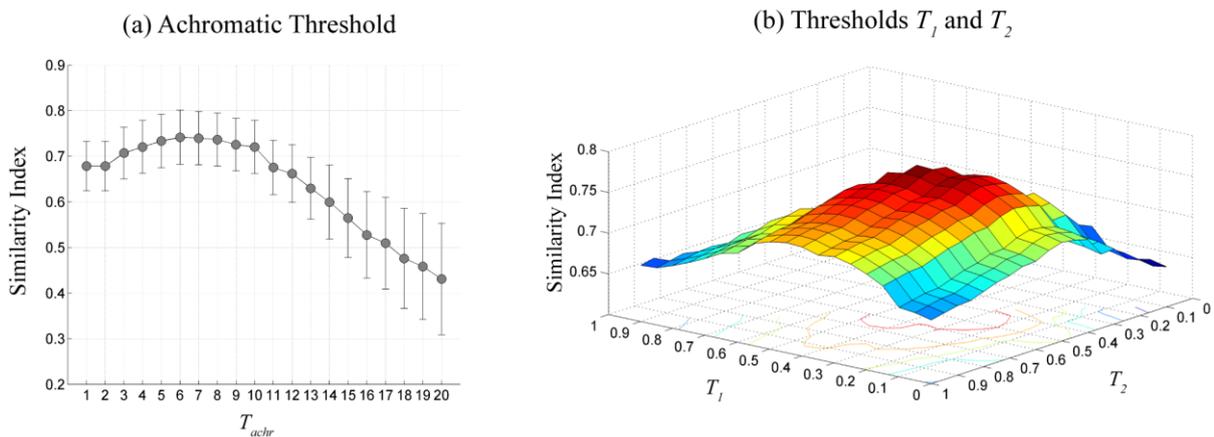
Figure 6.12 Influence of parameters on the accuracy of CCCP. Panel a: accuracy of the similarity index according to the value of the achromatic threshold T_{achr}. Black error bars indicate variation over thresholds T_1 and T_2. Panel b: Variation of the similarity index according to values of T_1 and T_2 once T_{achr} was fixed and equal to 10. See text for further details on how the similarity index was computed.

Test 2:

Our second test studied the shape of the categorical regions generated by the CCCP. To do so, we considered a partition of the central part of the CIELab colour space (due to the bounded nature of chromatic settings) and applied to these points our categorical prediction model. The first row of each panel in Figure 6.12 shows CCCP predictions for three constant-lightness cross sections (L*=40, 55 and 60) for one particular observer (JR). From the top panels of Figure 6.12 we can observe the

following: (1) BCT chromatic regions have a radial distribution, as expected from the particularities of the model's definition; (2) the sizes of chromatic regions agree with Chapter 3 results (green and blue are the largest regions); (3) there are marked boundary regions such as green-yellow, green-brown, green-blue, etc; (4) there is a small central region for grey. Bottom panels of Figure 6.13 represent the same region of colour space as top panels but with categorizations assigned following the fuzzy-sets model of colour naming (Benavente, Vanrell et al. 2008). Visual inspection of the top and bottom rows for each lightness level in Figure 6.13 confirms the consistency of our approach, the main difference being the size of the border regions. These results agree with the results of Chapter 3 which reported the existence of larger border regions.
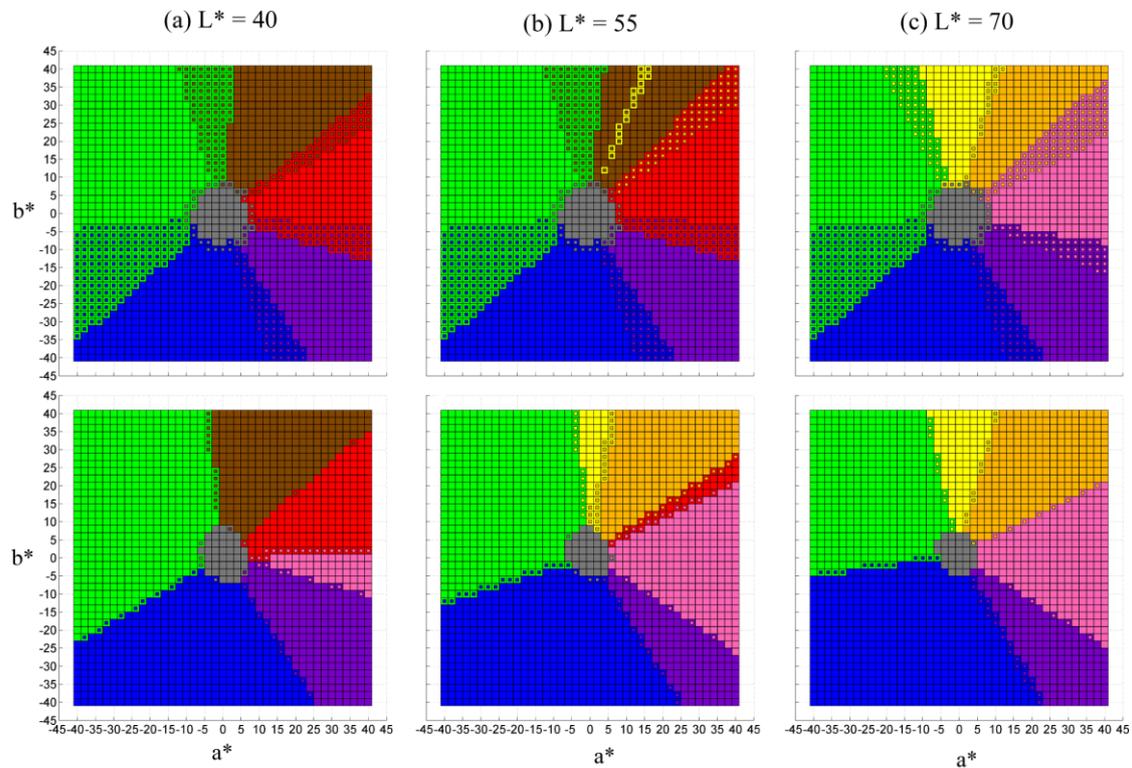
Figure 6.13 Categorical colour predictions for observer JR. We sampled the inner part of CIELab colour space in steps of 2 JND units for all three dimensions and predicted its categorical colour appearance using CCCP. Top plots correspond to observer JR under *D65* illumination and background Type III. Top and bottom panels compare categorical classifications from CCCP and the fuzzy-sets colour naming model of Benavente *et al* (Benavente, Vanrell et al. 2008). Markers are colour-coded according to the categorical colour prediction results for each point. Single-coloured markers correspond to one BCT and two-coloured markers correspond to two BCTs. Each panel shows a different lightness cross-section of colour space.

The availability of only nine chromatic settings in colour space restricts the theoretical framework of CCCP, something that constrains the shape of the regions in Figure 6.13. Interestingly, a more

130

complex approach such as that of Benavente *et al* (Benavente, Vanrell et al. 2008), did not produce remarkable qualitative differences except perhaps in the border regions. However, since our model is customised to each observer, quantitative differences exist in the hue locations and also regarding the extension of BCT regions.

6.6.2 Categorical constancy and inconstancy in colour adaptation

Section 6.5.1 showed that observers were able to perform the chromatic settings with considerable precision; with an average inter-session error of 2.22 ΔE* and intra-session error of 3.2 ΔE*. Table 6.11 summarizes relevant results from the previous sections. The first row of Table 6.11 contains the magnitude of the chromaticity shift in ΔE* units for each illuminant (see Table 6.2), the following rows contain results from previous analysis from the chromatic settings: the proportion of structural deformation (see Table 6.5), displacement of hues in degrees (see subsection 6.5.3), categorical similarity from the colour naming test (see Table 6.7), distances between reference and adapted colours (see Table 6.8) and the CCCP predictions for the achromatic axis (see Table 6.9). Notice how all these measurements, except for the proportion of structural disruption, correlate with the magnitude of the illuminant shift. As Table 6.9 shows, increasing the magnitude of the illuminant shift implies larger hue disruption, bigger distances between adapted and reference colours, and less categorical similarities between tested and simulated colours.

Result / Illum.	*Greenish*	*Yellowish*	*Bluish*	*Unsaturated greenish*	*Unsaturated yellowish*
Illuminant shift (ΔE*) (Table 6.2)	43.5	48	20.5	24.4	26.6
Proportion of structural disruption (Table 6.5)	0.14	0.16	0.14	0.12	0.18
Hue displacement (deg) (Subsection 6.5.3)	11.8	8.9	5.6	9.9	8.5
Categorical similarity from the colour naming test (Table 6.7)	0.59	0.56	0.72	0.65	0.77
Adapted distances (ΔE*) (Table 6.8)	12.17	11.6	5.97	6.44	6.65
Categorical similarity from simulated achromatic axis (Table 6.9)	0.32	0.24	0.93	0.52	0.56

Table 6.11 Summary of the results discriminated according to illuminant

Colour constancy indices from subsection 6.5.2 indicated a high degree of adaptation (83%) and only small variations across colour categories. Also, the proportion of structural differences revealed a high

degree of stability among chromatic settings' inter-distances (85%) when illumination was changed, and small differences according to background types. Observers in the colour naming test used mostly [BCT] or [BCT, BCT-ish] descriptors, which suggests that colour appearance after adaptation was described mostly by basic colour terms, revealing an overall 66% of categorical colour coincidence between coloured samples perceived under the reference and test illuminations.

Despite that the previous summary of results may suggest a high degree of categorical constancy, there is a complementary set of results which suggests categorical changes. Categorical constancy in the colour naming test corresponds mostly to the green (C_1 and C_4) and blue (C_2 and C_5) colours, the remaining inconstancy (33%) corresponds to the other colours (C_3 and C_6) which changed systematically accordingly to the illumination. Also, structural differences show different tendencies according to the illumination, confirming that under *yellowish* or *unsaturated yellowish* illuminants, the inter-distances among chromatic settings were slightly more disrupted than under *greenish* or *unsaturated greenish* illuminants. Not all tested illuminations achieved the same colour appearance as under the *D65* reference illumination: *bluish* illumination was better compensated than *unsaturated greenish* and *unsaturated yellowish* illuminations. This is supported by the values in the *bluish* column of Table 6.11.

6.6.3 Categorical colour constancy

According to Webster (Webster 1996) colour adaptation is composed by two processes: (1) an initial adaptation to the average scene colour which is performed by low-level mechanisms and (2) a contrast adaptation which adjusts to the distribution of colours and is lead by cortical mechanisms. There is evidence that the colour constancy phenomenon is the product of this two processes (Hurlbert and Wolf 2004).

Our experiments show that the "grey" chromatic settings are always close to their correspondent illuminant chromaticity (see Figure 6.6), indicating that the adaptation of our observers followed the direction of the illuminant shift. This effect is explained by the activation of the global contrast mechanisms (Webster 1996; Hurlbert and Wolf 2004) and if complete, we should expect successful colour constancy. However, our results indicate clearly that categorical colour constancy does not hold for one of the colours present in the three-coloured backgrounds. Furthermore, our analysis showed an asymmetry between the adapted colours and expected colours; as if the perception of green "lead" the adaptation of the other two colours.

Figure 6.10 shows the results obtained when we applied CCCP to our measured adaptation data. Categorical prediction on the achromatic axis suggest large failures of categorical constancy for the

yellowish and *greenish* illuminants, moderate failures for *unsaturated greenish* and *unsaturated yellowish*, and only small failures for the *bluish* illuminant.

Figures 6.10 and 6.11 allow a direct comparison between the CCCP predictions for the test illumination and for the reference illumination, i.e., colours perceived if colour constancy was complete. Figure 6.14 provides a summary of these comparisons for the results of the colour naming test (reported in Table 6.6), also for the simulation on the achromatic axis (Table 6.9), and for the simulation on the Munsell circle (Figure 6.11). The x axis of Figure 6.14 shows the average degree of categorical coincidence ($<I_1, A_i>$) between CCCP predictions for the A_i adaptations to the various illuminants and CCCP predictions for the reference illuminant I_1, i.e., the expected colours if colour constancy was complete. The y axis shows the average degree of categorical coincidence ($<I_1, I_i>$) between CCCP predictions for the various illuminants and CCCP predictions for the test illuminant.
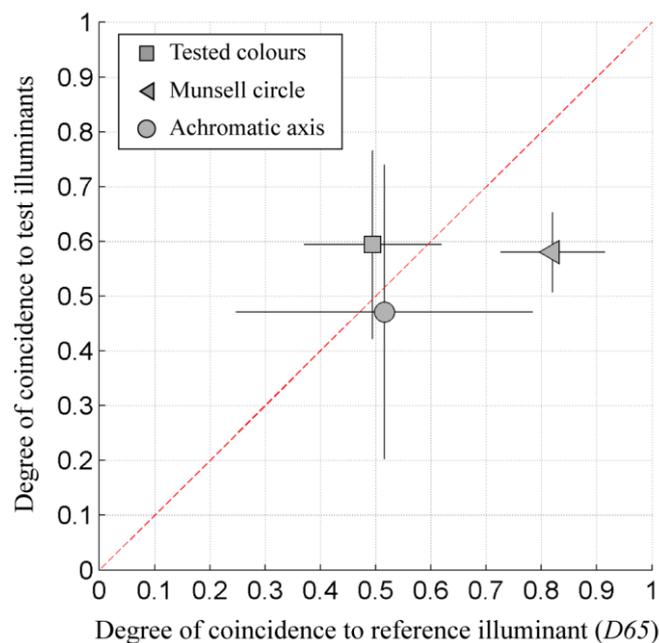
Figure 6.14 Average degree of categorical coincidence ($<I_1, A_i>$) between CCCP predictions for the A_i adaptations to the various illuminants and CCCP predictions for the reference illuminant I_1 (x-axis) and average degree of categorical coincidence ($<I_i, A_i>$) between CCCP predictions for the various illuminants and CCCP predictions for the test illuminant (y-axis). The square markers correspond to the average of CCCP results when applied to colour naming test, the circles correspond to the average of CCCP results when applied to achromatic axis points, and the triangles correspond the average of CCCP results when applied to Munsell circle colours. Crosses indicate standard deviation (SD).

Notice that values in Figure 6.14 are clustered near the central part of the plot, indicating that after adaptation, observers perceive a mixture of the colour categories perceived under the reference and test illuminations.

In order to explain these results we could speculate that colour constancy mechanisms consist of two processes: the first one tries to keep the ongoing categorical perception while adaptation to the average scene chromaticity takes place, and the second one tries to generate a definitive visual scene where colours are consistent with perception under a *canonical* illuminant.

6.7 Conclusions

We studied categorical colour constancy under six illuminants, by means of a new psychophysical experiment which combined chromatic setting with a colour naming task. Although our chromatic setting results indicate that adaptation to the average scene chromaticity was almost complete, the colour naming task shows that minor but systematic categorical changes occurred. Chromatic settings also revealed that the perception of green was stable for all illuminations, contrary to that of the other two colours tested. In order to increase our knowledge about categorical changes we developed a categorical colour prediction model (CCCP) from the chromatic settings, which allows to predict categorical colours according to each particular observer and adaptation state. We validated the model using several complementary techniques and applied it to predict categorical changes for other (not measured) colours. Finally, we compared the categorical similarity between the colours perceived after adaptation and the same colours perceived under the reference and test illuminations. Our results show that the perception of colour categories under chromatic illuminants was a mixture of the colours perceived under the test and reference illuminants. Several effects were reported according to the illuminant, which suggests that the chromatic properties of the illumination may constrain the final perception of colours.

Chapter 7 Conclusions and further work

This final chapter summarizes the main findings of the previous chapters in relation to the hypotheses stated in Chapter 1. It begins with a brief summary of the four experiments performed, the methods developed for the data analysis and the results obtained. Finally, we propose future lines of research in human colour constancy.

7.1 Conclusions

Colour constancy is a fundamental property of colour vision and it plays a central role on the colour stability of objects in everyday life. Object recognition is partly achieved by recognizing the colour of surfaces and it may be adequately accomplished by colour category matching. Therefore, categorical colour constancy may play a relevant role for object identification to be successful. The degree of colour constancy is typically quantified using colour constancy indices which mostly measure the adaptation to the average scene chromaticity and may neglect changes in categorical colour perception.

The aim of this work was to study whether categorical colour perception influences colour adaptation, constraining the outcome of the colour constancy process. To approach this issue we felt the need to develop new methodological paradigms and analysis techniques which complemented the existing methods and for this reason we split the contributions of this work in two types: methodological and scientific.

The main methodological contributions of this work are:

- A new method to analyse the categorical structure of 3D colour space based on colour solids, which allows us to characterize individual categorical colour perception as well as quantifying inter-individual variations in terms of shape and centroid location of 3D categorical regions.

- A new colour constancy paradigm, termed *chromatic setting*, which allows us to measure the precise location of nine categorically-relevant points in colour space under immersive illumination.

- We derived from these chromatic setting measurements a new colour constancy index (SCI) which takes into account the magnitude and orientation of the chromatic shift, memory effects and the interrelations among colours.

- Finally, we developed a model of categorical colour prediction (CCCP), which is tuned to each observer/adaptation state and allows us to predict the categorical appearance of any region of colour space.

Using these experimental and analytical methods, we conducted a series of experiments that investigated aspects of colour perception related to colour constancy and categorical colours.

Our first experiment studied the extension of inter-individual variations in the 3D categorical structure of colour space. In this experiment, we employed several colour naming tests on a large set of coloured samples and characterized the 3D structure of the individual categorical colour space with a reduced set of indices. Our results indicate that despite the population of normal trichromat observers tested had broadly similar categorical structures; there was significant variation amongst observers in the centroid locations of categories. This implies that if categorical information has to be included in the stimuli to study its influence on colour constancy, then this information must be tuned according to the particularities of each observer.

Our second experiment tested whether colour constancy, under immersive illumination conditions, is best measured using multiple points in colour space. To do so, we first demonstrated the feasibility of the chromatic setting paradigm which allows us to measure the precise location of nine categorically relevant points in colour space. The paradigm was applied to 2D Mondrian stimuli under three different illuminants, and the results were used to fit a set of linear colour constancy models. Using multiple points improved the precision of more complex linear colour constancy models and suggested that mechanisms other than cone gain might be best suited to explain colour constancy.

The third experiment investigated the perceptual interrelations of coloured surfaces under illuminant changes. Since the chromatic setting paradigm provided the precise location of nine categorically

relevant points, we developed a method to quantify the degree of deformation among the interrelations of these chromatic settings when illumination was changed. Our results show that these interrelations remained mostly constant under an illuminant shift, suggesting that categorical colour perception may be used to guide colour constancy.

We first tested our main hypothesis by studying whether colour constancy is assisted by the categorical information content of the scene, maximizing the categorical colour perception under a reference (near *achromatic*) illumination. Our previous experiments used two categorically different backgrounds under several test illuminations, designed to contain maximal and minimal categorical information under an *achromatic* illumination, and customised for each observer. Our results do not show any quantitative or qualitative differences regarding the two types of backgrounds tested.

The fourth experiment further tested our main hypothesis by studying whether the chromatic properties of the illuminant influence categorical colour constancy. Our results indicate that adaptation to the average scene chromaticity was almost complete for all subjects, but also that minor but systematic categorical changes occurred. Chromatic settings allowed us to predict the location of perceived colours in colour space and revealed that adaptation was more stable for the perception of green under all illuminations than for other colours. We compared the categorical similarity between the colours perceived under a coloured (test) illuminant, colours perceived under the neutral (reference) illuminant and colours simulated under the test illuminant (no colour constancy), these results revealed that the final categorical perception of colours was a mixture of the colours perceived under the test and reference illuminations. Several differences were reported according to the adapted illuminant, which suggested that the chromatic properties of the illumination could constrain the final categorical perception of colours.

7.2 Further work

We have encountered several ideas for further work, which may improve our understanding of the complex and fascinating phenomenon of colour constancy. Some future lines of work may include:

- To explore more contextual properties through the chromatic setting paradigm. In particular to study different geometrical distributions of colours in the background under a more extensive set of illuminants.

- To get a more reliable tool to quantify the magnitude of the colour constancy phenomenon, most probably by developing multidimensional indices to describe it.

- To produce a review of empirical results in colour constancy. Up to now there are only qualitative reviews of the phenomenon. This implies to gather all empirical data, from different experiments and paradigms into a single colour space, then define a vector field that summarizes this information and models it with differential equations. In this way the empirical behaviour of the phenomenon would be best summarized.

Also there are some interesting lines to follow in the field of computer vision:

- Most illuminant estimation approaches estimate the illuminant's chromatic coordinates in a 3D colour space and they use a diagonal matrix to correct the image (Ebner 2007). From conclusions in Chapter 4 we know that also other affine models should be tested to increase the accuracy of the illuminant estimation.

- The conclusions in Chapter 6 about adaptation in scenes with low categorical diversity suggests a new criteria for illuminant selection when working with the colour category hypothesis.

Appendix A: CIELab coordinates of *surface* and light *samples* in Experiment I

Num	L*	a*	b*	Num	L*	a*	b*	Num	L*	a*	b*
1	85,67	26,41	6,98	115	81,89	-21,26	66,50	229	83,97	-4,92	-22,89
2	115,48	25,69	8,36	116	112,28	-23,07	69,25	230	113,10	-6,57	-19,93
3	144,55	26,50	10,97	117	143,94	-24,11	70,11	231	144,14	-7,88	-20,52
4	85,92	38,90	9,25	118	160,12	-24,16	72,23	232	83,04	-5,98	-34,08
5	114,39	38,01	11,84	119	145,02	-35,62	118,15	233	114,21	-8,42	-33,08
6	145,07	39,96	16,11	120	79,57	-19,29	37,71	234	143,91	-8,98	-29,09
7	86,42	63,52	15,45	121	112,25	-21,65	43,53	235	83,56	-4,87	-60,79
8	114,42	63,17	19,28	122	143,56	-21,39	47,18	236	113,04	-9,73	-58,10
9	85,65	76,81	19,06	123	159,56	-23,01	50,70	237	81,93	0,42	-23,61
10	114,57	77,98	25,85	124	80,83	-29,51	57,65	238	113,54	-0,73	-19,99
11	85,64	91,79	22,32	125	112,69	-30,05	61,96	239	143,45	-2,97	-20,90
12	85,40	27,20	11,36	126	143,70	-31,29	64,61	240	83,05	0,61	-36,68
13	114,71	25,07	10,36	127	143,99	-47,05	103,65	241	113,37	-0,88	-33,82
14	144,68	26,80	14,21	128	79,86	-26,64	34,22	242	144,57	-2,61	-29,36
15	85,99	37,28	14,80	129	112,87	-26,21	34,90	243	83,65	4,60	-59,17
16	114,32	38,68	16,93	130	144,41	-29,04	37,59	244	112,87	0,37	-56,68
17	144,32	37,52	19,19	131	160,90	-28,35	39,24	245	83,36	6,51	-71,82
18	85,68	62,37	26,77	132	78,72	-39,22	45,29	246	81,91	7,76	-23,40
19	114,18	64,53	30,69	133	112,31	-37,98	51,69	247	113,20	5,76	-20,85
20	85,16	74,05	30,24	134	144,21	-41,07	55,89	248	143,33	4,17	-21,29
21	114,71	76,05	32,74	135	113,15	-64,03	85,97	249	83,10	11,02	-34,59
22	86,60	89,33	35,95	136	143,60	-65,28	88,84	250	114,48	9,43	-34,40
23	85,99	24,38	13,41	137	81,93	-26,52	23,11	251	82,37	21,59	-58,20
24	114,03	23,81	15,48	138	113,56	-29,38	26,68	252	113,30	15,82	-52,18
25	144,07	24,53	18,14	139	144,71	-31,28	31,25	253	82,78	24,90	-66,30
26	85,02	34,83	18,77	140	160,08	-32,82	33,52	254	84,23	13,23	-23,11
27	114,58	36,66	22,67	141	81,82	-40,92	33,59	255	114,62	10,28	-19,34
28	86,05	60,45	34,33	142	113,62	-44,07	39,12	256	144,56	8,90	-19,86
29	114,89	59,96	38,41	143	144,59	-48,25	42,88	257	83,84	18,97	-34,66
30	86,53	68,52	36,19	144	114,73	-73,86	63,33	258	113,48	16,42	-31,87
31	114,78	71,65	45,07	145	113,13	-86,56	73,28	259	84,37	33,41	-54,81
32	85,82	82,32	46,27	146	88,63	-24,11	12,78	260	113,19	26,22	-50,09
33	85,11	23,01	17,34	147	116,74	-29,77	17,64	261	82,69	38,37	-61,13
34	114,59	22,73	20,86	148	144,97	-34,04	22,04	262	83,55	17,27	-21,88
35	144,07	22,48	23,39	149	159,80	-30,07	21,08	263	114,20	14,35	-18,97
36	85,27	32,56	24,61	150	89,16	-36,30	18,10	264	144,52	13,19	-18,59
37	114,04	35,77	31,73	151	116,73	-43,98	24,32	265	82,83	24,45	-30,17
38	144,61	35,54	36,12	152	145,41	-50,47	30,74	266	112,68	22,50	-29,68
39	86,06	57,16	44,16	153	89,02	-61,16	27,72	267	82,90	41,13	-49,79
40	114,42	54,90	48,13	154	117,35	-72,71	38,44	268	81,86	45,97	-52,83
41	85,87	65,72	51,94	155	87,80	-24,28	8,96	269	83,02	18,69	-19,03
42	114,43	65,41	58,44	156	116,63	-30,51	12,61	270	112,70	17,17	-17,28
43	114,49	73,90	69,89	157	145,00	-34,17	15,83	271	143,38	16,14	-16,20
44	85,29	22,22	21,94	158	88,62	-37,23	12,35	272	82,81	28,44	-27,75
45	114,97	20,70	25,07	159	117,63	-44,87	17,58	273	113,24	24,89	-26,24
46	144,71	20,70	28,04	160	144,87	-51,81	22,68	274	82,23	45,60	-43,39
47	86,77	32,23	33,10	161	89,45	-62,91	17,58	275	83,05	51,48	-47,71
48	114,75	31,52	37,77	162	116,95	-72,68	24,97	276	83,52	21,13	-15,63
49	85,76	47,85	50,42	163	87,79	-25,47	6,17	277	113,89	20,43	-13,12
50	114,63	52,77	62,91	164	116,08	-32,60	9,60	278	143,79	19,35	-11,67
51	114,33	60,00	75,02	165	144,99	-35,89	12,08	279	83,43	31,46	-22,19
52	115,57	67,90	90,59	166	86,71	-39,98	8,47	280	113,81	31,14	-19,84
53	85,81	18,07	22,24	167	115,31	-45,08	11,45	281	143,31	30,30	-16,50

54	114,47	19,80	28,07	**168**	144,94	-51,39	15,81	**282**	83,24	50,35	-36,53
55	144,46	18,01	31,29	**169**	87,19	-64,21	11,93	**283**	113,35	45,44	-32,66
56	85,27	27,99	35,09	**170**	115,54	-73,91	16,57	**284**	83,79	55,63	-40,48
57	114,81	27,92	41,64	**171**	88,34	-25,60	3,40	**285**	84,29	24,31	-12,73
58	144,01	29,27	48,43	**172**	116,49	-31,68	5,16	**286**	113,74	21,83	-10,28
59	114,22	43,99	70,21	**173**	145,02	-34,00	8,22	**287**	144,07	21,87	-7,53
60	114,49	47,61	79,95	**174**	87,55	-38,39	3,80	**288**	84,34	34,38	-18,07
61	115,49	55,49	95,86	**175**	116,82	-46,77	7,45	**289**	113,45	34,29	-15,32
62	86,05	17,72	28,72	**176**	144,57	-50,29	10,01	**290**	144,43	34,70	-14,49
63	114,88	15,76	31,86	**177**	87,68	-51,25	4,88	**291**	83,64	55,78	-29,23
64	144,80	14,76	35,72	**178**	115,41	-68,08	9,12	**292**	113,77	56,86	-26,70
65	87,32	21,20	38,93	**179**	88,42	-25,06	-0,07	**293**	82,25	66,40	-35,73
66	115,85	22,35	48,22	**180**	115,28	-31,62	2,47	**294**	84,37	26,96	-8,43
67	144,52	20,30	52,09	**181**	145,62	-33,98	4,64	**295**	114,14	22,78	-5,75
68	115,23	34,11	76,79	**182**	88,10	-37,43	-0,44	**296**	144,36	23,08	-3,63
69	116,31	37,60	93,30	**183**	115,39	-46,14	3,02	**297**	84,02	37,87	-13,01
70	115,43	39,99	98,32	**184**	144,40	-48,77	5,70	**298**	114,08	36,83	-9,41
71	78,84	11,09	37,32	**185**	115,54	-68,02	1,96	**299**	143,28	35,70	-6,39
72	112,87	10,01	39,79	**186**	88,17	-24,12	-4,11	**300**	83,65	61,24	-20,14
73	144,14	9,55	43,76	**187**	116,37	-28,60	-2,05	**301**	113,54	61,52	-17,15
74	80,45	18,42	57,89	**188**	144,46	-33,40	-0,08	**302**	83,63	72,86	-24,14
75	112,07	17,64	58,86	**189**	87,83	-36,44	-6,10	**303**	113,38	73,07	-21,18
76	143,46	12,82	60,63	**190**	114,59	-42,91	-3,24	**304**	84,20	27,25	-2,24
77	112,35	22,89	99,58	**191**	115,60	-60,93	-7,08	**305**	113,84	24,96	-0,04
78	79,88	5,58	41,83	**192**	88,82	-22,01	-7,64	**306**	143,57	24,04	0,41
79	111,63	5,10	42,30	**193**	116,57	-25,51	-6,48	**307**	83,33	39,89	-5,77
80	144,00	4,47	44,64	**194**	144,54	-30,02	-3,99	**308**	114,03	37,65	-2,96
81	78,93	8,09	58,82	**195**	87,99	-33,67	-12,34	**309**	143,76	37,03	1,54
82	112,07	7,32	65,34	**196**	115,87	-40,80	-10,18	**310**	83,87	62,73	-8,08
83	143,77	6,52	66,76	**197**	87,45	-20,62	-10,69	**311**	113,96	62,86	-5,36
84	144,18	12,60	108,28	**198**	116,50	-26,07	-9,48	**312**	84,08	76,88	-11,99
85	79,88	1,00	45,82	**199**	145,30	-30,53	-8,31	**313**	113,59	77,66	-5,87
86	112,37	-1,93	46,84	**200**	87,93	-31,82	-17,55	**314**	84,59	27,16	-0,17
87	144,56	-3,47	48,39	**201**	116,55	-37,27	-15,58	**315**	113,76	24,68	1,87
88	159,74	-4,26	54,26	**202**	88,45	-17,16	-13,89	**316**	144,23	24,78	5,00
89	80,36	-0,14	62,97	**203**	115,83	-22,80	-13,04	**317**	84,14	41,09	1,03
90	112,66	-1,50	69,52	**204**	145,14	-27,50	-12,40	**318**	113,46	40,65	3,73
91	144,31	-4,09	71,34	**205**	87,77	-27,28	-21,51	**319**	143,61	39,00	5,61
92	144,20	-2,63	115,23	**206**	116,41	-34,03	-21,32	**320**	84,20	65,34	0,61
93	80,89	-5,51	47,95	**207**	88,02	-16,00	-16,94	**321**	113,38	66,27	4,96
94	112,19	-6,37	48,99	**208**	114,91	-18,70	-16,20	**322**	83,37	75,16	-1,23
95	144,38	-7,76	51,01	**209**	145,13	-21,89	-15,97	**323**	114,13	75,98	4,82
96	160,40	-7,52	53,22	**210**	88,38	-21,64	-26,09	**324**	85,10	28,12	4,86
97	80,66	-5,82	63,86	**211**	116,38	-27,46	-24,72	**325**	115,05	24,37	4,97
98	111,79	-8,93	69,83	**212**	87,75	-29,19	-43,55	**326**	144,15	26,18	7,36
99	143,52	-10,29	72,06	**213**	115,42	-39,55	-41,27	**327**	85,94	39,87	6,41
100	159,97	-11,02	77,15	**214**	86,66	-12,57	-18,06	**328**	114,51	37,95	7,11
101	144,49	-13,91	118,07	**215**	115,82	-14,49	-16,96	**329**	144,48	39,60	9,23
102	79,10	-10,15	47,24	**216**	145,29	-17,85	-16,87	**330**	84,86	64,05	8,30
103	112,32	-10,75	47,76	**217**	87,25	-16,65	-27,98	**331**	115,01	63,39	11,17
104	143,77	-11,39	51,43	**218**	115,94	-21,19	-28,26	**332**	84,47	78,69	10,58
105	159,33	-12,21	55,45	**219**	87,66	-22,74	-50,14	**333**	113,86	75,46	14,51
106	79,74	-13,35	68,95	**220**	115,35	-32,44	-49,54	**334**	84,08	87,26	11,41
107	112,39	-15,24	74,02	**221**	87,51	-8,60	-19,66	**335**	48,62	0,36	-0,15
108	143,70	-16,28	72,04	**222**	115,36	-9,76	-17,88	**336**	52,81	0,01	-0,20
109	159,94	-18,29	77,97	**223**	145,01	-12,35	-17,41	**337**	56,00	0,34	-0,63
110	144,36	-22,95	118,18	**224**	87,98	-11,69	-29,87	**338**	60,75	-0,01	-0,43
111	78,97	-14,42	42,91	**225**	116,36	-15,46	-29,63	**339**	71,45	-0,25	-0,03
112	112,55	-16,16	45,62	**226**	143,84	-18,97	-28,60	**340**	85,48	0,01	0,34
113	144,46	-16,96	49,01	**227**	87,96	-16,47	-58,21				
114	159,46	-17,83	53,30	**228**	116,12	-22,77	-53,97				

Table A.1 CIELab coordinates of light samples in Experiment I.

140

Num	L*	a*	b*	Num	L*	a*	b*	Num	L*	a*	b*
1	190,38	0,23	1,80	48	129,97	45,63	37,12	95	51,76	-0,25	-3,50
2	42,94	-0,54	-3,74	49	50,35	-28,67	6,39	96	163,83	0,06	-0,24
3	103,00	0,04	-0,16	50	89,41	0,08	0,39	97	118,07	39,80	40,91
4	179,91	-0,60	2,36	51	46,32	-0,20	-2,91	98	119,51	33,48	41,29
5	37,40	-0,03	-3,75	52	130,82	35,40	29,00	99	110,02	20,96	75,23
6	96,11	0,04	-0,15	53	83,12	27,96	-82,19	100	162,99	-1,29	2,76
7	169,24	-0,56	2,23	54	108,82	-74,02	60,61	101	109,68	0,04	-0,17
8	33,56	-0,05	-3,73	55	151,58	0,82	0,06	102	169,76	17,23	31,09
9	89,65	-0,41	0,21	56	69,37	-0,10	-1,83	103	173,10	-28,24	11,71
10	158,96	-0,53	2,11	57	119,31	20,57	39,13	104	162,14	9,93	-9,93
11	107,59	0,56	-0,58	58	117,06	20,37	26,16	105	158,07	-19,77	-13,67
12	74,93	82,37	-20,53	59	109,05	-62,40	11,32	106	133,28	0,05	-0,20
13	124,25	45,10	-31,19	60	32,70	-0,64	-2,68	107	89,89	0,04	-0,14
14	68,99	80,52	-64,71	61	195,34	-0,64	2,55	108	50,96	-0,11	-1,77
15	100,40	-10,48	-78,51	62	103,95	-3,16	-41,39	109	110,87	-28,83	86,19
16	115,67	-38,46	-41,98	63	100,59	88,33	29,66	110	89,73	-0,36	0,74
17	50,56	-26,29	-12,57	64	81,63	102,71	52,61	111	46,32	-0,74	-2,48
18	110,31	-55,79	-20,43	65	127,39	0,05	-0,19	112	58,36	51,96	8,07
19	49,73	0,93	9,00	66	84,39	-0,04	-1,14	113	90,29	111,23	68,31
20	88,51	0,08	0,39	67	86,92	22,36	39,29	114	119,41	60,06	4,86
21	45,86	-0,78	-3,09	68	117,87	24,28	24,59	115	118,42	57,91	26,91
22	62,96	27,23	-28,63	69	110,36	-51,31	65,07	116	115,10	81,82	116,21
23	91,01	37,34	-63,78	70	162,99	-1,29	2,76	117	132,47	-26,14	138,22
24	56,69	8,85	-55,67	71	109,97	0,04	-0,17	118	112,96	-2,19	88,29
25	118,62	-21,28	-51,40	72	89,65	-33,11	48,08	119	125,61	-49,46	111,82
26	52,72	-12,99	-33,88	73	61,34	43,11	-40,93	120	33,00	-0,85	-2,52
27	113,93	-2,58	-50,84	74	155,96	-4,91	150,56	121	108,47	0,04	-0,17
28	96,54	-62,74	-17,14	75	100,66	-0,45	0,23	122	92,25	103,32	13,80
29	109,57	-59,88	29,85	76	117,21	0,60	-0,63	123	56,00	42,32	-11,68
30	32,62	-0,41	-2,59	77	123,79	22,00	24,78	124	82,66	104,63	43,22
31	193,92	0,23	1,83	78	118,42	24,88	26,41	125	100,99	109,96	69,87
32	168,98	-1,06	-14,22	79	96,19	-78,90	66,51	126	146,96	30,32	140,17
33	165,76	25,09	0,07	80	90,05	0,08	0,39	127	145,59	0,53	149,71
34	163,83	-30,54	-0,82	81	45,81	-0,20	-2,88	128	127,39	-34,19	116,41
35	159,55	21,92	10,97	82	111,82	20,07	-46,35	129	47,97	17,19	16,92
36	155,35	-16,47	40,85	83	142,78	-48,31	105,93	130	88,34	-0,40	0,21
37	114,20	46,17	55,74	84	99,93	92,04	-26,45	131	190,38	-0,63	2,49
38	116,34	30,87	25,61	85	78,80	-0,04	-1,08	132	104,69	0,04	-0,16
39	94,44	-75,23	20,33	86	141,34	0,77	0,05	133	42,78	-0,52	-3,43
40	161,86	-1,28	2,74	87	87,76	40,41	54,86	134	178,97	-0,59	2,35
41	110,25	0,04	-0,17	88	73,27	20,61	36,56	135	97,39	0,56	0,04
42	79,10	27,49	29,14	89	111,43	-81,22	68,12	136	37,27	-0,70	-2,93
43	123,24	61,99	110,31	90	33,56	-0,42	-2,64	137	167,92	0,27	2,53
44	56,69	40,99	-89,57	91	194,94	-0,64	2,54	138	91,01	0,08	0,39
45	180,15	-0,60	2,37	92	143,49	-54,64	-1,21	139	32,51	-1,33	-2,90
46	38,94	-0,25	-3,39	93	139,87	26,06	131,14	140	158,37	-0,53	2,10
47	138,75	31,55	35,35	94	101,38	-39,33	-50,44				

Table A.2 CIELab coordinates of surface samples in Experiment I.

Appendix B: Shape of ACNS regions in Experiment I
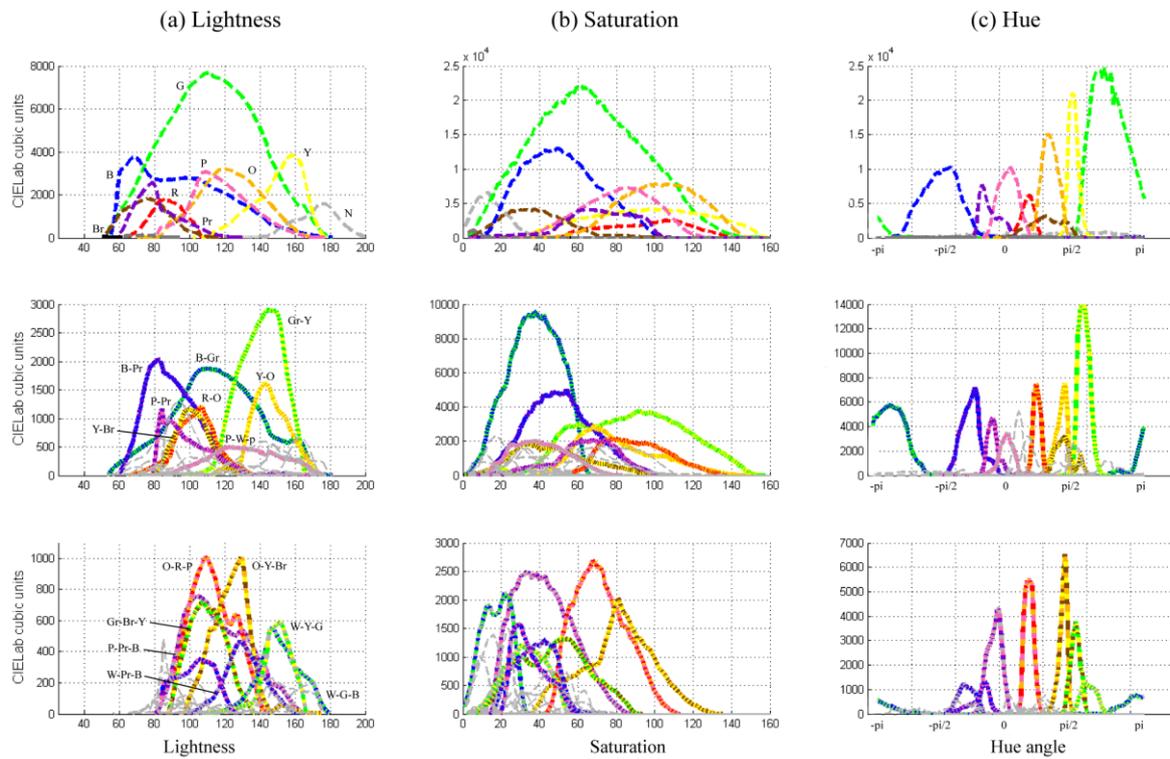
Figure B.1 Structure of the average colour naming space (ACNS). Volume proportion histogram of *basic* (coloured lines; first row) and *borders of two* (two-coloured lines; second row) and *borders of three* (three-coloured lines; third row) and higher order regions (grey coloured lines) regions according to their hue, lightness and saturation (columns).

Appendix C: Long term memory in Experiment II

Given that our experiments were conducted over a few weeks, we tested whether the uncertainty introduced by longer-term memory was significantly larger than the uncertainty present in a typical 25-minutes session. We did this by repeating the same measures over different days using two experienced subjects. They were required to select 4 SR (green, purple, orange and grey) and to reproduce the same colours 7 days later. To collect more data, the selection of SRs was repeated forty times for each colour. Figure D.1 shows the variability of our measures for these control sessions: the small darker points correspond to results for the first session and the small lighter points to the second session. Squares and triangles represent the corresponding averages. The lightness variability results followed a similar trend and were omitted from the plots for clarity's sake. To determine if both distributions of points are the same, we computed the statistic D, the maximum difference of the integrated probabilities of the two distributions, developed by Fasano and Franceschini and others (Peacock 1983; Fasano and Franceschini 1987). Our results showed that, predictably there were memory effects in all cases except two. However, D was comparatively small, i.e. the mean's difference between the light and dark points was always smaller than the standard deviation (itself about 1 ΔE^*) of either the light or the dark point distributions.
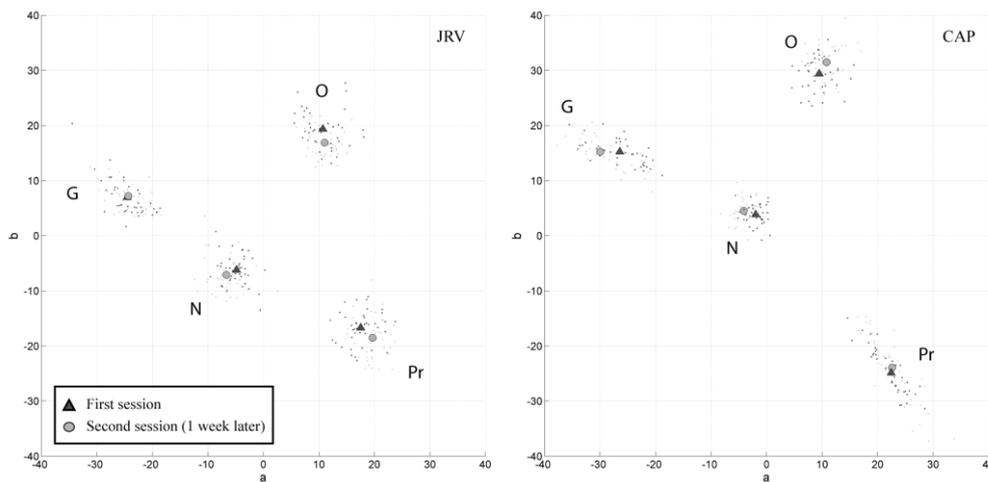
Figure C.1 Results of the long-term memory control experiment for two subjects. Four categories were tested (40 trials each). Dark and light dots were measured with a 7-day time difference. Averages are represented by triangles (first session) and squares (second session). The results clustered near the origin, are equivalent to those of a typical achromatic setting experiment.

145

Appendix D: Akaike Information Criterion for model selection

The *Akaike Information Criterion (AIC)* method is based on Information Theory and it is widely used for model selection, i.e., given several candidate models the method selects the model which minimize the loss of information when approximating the reality. In order to test our four models we used the AIC version adapted to small sets of samples (AIC_c) and the residual sum of squares (*RSS*) as detailed in Equation 9, where n corresponds to the number of data points and k to the number of variables plus the error term (Burnham and Anderson 2002).

$$AIC_c = nln\left(\frac{RSS}{n}\right) + 2k + \frac{2k(k+1)}{n-k-1} \tag{E.9}$$

Notice that AIC_c formulae depends exclusively on the dimensions of the multivariate system resulting from Equation 5, because this approach does not reflect the number of free parameters existent in our tested models we rearranged the system into an equivalent univariate system. In order to apply the AIC_c we assumed that our prediction errors followed a normal distribution.

The model with the lowest AIC value is the best model among all models specified. However, AIC values become interesting when compared to the AIC value of a series of models. Two measures associated with AIC can be used to compare models: (i) the difference between the model with the lowest AIC and the rest ($\Delta_i = AIC_i - min(AIC_i)$) and (ii) the *Akaike weights* which quantify the plausibility of each model as being the best ($w_i = exp(-0.5\Delta_i)/\sum_{r=1}^{R} exp(-0.5\Delta_r)$). As a rule of the thumb, a $\Delta_i < 2$ suggests substantial evidence for the model, values between 3 and 7 indicate that the model has considerably less support, whereas $\Delta_i < 10$ indicate that the model is very unlikely (Burnham and Anderson 2002).

Table 6 contains the values of the RSS, AIC_c, Δ_i and w_i when applied to our data according to the model, number of fitting points and illumination used. Notice that the reported RSS values do not correspond to the minimization ones in Figure 6, this is because we took as RSS value the accumulative error of the fitting points that participated in the minimization process only. In practice, RSS values were not obtained by linear regression but from the minimization process described in

Equation 6, however the target value of the minimization is equivalent. Also the RSS values used in Table 6 resulted from the average over all subjects and backgrounds.

			Greenish				*Yellowish*			
Model	**3n**	**k**	**RSS**	**AIC$_c$**	**Δ$_i$**	**w$_i$**	**RSS**	**AIC$_c$**	**Δ$_i$**	**w$_i$**
D	15	4	293.18	56.59	0	0.96	535.1	65.62	0	0.99
DT	15	7	134.79	62.93	6.34	0.04	298.2	74.84	9.23	0.01
L	15	10	109.74	104.85	48.26	0	218.6	115.18	49.57	0
A	15	13	40.96	405.01	348.48	0	81.1	415.31	349.69	0
D	18	4	367.05	65.35	0	0.77	671.6	76.22	0	0.95
DT	18	7	187.90	67.42	2.07	0.26	425.9	82.15	5.92	0.05
L	18	10	164.36	91.24	25.89	0	328.4	103.70	27.47	0
A	18	13	82.27	144.36	79.00	0	170.5	157.47	81.25	0
D	21	4	440.95	74.43	0.44	0.45	808.3	87.16	0	0.88
DT	21	7	242.55	74.00	0	0.55	547.2	91.08	3.92	0.12
L	21	10	218.61	91.20	17.20	0	434.7	105.63	18.47	0
A	21	13	123.84	115.26	41.27	0	270.1	131.63	44.48	0
D	24	4	514.87	83.69	2.27	0.24	945	98.26	0	0.83
DT	24	7	297.47	81.41	0	0.76	685	101.45	3.19	0.17
L	24	10	273.25	95.30	13.88	0	550	112.07	13.81	0
A	24	13	169.47	109.31	27.90	0	381.9	128.81	30.55	0
D	27	4	588.79	93.04	3.72	0.13	1081.8	109.46	0	0.70
DT	27	7	353.17	89.31	0	0.86	794	111.19	1.72	0.29
L	27	10	327.65	101.14	11.83	0	617	118.23	8.77	0.01
A	27	13	226.17	111.39	22.07	0	495.3	132.55	23.10	0

Table D.1 Akaike Information Criterion applied to our data. Each row corresponds to the model case considered and the columns correspond to the number of fitting points used, the number of free parameters in each model, the RSS, and the Akaike results: AICc, Δi and wi. Note that the multivariate system was rearranged into an equivalent univariate system, therefore the 3n factor in the second column. See details on how these values were computed in the main text.

Δ_i and w_i values in Table 6 indicate that the Diagonal and Diagonal plus Translation models are the ones that best model the data, and thus suggesting that the Linear and Affine models overfits. The small differences in Δ_i between D and DT are not conclusive about which is the best model, however there is a clear tendency as we add more fitting points; the DT model becomes better than D. From one to four fitting points the AIC indicates that the best model is the D, DT, L and A as expected due to the coincidence between the number of fitting points and the free parameters.

Appendix E: Repeatability sessions in Experiment IV

Figure E.1 shows the results of the two repeatability sessions without the Bounding Cylinder in Experiment IV.
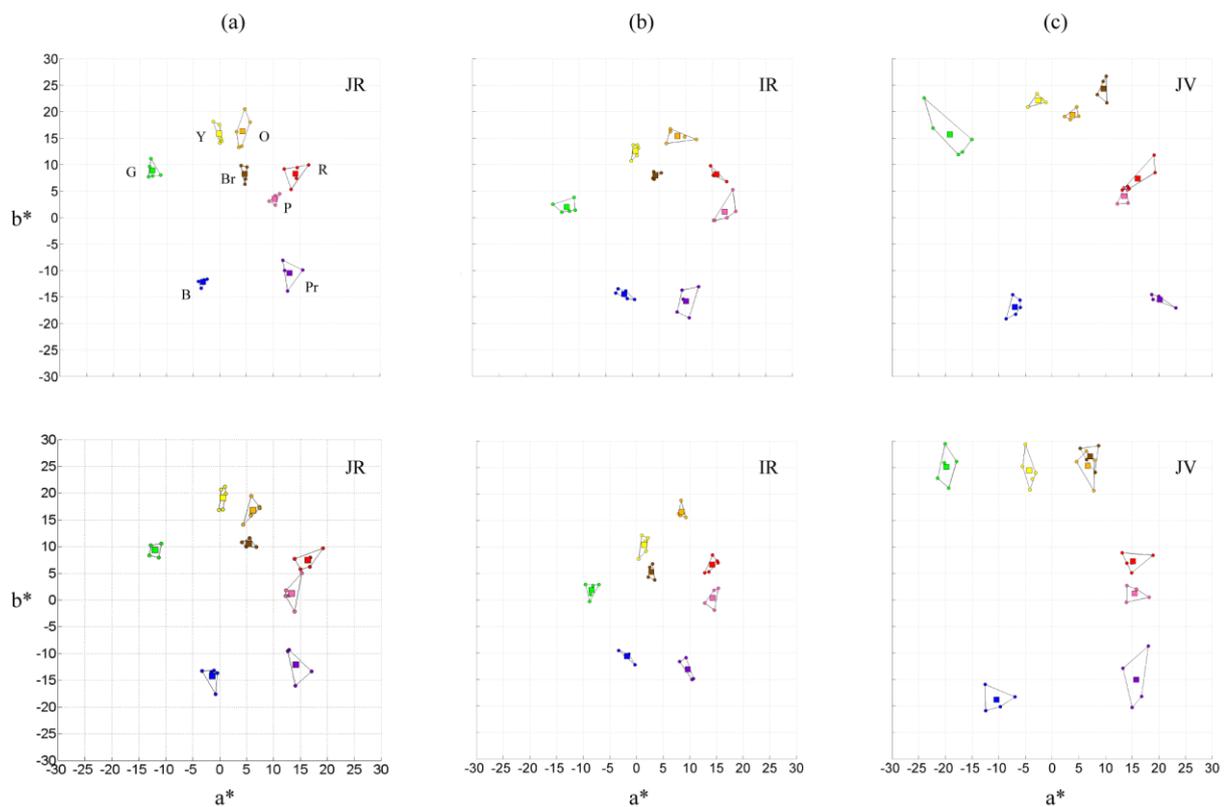
Figure E.1 Chromatic settings and their repeatability in Experiment IV. Each column corresponds to one observer and each row to one of the repeatability sessions without the Bounding Cylinder. Square markers represent the average of individual trials (small dots joined by lines) and they are colour coded according to the colour category which represent. Results are plotted in the a*b* chromaticity plane with D65 as a reference white point.

Appendix F: Linking lost correspondences in the colour naming test

Since the background stimuli introduced in Chapter 6 had only three colours, observers named three colour terms in each session. Correspondences between colours across sessions were confounded because these colour terms were recorded in an arbitrary sequence by observers. In order to study possible categorical changes across different adaptations we need to recover these correspondences from the recorded data. Consequently, we need to find a bijective correspondence between the list of three colour terms recorded by the subject in a given session and the list recorded in another session. In this appendix, we present a simple algorithm to achieve this purpose.

The algorithm has two steps. First, the algorithm translates each list into a 3x9 matrix with coefficient values 0, 0.5 and 1. In that matrix each row corresponds to one of the three colours of the background. The colours are coded into chromatic basic terms plus grey (8 + 1 terms) with each column indicating membership to a particular basic term (column 1 for red; 2 for green; 3 for yellow; and so on). Initially all matrix coefficients are zero. For instance, if the first term of the list is green then the coefficients corresponding to the first row of the matrix and second column will be updated with value of 1. If the second term is green-yellow then the coefficients corresponding to the second row and columns 2 and 3 will be updated with value 1. If the third term is green-yellowish then the previous coefficient corresponding to yellow will be updated to 0.5 instead of 1. All other combinations of basic colour terms and/or their adjectival forms follow from these examples. The same process applies to the second list and gives a second matrix. The next step of the algorithm takes one of the matrices and considers the resultant matrices from all possible row permutations. Then, for each permutation it computes the accumulated sum of the difference between the permutated matrix and the other matrix. The bijective correspondence is determined by the permutation that gives the lowest value to this accumulated sum.

References

Amano, K. and D. H. Foster (2008). Tracking Categorical Surface Colour Across Illuminant Changes in Natural Scenes. CGIV 2008 / MCS/08 - 4th European Conference on Colour in Graphics, Imaging, Terrassa – Barcelona, España, Society for Imaging Science and Technology

Amano, K., D. H. Foster, et al. (2012). "Visual search in natural scenes explained by local color properties." Journal of the Optical Society of America a-Optics Image Science and Vision 29(2): A194-A199.

Arend, L. (1993). "How much does illuminant color affect unattributed colors?" Journal of the Optical Society of America A 10: 2134-2147.

Arend, L. and A. Reeves (1986). "Simultaneous Color Constancy." Journal of the Optical Society of America a-Optics Image Science and Vision 3(10): 1743-1751.

Arend, L., A. Reeves, et al. (1991). "Simultaneous Color Constancy - Papers with Diverse Munsell Values." Journal of the Optical Society of America a-Optics Image Science and Vision 8(4): 661-672.

Baronchelli, A., T. Gong, et al. (2010). "Modeling the emergence of universality in color naming patterns." Proceedings of the National Academy of Sciences of the United States of America 107(6): 2403-2407.

Belpaeme, T. and J. Bleys (2005). "Explaining universal color categories through a constrained acquisition process." Adaptive Behavior 13(4): 293-310.

Benavente, R., M. Vanrell, et al. (2008). "Parametric fuzzy sets for automatic color naming." Journal of the Optical Society of America A 25(10): 2582-2593.

Berg, M. d. (2008). Computational geometry : algorithms and applications. Berlin, Springer.

Berlin, B. and P. Kay (1969). Basic color terms: their universality and evolution. Berkeley ; Oxford, University of California Press, 1991.

Birch, J. and C. M. Chisholm (2008). "Occupational colour vision requirements for police officers." Ophthalmic and Physiological Optics 28(6): 524-531.

Boynton, R. M. and C. X. Olson (1987). "Locating Basic Colors in the Osa Space." Color Research and Application 12(2): 94-105.

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.

Brainard, D. H., W. A. Brunt, et al. (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.

Brainard, D. H. and W. T. Freeman (1997). "Bayesian color constancy." Journal of the Optical Society of America a-Optics Image Science and Vision 14(7): 1393-1411.

Brainard, D. H. and B. A. Wandell (1986). "Analysis of the Retinex Theory of Color-Vision." Journal of the Optical Society of America a-Optics Image Science and Vision 3(10): 1651-1661.

Brainard, D. H. and B. A. Wandell (1992). "Asymmetric Color Matching - How Color Appearance Depends on the Illuminant." Journal of the Optical Society of America a-Optics Image Science and Vision 9(9): 1433-1448.

Brainard, D. H. K. J. M. L. P. (2003). Colour constancy: developing empirical tests of computational models. Colour Perception.

Brown, A. M., D. T. Lindsey, et al. (2011). "Color names, color categories, and color-cued visual search: Sometimes, color perception is not categorical." Journal of Vision 11(12).

Brown, R. O. and D. I. A. MacLeod (1997). "Color appearance depends on the variance of surround colors." Current Biology 7(11): 844-849.

Buchsbaum, G. (1980). "A Spatial Processor Model for Object Color-Perception." Journal of the Franklin Institute - Engineering and Applied Mathematics 310(1): 1-26.

Burnham, K. P. and D. R. Anderson (2002). Model selection and multimodel inference : a practical information-theoretic approach. New York, Springer.

Burnham, R. W., R. M. Evans, et al. (1957). "Prediction of Color Appearance with Different Adaptation Illuminations." Journal of the Optical Society of America 47(1): 35-42.

C. S. McCamy, H. M., and J. G. Davidson (1976). "A Color-Rendition Chart." Journal of applied Photographic Engineering 2(3): 95-99.

Cole, B. L. and J. D. Maddocks (1998). "Can clinical colour vision tests be used to predict the results of the Farnsworth lantern test?" Vision Research 38(21): 3483-3485.

Chichilnisky, E. J. and B. A. Wandell (1999). "Trichromatic opponent color classification." Vision Res 39(20): 3444-3458.

Delahunt, P. B. and D. H. Brainard (2004). "Color constancy under changes in reflected illumination." Journal of Vision 4(9): 764-778.

Delahunt, P. B. and D. H. Brainard (2004). "Does human color constancy incorporate the statistical regularity of natural daylight?" J Vis 4(2): 57-81.

Dzmura, M. (1991). "Color in Visual-Search." Vision Research 31(6): 951-966.

Ebner, M. (2007). Color constancy. Chichester, John Wiley.

Fairchild, M. D. (2005). Color appearance models. Chichester, West Sussex, England ; Hoboken, NJ, J. Wiley.

Farnsworth, D. (1957). Manual: The Farnsworth Munsell 100 Hue test for the examination of colour discrimination. B. M. C. C. Inc.

Fasano, G. and A. Franceschini (1987). "A Multidimensional Version of the Kolmogorov-Smirnov Test." Monthly Notices of the Royal Astronomical Society 225(1): 155-170.

Finlayson, G. D., S. D. Hordley, et al. (2001). Color by correlation: A simple, unifying framework for color constancy. IEEE.

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

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

Foster, D. H. (2011). "Color Constancy." Vision Research Volume 51(7): 674-700.

Foster, D. H. and S. M. Nascimento (1994). "Relational colour constancy from invariant cone-excitation ratios." Proceedings of the Royal Society of London B 257(1349): 115-121.

Fu, C., K. Xiao, et al. (2011). "Investigation of unique hue setting changes with ageing." Chin. Opt. Lett. 9(5): 053301.

Gärdenfors, P. (2000). Conceptual spaces : the geometry of thought. Cambridge, Mass. ; London, MIT Press.

Gegenfurtner, K. R. (2003). "Cortical Mechanisms of Colour Vision." Nature Neuroscience 4: 563-572.

Gegenfurtner, K. R. and D. C. Kiper (2003). "Color vision." Annual Review of Neuroscience 26: 181-206.

Gilbert, A. L., T. Regier, et al. (2006). "Whorf hypothesis is supported in the right visual field but not the left." Proceedings of the National Academy of Sciences of the United States of America 103(2): 489-494.

Golz, J. and D. I. A. MacLeod (2002). "Influence of scene statistics on colour constancy." Nature 415(6872): 637-640.

Gross, J. L. and J. Yellen (2004). Handbook of graph theory. Boca Raton, Fla. ; London, CRC Press.

Guest, S. and D. Van Laar (2000). "The structure of colour naming space." Vision Research 40(7): 723-734.

Hansen, T., M. Olkkonen, et al. (2006). "Memory modulates color appearance." Nature Neuroscience 9(11): 1367-1368.

Hansen, T., S. Walter, et al. (2007). "Effects of spatial and temporal context on color categories and color constancy." Journal of Vision 7(4): -.

Hardin, C. L. (2005). "Explaining basic color categories." Cross-Cultural Research 39(1): 72-87.

Hedrich, M., M. Bloj, et al. (2009). "Color constancy improves for real 3D objects." Journal of Vision 9(4): -.

Heider, E. R. (1972). "Universals in color naming and memory." J Exp Psychol 93(1): 10-20.

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

Hordley, S. D. (2006). "Scene illuminant estimation: Past, present, and future." Color Research and Application 31(4): 303-314.

Hubel, D. H. (1988). Eye, brain, and vision. New York, Scientific American Library : Distributed by W.H. Freeman.

Hurlbert, A. (1996). "Colour vision: Putting it in context." Current Biology 6(11): 1381-1384.

Hurlbert, A. (1997). "Colour vision." Current Biology 7(7): R400-R402.

Hurlbert, A. (1998). Computational models of color constancy. Perceptual constancy: Why things look as they do. W. V. K. J., Cambridge University Press: 283-322.

Hurlbert, A. (2007). "Colour constancy." Current Biology 17(21): R906-R907.

Hurlbert, A. and K. Wolf (2004). "Color contrast: a contributory mechanism to color constancy." Roots of Visual Awareness 144: 147-160.

Hurlbert, A. and K. Wolf (2004). "Color contrast: a contributory mechanism to color constancy." Prog Brain Res 144: 147-160.

Ishihara, S. (1917). "Tests for color-blindness." Handaya, Tokyo, Hongo Harukicho.

Ishihara, S. (1972). Tests for colour-blindness. Tokyo, Kyoto, Kannehara Shuppan Co., LTD.

Jäger, G. (2010). Natural Color Categories Are Convex Sets. LNAI. Amsterdam Springer-Verlag Berlin Heidelberg: 11-20.

Jameson, D. and L. M. Hurvich (1955). "Some Quantitative Aspects of an Opponent-Colors Theory .1. Chromatic Responses and Spectral Saturation." Journal of the Optical Society of America 45(7): 546-552.

Jameson, D. and L. M. Hurvich (1964). "Theory of Brightness and Color Contrast in Human Vision." Vision Research 4(1-2): 135-154.

Jameson, D. and L. M. Hurvich (1989). "Essay Concerning Color Constancy." Annual Review of Psychology 40: 1-22.

Judd, D. B. (1940). "Hue, saturation, and lightness of surface colors with chromatic illumination." Journal of the Optical Society of America 30(1).

Kay, P. and T. Regier (2003). "Resolving the question of color naming universals." Proc Natl Acad Sci U S A 100(15): 9085-9089.

Kay, P., W. T. Siok, et al. (2009). "Language regions of brain are operative in color perception." Proceedings of the National Academy of Sciences of the United States of America 106(20): 8140-8145.

Kraft, J. M. and D. H. Brainard (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.

Kuehni, R. G. (2001). "Determination of unique hues using munsell color chips." Color Research and Application 26(1): 61-66.

Kuehni, R. G. (2004). "Variability in unique hue selection: A surprising phenomenon." Color Research and Application 29(2): 158-162.

Kulikowski, J. J. and H. Vaitkevicius (1997). "Colour constancy as a function of hue." Acta Psychol (Amst) 97(1): 25-35.

Lammens, J. M. G. (1995). A computational model of color perception and color naming. Buffalo, NY, State University of New York at Buffalo Ph.D. Dissertation.

Land, E. H. (1964). "Retinex." American Scientist 52: 247-264.

Land, E. H. and J. J. McCann (1971). "Lightness and retinex theory." Journal of the Optical Society of America 61(1): 1-11.

Lindsey, D. T. and A. M. Brown (2006). "Universality of color names." Proc Natl Acad Sci U S A 103(44): 16608-16613.

Lindsey, D. T. and A. M. Brown (2009). "World Color Survey color naming reveals universal motifs and their within-language diversity." Proceedings of the National Academy of Sciences of the United States of America 106(47): 19785-19790.

Ling, Y. and A. Hurlbert (2008). "Role of color memory in successive color constancy." Journal of the Optical Society of America A: Optics and Image Science, and Vision 25(6): 1215-1226.

Linhares, J. M. M., P. D. Pinto, et al. (2008). "The number of dicernible colors in natural scenes." Journal of the Optical Society of America A 25: 2918-2924.

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

Lorensen, W. E. and H. E. Cline (1987). "Marching cubes: A high resolution 3D surface construction algorithm." SIGGRAPH Comput. Graph. 21(4): 163-169.

Loreto, V., A. Mukherjee, et al. (2012). "On the origin of the hierarchy of color names." Proceedings of the National Academy of Sciences.

Lucassen, M. P., A. Gijsenij, et al. (2008). "Chromatic distribution affects colour constancy." Perception 37: 147-147.

Lucassen, M. P. and J. Walraven (1996). "Color constancy under natural and artificial illumination." Vision Research 36(17): 2699-2711.

Malo, J. and M. J. Luque (2002). COLORLAB: a color processing toolbox for Matlab.

Maloney, L. T. (2002). "Illuminant estimation as cue combination." Journal of Vision 2(6).

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

Margrain, T. H., J. Birch, et al. (1996). "Colour vision requirements of firefighters." Occupational Medicine-Oxford 46(2): 114-124.

Marr, D. (1982). Vision: a computational investigation into the human representation and processing of visual information. San Francisco.

McCann, J. J., S. P. Mckee, et al. (1976). "Quantitative Studies in Retinex Theory - Comparison between Theoretical Predictions and Observer Responses to Color Mondrian Experiments." Vision Research 16(5): 445-&.

Menegaz, G., A. L. Troter, et al. (2007). "A discrete model for color naming." EURASIP J. Appl. Signal Process. 2007(1): 113-113.

Murray, I. J., A. Daugirdiene, et al. (2006). "Almost complete colour constancy achieved with full-field adaptation." Vision Research 46(19): 3067-3078.

Neitz, M. and J. Neitz (2001). "A new mass screening test for color-vision deficiencies in children." Color Research and Application 26: S239-S249.

Nemes, V. A., D. J. McKeefry, et al. (2010). "A behavioural investigation of human visual short term memory for colour." Ophthalmic and Physiological Optics 30(5): 594-601.

Nolan, J., S. Riley, et al. (2008). "Color naming based on clinical visual condition: A surprising interaction." Journal of Vision 8(6): 579.

Olkkonen, M., T. Hansen, et al. (2009). "Categorical color constancy for simulated surfaces." Journal of Vision 9(12): 6 1-18.

Olkkonen, M., C. Witzel, et al. (2010). "Categorical color constancy for real surfaces." Journal of Vision 10(9).

Otazu, X., C. A. Parraga, et al. (2010). "Towards a unified model for chromatic induction." Journal of Vision 10(12): 5 1-24.

Paggetti, G., G. Bartoli, et al. (2011). "Re-locating colors in the OSA space." Attention Perception & Psychophysics 73(2): 491-503.

Parraga, C. A., J. Roca-Vila, et al. (2011). "Do basic colors influence chromatic adaptation?" Journal of Vision 11(11): 349.

Peacock, J. A. (1983). "Two-Dimensional Goodness-of-Fit Testing in Astronomy." Monthly Notices of the Royal Astronomical Society 202(2): 615-627.

Pointer, M. R. and G. G. Attridge (1998). "The number of discernible colours." Color Research & Application 23: 52-54.

Raskin, L. A., S. Maital, et al. (1983). "Perceptual categorization of color: a life-span study." Psychol Res 45(2): 135-145.

Roberson, D., I. Davies, et al. (2000). "Color categories are not universal: replications and new evidence from a stone-age culture." Journal of Experimental Psychology: General 129(3): 369-398.

Roca-Vila, J., A. Owen, et al. "A compact description of colour naming ability: Quantifying individual variations." In preparation.

Roca-Vila, J., A. Owen, et al. (2011). "Inter-individual variations in color naming and the structure of 3D color space." Journal of Vision 11(11): 386.

Roca-Vila, J., C. A. Parraga, et al. "Chromatic settings and the structural color constancy index." Submitted to Journal of Vision (2^{on} revision).

Roca-Vila, J., C. A. Parraga, et al. (2012). "Predicting categorical colour changes in successive colour constancy." Perception(ECVP Abstract Supplement).

Roca-Vila, J., C. A. Parraga, et al. (2012). What is constant in color constancy? CGIV 2012 Sixth European Conference on Colour in Graphics, Imaging, and Vision. IS&T. Amsterdam: 337-343.

Roca-Vila, J., M. Vanrell, et al. (2011). "Categorical focal colours are structurally invariant under illuminant changes." Perception 40(ECVP Abstract Supplement): 196.

Rushton, W. A. (1972). "Pigments and signals in colour vision." J Physiol 220(3): 1P-P.

Schultz, S., K. Doerschner, et al. (2006). "Color constancy and hue scaling." Journal of Vision 6(10): 1102-1116.

Shevell, S. K. (2003). The science of color. Amsterdam ; London, Elsevier ; [United States] : Optical Society of America.

Shevell, S. K. and F. A. A. Kingdom (2008). "Color in complex scenes." Annual Review of Psychology 59: 143-166.

Shevell, S. K. and J. P. Wei (2000). "A central mechanism of chromatic contrast." Vision Research 40(23): 3173-3180.

Smith, V. C. and J. Pokorny (1975). "Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm." Vision Research 15: 161-171.

Smithson, H. and Q. Zaidi (2004). "Colour constancy in context: roles for local adaptation and levels of reference." J Vis 4(9): 693-710.

Smithson, H. E. (2005). "Sensory, computational and cognitive components of human colour constancy." Philos Trans R Soc Lond B Biol Sci 360(1458): 1329-1346.

Snowden, R. J., P. Thompson, et al. (2006). Basic vision : an introduction to visual perception. Oxford, Oxford University Press.

Spalding, J. A. (1999). "Colour vision deficiency in the medical profession." Br J Gen Pract 49(443): 469-475.

Speigle, J. M. and D. H. Brainard (1997). Is color constancy task independent? Recent progress in color science, Springfield, The Society for Imaging Science and Technology.

Speigle, J. M. and D. H. Brainard (1999). "Predicting color from gray: the relationship between achromatic adjustment and asymmetric matching." Journal of the Optical Society of America a-Optics Image Science and Vision 16(10): 2370-2376.

Stockman, A. a. D. H. B. (2010). Color vision mechanisms. OSA Handbook of Optics. M. B. a. V. N. Mahajan, McGraw-Hill.

Sturges, J. and T. W. A. Whitfield (1995). "Locating Basic Colors in the Munsell Space." Color Research and Application 20(6): 364-376.

Troost, J. M. and C. M. de Weert (1991). "Naming versus matching in color constancy." Perception and Psychophysics 50(6): 591-602.

Uchikawa, K. (2008). "Trichromat-like Categorical Color Naming of Dichromats." VISION 20(2): 62-66.

Uchikawa, K., Y. Emori, et al. (2002). "Color constancy in categorical color appearance." Journal of Vision 2(7): 548.

Vanleeuwen, M. T., C. Joselevitch, et al. (2007). "The contribution of the outer retina to color constancy: a general model for color constancy synthesized from primate and fish data." Vis Neurosci 24(3): 277-290.

Vazquez-Corral, J., C. A. Parraga, et al. (2009). "Color Constancy Algorithms: Psychophysical Evaluation on a New Dataset." Journal of Imaging Science and Technology 53(3).

Vazquez-Corral, J., M. Vanrell, et al. (2011). "Color Constancy by Category Correlation." IEEE Trans Image Process.

von der Twer, T. and D. I. MacLeod (2001). "Optimal nonlinear codes for the perception of natural colours." Network 12(3): 395-407.

Von Kries, J. (1905). Influence of adaptation on the effects produced by luminous stimuli (Handb. Physiol. Menschen). Sources of Colour Science. D. L. MacAdam. Cambridge, MA (1970), MIT Press: 109-288.

Walsh, V. and J. J. Kulikowski (1998). Perceptual constancy : why things look as they do. Cambridge, UK ; New York, NY, USA, Cambridge University Press.

Wandell, B. A. (1995). Foundations of vision. Sunderland, Mass., Sinauer Associates.

Webster, M. A. (1996). "Human colour perception and its adaptation." Network-Computation in Neural Systems 7(4): 587-634.

Webster, M. A. and P. Kay (2007). Individual and population differences in focal colors. Anthropology of color : interdisciplinary multilevel modeling. R. E. MacLaury, G. V. Paramei and D. Dedrick. Amsterdam ; Philadelphia, J. Benjamins Pub. Co.: 29-54.

Webster, M. A. and D. I. A. Macleod (1988). "Factors Underlying Individual-Differences in the Color Matches of Normal Observers." Journal of the Optical Society of America a-Optics Image Science and Vision 5(10): 1722-1735.

Webster, M. A., E. Miyahara, et al. (2000). "Variations in normal color vision. II. Unique hues." Journal of the Optical Society of America a-Optics Image Science and Vision 17(9): 1545-1555.

Webster, M. A., S. M. Webster, et al. (2002). "Variations in normal color vision. III. Unique hues in Indian and United States observers." Journal of the Optical Society of America a-Optics Image Science and Vision 19(10): 1951-1962.

Wolf, K. (2011). The Modulation of Simultaneous Chromatic Contrast PhD Thesis, Newcastle University.

Wyszecki, G. n. and W. S. Stiles (1982). Color science : concepts and methods, quantitative data and formulae. New York ; Chichester, Wiley.

Xiao, K., S. Wuerger, et al. (2011). "Unique hue data for colour appearance models. Part I: Loci of unique hues and hue uniformity." Color Research & Application 36(5): 316-323.

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

Yokoi, K. and K. Uchikawa (2005). "Color category influences heterogeneous visual search for color." Journal of the Optical Society of America a-Optics Image Science and Vision 22(11): 2309-2317.

Zaidi, Q. and M. Bostic (2008). "Color strategies for object identification." Vision Research 48(26): 2673-2681.

Zaidi, Q. and H. Smithson (2004). "Roles for local adaptation and levels of reference in colour constancy." Perception 33: 24-24.