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.

Jordi Roca i Vilà

Bellaterra, September 2012

Supervisors:

Dr. Maria Vanrell

Universitat Autònoma de Barcelona

Dept. Ciències de la Computació & Computer Vision Center

Dr. Alejandro Párraga

Universitat Autònoma de Barcelona

Dept. Ciències de la Computació & Computer Vision Center

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

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 cotidianitat dels anys en què aquest projecte s'ha extès. 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 conectat malgrat les meves intenses tribulacions.

Seguidament voldria expressar el meu agraïment a la Maria i l'Alejandro, els meus directors de tesi. Primer de tot, per donar-me la oportunitat d'empendre un projecte tant definitori en el procés formatiu d'una persona com n'és la realització d'un doctorat. També, la 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 mutu 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 al seu laboratori i em va permetre complementar i extendre 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 als 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.

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. Addicionalment, 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

- *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. Journal of Vision (Under review).

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

- *Categorical colour perception in successive colour constancy*. Roca-Vila J., Parraga C. A., and Vanrell M. (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

xiii

List of Figures

xv

List of Tables

xvii

Chapter 1 Introduction

This first chapter begins with the introduction to the main framework of the thesis, the human visual system, and in particular with a broad description of the basic concepts of human vision, colour vision, colour constancy and categorical colour perception. Next we motivate and introduce the main goal of the thesis; the study of human categorical colour perception under illumination changes. The secondary goal of this work is to contextualize its contributions in the computer vision field, something that keeps on the long tradition of knowledge flow between the fields of biological and artificial vision. This point will be stressed in the last chapter where the thesis conclusions will suggest new approaches to solve the colour constancy problem in computer vision.

1.1 Background

Vision

The first and primary component for human vision to take place is the existence of light, without light there is no vision. The back layer of the human eye, called retina, has several types of photosensitive cells with a specific sensitivity range to the spectral properties of light; it responds to wavelengths from about 400 to 750 nm. Thus, light is only processed when its spectral wavelength is in that particular range, light with such properties is called visible light. 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, 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 and psychophysics.

Seeing starts when visible light is focused by the lens 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 (Hubel 1988). 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. Rods are active in dim light and provide black-and-white 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 output from 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. 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). 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. 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 included: 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.

(a) The eye

(b) The HVS in the brain

(c) FMRI of JR observer

(d) JR V4 area in a normalized brain
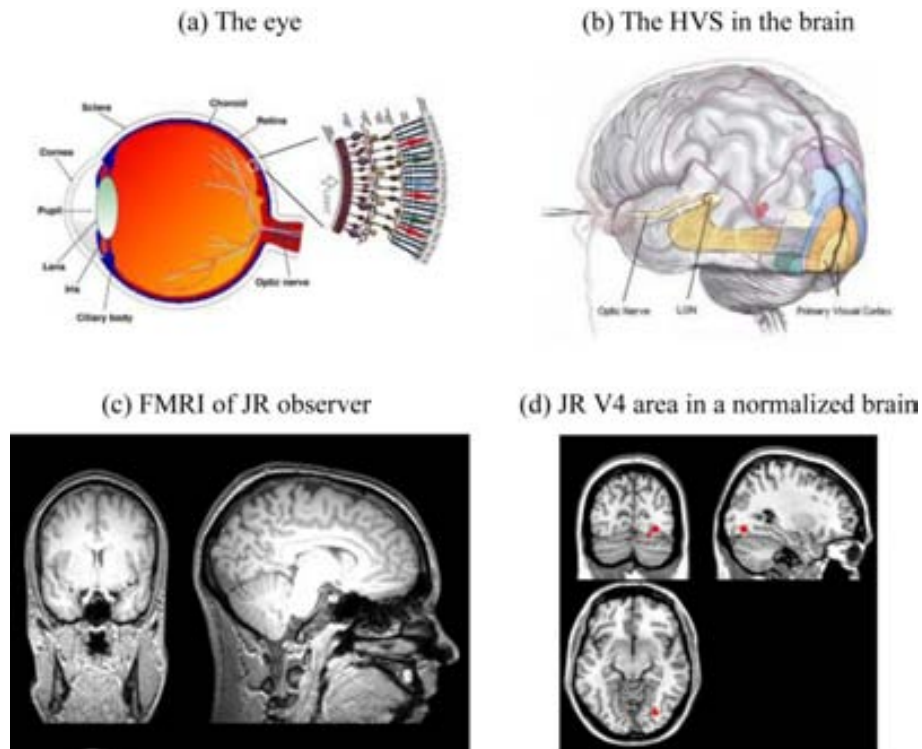
Figure 1.1 The Human Visual System (HVS).Panel a: it contains a schematic description of the main components of the eye. Panel b: it visually describes the location of the visual pathway in the brain. Panel c: FMRI brain scan of the author (JR) done when he participated as a subject in a colour perception experiment at the Institute of Neuroscience (UK). Images courtesy by Anya Hurlbert Professor of Visual Neuroscience (Newcastle University, UK). Panel d: Normalized brain with JR V4 area highlighted in red.

Colour Vision

The existence of three physiologically different cones in the retina makes possible the perception of colour. These cones are distinguished by the type of photo pigment molecules they contain and they are called L, M and S which are selective for long, middle and short wavelengths respectively. This is called the *trichromacy* principle and in practical terms implies that any visible light can be 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) and in practical terms; 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 sensitivity 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. This 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.
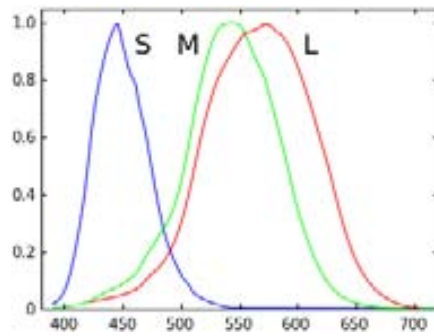
Figure 1.2 Spectral absorption functions of the three different types of cone photoreceptors.

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, which 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 difference of L and M cones is taken; and a blue-yellow opponent axis where the S cone signal is differenced with the sum of L and M cone signals. Notice that despite the colour term nomenclature used to define these colour opponent axis, its directions do not align with the axis predicted by traditional colour appearance theories of unique hues (Those four primary hues that do not contain perceptual components of other hues, e.g., a unique red hue is neither yellowish nor bluish (Kuehni 2001)) (Hurlbert 1997; Gegenfurtner and Kiper 2003).

Despite this simple explanation on 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 colour perception (Hurlbert 1997; Gegenfurtner and Kiper 2003). For instance, recent studies has 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 in the objects of the visual scene serves several functionalities; coloured objects are more easily detected, more easily identified, more easily grouped, and more easily remembered than black-and-white 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 theories of colour vision used psychophysical experiments with stimuli focused on isolated patches of light but the colour appearance of objects/surfaces depends strongly on the light from other

4

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 an isolated patch of light. The simplest context used in laboratory conditions is an uniform background field, which has been used extensively over the last century. However, has been demonstrated that this simple stimuli cannot account for appearance shifts caused specifically by chromatic variegation within the scene (Shevell and Kingdom 2008).

Colour constancy

A green apple looks green to us at midday, when the main illumination is white sunlight and also at sunset, when the main illumination is red, this property helps us to identify objects. In general the colour of perceived objects and surfaces tends to stay the same over changes in illumination, this phenomenon is known as colour constancy (Land 1964; Smithson 2005; Foster 2011). It is a fundamental property of visual perception and it is achieved by multiple mechanisms located at several locations in the HVS (Hurlbert and Wolf 2004). The simultaneous activation or inhibition of such mechanisms depends on the visual cues present in the visual environment. Figure 1.2 shows a schematic example of colour constancy; panel a contains a coloured cube under an "achromatic" illuminant and panel b contains the same cube under a "greenish" illuminant. If the visual scene corresponding to the panel b was seen in isolation then colour constancy mechanisms would tend to compensate the greenish illumination to recover somehow a "normalized" perception, similar to the one of panel a. Since humans are not perfectly colour constant, a natural goal for experimentation is to explore up to what extent humans are colour constant. Furthermore, to find principles and to develop models that allow us to predict colour appearance in complex scenes (Brainard, Brunt et al. 1997). Colour constancy is not a property of objects; it is a perceptual phenomenon resulting from mechanisms in the eye and brain (Hurlbert 1996).

Figure 1.3 Schematic example of colour constancy. Panel a and b corresponds to the same object illuminated under different illuminations, an "achromatic" and "greenish" illumination respectively.

The physics of the retinal image formation rely on two factors: the illuminant's spectral power distribution, $E(\lambda)$, which specifies the amount of power in the illuminant at each wavelength; and 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, $C(\lambda)=E(\lambda)S(\lambda)$. The colour signal confounds the illuminant's spectral power distribution and the surface's reflectance function, then 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. Several possible strategies have been proposed to make colour constancy possible. These include restrictions on the number and dimensionality of the spectral reflectances 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).

Categorical colour perception

Categorization is a fundamental human attribute and humans tend to classify colours into several categories. An important method of categorization is colour naming, which may assign the same colour term to two coloured samples with different LMS responses. Names of some commonly used colour categories include the unique hues: red, green, blue and yellow. This categorization divides the 3D colour space into several regions each linked to a different colour name. 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.

There are two main proposals to explain our ability to categorize colour perception. First, the Sapir-Whorf hypothesis, which states that colours are perceived categorically only because they happen to be named categorically and vary across cultures and languages (Roberson, Davies et al. 2000). Second, Berlin and Kay showed that most cultures and languages subdivide and name the colour spectrum in the same way (Berlin and Kay 1969), and so colour names are coded in some innate human feature. 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).

Thus, the explanation of colour categorization is framed in terms of the old nature-nurture debates (Hardin 2005), i.e., if colour categorization reside on childhood learned patterns or it is physiologically based. There is solid evidence of the physiologically based explanation (Kay and Regier 2003; Lindsey and Brown 2006), however 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 when the only constraints in their formation are fundamental sensory discrimination abilities (Belpaeme and Bleys 2005; Baronchelli, Gong et al. 2010). Thus, a combination of these two factors seems to be the most plausible answer.

1.2 Motivation

The main aim of this work is to study the relationship between colour constancy and categorical colour perception. In practical terms, the perceptual phenomenon studied can be grasp from the example of Figure 1.2. Suppose that you are immersed in a visual scene equivalent to the one in panel a, i.e., you are perceiving a single object with multiple patches of colours in its surface and it is diffusely illuminated with an "achromatic" illuminant. Next, the chromaticity of this overall illumination is changed to be highly chromatic, for instance greenish, as shown in panel b. Immediately after this illumination change, the colour adaptation processes in the HVS will be activated and the colour

perception of the patches in the object surface will change and it will become stable after a while, probably several minutes (usually a minimum of 5 minutes). Some interesting questions arise from this situation. Are the colours perceived in both scenes the same? If it is so, we will say that colour adaptation mechanisms achieved colour constancy. If not, and this is most probable since colour constancy is not perfect and specially for high chromatic illuminants, which are the differences? Quantitative differences according to the number of perceived colours are most likely to appear since adaptation moves toward the centre of colour space. Also, qualitative differences could appear, i.e., are perceived colours more categorical? Even we can study if their distribution in colour space has been altered. Interestingly, we can also modify the initial stimuli in terms of the same variables; number of colours, spatio-chromatic distribution of colours; and categorical quality of colours. For instance, if all the colour patches of the object belongs to a colour category, or any of them, would make a difference in the adaptation? A particular case of the previous example is when all coloured patches are different achromatic samples, then according to the Helson-Judd effect, after adaptation the lighter ones are perceived towards the illumination chromaticity, the mid ones as greyish and the dimmer ones as the chromaticity opposed to the illumination one. Thus, the Helson-Judd effect reveals how in the simplest case of using only achromatic colours the quantity and quality of perceived colours is altered by illumination changes.

Now we state the previous discussion in formal terms; to study if exist, or not, any influence of categorical colour perception in the processes of colour adaptation. In order to answer this question we structured our approach around the study of the next two points:

- Categorical colour perception is maintained. Categorical colours just perceived after the illumination change and over the first stages of adaptation, while adapting to the global contrast, constraint the adaptation process and the final categorical colour perception.
- Categorical colour perception is seek. The probability of categorical recovery under a canonical illumination is what constraints the adaptation process.

The first question to answer is if any relation exists and if it so how it is articulated. We started our approach by measuring the categorical colour perception of individuals in detail, specially at inter categorical regions. Next we used a psychophysical paradigm that provided precise measures in colour space as well as categorical information about them. This allowed to study with precision how categorical colour information present in the stimuli influenced colour perception, i.e., characterizing the whole categorical structure of colour space. In order to isolate this property we performed the experiments in laboratory conditions with two dimensional simple images, i.e., a collage of coloured rectangles whose surfaces followed a Lambertian model (uniformly illuminated and without

specularities and/or multiple reflectances). Then we varied the number of colours and its categorical quality while each stimuli was particularly tuned to each observer in order to keep assure a high degree of precision in the stimuli design when selecting categorical colours.

A further understanding of this relations could lead to understand better the HVS and in particular colour adaptation processes and the colour constancy phenomenon, thus making available accurate models to predict colour appearance.

1.3 Colour constancy in Computer Vision

Computer Vision is the scientific field that is concerned with the acquisition, processing, analyzing and ultimately understanding of images. A principal line of research within this field has been to replicate the abilities of human vision, this propitiated a flow of inspirational ideas from theories of visual perception. Also, from the other way round, methods and algorithms developed in the Computer Vision field have been a source of debate in visual perception. There is a flow of knowledge between both scientific fields, especially since Marr contributed to this schema in his influential book Vision (Marr 1982; Brainard 2003). This thesis is deeply framed in this dialog.

Colour constancy is a particular relevant example of this interaction. Images acquired with cameras by computer vision systems face the same challenge as the human visual system, i.e., in a highly variable environment where spectral information changes frequently they need to provide a constant output which can be interpreted as stable colours. Computer vision algorithms has a long tradition of dealing with the colour constancy phenomenon, however the aims of this algorithms do not necessarily coincide with the same. For instance, the output of a commercial camera needs to be highly similar to the human colour perception while the output for a industrial camera in a quality control process may only require some stability.

Computational colour constancy seems to be a relevant example to illustrate a major approach of Computer Vision to solve the colour constancy problem. According to Hurlbert the computational colour constancy problem is stated as: "*to recover the invariant spectral reflectance properties of object surfaces from the image irradiance, in which reflectance is entangled with surface illumination*" (Hurlbert 1998). However, Jameson and Hurvich state 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 the normalization of the image. However computer vision researches should keep in mind the final aim of their algorithms, and if it is to mimic the HVS, and in particular colour adaptation processes, they should evaluate their algorithms according to perceptual and not physical properties (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 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 about the probability with which different colours are observed under different lights. Finally, a threshold procedure can return the most likely illuminant or set of illuminants. A notable feature of Finlayson approach is that it incorporates several previous colour constancy algorithms in its theoretical framework.

Following this approach a new study has focused on the selection of illuminants according to the effects they produce on the final categorical properties of colours, i.e., to include perceptual constraints that are computed on the corrected images. In detail, 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 (Vazquez-Corral, Vanrell et al. 2011).

The research line of this thesis is closely linked to the previous described approach in Computer Vision, i.e., it aims to answer if the HVS may also use this categorical information to achieve colour constancy.

1.4 Thesis outline

This thesis is conformed of seven chapters. In Chapter 1 we have described the main framework where the thesis develops and introduced its principal aim. Each of Chapters 3,4,5 and 6 reports a new psychophysical experiment, and they are labelled as Experiment I, II, III and IV respectively.

Chapter 2 contains a bibliographical review of the two main topics involved in this thesis. First, the topics of categorical colour perception and colour naming are introduced. Second, and more

extensively, the topic of colour constancy is introduced and special attention is devoted to previous studies which tackled both issues at the same time.

Chapter 3 studies the structure of individual categorical colour space through an extensive colour naming experiment, Experiment I. The results from this experiment on the structure of categorical colour space set the constraints for the newly developed paradigm to measure colour constancy introduced in the following chapter.

Chapter 4 introduces a new paradigm, termed chromatic setting, to measure colour constancy. It also contains a new experiment, Experiment II, with several different simple stimuli in order to test the new paradigm. The results demonstrate that the chromatic setting paradigm is a reliable method to study colour constancy.

Chapter 5 applies the chromatic setting paradigm in a new experiment, Experiment III, to study the stability of the categorical colour space structure. The experiment used Mondrians with a reduced number of categorically balanced colours under five different illuminations. The results suggest that most of the categorical structure is preserved.

Chapter 6 studies categorical colour changes in colour constancy through Experiment IV which is conformed of the chromatic setting paradigm and a constrained colour naming task. In order weaken colour constancy, i.e. to maximize possible categorical changes, the number of colours in the stimuli was reduced to only three. In order to quantify categorical changes, it also introduces a new categorical colour prediction model which uses the categorical information contained in chromatic settings as well as its precise location.

Finally, Chapter 7 contains a summary from the results of the previous chapters. A general discussion states the main contributions of the thesis and future lines of research are proposed, and its conclusions linked to the Computer Vision field.

Chapter 2 Review of related research

This chapter contains a bibliographical review of the two main topics involved in this work; categorical colour perception and colour constancy. First, categorical colour perception is introduced; its main definitions and features. Second, the colour constancy topic is introduced and described common definitions and its main features, also how its measured, quantified and modelled.

2.1 Categorical colour perception

The colour experience can be accessed at two levels: discriminative and categorical. Humans have the ability to assign categories to many concepts, the same is true for the perception of colour. Colours are communicated through words which clearly indicate the candidate categories in which to divide the colour experience.

2.1.1 Colour naming

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 threatened 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). 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 (Kuehni 2004) (Raskin, Maital et al. 1983). 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).

14

2.1.2 Basic Colour Terms

Here we state in detail the work of Berlin and Kay and its main conclusions (Berlin and Kay 1969). Berlin and Kay observed that colour words translated easily between different languages and studied systematically the colour terms usage in several languages. Their approach consisted in a simple unconstrained colour naming task where observers assigned a colour word to a coloured surfaces extracted from a subset of the Munsell colour space (Fairchild 1998). Observers were also asked to specify the best example of the colour category, and the region of chips on the array that would assign the same colour name. Basic colour terms were defined as colour names with the following properties:

1. It is monolexemic, a single lexical term.
2. They refer to the colour of objects.
3. Colours are applicable to a wide range of objects.
4. Colours are in frequent use.

The experiment used 20 subjects and was completed with information from published works of other 78 languages. They reach the main conclusions: First, 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; Second, 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.

Furthermore, in their research they also studied which basic colour terms describes best any particular colour. They obtain 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 specified colour term, or ask to identify the colours that best describe a particular colour term. Participants gives best and faster answers to the second task and this implies that colour categories are structured around focal colours, rather than around colours at the boundaries of colour terms (Heider 1972).

These and other results are a considerable evidence that focal colour play a crucial role in our internal representation of colour categories.

2.2 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 relatively 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).

2.2.1 Definitions and main features

Perceptual constancy is a functional feature of the HVS which aims to keep the perception of objects stable (size, shape, colour and location) regardless of changes in the viewing angle, distance or illumination (Brainard,1993 CC: From physics to appearance) (Smithson 2005). This feature takes part in the general process of mental reconstitution of the *real image* from retinal images. However, according to Jameson and Hurvich (Jameson and Hurvich 1989) the HVS is likely to have evolved a design that provides perceptual information about constancy as well as change (about light, weather or time of the day) (Jameson and Hurvich 1989). Colour constancy is a major example of perceptual constancy and has been typically studied only when the illumination variable is changed (ref).

In order to keep the colour perception of objects stable the HVS has two main challenges. First, the information from the object spectral reflectance function is mixed with the spectral composition of the illumination and thus the spectral information reaching the cones in the retina confounds both sources of information. Second, the colour signal reaching the retina is further simplified to only three different signals produced by the L, M and S cones. Colour constancy mechanisms deal with several physical and physiological constraints. Thus colour constancy has not a unique solution and then restrictions on the surface spectral reflectance and spectral properties of illuminants are needed to achieve successful colour constancy (Maloney and Wandell 1986).

The most common definition of colour constancy states the problem as the ability of the HVS to keep a stable colour perception of objects and surfaces despite changes in the illumination (Foster and Nascimento 1994; Smithson 2005). However, the richness of situations in the real world allows to more concise definitions of colour constancy according to how light can vary or more specific conditions. Successive colour constancy refer to illumination changes that occur over time, e.g., the light spectrum at dawn differs from that at noon. Another possibility is light changes across space, for example when the same scene is illuminated by two different lights, it is called simultaneous colour constancy (Brainard, Brunt et al. 1997; Ling and Hurlbert 2008). However, both types of colour constancy are interrelated since simultaneous thorough time becomes successive. The key differences lies in the temporal scale, simultaneous much shorter than successive. Atmospheric colour constancy refers to the change of the viewing medium, nor the light source or the object's surface properties, for

example fog, mist or smoke (Foster 2011). A different approach is relational colour constancy which stands only for keeping correspondences with object surfaces despite losing the perception of colour, something that follows from the property of cone ratio invariance (Foster and Nascimento 1994). This property states that the ratios of cone excitations generated from pairs of surfaces or several surfaces, adjacent or not, are nearly invariant under illumination changes.

Main empirical features

Previous literature studied the extent and quality of the colour constancy phenomenon according to features of the visual scene and derived conclusions about the HVS (Smithson 2005; Foster 2011). The following points summarize the main empirical features of successive colour constancy:

- It is not perfect, i.e., adaptation mechanisms in the HVS compensate for illuminations changes up to certain degree according to visual cues (Brainard, Brunt et al. 1997; Hurlbert 2007).

- It is a contextual phenomenon, i.e., there is no colour constancy without a context (Hurlbert 2007). The phenomenon can be studied according to the number of visual cues present in the visual scene. In general, an increasing number of consistent visual cues implies a higher degree of achieved colour constancy (Kraft and Brainard 1999).

- The resultant adaptation state is always shifted towards the illumination chromaticity and close to the line joining the reference and illuminat coordinates in an appropriate colour space. Common colour constancy indices are based on this property.

2.2.2 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). The most popular colour constancy paradigms are *achromatic setting*, *asymmetric colour matching* and *colour naming* (Smithson 2005; Foster 2011).

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 will be determined by the pattern of the eye movements across the two halves of the scene. Successive matching allows experimental control of adaptation to the two illuminants, but performance will additionally 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 focals 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 last 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).

There are several common methodological factors that can influence the outcome of the previous paradigm measurements:

- Nature of the stimuli, if using laboratory or real conditions. Most of the traditional studies used simplified stimuli (e.g. Mondrians in Figure 2.1) which allowed to discriminate among several properties that in natural images are entangled (Kraft and Brainard 1999). This leads to the question: which cues present in the natural world are exploded by colour constancy mechanisms?

- Question to ask the observer (Troost and de Weert 1991).

- Effects of familiarity and memory (Hansen, Olkkonen et al. 2006; Ling and Hurlbert 2008).

- The number of surfaces in the visual scene (Linnell and Foster 2002) .
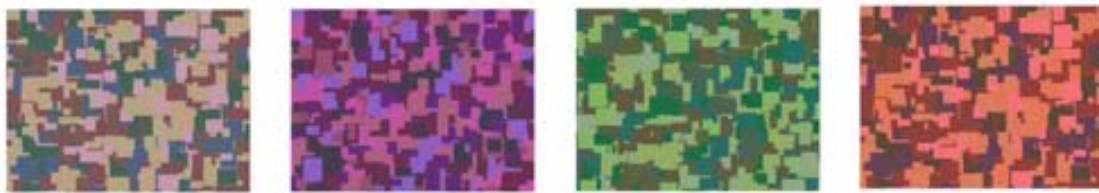
Figure 2.1 Mondrian images used as stimuli. The image on the left contains eight different colours under D65 illumination and the following images from left to right correspond to three different simulated illuminations on the left image.

2.2.3 Composing mechanisms and visual cues

Colour constancy is composed by multiple mechanisms that span multiple levels in the HVS. This mechanisms are activated simultaneously or exclusively according to the visual cues elicited (Hurlbert and Wolf 2004; Smithson 2005; Foster 2011). According to Hurlbert et al (Hurlbert and Wolf 2004), this mechanisms can be grouped according to the type of computational mechanism and neural level into *sensory*, *perceptual* and *cognitive*. Each one responsible to a certain extent of the phenomenon.

Sensory mechanisms are represented by linear transformations of the photoreceptor responses. Traditional Von kries adaptation is based on that approach when normalizing each photoreceptor signal by its average cone type over the image (Von Kries 1905). However, current approaches do not support Von Kries adaptation, take for instance inter connections among cone types as early as in the outer retina between the different cone types through the horizontal cells (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 the visual cues from which estimate the illuminant of the scene corresponding to that neural level.

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).

	Sensory	Perceptual	Cognitive
Computational mechanism	Chromatic adaptation		
Neural level	Retina, LGN	Cortex VI and V2	Brain, Cortex V4
Requirements	Light	Scene segmentation	Object identification
Visual cues	Chromatic adapt. to the mean (1) Chromatic adapt. to the variance (2) Colour contrast (3)	3D information Mutual reflectance Chroma variance Specularities Scene movement	Memory colours Consciousness of illumination change Familiar objects
Temporal scale	(1) 60 sec (2) minutes (3) milliseconds	Several minutes	Days, years

Table 2.1 Visual cues in colour constancy.

According to Webster (Webster 1996), human colour adaptation has two main processes: (i) retinal mechanisms of light adaptation adjust for changes in the average colour across scenes and (ii) cortical mechanisms of contrast adaptation adjust for changes in the distribution of colours. These two processes are in accordance to the previous division into sensorial, perceptual and cognitive factors of colour constancy.

2.2.4 Quantifying colour constancy

The extent of colour constancy achieved by observers is traditionally measured through a one dimensional index, i.e., a normalized measure which compares measurements done under the reference and test illuminations (Smithson 2005; Ling and Hurlbert 2008; Foster 2011). Values of colour constancy indices ideally range between 0 and 1; values lower than one indicate a colour constancy failure and values over one indicate an overcompensation. However, colour constancy indices are a simplification of the phenomenon and must be interpreted according to its particular experimental context.

Obtained measurements under the reference and test illumination are interpreted in some colour space, and computations are applied on their colour space coordinates. Each colour space has its own characteristics and its selection must be in accordance to the particular features to be extracted from data. Two dimensional colour spaces as CIE1931 xy and CIE1976 uv are widely used and one of its advantages is the lack of a reference white point specification and thus not incorporating any further chromatic transform of data. CIE1976 uv also has the advantage of being perceptually uniform,

meaning that distances are more meaningful than in CIE1931 xy (Brainard, Brunt et al. 1997; Ling and Hurlbert 2008). Also three dimensional colour spaces are used, for instance CIELab and CIELuv, or its projections in their chromaticity planes (Foster 2011).

Equations 2.1 and 2.2 details 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 *test patch* the *ideal match* and the *observer match*. The test patch corresponds to the coordinates of the patch to study its constancy, the ideal match corresponds to the coordinates of the test patch under the illuminant and the observer match correspond to the coordinates of the resulting experimental measurement.

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

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

The implicit idea behind CI and CR assume that observer adaptation lies somewhere in the joining line between the test and reference illuminant coordinates. Something that is not always the case as shown by Figure 2.2 where the BR takes the same value for different observer match locations. Subsection 4.5.3 follows the discussion on colour constancy indices while introducing a new index.

When using achromatic setting to perform the experimental measurements the test and ideal match corresponds to the coordinates of the respective test and reference illuminants.
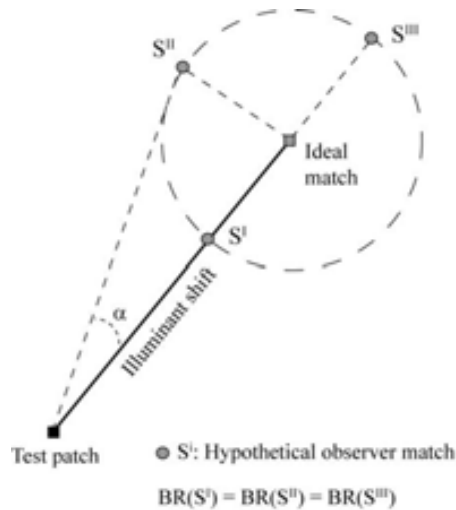
Figure 2.2 Accuracy deviation of traditional colour constancy indices. The Figure illustrates how several different measurements from the observer match (S^i) derive in the same value of the Brunswich Ratio index.

2.2.5 Predicting colour constancy

A general goal of colour constancy studies is to develop principles and models that allow the prediction of colour appearance under complex scenes (Burnham, Evans et al. 1957; Jameson and Hurvich 1964; Brainard, Brunt et al. 1997), and in extension to our everyday environments. In particular, to predict the colours seen under a particular state of adaptation, that were not directly measured under a test illumination. Notice that some colour appearance spaces incorporate versions of these colour constancy models (Fairchild 1998). Three are two main common approaches to predict colour constancy effects: (i) Colour conversions expressed in terms of a limited set of linear basis functions, (ii) Illuminant estimations from the computational approach and (iii) Modelling in the LMS colour space.

As previously exposed colour constancy is a contextual property and some studies linked this context to the statistics existing in the natural world. From this standpoint the idea has been to restrict the number of possible illuminations and surface reflectance's as the combinations of limited set of basis function in a Banach space. This is a set of functions with embedded in the structure of a vector space. Then, once the whole problem is reduced to find the components of the vector that represents the illumination or the surfaces spectral reflectance. The second method is to model via an affine matrix changes in the LMS cones. Typical Von Kries adaptation is a simple diagonal matrix.

Chapter 3 Quantifying individual variations in categorical colour perception

Previous studies reported large variations in the categorical colour perception according to different individuals. So our first step is to describe the individual categorical colour structure in an adequate colour space and quantify its inter-individual variability. This chapter quantifies this variability through the colour naming ability of individuals. We performed a psychophysical experiment (*Experiment I*) which used normal trichromat observers and a series of custom colour naming tasks under D65 illumination and achromatic uniform backgrounds. The study of the colour naming abilities of our observers allowed a direct link to the categorical colour structure of their internal colour spaces. Six indices are developed to compactly measure their colour naming abilities, and from its values conclusions are derived. Results indicate that the categorical structure of colour space is broadly uniform across observers: similar when close to the basic categorical centroid regions but quite different close to its 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, that this vocabulary needs to be expanded into a few more several terms at least. Finally the suggestion that the inner part of the colour space are the most susceptible to deviations from colour constancy. These findings sets the constraints for the new paradigm and experimental design that will be developed in the following chapter.

Some of the work described in this chapter has been published in a paper (currently submitted) to which other co-authors contributed. The original contribution of the author consisted into the analysis

of the experimental data, the development of new methods, the computational analysis and its discussion.

3.1 Introduction: Categorical colour perception from colour naming ability

Colour naming ability is 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 compactly 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, 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, 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 custom-made 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; $\sigma = 13.24$; 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%.
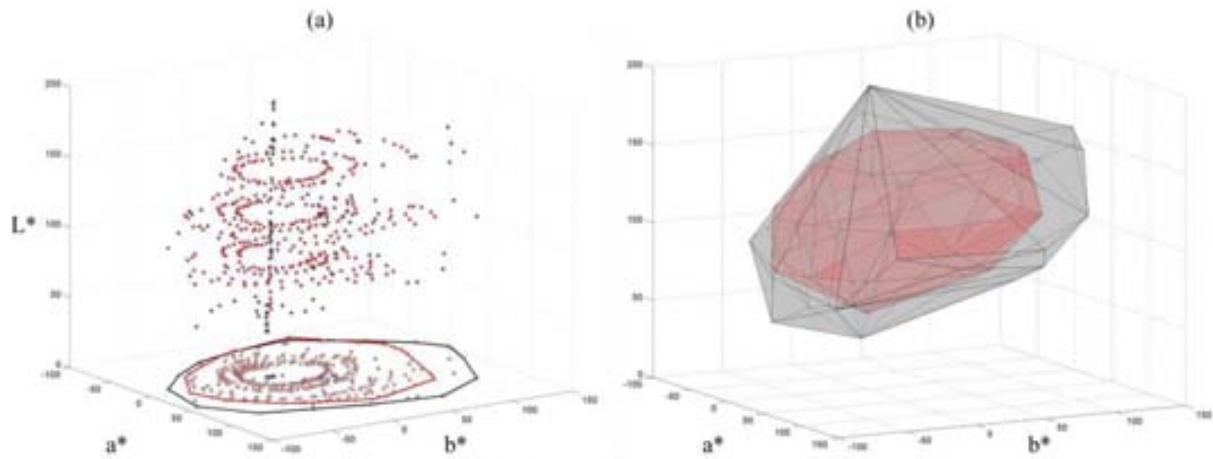
Figure 3.1 Tested coloured samples in CIELab colour coordinates. Panel a: Red dots and grey squares correspond to light and surface tested samples. Panel b: Volumetric contribution of each experimental set.

3.3 Data analysis

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 of the individual colour solid

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 $\mathcal{T}_i^j$ as the corresponding set of response times, where each t_k is the time spent classifying p_k in seconds.

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

$$\mathcal{T}_i^j = \left\{ t_{k_1}, \dots, t_{k_{M_i^j}} \right\} ; \ t_{k_l} \in \mathbb{R} \ l = 1, \dots, M_i^j \ ; \ j = 1, \dots, \bar{n} \ and \ i = 1, \dots, 11 \qquad (3.2)$$

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.
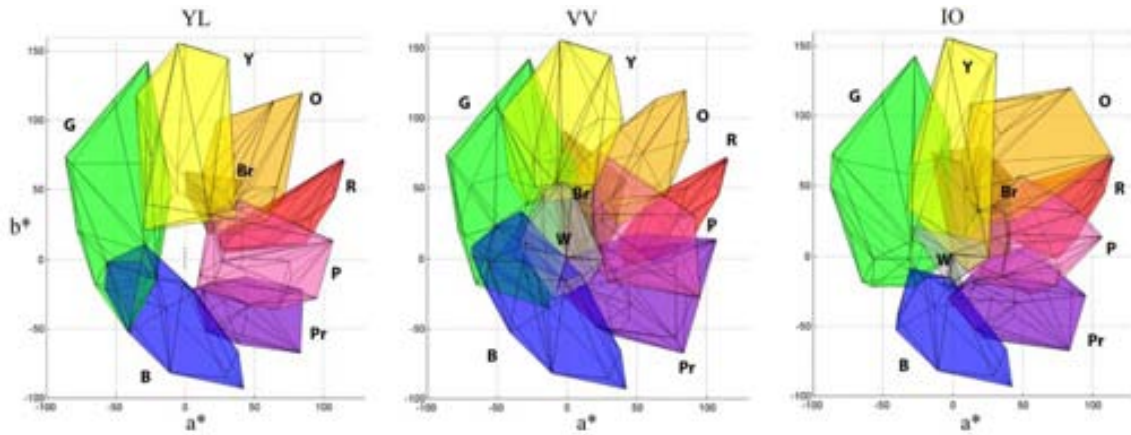
Figure 3.2 Computed colour solids for three normal trichromatic observers(YL, VV and IO). Each colour solid (set of 11 convex hulls corresponding to the basic colour categories), constructed from light and surface naming tasks as described in the text, is shown projected onto the chromaticity plane in 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 Quantifying the individual colour solid

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.

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

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.

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

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).

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

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:

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

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

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.

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

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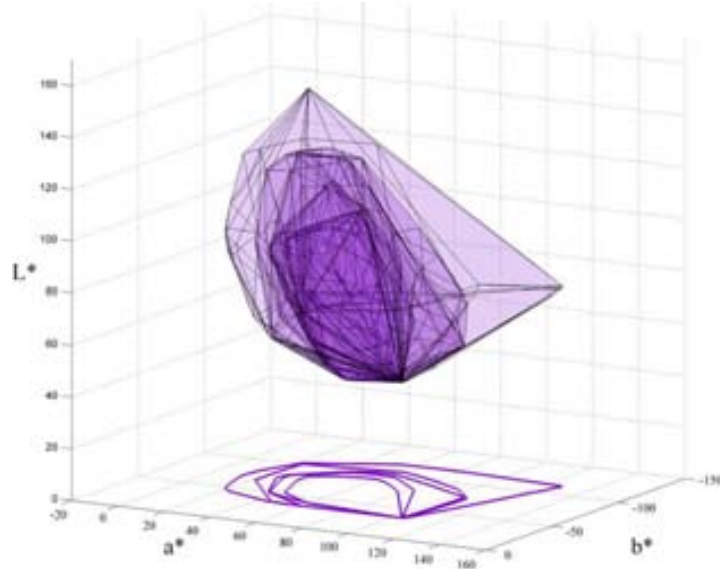
Figure 3.3 Purple layered category from 23 NT in CIELab. It was used in the computation of the Structure Deviation index.

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.

$$F_i^j = \frac{1}{M_i^j} \sum_{l=k_1}^{k_{M_i^j}} p_l \ \ where \ p_l \in C_i^j \ as \ defined \ in \ Equation \ (3.1) \tag{3.12}$$

$$T(i,c) = \frac{1}{N} \sum_{j=k_1}^{k_N} \left\| F_i^j - F_c^i \right\|_2 \quad where \ i,c = 1, \dots, 11 \tag{3.13}$$

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

34

$$CD_i^j = \max_{c=1,\ldots,11}\left\{\left|1 - \frac{\left\|F_i^j - F_c^j\right\|_2}{T(i,c)}\right|\right\} \tag{3.14}$$

	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 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.
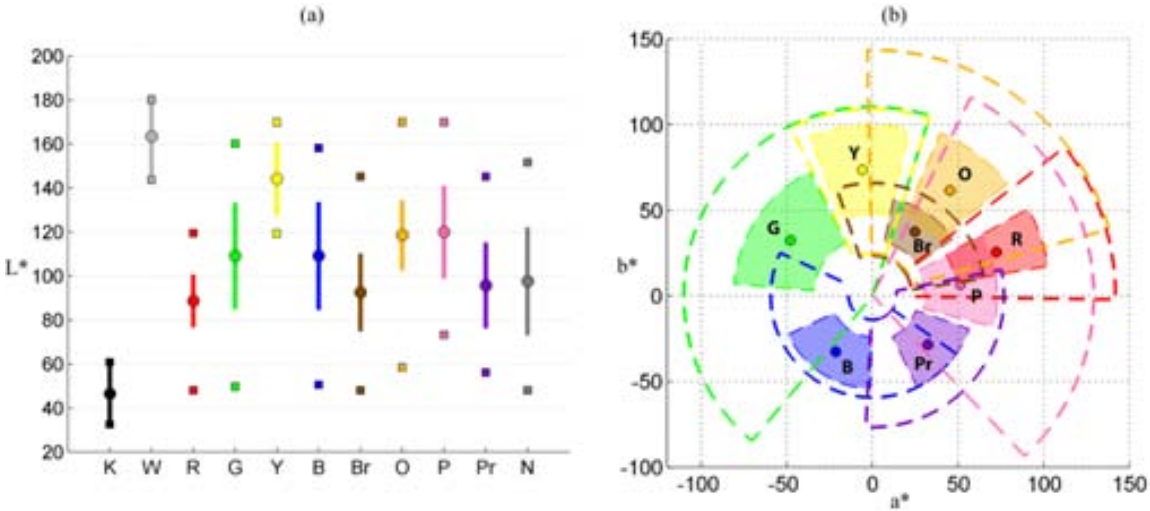
Figure 3.4 CIELab coordinate locations of mean and range of categorized samples in each basic colour category, calculated from all naming data (light and surface sets) across all observers. Panel a: Mean (open circle), standard deviation (coloured bar), and maximum and minimum values of L* (open squares) 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.

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). 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).
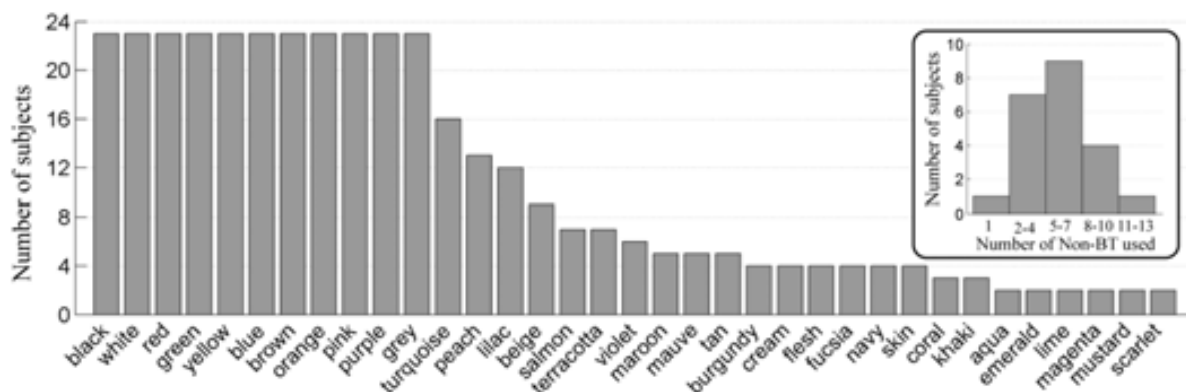
Figure 3.5 *Free naming* test results. The main histogram indicates the number of observers (total of 23) that used each particular term given on the horizontal axis. The inset histogram indicates the number of non-basic terms used (omitting one observer who used 21 different terms).

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

37

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.)
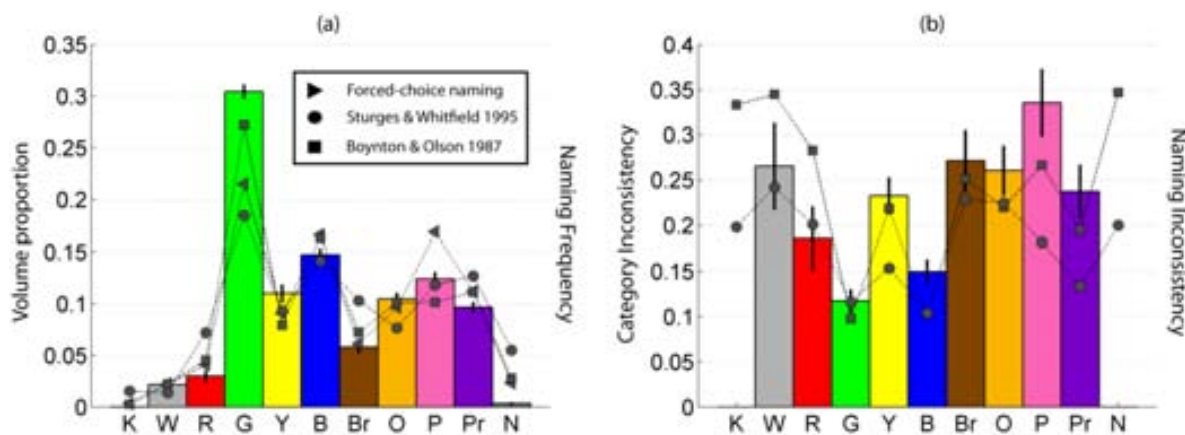
Figure 3.6 Volume and Category Inconsistency indices (coloured bars) compared to previous literature results (triangles correspond to the forced-choice naming test, squares to B&O1987 and circles to S&W1995). Black lines indicate the standard error. Panel a: Bars indicate the Volume index values normalized 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.

Time. The *Time* index was calculated for data from forced choice naming tests. The mean T was 1.85 seconds (σ = 0.45s) 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.

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 (σ=0.13), equating to roughly one-quarter of the category volume. Green and blue have the lowest CI values (mean=0.13; σ= 0.05), significantly lower than brown and pink (mean=0.27; σ= 0.16) ($t_{(22)}$=-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 (σ=0.19), 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 (σ =0.07). 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 (σ=0.09), 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 (σ =0.05). The low values overall (around 10% maximum deviation from normal focal distances) suggests relative stability in centroid locations across observers.
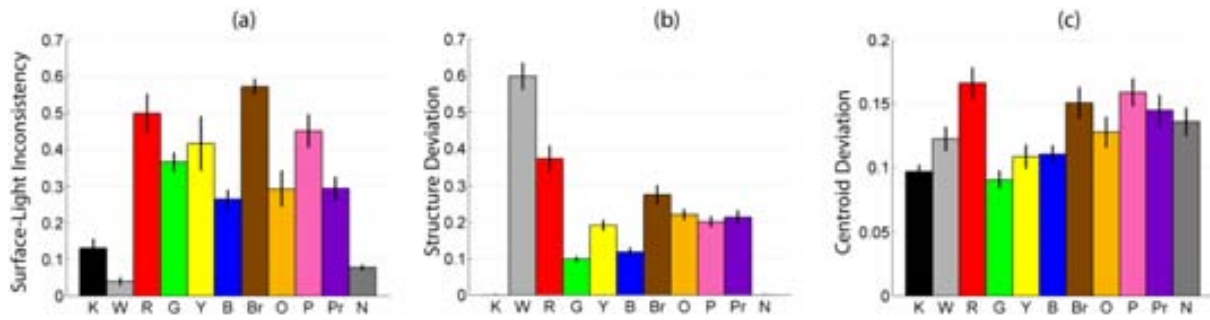
Figure 3.7 The Surface-Light Inconsistency, Structure Deviation and Centroid Deviation index values(panel a, b and c respectively). For each panel coloured bars and black lines indicate observer averaged values and the standard error, respectively. As before, the K and N structure deviation indices are 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).

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	SD	CD	Mean
Black	-	0.31	-	0.93	-	0.27	0.50
White	1.07	0.33	0.87	1.37	0.29	0.37	0.72
Red	0.61	0.24	0.92	0.52	0.47	0.35	0.52
Green	0.07	0.18	0.43	0.34	0.34	0.35	0.28
Yellow	0.22	0.23	0.37	0.86	0.37	0.42	0.41
Blue	0.16	0.19	0.38	0.46	0.49	0.30	0.33
Brown	0.45	0.28	0.40	0.18	0.46	0.39	0.36
Orange	0.22	0.23	0.46	0.78	0.32	0.45	0.41
Pink	0.22	0.25	0.48	0.49	0.37	0.33	0.36
Purple	0.28	0.17	0.50	0.52	0.44	0.40	0.38
Grey	-	0.28	-	0.53	-	0.39	0.40
Mean	0.37	0.24	0.53	0.63	0.39	0.37	0.42

Table 3.3 *Coefficient of variability* according to colour categories and indices. Note that values for black and grey categories on the V, CI, and SD indices have been disabled due to their convex hull existence restrictions.
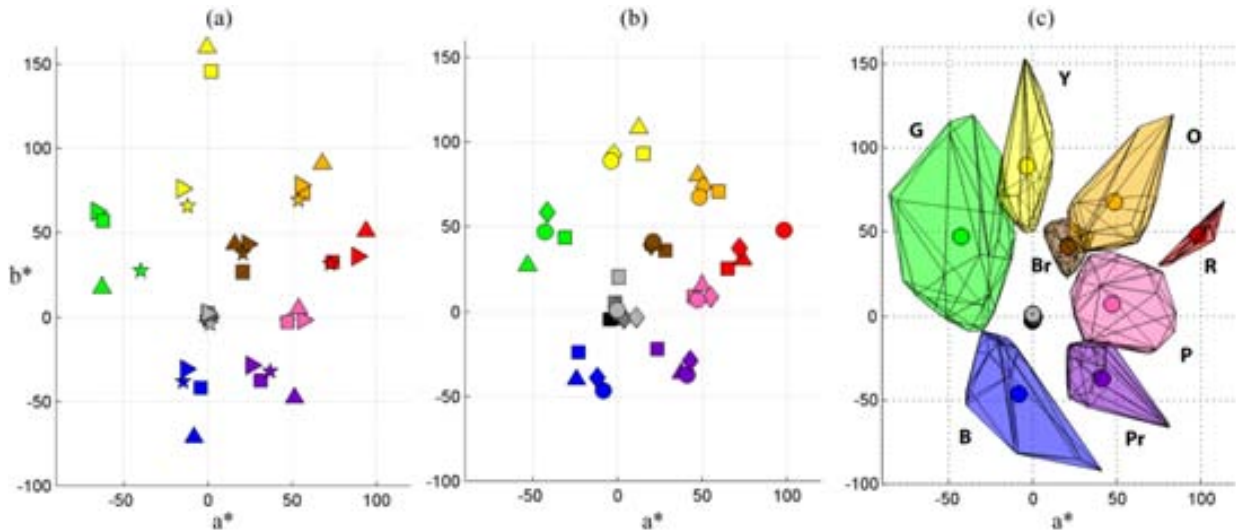
Figure 3.8 Colour category regions location in the a*b* chromaticity plane. Panel a: focal colour locations for forced-choice naming data (pentagon), pick-best data (rotated triangle), B&O (triangles) and S&W (squares). Panel b: centroid locations for all our categorized samples (rhomboids), B&O (triangles), 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).

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.
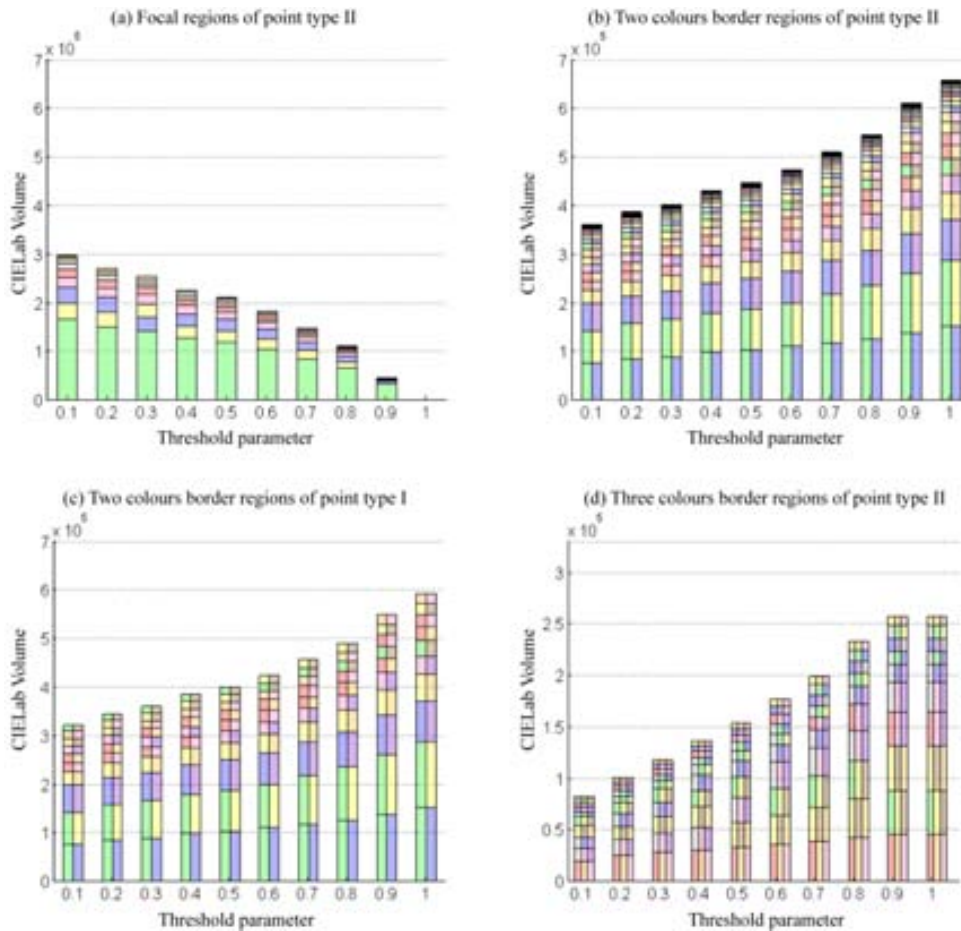
Figure 3.9 Influence of threshold parameter in ACNS regions volume.

Our single variable multidimensional approach allowed easily to get 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

44

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.
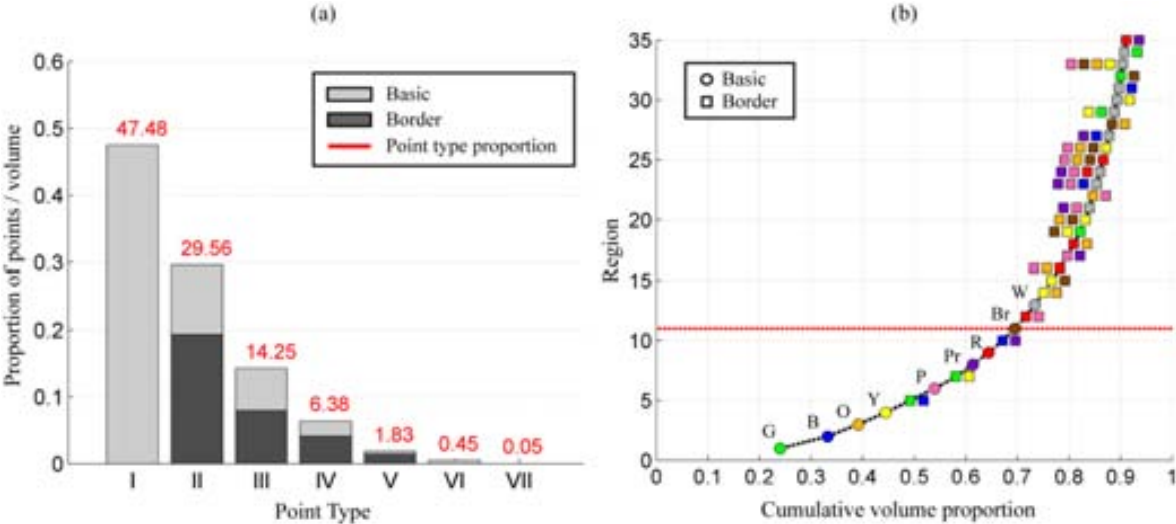
Figure 3.10 Composition of the Average Colour Naming Space. Panel a: Point type histogram. Light and dark grey bars indicate basic and border regions respectively as described in the text. Panel b: Cumulative proportion of total volume of the average colour naming space spanned by basic and border category 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.

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.
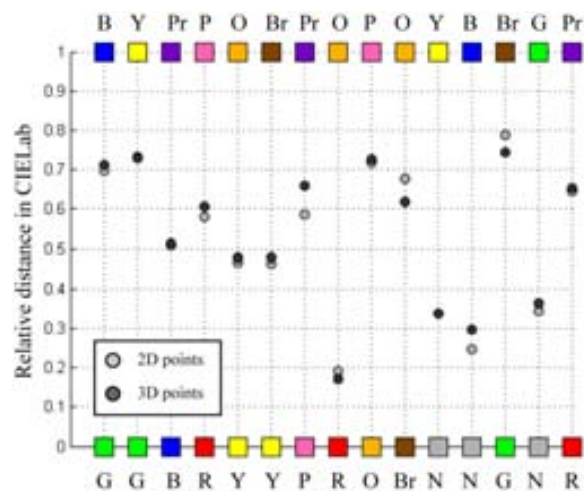
Figure 3.11 Relative distance location between ACNS region centroids of two colours and their constituent basic colour categories. Light grey markers indicate computations using only a* and b*, and dark grey markers indicate using all three dimension of CIELab space. See text for further computation 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.
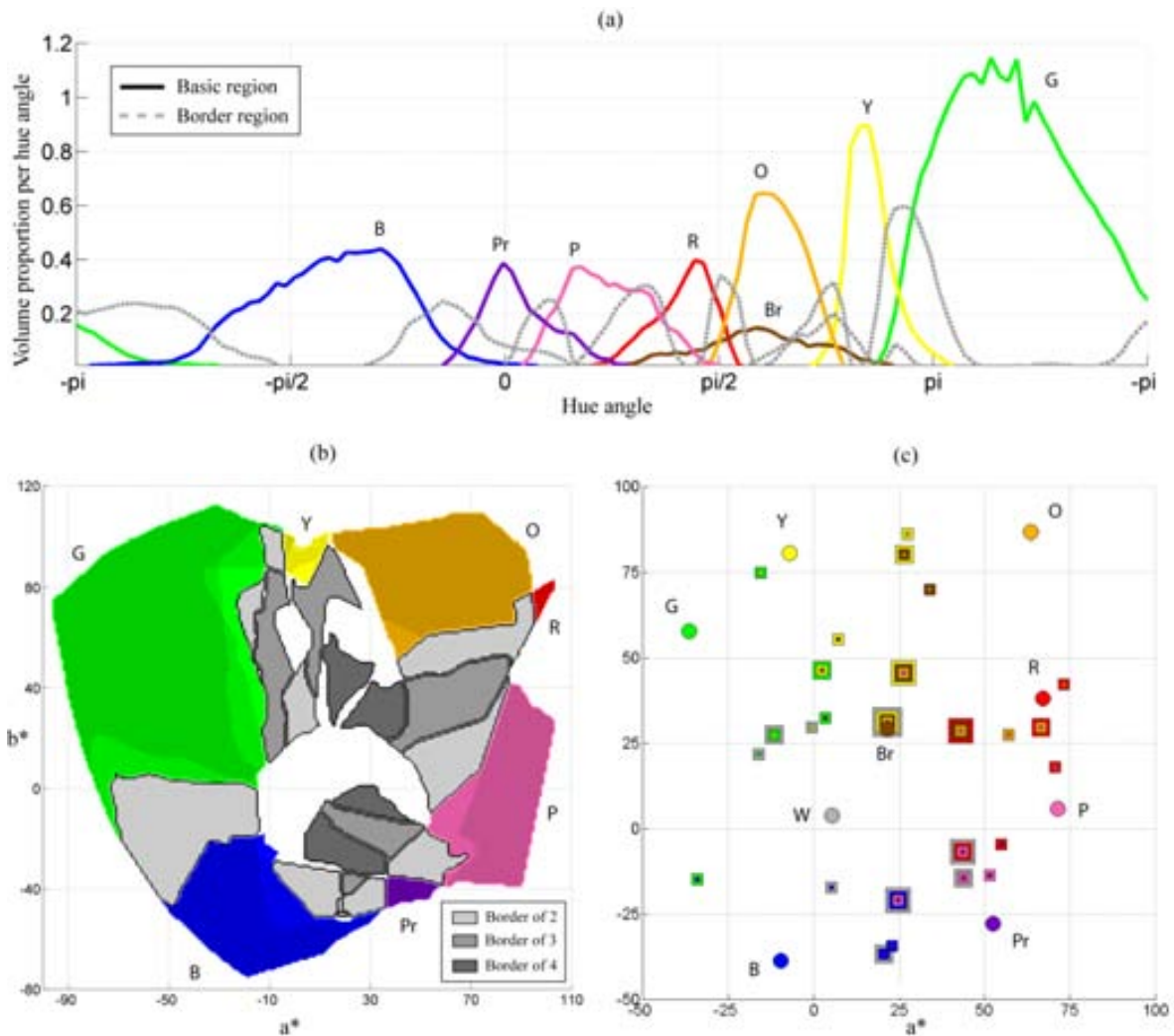
Figure 3.12 Structure of the average colour naming space (ACNS). Panel a: Volume proportion histogram of *basic* (coloured lines) and *borders of two* (grey dotted lines) regions according to their hue. The same analysis was performed for Saturation and Luminance and similar shapes were obtained. Panel 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.

3.4.6 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.
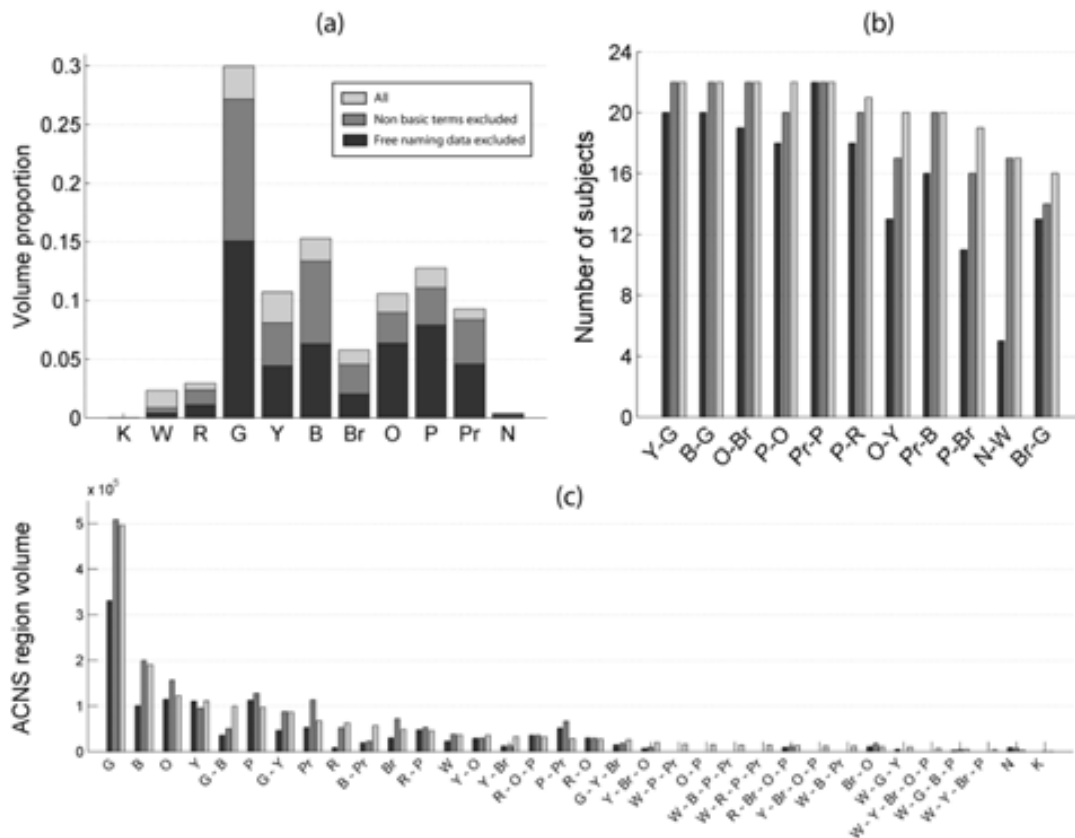
Figure 3.13 Influence of the coding of non basic terms in our main results.

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 1998). 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.

51

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	-.104	-.207	-.191	-.217	-.116
	Volume (V)	-.104	1	.858**	.478*	.399	.465*
Naming Behaviour	Category Inconsistency (CI)	-.207	.858**	1	.436*	.423*	.360
	Surface-Light Inconsistency (SLI)	-.191	.478*	.436*	1	.174	.321
Category Geometry	Centroid Deviation (CD)	-.217	.339	.423*	.174	1	.191
	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.

Chapter 4   The chromatic setting paradigm

Chapter 3 conclusions stated large individual variations in the categorical colour structure of colour space. Also, that colour samples in colour space were mostly represented by the eleven basic colour terms and despite the categorical structure of centroid regions was mostly stable, in between regions had a higher potential to fail in colour communication and thus to not coincide in categorical colour perception. Then, if we are interested into measure categorical properties of colours under illumination changes we should restrict to study the changes among the basic categories and also tune our measuring paradigm and stimuli to each particular observer. The current chapter introduces a newly developed psychophysical paradigm and experiment to measure colour appearance under long periods of immersive adaptation to several illuminations. Finally, as usual in colour constancy studies, the measurements under two different illuminations are compared and conclusions derived. The new paradigm can be readily explained as an extension of the chromatic setting paradigm where instead of adjusting the internal "grey" observers are asked to adjust close versions of the nine chromatic basic colour terms. So the new paradigm is called *chromatic setting* and was applied to a new experiment, *Experiment II*, which tested simple stimuli under three different illuminations. Also, Experiment II included a methodology to validate the new paradigm. Experimental results indicated that the paradigm was feasible. Since our paradigm relies strongly on colour memory ability, we studied its validity by testing the stability of these internal references along the experiment. To test whether our paradigm provides a more comprehensive measure of the colour constancy phenomenon, we have applied linear models to study its behaviour, e.g. to what extent these models are capable of absorbing the growing data complexity that results from the addition of extra measurements. 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.

Chapter 5 and 6 exploded the chromatic setting paradigm to an extended set of stimuli and demonstrate its usefulness when studying categorical changes due to illumination changes.

4.1 Introduction: towards a new paradigm

4.1.1 Constraints on the experimental design

The design of the experiment (laboratory conditions, stimuli selection, number of observers and selection,...) is as important as the selection of the measuring paradigm. Some constraints follow from Chapter 3 observations on the categorical structure of individual colour space.

Because we are only interested into study the influence of categorical colour perception in adaptation we should restrict to simple flat stimuli without any high level information as tridimensional shape, specular highlights, shadows, known objects, etc. So our stimuli will be simple Monrians, i.e., a geometrical composition of squares following a Lambertian shcema which also facilitates a feasible simulation of illumination on it.

Since we are interested into categorical colour perception and it has large individual variations, our stimuli must be tuned to each particular observer. Otherwise, a common stimuli for all observers could not *awake,* if any, effects on the categorical colour perception through adaptation. Then our stimuli will be different for each observer and its colours will be based in categorical perception of each observer.

The colours present in the stimuli also have some constraints: number, categorical quality and location in colour space. The number of scene colours has been proved (ref) to provide qualitative information to colour constancy mechanisms, it ranges from a small number of colours (ref) to large distributions of colours changing gradually in a surface (ref anya). In our case we will test small sets of colours from on to eight. However, the categorical distribution of this colours influences the quality of the information provided to the observer about the illumination, it is not the same using say eleven colours categorized as green than eleven colours where each one belongs to a different basic colour category. the latter case provides much more information about the illumination. Also from Chapter 3 we have seen that the colour space is mostly categorized into the eleven basic colour categories and also that the inner part of the colour space (near the achromatic locus but not so close) is the one that is more interesting in terms of possible categorical changes when the illumination changes, this is due to the increasing proportion of border points compared to inner points as we go to the centre of colour space.

So the colours present in our stimuli will be based on the location of the centroids of the basic colour categories for each observer and will be close to the inner part of the colour space.

Colour contrast effect has been divided into two main effects: simultaneous colour contrast which is unavoidable and global colour contrast which depends on the geometrical statistical properties of the visual scene. Despite that both effects has been linked to colour constancy, in this study we are not interested into study them so we will rule out of our analysis using several techniques embedded in our experimental analysis. Global contrast was avoided by using the traditional geometrical randomizations of Mondrians squares composition for each session trial, and a large number of trials ensured cancelled influences. Also, the Mondrians were chromatically centred to D65, this means that the average chromaticity of its pixels was of D65. A new feature was included in the experimental design which consisted into changing the single test patch by a multiple test patch version where all of its patches changed simultaneously and their location in the Mondrian was randomized each trial. The proportion of the patches in the multiple test patch regarding the overall number of Mondrian patches was kept low. Also the average size of Mondrian patches was kept large enough to avoid contrast effects.

In order to understand possible categorical colour influences we need to measure multiple points in colour space which reflect the categorical structure for each particular observer. In addition to using internal grey as a reference, some researchers have included multiple colour references to study colour constancy (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 (Smithson and Zaidi 2004; Benavente, Parraga et al. 2009). Some of these studies have measured directly the colour appearance of several coloured patches under different illuminants (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, immediate 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). Their conclusions were 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.

4.1.2 Discarding traditional paradigms

As reviewed in Chapter 2 there are several paradigms to measure colour constancy. However when testing a particular hypothesis the selection of the paradigm and the stimuli becomes critical. Here we review the most commonly used paradigms under the light of our needs.

The main handicap in colour naming paradigms is that tested coloured samples are assigned to only one colour term from a small set of terms. In practical terms this means that observers has to perform an initial categorizing step from the "real" colour perception and this difference remains embedded in the measurements. In Chapter 3 we have seen how our observers used a small set of colour terms to perform the free naming task and all of them used mostly only the eleven basic terms. Then the method's precision can be quite low unless using large amounts of testing samples, and then focusing on studying the overall features of the categorical classifications instead of the isolated categorizes samples itself. Take for instance the work of Hansen et al where they modelled the boundaries of colour categories and studied the transformations of this boundaries in an equiluminant plane. Also, the work of Olkkonen et al used overall features, concisely the displacement of the centroid location for each categorical region. Thus, colour naming is a too coarse method to study categorical colour changes around the categorical border locations of basic colour terms.

The main challenge when using asymmetric matching to measure colour constancy is that the method switches the adaptation state from one illumination to the other and thus not enabling immersive states of adaptation. However we could solve this issue by performing different adaptation to each eye but in this case the categorical information will be confounded by the different information provided by each eye. Furthermore, matching between arbitrary test samples do not inform us about the categorical perception of the observer and so any information is provided in the location of the basic regions in colour space.

The achromatic setting paradigm has a high degree of precision and also provides categorical information on their measurements, however only one point in colour space is measured.

4.2 The Chromatic Setting paradigm

With this in mind, we have developed a colour constancy paradigm that tries to minimize the weaknesses while keeping the strongest points of previous paradigms. In our method, the measurements are done under a permanent state of adaptation due to a long period of illumination, thus avoiding potential illuminant-switching issues. Our new paradigm 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 the effect of language (Heider 1972; Kay, Siok et al. 2009) and memory (Hansen, Olkkonen et al. 2006; Ling and Hurlbert 2008; Nemes, McKeefry et al. 2010) on their perceptual categorization. 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 summary, we present a new psychophysical paradigm to determine colour constancy under immersive illumination which measures the perception of nine colours under different states of adaptation by using CRT-based stimuli. The stimuli were combinations of three different two-dimensional Mondrians and three different illuminants. The subject's task was to select and then reproduce a particular colour from memory. In order to rule out memory failings the experimental procedure included a series of repeatability tests.

4.2.1 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*.
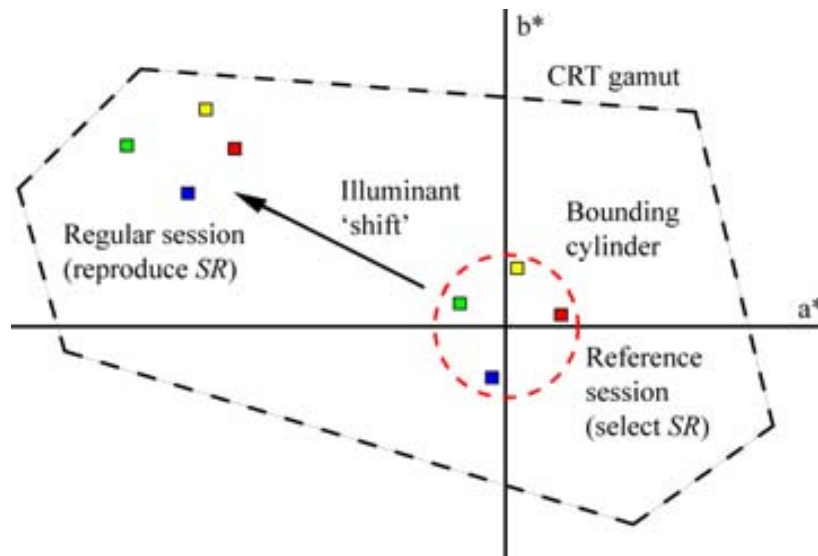
Figure 4.1 Schematics of the chromatic setting paradigm in the a*b* plane of CIELab colour space. The black broken lines represent the boundary of the CRT gamut. The squares inside the red circle represent the colours selected in the reference session. The squares outside the red circle represent the colours reproduced once adapted to the new illuminant in a regular session. The arrow represents the chromatic shift induced by the illumination.

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.

4.2.2 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.

4.3 Methods

4.3.1 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.

4.3.2 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* colourimeter 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.

4.3.3 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.

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.

4.3.4 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 (see Appendix C for detailed values).
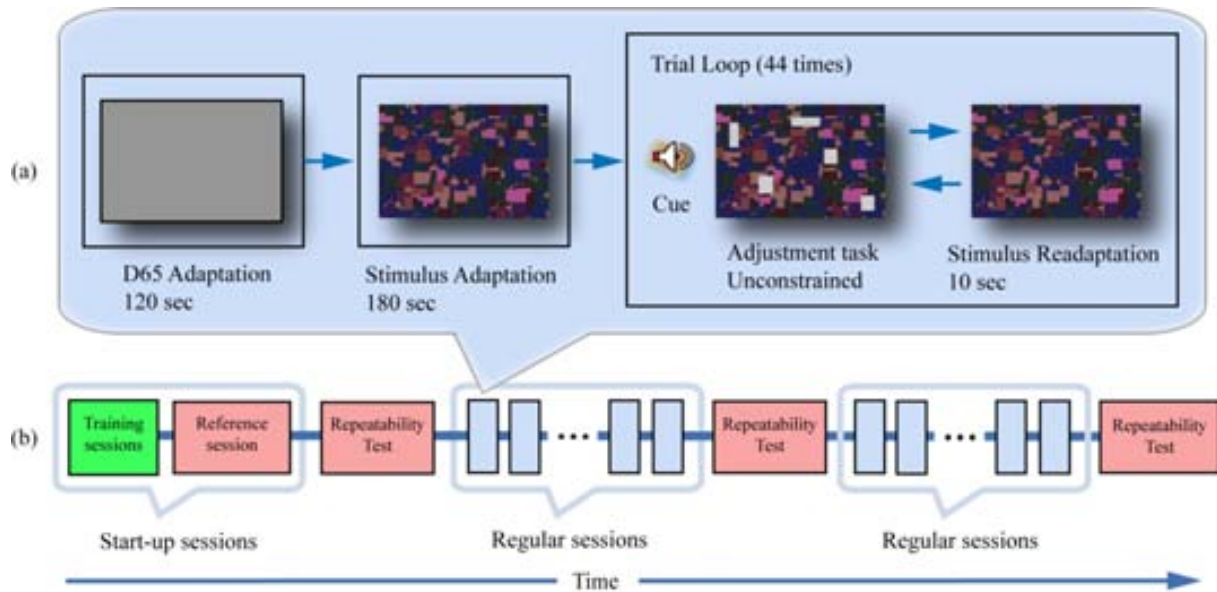
Figure 4.2 Temporal sequence of the chromatic setting paradigm .Panel a shows the common schematics for a reference or regular session. Panel b illustrates the setup of the whole experiment. Start-up sessions consisted in both training and reference sessions. In a reference session, subjects selected 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.

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 StDev) for the Type I backgrounds and 4.1 (0.75 StDev) 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.

4.4 Results

4.4.1 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.
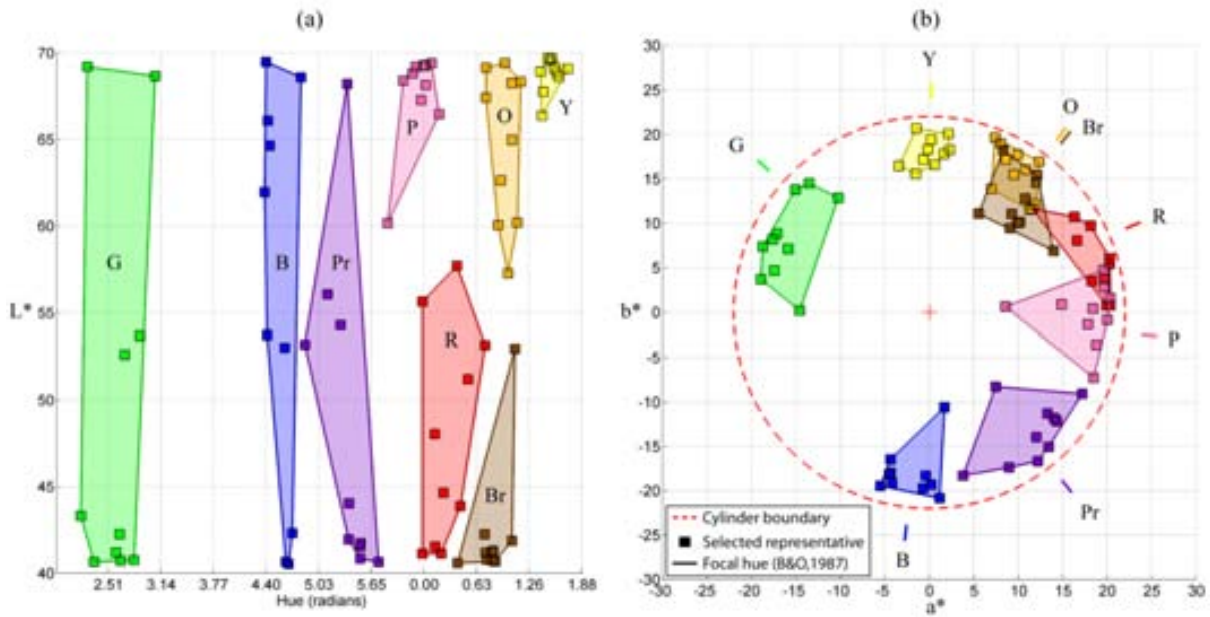
Figure 4.3 CIELab locations of the selected representatives adjusted in the reference sessions by all 10 subjects. Square markers in both panels indicate the average location (5 trials) of each colour category and subject. Colour categories are labeled and colour-coded with their representative colours (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.

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.
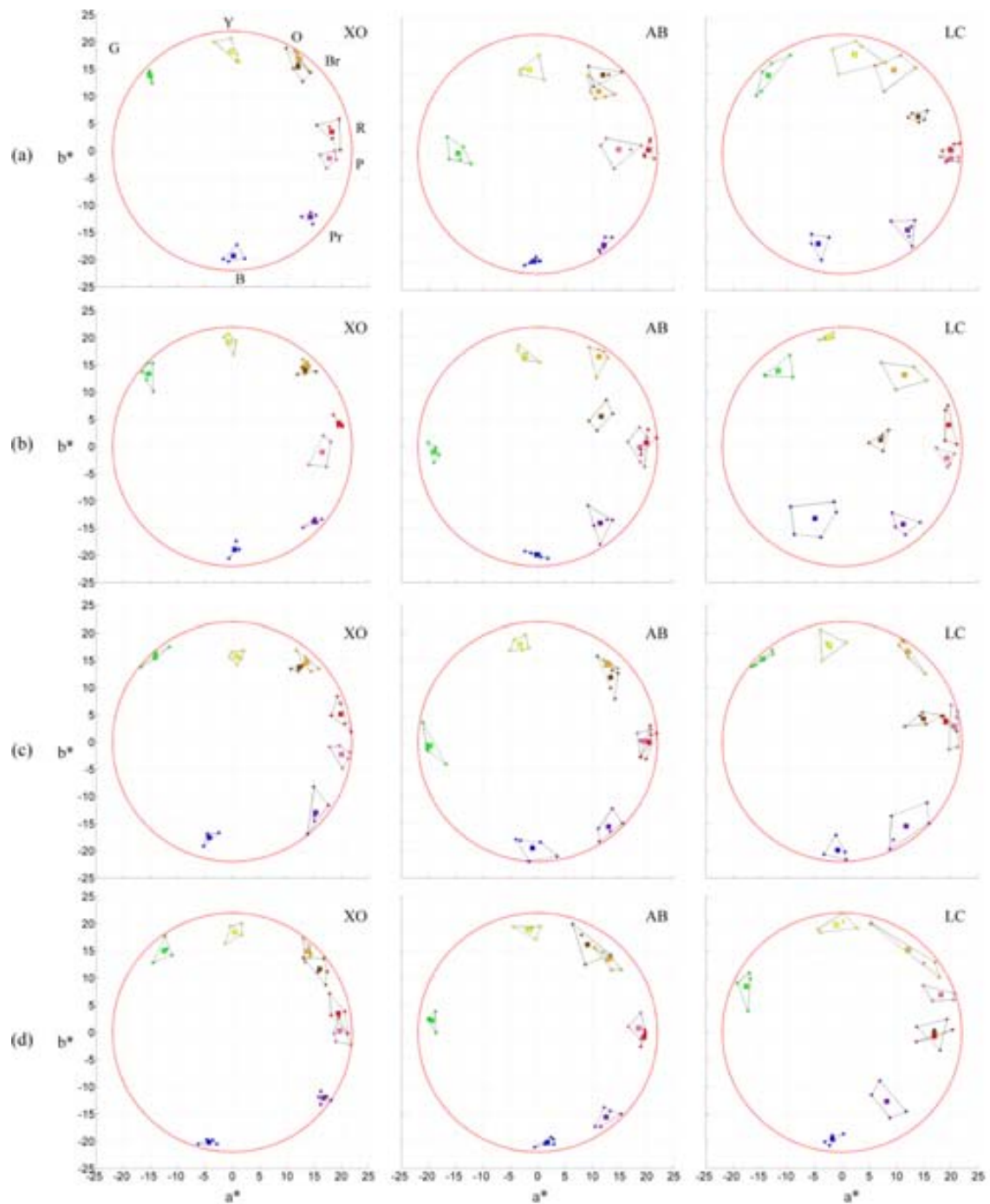
Figure 4.4 Chromatic settings from the reference session and the repeatability sessions. Row a shows the selected representatives chosen by three subjects in the reference session. Rows b, c and d show the corresponding settings for the three subsequent repeatability tests. Square markers represent the average of individual trials (small dots joined by lines) and the large red circle corresponds to the Bounding Cylinder in a*b* chromaticity plane.

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'

69

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.
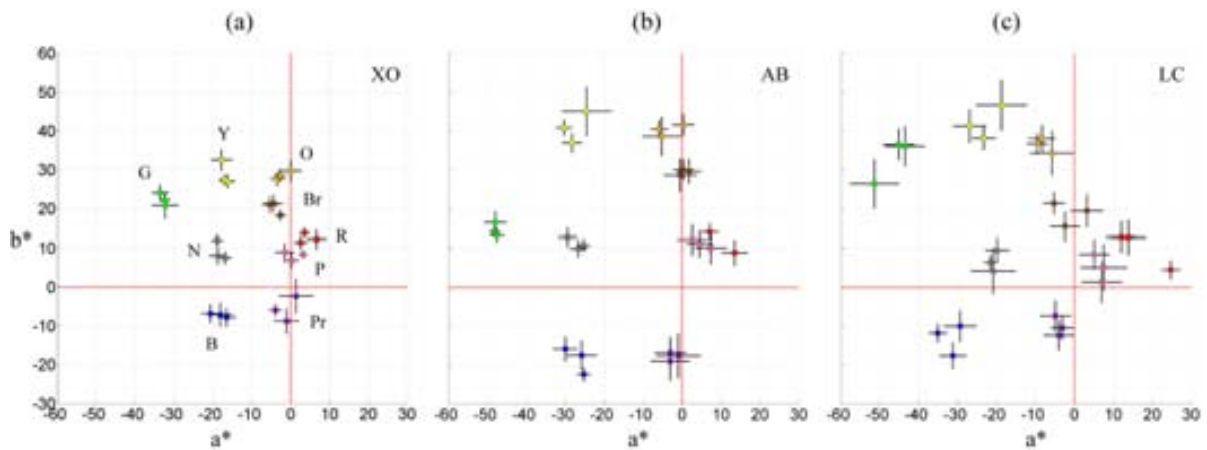
Figure 4.5 Chromatic settings for repeatability sessions. Results include settings for three subjects, Type II background, greenish illuminant and no Bounding Cylinder. Each point represents the average of 5 trials (4 for grey) and it was produced in different days over the experiment lifespan. Error bars show the StDev. 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.

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 D.

4.4.2 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.

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.
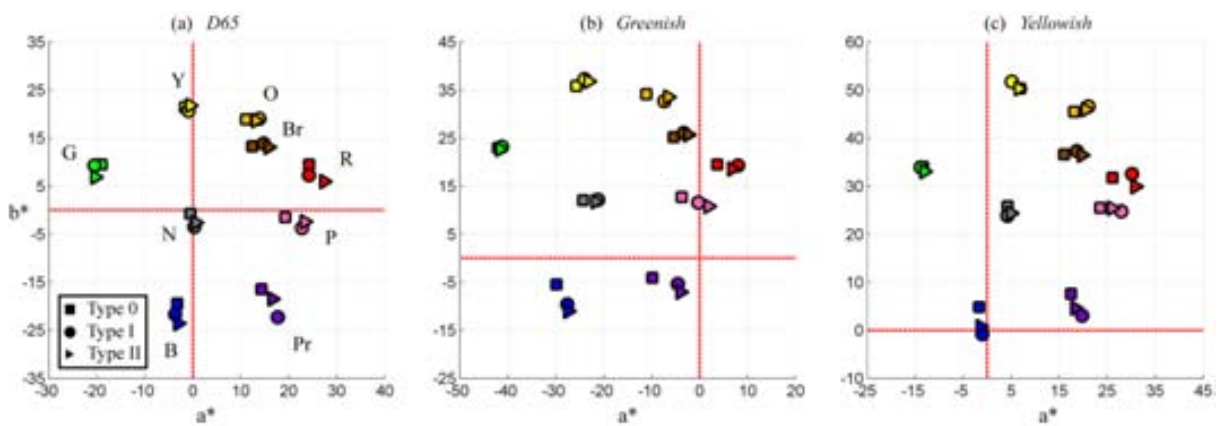
Figure 4.6 Chromatic settings of the selected representatives in regular sessions. The symbols show the chromatic settings for each background type: squares for Type 0, circles for Type I and triangles for Type II. Points were computed by averaging the corresponding SR for all subjects, for each particular 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 StDev) for the reference sessions and 4.60 ΔE^* (2.06 StDev) 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.
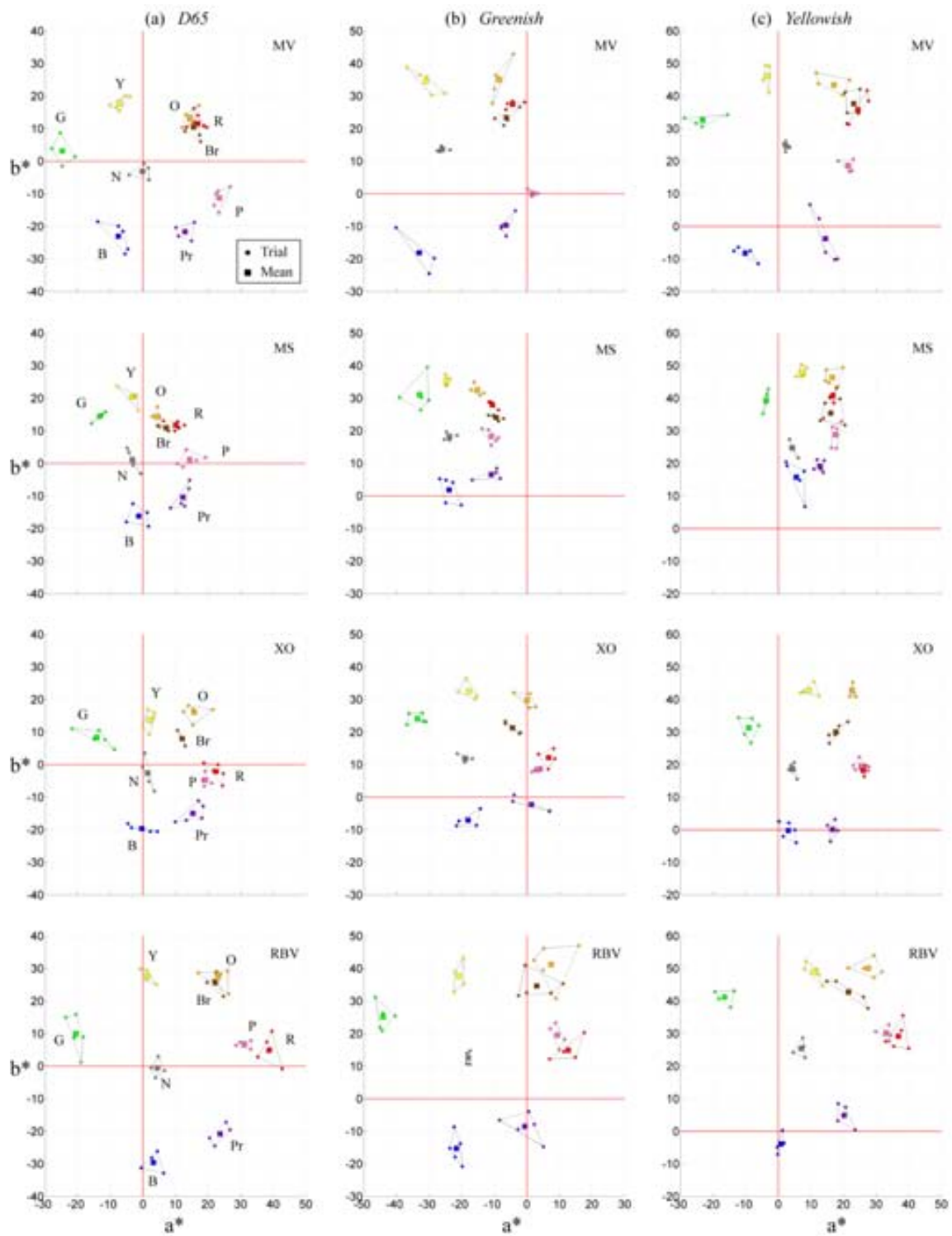
Figure 4.7 Typical chromatic settings from four different subjects for regular sessions. Column a: under D65 illuminant; column b: under greenish illuminant; and column c: under yellowish illuminant. The background was Type II in all cases. Individual trials are represented by small dots joined by lines and their average is represented by a colour-coded square.

73

A balanced one way ANOVA F-test was performed to test for variability differences among colour categories. The results show that mean variability for colour categories is significantly different, $F(8,801)=3.54$ and $p<0.01$. Post hoc analysis using multiple comparison procedures determined which categories differ: red (mean=4.03, 1.25 StDev) and grey (mean=4.05, 1.49 StDev) have significantly lower variability than purple (mean=5.29, 1.15 StDev) and orange (mean=5.02, 1.22 StDev). The same analysis was performed to test for variability differences among background types and there were no significant effects, $F(2,807)=2.98$ and $p>0.05$. The different illuminants significantly affected $F(2,807)=32.52$, $p<0.01$, the variability of our measures: D65 illuminant has the lowest variability (mean=3.83, 1.52 StDev), followed by greenish (mean=4.81, 2.08 StDev), and yellowish (mean=5.16, 2.27 StDev) illuminants.

	R	G	B	Y	N	Pr	P	O	Br	Mean
JRV	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 StDev) sec for the reference sessions and 20.7 (6.2 StDev) sec for the regular sessions. Also, there were no significant time differences in the regular sessions according to illuminants $F(2,807)>0.05$, $p>0.05$ and backgrounds $F(2,807)=0.17$, $p>0.05$, but there were significant differences according to colour categories, $F(8,801)=13.4$ and $p<0.01$. Multiple comparison procedures revealed that grey was the longest to adjust (mean=25.1, 7.6 StDev), followed by brown (mean=22.1, 5.6 StDev) which took significantly longer than blue (mean=18.6, 5.7 StDev), purple (mean=18.7, 4.5 StDev) and pink

(mean=17.9, 4.7 StDev). Red (mean=21.1, 7.1 StDev) and yellow (mean=20.7, 6.2 StDev) took longer time than pink which was the fastest to adjust.

4.4.3 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.

In order to find significant differences in the mean values of the indices, we grouped our data according to colour categories before applying statistical tests. All index values had significant differences when grouped according to colour categories and illuminants: CI greenish with $F_{(8,261)}=3.49$ and $p<0.01$, CI with yellowish $F_{(8,261)}=3.78$ and $p<0.01$, $\overline{BR}$ with greenish $F_{(8,261)}=10.88$ and $p<0.01$, $\overline{BR}$ with yellowish $F_{(8,261)}=12.18$ and $p<0.01$, CCI with greenish $F_{(8,261)}=11.22$ and $p<0.01$, and CCI with yellowish $F_{(8,261)}=4.83$ and $p<0.01$. 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/Ind	CI		$\overline{BR}$		CCI		Mean
Cat/Ill	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.

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.

4.4.4 Linear colour constancy models

As Table 4.3 indicates, the chromatic settings of our subjects were different for different illuminants. We modeled 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).

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

The model is represented by the matrix $\mathbf{M}$ 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 $\mathbf{M}$ 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 $\mathbf{X}$ contains the LMS coordinates of n colours $\mathbf{x}_i$ under reference illuminant and matrix $\mathbf{Y}$ contains the settings of those same colours, $\mathbf{y}_i$, under test illuminant.

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

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, 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**.

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

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

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

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.
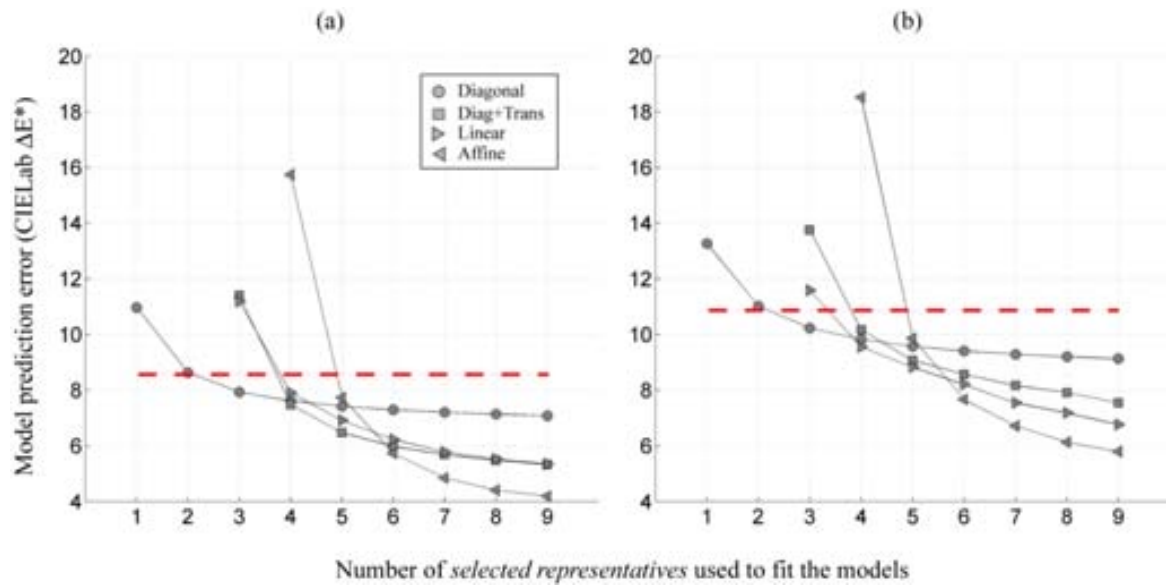
Figure 4.8 Model prediction error according to the number of colours used to estimate their parameters. Panel a corresponds to greenish test illuminant and panel b to yellowish. Each point corresponds to a particular model (circles for the Diagonal, squares for the Diagonal plus Translation, right-pointing 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.
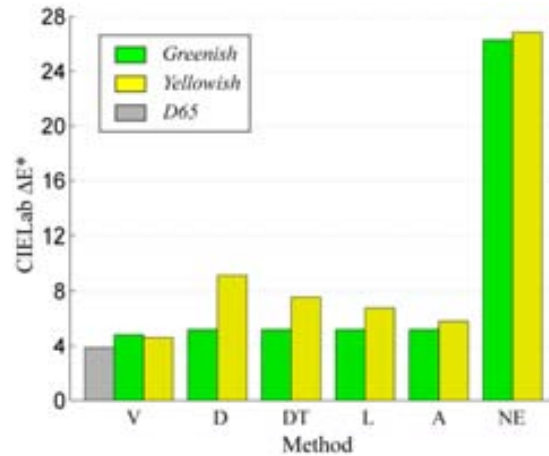
Figure 4.9 Models' prediction error when all nine SRs points were included. The first column (V) contains the subject average variability in the trials (see Table 2) and the last column (No Effect – NE) is a quantitative measure of the illuminant shift computed without any predictive model. The groups of bars labeled as D, DT, L and A correspond each to the Diagonal, Diagonal plus Translation, Linear and Affine models respectively.

4.5 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.

4.5.1 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 E). 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).

4.5.2 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).

4.5.3 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,$\overline{\text{BR}}$	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:

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

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 $\mathbf{A_{percep}}$ and $\mathbf{A_{phys}}$ 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 $\mathbf{A_{percep}}$ 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 $\mathbf{A_{phys}}$ 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 $\mathbf{A_{percep}}$ and $\mathbf{A_{phys}}$ 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 $\mathbf{A_{percep}}$ and $\mathbf{A_{phys}}$. 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 $\mathbf{rs}$ 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. 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.
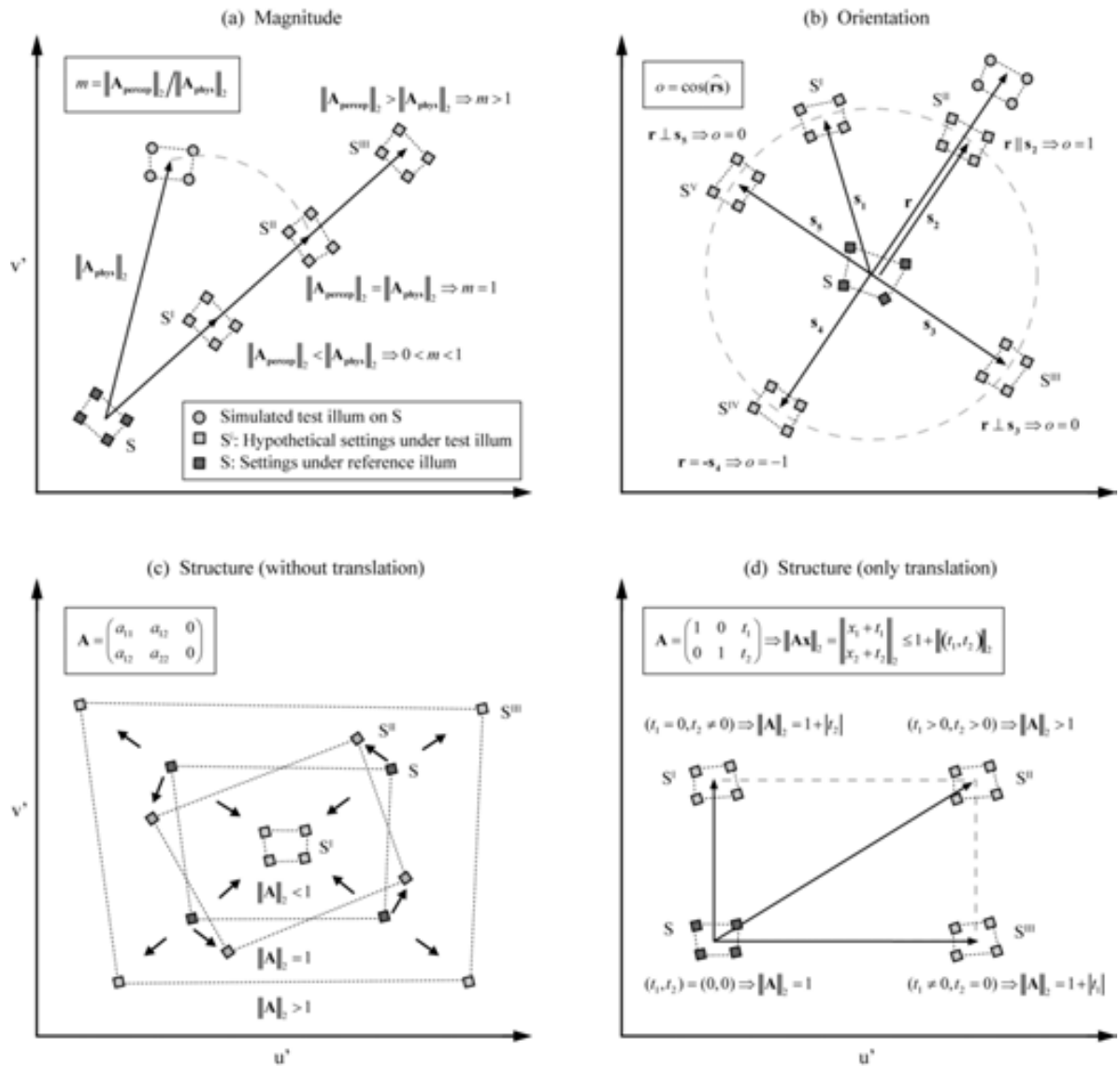
Figure 4.10 Hypothetical cases of chromatic settings and their contribution to SCI values in the CIE1976 uv colour space. Each panel illustrates the contribution of a particular feature of our index. Dark squares correspond to chromatic settings made under the reference illuminant, light squares correspond 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.

Table 4.5 shows the average values obtained from applying four colour constancy indices ($\overline{\text{BR}}$, 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 $\overline{\text{BR}}$ were generally lower and SCI was the highest. The differences between popular indices such as $\overline{\text{BR}}$ 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
$\overline{\text{BR}}$	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.

4.5.4 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 neighboring 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.

4.5.5 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.
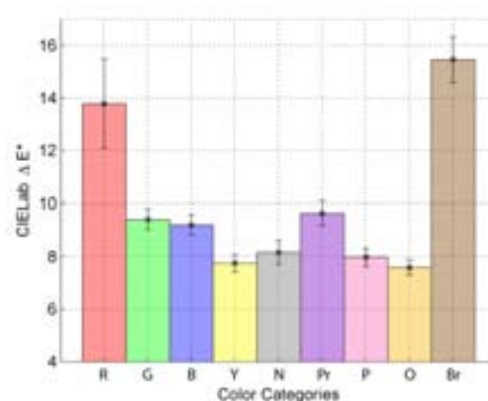
Figure 4.11 The Diagonal model's error for each *selected representative* used to set model's parameters. Results are clustered around two groups, the red-brown with higher error values and the rest.

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.
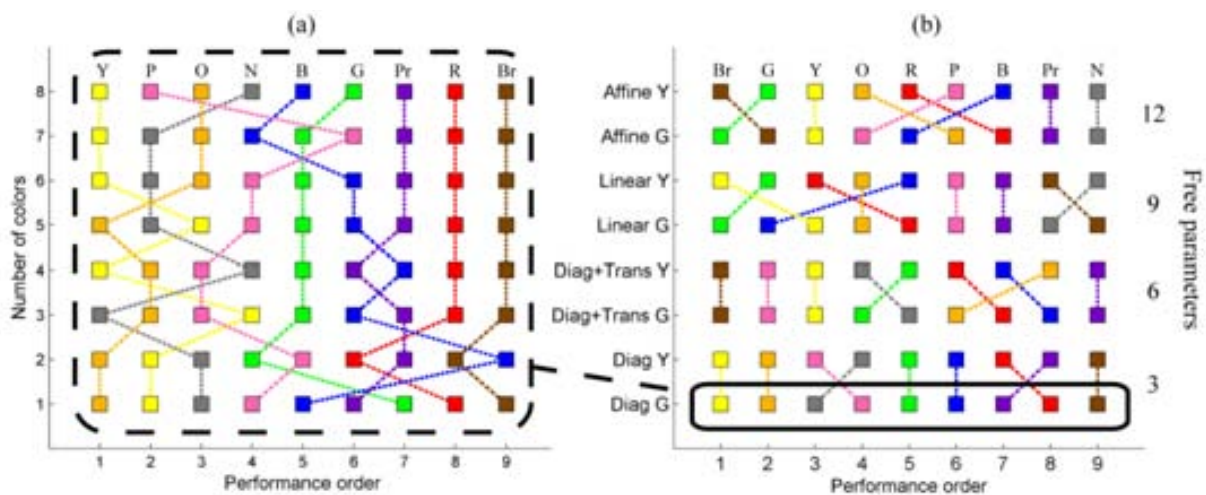
Figure 4.12 Selected representative ranking. The horizontal axis shows a ranking of the colours according to their contribution to the total error (largest contribution to the right). Panel a shows results under *greenish* illumination (all other variables were averaged). Panel b shows the results for 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).

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).

Chapter 5 Inter-relations among chromatic settings when illumination is changed

Chapter 6 introduced a new psychophysical paradigm and demonstrated its feasibility to measure multiple points in colour space under long periods of immersive adaptation. The measured points correspond to relevant categorical samples for each of the chromatic basic colour categories, especially in terms of hue. In this chapter we performed a new psychophysical experiment similar to Experiment II which expanded the number of illuminations tested. As in Experiment II the stimuli colours were specially tuned to each observer due to the large variations found in the categorical perception in Chapter 3. The new experiment used the chromatic setting paradigm and the results were studied in terms of the interrelations among the nine chromatic settings when illuminations were changed, particularly if they remain stable. The chromatic setting paradigm allows us to make use of more sophisticated techniques for describing successive colour constancy. From our measures we defined a perceptual structure descriptor that is up to 87% stable over all conditions, suggesting that categorical colour features could be used to predict the overall behaviour of the colour constancy phenomenon. Interrelations among chromatic settings were modelled as a *graph*, which allowed to measure with precision their changes when illumination was changed and to state conclusions about the stability of the categorical colour perception. Our conclusions are in agreement with previous studies on the stability of the categorical colour space (Hansen, Walter et al. 2007; Olkkonen, Hansen et al. 2009) and also formalized the observations done in Chapter 4. Despite the structural colour constancy index had high values there was still a remaining 13% of inconstancy which may allow remarkable categorical changes when illumination is changes specially for those samples near the

borders of basic colour categories. This fact will be explored in detail in Chapter 6 where categorical changes between basic colour terms has been studied systematically for different adaptation states.

5.1 Introduction: Structural colour constancy

When studying colour constancy, few studies have measured more than one point 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 the question of 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 (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 tridimensional Munsell samples), which constraints 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 52.000 different (volume of the Bounding Cylinder expressed in JND spaced units) CIELab samples which look continuous to the subjects. This paradigm allows us to investigate changes in the perceptual categorical "structure" of colours after adaptation to illumination changes by simultaneously measuring several "key" points in colour space. We study the structure of colour space through its categorical distribution in basic colour categories as described in Chapter 3 and the chromatic setting paradigm allows to set correspondences between nine points in colour space under two different illuminations.

5.2 Methods

Here we introduce a new psychophysical experiment (*Experiment III*) were observers performed chromatic settings under backgrounds Type I and II, and six different illuminations. In more detail: observers were presented on the screen with the written name of a focal colour and asked to match it

to their own internal representations by manipulating the colour of a patch 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 and their colour vision was normal (or corrected to normal) as tested by the Ishihara coloured plates and the Farnsworth-Munsell D15 Hue Test. Of these, two were volunteers and naïve to the experiment's purpose, and the other two were JRV and CAP.

5.2.2 Experimental setup and procedure

Because the chromatic setting paradigm was applied to measure the state of adaptation of our observers 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 the same as used in Experiment II which are described in subsection 4.3.3. However now we used only Type I and II backgrounds and three different illuminations were added. Follows a brief reminder:

The spatial structure of our stimuli was a "Mondrian" pattern and consisted 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. We defined three types of Mondrian stimulus:

Type I was constructed from the colours selected to be the best representative available of each category.

Type II was constructed by selecting hues in-between *selected representatives* but similarly saturated.

Type I and *Type II* were different for each subject and did not contain grey patches to avoid cueing the observer on the illuminant (Foster 2011). All Mondrians were in turn "illuminated" by performing the spectral product of each patch's reflectance times one of five simulated illuminations assuming a Lambertian reflectance model (Walsh and Kulikowski 1998). The illuminants were chosen so that the

final product (the illuminated Mondrian chromaticities) 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 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.

Illum	x	y
D65	0.312	0.329
Purple	0.316	0.228
Green	0.296	0.453
Yellow	0.453	0.434
Orange	0.437	0.343

Table 5.1 CIExy chromaticity for the illuminants used in Experiment II.

5.3 Data analysis: graph definition from chromatic settings

The chromatic setting paradigm provided us with a set of nine measured points for each adaptation state. Notice how this measures were not random in colour space but had a relevant categorically location, especially in terms of hue. To formally describe the overall interrelations among sets of measures we modelled the CIELab coordinates from each set of measured data. Each set of nine chromatic settings is modelled as a *graph* (Equations 5.1 to 5.3), where each node is represented by the CIELab coordinates of one chromatic setting and edges were defined for all possible node combinations. Figure 5.1 illustrates schematically the approach.

$$\textbf{\textit{node}}: n_i = Chromatic\ setting\ i \quad where\ i=1,...,9 \tag{5.1}$$

$$\textbf{\textit{edge}}: e(i,j) = \overline{e_i e_j} \tag{5.2}$$

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

And when all edges were considered, a graph distance matrix was defined. In order to extract inter-subject comparisons we normalized this distance matrix by the mean distance from all nodes to the

node corresponding to grey. This allowed us to produce a distance matrix (Equation 5.4 and 5.5) which enclosed the distance proportions to each node in terms of the distance to the central node.

$$\bar{w} = \frac{1}{8} \sum_{\substack{i=1 \\ i \neq j \text{ and } j = \prime grey \prime}}^{N} w_{i,j} \tag{5.4}$$

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

In summary, each session produced nine SRs which were modelled as a graph, whose nodes were the corresponding measured CIELab coordinates and its distance matrix was build from the Euclidean distances among these CIELab coordinates. In this way the perceptual categorical structure of our subjects was captured by a graph allowing us to get a reliable comparison among adaptation conditions, i.e., we defined the distance between two graphs (see Equation 5.6) by the mean absolute difference between the corresponding graph distance matrices:

$$dist\big(G_1(N_1, E_1), G_2(N_2, E_2)\big) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} b_{i,j} \tag{5.6}$$
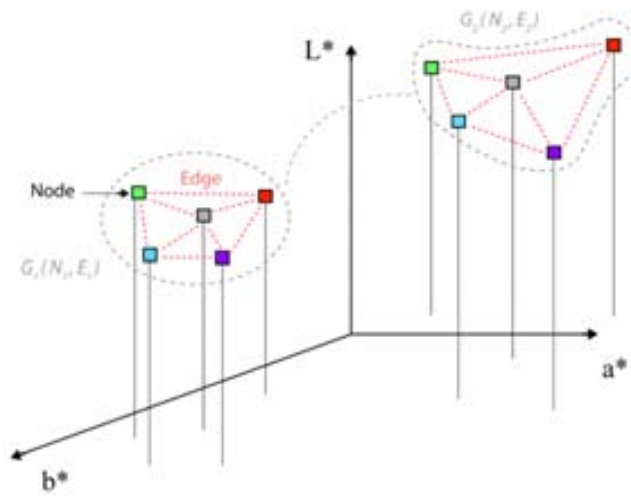
Figure 5.1 Shcema of the graph definition for two sets of chromatic settings.

5.4 Results

Data presented in this section comes from Experiment II and III. The first one was done six months before and used 10 subjects, three illuminations (*D65*, *yellowish* and *greenish*) and three backgrounds (*Type* 0, I and II). The second one used 4 subjects, two backgrounds (*Type* I and II) and two different illuminations added (*purplish* and *orange*). Our results reveal the invariance of perceived colour interrelations under different illuminations.

5.4.1 Chromatic settings under six different illuminations

Subsection 4.4.1 demonstrated that subjects are able to remember the *selected representatives* over the experimental period of several weeks. Figure 5.2 shows all gathered data alongside the CRT gamut (all subjects, illuminants and backgrounds: 999 colour measurements obtained from adjusting 4880 settings over 111 regular sessions). Each point in Figure 5.2 is the average of 5 different trials. The experimental error was computed as the standard deviation from the mean for each colour category, averaged over subjects and sessions. The overall standard deviation in Experiment III was $4\Delta E^*$ units. In order to ensure that subjects did not use the CRT gamut as a reference when doing the adjustments requested, we computed for each of the 999 colour samples the distance to the CRT gamut boundary and averaged this information at the top of each histogram bar in Figure 5.2. Our results show that only 1.4% of the 999 points were closer than $5\Delta E^*$ units from the CRT gamut boundary (only 16.3% were between 5 and $10\Delta E^*$ units).
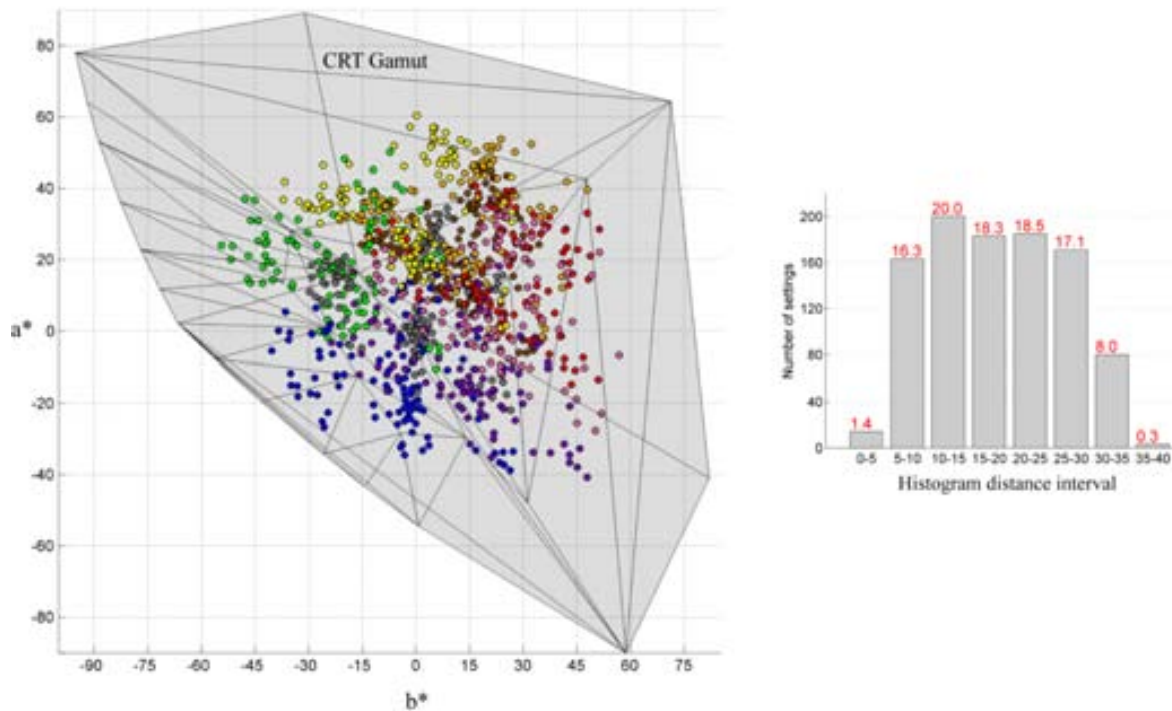
Figure 5.2 Chromatic settings and the CRT gamut in Experiment II and III.Left: CIELab chromaticity plane where the CRT convex hull gamut is projected and 999 coloured dots where each one corresponds to one different colour measurement under different conditions. Each dot was computed as the average of 5 settings and its colour indicates the perceived colour by the subject when adjusted. Right: Histogram where the number of measured colours is divided according to their distance to the convex hull CRT gamut boundary.

Figure 5.3 illustrates the *selected representatives* chosen by three different subjects on the *reference session*. Each coloured circle shows the SR for the corresponding category and the joining lines help to illustrate their geometrical interrelations (Euclidean distances). Notice how this interrelations are different for each subject, for instance, subject JRV SR blue has high luminance while the other two subjects selected colours with low luminance values. Also, when comparing subject CAP and XO, notice how the selection of orange and red differs also in luminance level: subject CAP selected higher luminance than XO. Finally, notice how the red and pink SRs of CAP were different in hue and luminance 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, but at the same time the pattern conformed by those remained approximately invariant. In the following sections we studied whether this pattern was also invariant under different illuminant adaptations (over different stimuli, background and illuminant conditions) occurred.
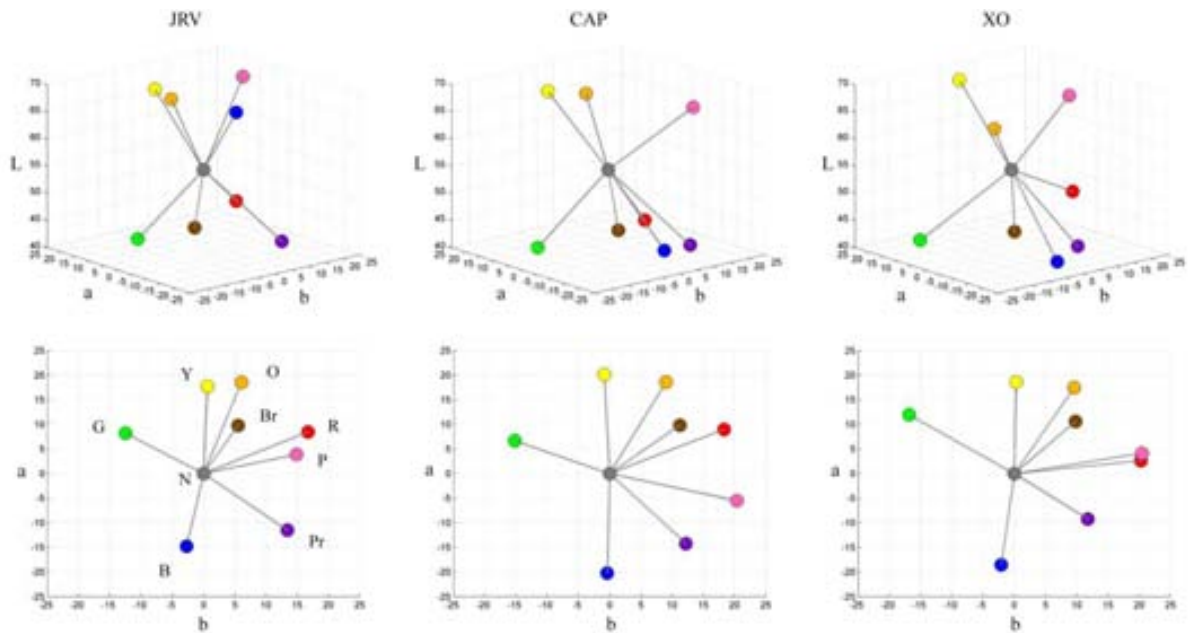
Figure 5.3 Selected representatives for three subjects.

Each column corresponds to one subject and it contains two views of the selected representatives in the reference session. Results are plotted in CIELab colour space: the first row contains the isometric view and the second row the projection in 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).

5.4.2 Interrelations among chromatic settings quantified

The psychophysical paradigm provided us with a set of 9 measures for each adaptation state. The first row of Figure 5.4 illustrates these measures for two subjects over the five illuminants tested. When considering a particular subject, notice how the same groups of measures (linked by coloured lines) seem to keep their structures stable over illuminants changes. To formally describe the overall interrelations among sets of measures we modeled the CIELab coordinates for each set of chromatic settings obtained from each observer and adaptation state. The top plots of Figure 5.4 show the graphs corresponding to the colours *perceived* by two subjects after a change of illumination. The bottom plots represent the *calculated* CIELab coordinates of the same stimuli after illumination (the *physical* colours). Looking at Figure 5.4 and comparing the top and bottom plots we can identify two trends: first, the perceptual representations of 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. Second, the region spanned by the perceptual measures is more compact than the region spanned by the physical ones. The latter observation is wholly

captured by standard colour indexes (which measure distances between perceptual and physical grey) (Foster 2011) while the former observation can only be captured by the graph representation approach.
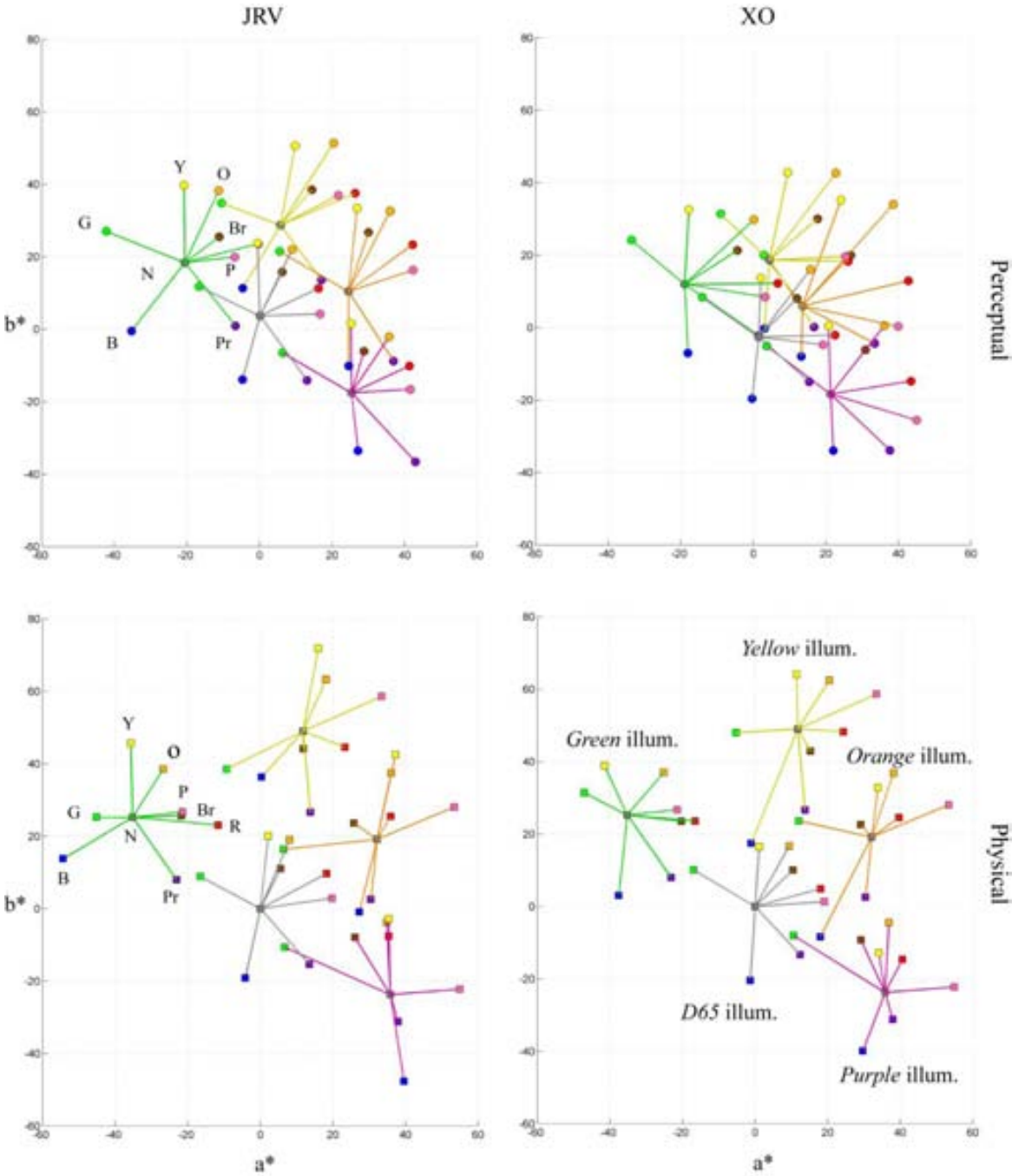
Figure 5.4 Selected representatives obtained for two subjects and five illuminants.

Each plot shows the measures obtained for 5 illuminants (each line colour corresponds to one illuminant). The top row shows the colours (SR) selected by two observers and the bottom row shows the corresponding projections of the original colours after illumination. Left plots correspond to subject JRV with type I background and right plots to subject XO with the type II background. All measures are shown as projections on the CIELab ab-chromaticity plane.

Figure 5.5 contains four sets of bars, each corresponding to one comparison between conditions under illuminant D65 and other illuminants indicated by the x axis label. They were produced averaging the results for all subjects. For instance, looking back at the top left plot of Figure 5.4, each of its five sets of measured data were modeled as graphs, and graph distances were computed from the graph centered at the achromatic locus (D65 illumination). When grouping over subjects and averaging graph distances we obtain the bar height and StDev shown in Figure 5.5. Notice how the values corresponding to the physical graph distances are significantly higher than the perceptual ones: approximately 23% for the physical distances versus 13% for the perceptual distances in average.
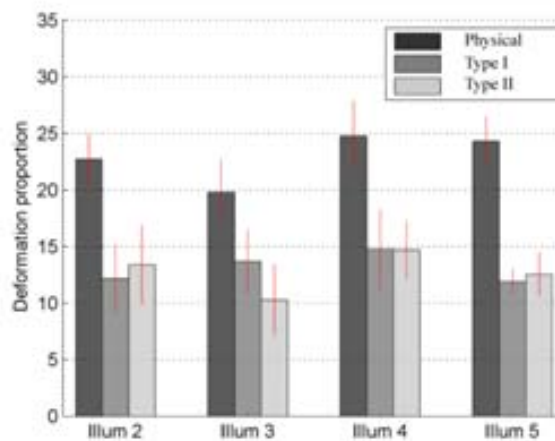
Figure 5.5 Structural deformation of chromatic settings.

Each group bar indicate the average distance, with D65 and the indicated illuminant, of the graphs used to assess the interrelations of measured data under each condition. Data is averaged over subjects and compared according to backgrounds and perceptual/physical.

5.5 Discussion

5.5.1 Issues related to the structural constancy of chromatic settings

Here we discuss briefly if the previous results on the structural constancy of chromatic settings could have been influenced by (i) the chromatic adaptation included in the CIELab formulae and (ii) the chromatic setting method error.

We know that perceptual distances are wrapped in CIELab colour space as we go away from the origin. So we could expect that if we plot all results (obtained under six different illuminations) using only D65 as a reference white point, then their CILab coordinates will have some degree of distortion since they will be in locations far away from the CIELab origin. However, our analysis is not using absolute distances but relative distances, all distances are normalized in the graph approach. So it is reasonable to expect that distances will be deformed equally at least for some group of points close enough which we thing that it is our case. This uniformity in the deformation will cancel the possible wrapping effects in the relative distances represented by the edge weights.
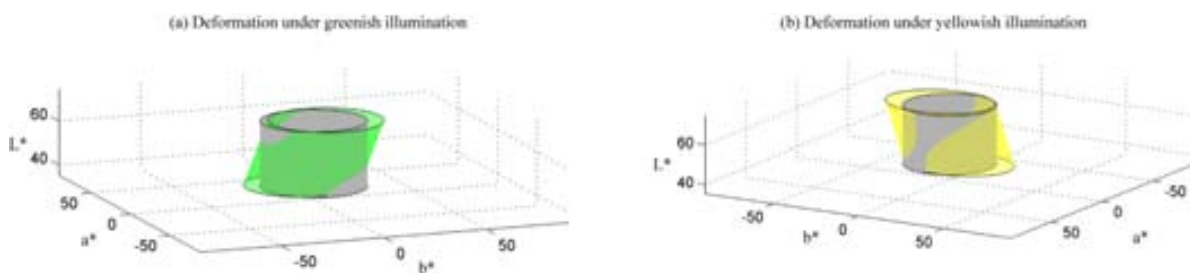
Figure 5.6 Deformation in CIELab colour space according to the greenish and yellowish illuminations.

The chromatic setting experimental error, which was mostly the product of the subjects' variability at producing SRs, was estimated to be about $4\Delta E^*$ in Experiment III. We propagated this error to our structural deformation index and obtained an averaged value of about 5.2%. The low magnitude of that value allows to discard this effect from our conclusions on the stability of chromatic settings interrelations under illuminations changes.

These results indicate that subjects tend to maintain the same perceptual interrelations between chromatic settings and thus categorical colours despite variations in the structure of physical colours under an illuminant change.

5.5.2 Comparison to previous work

Previous work has also focused on the colour appearance of multiple points under illumination changes. Some 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 colour appearance of multiple points indirectly by means of the deformation of categorical boundaries in colour space, concluding that the categorical structure of colour space remains roughly stable, which is in agreement with our results.

In order to further compare our results to others in the literature, we have considered pioneering work by McCann *et al* (Mccann, Mckee et al. 1976). They reported the Munsell coordinates of 17 matches between 5 different illuminants. Figure 6 shows CIELab colour space plots of MacCann *et al* results. Each colour indicates the illuminant under which the matching was done. Notice how each grey dot is linked by means of a black line to its corresponding illumination matching. 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%StDev) which is similar to our results and suggests a high degree of colour constancy (but not perfect colour constancy).
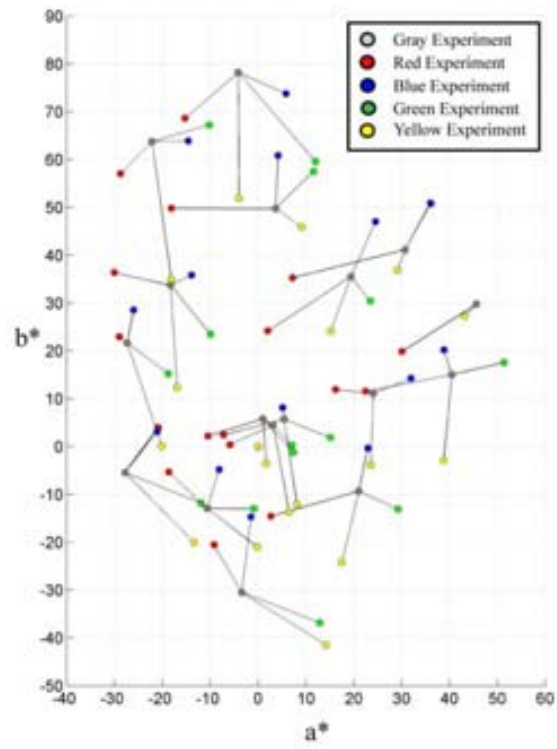
Figure 5.7 CIELab a*b* plane projection of MacCann et al reported results.

Each coloured circle corresponds to the CIELab coordinates of surface matches under a five different illuminants, which are linked by black lines. D65 was used as a reference white point.

Chapter 6 Categorical colour perception in colour constancy

This chapter studies the main topic of this thesis, the possible influence of the categorical colour perception to the colour constancy or the adaptation processes. A new experiment is introduced which uses similar stimuli but with a reduced number of colours to propitiate further deviations of expected adaptation. In order to measure the categorical colour perception changes under the six illuminations tested we used the chromatic setting paradigm which allowed to define a model of categorical colour perception and also a constrained naming task. Results are reported and interpreted in terms of categorical colour changes when the illumination changes. Two main hypothesis are considered in order to explain the possible changes of categorical colour: (i) influence of initial colours and (ii) influence of expected colours.

6.1 Introduction

Colour constancy is a perceptual phenomenon that seeks to keep stable the perception of objects colour under changing viewing conditions. Example of what we are trying to measure. Think of a case in successive colour constancy. If we take several points nearly grey or non saturated colours after illuminating them for a saturated illuminant we will perceive them at the first seconds much more uniform than if we allow the adaptation go on several minutes when some new colours will appear.

We developed an experimental procedure that is based in the categorical perception of each particular subject, thus tuning its particular stimuli to his/her particular colour categories. We have takes this

approach after realizing the large variations in colour categorization across individuals, and especially when dealing with non focal regions (border regions).

The main functionality to surface colour recognition is object identification and according to Jameson and Hurvich: "*such recognition is adequately accomplished by category matching and does not require precise matching-to-sample by the three colour variables of hue, brightness and saturation*"(Jameson and Hurvich 1989). Also Boynton and Olson: "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" (Boynton and Olson 1987) . Then when studying colour constancy we should focus on the changes of categorical colour perception of observers, something that is naturally measured by colour naming tasks but it is not obtained straightforwardly from others such achromatic setting. Also most of the colour constancy indices remain blind on categorical colour constancy and are only based on the samples coordinates in some colour space.

The Helson-Judd effect: "*Spectrally nonselective surfaces seen against a spectrally nonselective background all appear achromatic (white through grays to black) in white light. When illuminated by chromatic light, samples whose reflectances are near the background level continue to appear gray, those above the background level take on the hue of the illuminant, and those below take on a hue or with its complementary,...*" (Jameson and Hurvich 1989). It is a first demonstration that categorical colour perception may not be strictly linked to colour constancy.

The colour constancy phenomenon is composed by several mechanisms all gathered under this name due to its bonds to achieve its functional aim. One of the mechanisms that contributes most to the phenomenon is global contrast, i.e., adaptation to the chromatic mean of the visual scene. This process shift the achromatic locus of the observer in the illumination chromaticity direction, being a first consequence the maximization of the number of colour perceived. General movement under adaptation is towards grey, then growing the number of colours in the perceived scene. But are the number of perceived colour categories maximized as well? Is this perception influenced by the colours, categorical or not, present in the stimulus? Here we formulate two hypothesis:

H1: Illumination colour dependent (influence of departure categories): dependent the perception of some colour categories at the first miliseconds of adaptation constraints the process and makes that this categories are preserved. Ideally colour constancy happens in several stages, being the first simultaneous colour constancy which happens the first milliseconds. Then we wonder if this first categorical appreciation by the subject could influence the posterior adaptation and in particular the existence, or not, of some particular colour categories.

H2: Surface colour dependent (influence of arrival categories): object dependent the expectation or possibility to perceive some colours at the final temporal interval of the adaptation is what guides this adaptation. Are there any differences according to colour categories or the behaviour of all categories is symmetric? This is, if the colour constancy process is perfect then we should recover the original categories displayed under D65, then we wonder about the influence recovery of this particular colour categories.

From brainard shcema, studying colour constancy we must pursue two aims: describing the colour constancy extent and also try to model the phenomenon. In this paper we will aim to a new aim, predict the categorical colour. We would like to achieve this aim in order to predict the colour after adaptation(because we did not apply colour naming paradigm) and conclude if nameability has growth under adapation. Our aim is to predict categorical colour appearance of displayed colours from adjusted *Selected Representatives*.

In order to systematize the analysis of adaptation for each subject we could look for the intrinsic properties described by the adaptation matrix. This could be could be interesting to classify adaptation matrices according to its internal features. In particular studying the diagonal plus translation and affine cases. Can we extract a common matrix pattern for each adapted illuminant that holds at the same time for all subjects? Are the vep's and its vap's revealing this pattern? Then use this pattern to predict categorical colour appearance with simulated synthetic cases. Lemma: colour constancy mechanisms maximizes the number of colours. Lemma: colour constancy mechanisms maximize the number of colour categories. Here we show how high values of colour constancy indices may not be correlated by a high degree of categorical colour constancy.

6.2 Methods

Observers underwent a chromatic setting task followed by a constrained colour naming task for the adaptation state achieved in each experimental session. The main experiment consisted of 12 sessions and in each session observers achieved a particular adaptation state consequence of a long period of adaptation to an illuminated three-colour Mondrian background, displayed in a CRT. For each observer and adaptation state we defined his/her adapted colour naming model in CIELab colour space and from its parameters were derived categorical colour predictions of unmeasured samples.

6.2.1 Observers

Three observers (1 female and 2 males) participated, of mean age 29 (range 26-32). 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 observers were unpaid volunteers naïve to the experiment's purpose, the other was one of the authors. All were no native English speakers but had excellent English language skills.

6.2.2 Experimental setup and procedure

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 monitor and it 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, Lum = 100 Cd/m^2. The experimental setup was the same as the one detailed in subsection 4.3.2.

The whole experiment consisted of 16 sessions: 12 sessions from all combinations between 2 different Mondrian backgrounds and 6 different illuminations, and 2 reference sessions to reinforce the selected representatives, and 2 regular sessions with background Type 0 and D65 illumination to check for observers ability to reproduce the same colours throughout the experimental period. All sessions were done in two weeks. Each experimental session lasted approximately 25 minutes and each trial 30 seconds.

Observers performed two different tasks in each session. First, the observers performed a chromatic setting (see details in subsection 4.2) and just after, while keeping the same stimulus on the CRT screen, they were required to complete a constrained colour naming task for all the colours present on the screen. Colour names were recorded by hand by the observer itself to a white sheet of paper which was out of sight during the chromatic setting task.

Constrained-naming. Observers were required to assign a colour term as fast as possible to each colour present on the screen. In their colour naming task, observers were required to use the eleven basic terms (Green, Yellow, Blue, Red, Grey, Orange, Pink, Purple and Brown) or combinations of them, or complemented with the adjectival form of basic terms (Greenish, Yellowish, Bluish, Reddish, Greyish, Oranges, Pinkish, Purplish and Brownish).

6.2.3 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 70x70 pixels. There were two types of backgrounds:

Type III. It was built from the SRs chosen by each observer in the reference session. There were three colours in total and they were selected to lie in the SRs joining lines: in between yellow and green (25% away from green), in between yellow and brown (25% away from yellow), and in between purple and pink (50% from purple).

Type IV. It was built from the SRs chosen by each subject in the reference session. There were three colours in total and they were selected to lie in the SRs joining lines: in between blue and green (50% from green), in between yellow and brown (25% from purple), and in between purple and pink (50% from purple).

CIELab coordinates of the three colours displayed in backgrounds Type III and IV are detailed in Table 6.1. Also, their CIE 1976 u'v' coordinates are plotted in Figure 6.1. Notice that each observer had a different set of stimuli colours while keeping the same categorical information.

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

Table 6.1 CIELab coordinates of the three stimuli colours under D65 illumination for each observer. Reference white point D65 with L=100.

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 six illuminants, whose CIE *xy* chromaticities are shown in Table 6.2. Given the singular 3D shape of the CRT gamut we diminished (40%) the power intensity of the bluish illumination in order to ensure that all illuminated colours were displayable by the CRT.

Illuminant	x	y	Munsell coordinates	Saturation strength in ΔE*
D65	0.317	0.329	-	0
Greenish	0.296	0.454	2.5G 7/10	37.4
Yellowish	0.453	0.435	2.5Y 7/8	43.9
Bluish	0.242	0.285	7.5B 7/8	16.8
Unsaturated greenish	0.316	0.405	2.5G 7/6	21.2
Unsaturated yellowish	0.401	0.393	2.5Y 7/4	25.2

Table 6.2 Illuminant coordinates in CIE 1931 xy colour space used in Experiment IV. Also the last columns contain its Munsell coordinates.

The number and sizes of rectangles were manipulated so that the pixel average chromaticity of all background types under D65 illumination was close to D65, i.e., close to the origin in the a*b* plane. Table 6.3 reports these average values in CIELab coordinates. Background Type III was effectively centred around the origin but background Type IV was slightly shifted towards the negative part of the b* axis, for all three observers. This fact is probably due to the particular selection of colours selected in Type IV. Notice how in Figure 6.1b all triangles under D65 illumination do not include the D65 chromaticity, something that do not happen for background Type III.

Observer	Background Type III			Background Type IV		
	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 displayed Mondrians under D65 illumination in CIELab coordinates. D65 was used as a reference white point.

As specified in subsection 4.3.4 of Experiment II we used a multiple test patch to perform the chromatic setting task. In the current experiment the average number of patches was 11 (3 StDev) and their total average area was approximately a 7% of displayed images.
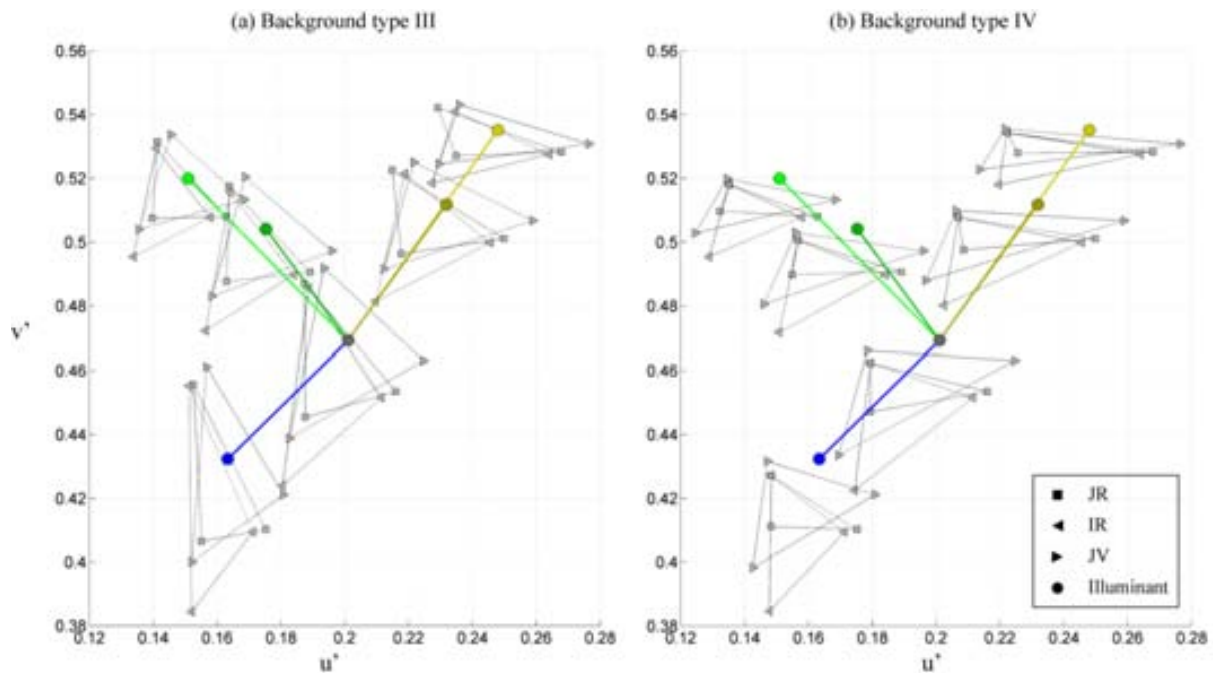
Figure 6.1 CIE 1976 u'v' coordinates of stimuli colours used in Experiment IV

. Panel a and b corresponds to the three-colour Mondrians (triangles joined by black dotted lines) under all tested illuminations for all observers, respectively Type III and IV backgrounds. Circle markers correspond to the u'v' coordinates of the illuminants and are colour coded according to its representative colour. Coloured lines link D65 to the rest of illuminants coordinates with the same colour codification.

6.3 Data analysis

6.3.1 Linking lost correspondences in the constrained naming test

During the constrained colour naming task each sample was classified by a single or compound colour term. Due that colour terms were recorded in an arbitrary order by observers, then the correspondences between colours across sessions were confounded. In order to study possible categorical changes across different adaptations we need to recover this correspondences from recorded data. Here we present a simple algorithm to achieve this purpose.

Since each stimulus had three different colours, then each session ended with a list of three basic or compound colour terms. Then, given the results of two different sessions, we need to find a bijective correspondence between the three elements of each list.

Our developed algorithm has two steps. First, for each list the three basic or compound terms are translated into a 3x9 matrix with coefficient values 0, 0.5 and 1. In that matrix each row corresponds to the codification of one list colour term into basic terms and each column corresponds to a particular basic term. Initially all matrix coefficients has zero value. For instance, if the first term of the list was blue-greenish then the first row of the matrix would be updated with value 1 in *ith* coefficient corresponding to blue and with 0.5 in *jth* coefficient corresponding to green. If the second term was pink-purple then the second row would be updated with value 1 in both coefficient positions corresponding to the pink and purple. If the last term of the list was yellow then the last row of the matrix would update with 1 the *kth* column coefficient for yellow. The same process applies to the second list and gives another matrix. The second step, takes one of both matrices and consider the resultant matrices from all possible row permutations. Then, for each permutation the absolute difference between the permutated matrix and the other matrix is computed. We choose the permutation which gives the lowest value of the accumulated sum of the difference between the permutated matrix and the other matrix. Thus, this technique set correspondences between two lists of basic or compound colour terms. It is based in maximizing the coincidences for all three colours at the same time.

6.3.2 A similarity index for compound basic colour terms

For each displayed colour the constrained colour naming test produced a colour term and also, through the algorithm introduced in the last subsection, a correspondence of colour terms under the six different illuminations was obtained. In order to evaluate systematically how similar are these sets of six colours we developed an index to evaluate the similarity between pairs of such colour terms. Our naming results indicate that observers did not use in any case more than two terms for each coloured sample, then possible combinations are: [basic term], [basic term, basic term] and [basic term, basic term adjective]. When combining these naming types we have six possible combinations as described in the first column of Table 6.4.

Case description	Discussion				Index	Example
I.$(n_1,-)\&(n_2,-)$	$n_1=n_2$				1	(B,-)&(B,-)
	$n_1\neq n_2$				0	(B,-)&(G,-)
II. $(n_1,m_1)\&(n_2,m_2)$	$n_1=n_2$	$m_1=m_2$			1	(B,G)&(B,G)
	$n_1=n_2$	$m_1\neq m_2$			0.5	(G,B)&(G,Y)
	$n_1\neq n_2$	$m_1=m_2$			0.5	(B,G)&(Y,G)
	$n_1\neq n_2$	$m_1\neq m_2$			0	(B,Pr)&(G,Y)
	$n_1=m_2$	$m_1=n_2$			1	(B,G)&(G,B)
	$n_1=m_2$	$m_1\neq n_2$			0.5	(B,G)&(Pr,B)
	$n_1\neq m_2$	$m_1=n_2$			0.5	(B,G)&(G,Y)
	$n_1\neq m_2$	$m_1\neq n_2$			0	(B,G)&(Y,O)
III. $(n_1,-)\&(n_2,m_2)$	$n_1=n_2$				0.75	(G,-)&(G,Y)
	$n_1\neq n_2$	$n_1=m_2$			0.75	(G,-)&(Y,G)
	$n_1\neq n_2$	$n_1\neq m_2$			0	(G,-)&(Y,O)
IV. $(n_1,-)\&(n_2,a_2)$	$n_1=n_2$				0.875	(G,-)&(G,Y-ish)
	$n_1\neq n_2$	$a_2 \simeq n_1$			0.125	(B,-)&(G,B-ish)
	$n_1\neq n_2$	$a_2 \not\simeq n_1$			0	(B,-)&(G,B-ish)
V. $(n_1,m_1)\&(n_2,a_2)$	$n_1=n_2$	$a_2 \simeq m_1$			0.875	(B,G)&(B,G-ish)
	$n_1=n_2$	$a_2 \not\simeq m_1$			0.5	(B,G)&(B,Pr-ish)
	$n_1\neq n_2$	$m_1=n_2$	$a_2 \simeq n_1$		0.875	(G,B)&(B,G-ish)
	$n_1\neq n_2$	$m_1=n_2$	$a_2 \not\simeq n_1$		0.5	(B,G)&(B,Pr-ish)
	$n_1\neq n_2$	$m_1\neq n_2$	$a_2 \simeq m_1$		0.125	(G,B)&(Pr,B-ish)
	$n_1\neq n_2$	$m_1\neq n_2$	$a_2 \simeq n_1$		0.125	(G,B)&(Pr,G-ish)
	$n_1\neq n_2$	$m_1\neq n_2$	$a_2 \not\simeq n_1$	$a_2 \not\simeq m_1$	0	(G,B)&(Pr,P-ish)
VI. $(n_1,a_1)\&(n_2,a_2)$	$n_1=n_2$	$a_1=a_2$			1	(G,B-ish)&(G,B-ish)
	$n_1=n_2$	$a_1\neq a_2$			0.75	(G,B-ish)&(G,Y-ish)
	$n_1\neq n_2$	$a_1=a_2$			0.25	(G,B-ish)&(Pr,B-ish)
	$n_1\neq n_2$	$a_1\neq a_2$			0	(G,B-ish)&(Pr,P-ish)
	$n_1\neq n_2$	$n_1 \simeq a_2$	$n_2 \simeq a_1$		0.75	(G,B-ish)&(B,G-ish)
	$n_1\neq n_2$	$n_1 \simeq a_2$	$n_2 \not\simeq a_1$		0.125	(G,B-ish)&(B,P-ish)
	$n_1\neq n_2$	$n_1 \not\simeq a_2$	$n_2 \simeq a_1$		0.125	(G,P-ish)&(B,B-ish)
	$n_1\neq n_2$	$n_1 \not\simeq a_2$	$n_2 \not\simeq a_1$		0	(G,Y-ish)&(B,P-ish)

Table 6.4 A similarity index between two compound colour terms

. The first column contains the prototype description of the compound colour term combination (n_1,n_2, m_1 and m_2 are BCT, and a_1 and a_2 are BCT adjectives also the $\not\simeq$ symbol indicates coincidence of BCT category), the following four columns contains the discussion of possible cases and the sixth column contains the index value. Finally, the last column contains and example.

The similarity index reported in Table 6.4 was assigned following a rationale derived of each case while keeping consistency throughout all six cases. Index values ranges between 0 and 1, indicating null and full categorical coincidence respectively. Notice that compound names cannot have both terms belonging to the same basic colour category. Case I corresponds to the comparison between two single basic terms and thus only two options are possible: coincidence or not, and thus its values are 1 or 0 respectively. Case II corresponds to the comparison between two compound basic terms without adjectives, then possible cases are: coincidence of both basic colour terms, coincidence of one basic colour term and any coincidence, and thus its values are 1, 0.5 and 0 respectively. Case III is even

simpler, only allowing to the coincidence of one basic term or no coincidence at all, then its values are 0.75 and 0 respectively. Notice here that we need to keep consistency with a previous case: take for instance the current case (G,-)&(G,Y) with 0.75 and the previous one (G,B)&(G,Y) which was assigned 0.5 of similarity, if we think in terms of geometrical distances then the green centroid is closer to the green-yellow centroid than the green-blue and green-yellow are, from here follows the higher value of the (G,-)&(G,Y) case. Case IV is similar to the previous one but instead of having in the compound term two basic colour terms it has a basic term plus a basic term adjective. Then its values are basic terms coincidence, adjectival term categorical coincidence to the noun and no coincidence at all, and thus its values are 0.875, 0.125 and 0 respectively. Notice here that we have a similar case of inter-case consistency, for instance the (G,-)&(G,Y-ish) needs to have a higher index value than (G,-)&(G,Y). Case V and VI follows the same rationale as in the previous cases, however there is a singular case: take for instance (G,B-ish)&(B,G-ish) where formal coincidence is null but the categorical quality of the colour sample is so close, we assigned 0.75 similarity in that particular case.

6.3.3 A categorical colour prediction model

Despite the constrained colour naming task provided categorical information for each displayed colour in each adaptation state, we developed a method to predict its categorical colour appearance under the reference illuminant (D65). The method is based on three steps applied to each displayed colour: (i) To predict its LMS coordinates under the reference illuminant, (ii) From LMS coordinates we assigned a two-category probability, and (iii) From the two-category probability we assigned a single or compound BCT.

Given a colour sample under a test illuminant (any of detailed in Table 6.2), the first step obtains its LMS coordinates under the reference illuminant. We measured the chromatic settings under the reference and test illumination while keeping the same background, then as described in subsection 4.4.4 we modelled the adaptation through a linear model of colour constancy. Here we used an affine matrix fitted with the nine correspondences between test and reference chromatic settings. Finally, the matrix was applied to the LMS coordinates of displayed colours under test illuminant and a prediction of the LMS coordinates of them under the reference illuminant was obtained.

Second, for the produced LMS coordinates we assigned a probability of belonging to two basic colour categories. Since the reference illuminant was D65 and the LMS coordinates represented the colour seen under the reference illuminant too we computed its CIELab coordinates with D65 as a reference white point, label them as **x**. Next, from the set of nine chromatic settings measured under the

reference illuminant and same background, we assigned a categorical classification in terms of only two basic colour categories.

Construction of a Hue-luminance map. The categorical classification discarded saturation and was based on the hue of x in relation to the location of x_N while keeping luminance information. The particular implementation makes use of a cylinder whose main axis is parallel to the L*axis but its origin coordinates in a*b* plane were that of x_N. Its L* values ranged between the minimum and maximum L* of chromatic settings and its cylinder radius was 22 ΔE*. The method uses the CIELab colour space because it is a 3D colour space and perceptually uniform. It uses three main steps and the CIELab colour space, because it is opponent and perceptually uniform, and three dimensional to differentiate amongst the categories yellow, brown, pink, red and purple.
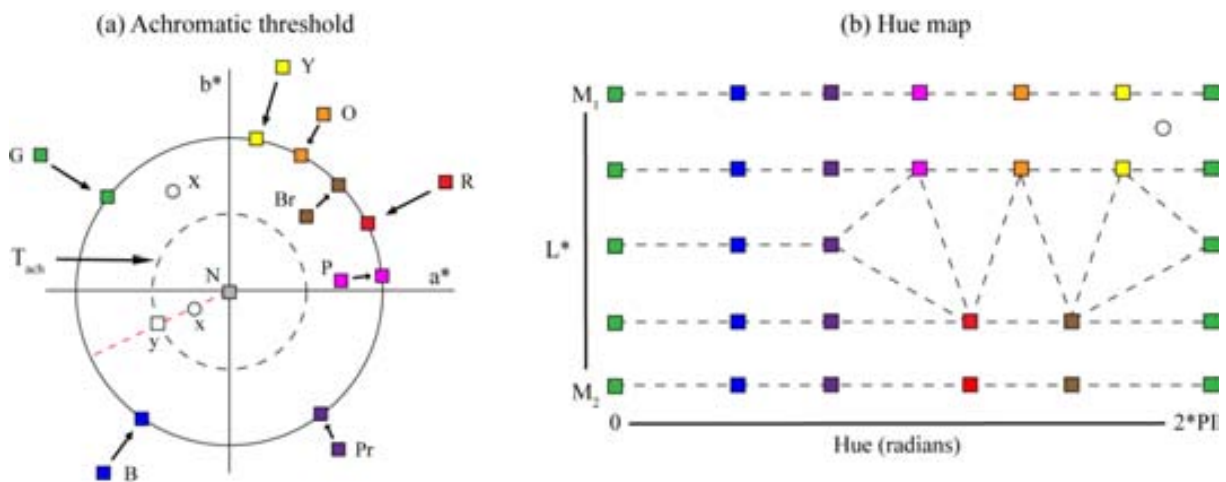
Figure 6.2 Schematics of the Hue-Luminance map. Panel a shows the first step in the Hue-Luminance map creation and two different hypothetical cases according to the achromatic threshold. Panel b shows how the cylinder surface is divided according to basic colour categories. See text for further details.

The procedure was divided in several instructions according to ΔE* distance between the grey chromatic setting and **x**, in the a*b* chromatic plane. If $\|x - x_N\|_2 \geq T_{achr}$ then:

1. Project **x** to the surface of the previously defined cylinder.
2. Assign a probability of belonging to two basic colour categories. In the hue map defined in Figure 6.2b compute the closest dotted line to the previous projection. The extremes of the dotted line set the two basic colour categories that characterize the colour appearance of that colour.
3. Assign a probability of belonging to each category from the relative distances between line extremes and the projection. Notice that this distances are computed on the cylinder surface and when the extremes has the same L* value then equals to use only the hue.

117

If $\|x - x_N\|_2 < T_{achr}$ then:

1'. Do (1) to (3) and select the category with the highest probability.
2'. Compute relative distances between y x and N (panel a of Figure 6.2).
3'. Assign a probability of belonging to grey and the other category.

Finally, the third step applied the function defined in Equation 6.1 to the probabilities output of previous step. Equation 6.1 summarizes all five possible cases and its classification depends on two threshold parameters:

$$\psi(p_1, p_2) = \begin{cases} (c_1, c_2) & |p_1 - p_2| < T_1 \\ (c_1, c_2 - ish) & T_1 < |p_1 - p_2| < T_2 \ and \ p_1 > p_2 \\ (c_1 - ish, c_2) & T_1 < |p_1 - p_2| < T_2 \ and \ p_1 < p_2 \\ (c_1, -) & |p_1 - p_2| > T_2 \ and \ p_1 > p_2 \\ (c_1, c_2) & |p_1 - p_2| > T_1 \ and \ p_1 < p_2 \end{cases} \qquad (6.1)$$

Threshold parameters are T_1 and T_2 and satisfy $0 < T_1 < T_2 \leq 1$. Threshold values rule when one colour is classified as belonging to only one category, to two categories, or complemented by basic colour terms adjectives. See discussion for an empirical discussion on our results. We tested its influence and set to 0.25 and 0.5 respectively, see subsection 6.5.1 for further details.

6.4 Results

We have developed a technique to study categorical changes under colour constancy. And if the technique works fine then we could claim that we have developed a method to predict categorical appearance under illumination changes. However the previous statement is provided that we demonstrate that the colour constancy is not categorically stable under some cases despite colour constancy indices suggest it.

6.4.1 Usage of compound basic colour terms

The constrained naming test provided a total of 108 named samples, 54 for each background type (3 colour names for each of 6 illuminants and 3 observers). Table 6.5 contains all assigned compound colour terms for each background, illuminant and observer. Rows highlighted in bold correspond to

the colour terms assigned under D65 illumination. As previously stated, correspondences between colour terms were confounded in the constrained colour naming task across sessions. Coded colours in Table 6.5 are sorted by the algorithm introduced in the previous subsection 6.3.1, i.e., colour terms were reordered in such a way that maximized the categorical coincidence with colours seen under the reference illuminant. For instance, the colours labelled as C1, C2 and C3 for the JR observer under illumination one were B, G and P-Pr-ish, then all three colour lists recorded under other illuminations were reordered such the similarity with those terms was maximized. The same procedure applied to all other cases.

		JR			IR			JV		
Back	Illum	C_1	C_2	C_3	C_1	C_2	C_3	C_1	C_2	C_3
	1	**B**	**G**	**P,Pr-ish**	**B**	**G**	**Pr,P-ish**	**B,R-ish**	**G,N-ish**	**P,Pr-ish**
	2	B,G	G,Y-ish	N	B,G	G	N,O-ish	B,O-ish	G	N,O-ish
III	3	N	G,Y-ish	P,R-ish	B,N-ish	G,Y-ish	P,O-ish	N	G	R-P
	4	B	G	Pr,N-ish	B	G	Pr,N-ish	B	G	Pr,N-ish
	5	B,G	G	Pr,P-ish	B	G	N,Pr-ish	B,R-ish	G	N,R-ish
	6	N,B-ish	G,Y-ish	P,R-ish	B	G	P	N,B-ish	G	P,R-ish
Back	Illum	C_4	C_5	C_6	C_4	C_5	C_6	C_4	C_5	C_6
	1	**B, N**	**G,N-ish**	**P,R-ish**	**B**	**G**	**P,N-ish**	**B**	**G,N-ish**	**R,P-ish**
	2	B,G	G	Pr,R-ish	B,G	G ,Y-ish	G	B,O-ish	G	R,N-ish
IV	3	N,Y-ish	G,Y-ish	R	B,N-ish	G	O,P-ish	B,N	N,B-ish	R,P-ish
	4	B	G,N-ish	Pr,R-ish	B	G	Pr,N-ish	B	G,B-ish	R,Pr-ish
	5	B,G	G,N	Pr,P-ish	B	G	Y,P-ish	B,O-ish	G	R,Br-ish
	6	N	G,N-ish	P,R-ish	B	G	P,O-ish	B,N-ish	G,N-ish	R,P-ish

Table 6.5 Naming results from the constrained naming task for our three observers.

Rows indicate results according to illuminants and background types. Columns indicated results according to observers and colours. In bold are the results obtained under D65 illumination. Notice that colours were sorted according to the method introduced in subsection 6.3.1.

Observers used at maximum two colour terms to describe perceived colours and its usage was reduced to the following cases: [basic term] or [basic term, basic term] or [basic term, basic term adjective]. The usage pattern of these colour terms is resumed in the histograms of Figure 6.4 where naming results are averaged over backgrounds and illuminants in panel a. Observers used an average of 8.8 times for [basic terms] case, 7.5 times for [basic term, basic terms adjective] and only 1.7 times for [basic term, basic term] case. Figure 6.4b contain the breakdown of the previous panel a histogram into different backgrounds and illuminants, top histogram for Type III and bottom for Type IV. Notice how for both background types the pattern N-NA-NN is similar for the first four illuminants but the last two had a tendency to a higher number of NA.
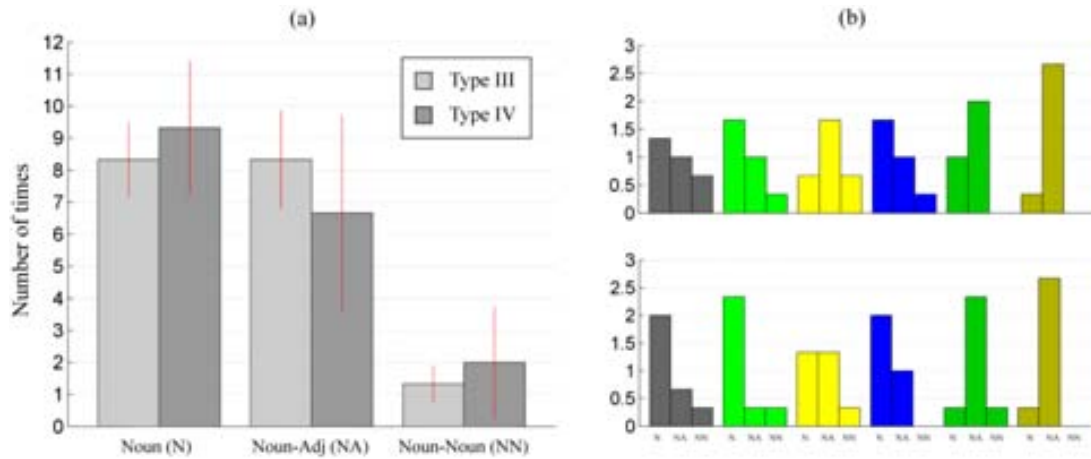
Figure 6.3 Pattern usage of compound colour terms in the constrained-naming task.

Panel a: usage averaged over all observers and illuminants. Panel b: usage breakdown according to illuminants and backgrounds type. Top corresponds to Type III and bottom to type IV.

If colour constancy was complete then the assigned colour terms in Table 6.5 would be the same in each column. We quantified these categorical variations using the colour terms similarity index introduced in the previous subsection 6.3.2, i.e., we computed the index of categorical coincidence for each pair of assigned colour terms under D65 and a test illumination. For instance, take the similarity index between the colour assignments done for observer IR under illumination 1 (D65) and C1, which is G, to the same but colour but under illumination 2, which is G and then the index is 1. Table 6.5 contains the resulting values of computing these similarities and also summarized in its last row for each background type (highlighted in bold).

		JR			IR			JV			Mean
Back	Illum	C$_1$	C$_2$	C$_3$	C$_1$	C$_2$	C$_3$	C$_1$	C$_2$	C$_3$	
	2	0.75	0.875	0	0.75	1	0	0.875	0.875	0	**0.57**
	3	0	0.875	0.75	0.875	0.875	0.125	0	0.875	0.5	**0.54**
III	4	1	1	0.125	1	1	0.75	0.875	0.875	0.125	**0.75**
	5	0.75	1	0.75	1	1	0.125	0	0.875	0	**0.61**
	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**

		JR			IR			JV			Mean
Back	Illum	C$_4$	C$_5$	C$_6$	C$_4$	C$_5$	C$_6$	C$_4$	C$_5$	C$_6$	
	2	0.875	0.5	0	0.875	0.75	0	0.75	1	0.875	**0.62**
	3	0.75	0.5	0.125	1	0.875	0.125	1	0.75	0.125	**0.58**
IV	4	1	0.75	0	1	1	0	0.75	1	0.75	**0.69**
	5	0.875	0.5	0.125	1	1	0.125	0.75	1	0.875	**0.69**
	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**

Table 6.6 Index of categorical term coincidence applied to constrained-naming data. Quantify the degree of categorical change from previous Table 6.4 through the index of categorical coincidence.

The average value of Table 6.6 is 0.71 and thus suggesting that categorical colour perception was mostly stable under illumination changes and over both background types. However, a close analysis into the conforming background colours suggested that this categorical stability only hold for colours C1 (0.60), C2 (0.92), C4 (0.91) and C5 (0.81), and not for colours C3 (0.32) and C6 (0.39) with a much lower index of categorical similarity. Then green and blue were the most stable colours when illumination was changed and colours C3 and C6 labelled as combinations of pink, red and purple under D65 illumination were the less stable.

6.4.2 Chromatic settings under six different illuminants

Figure 6.4 shows the averaged chromatic settings in CIELab colour space over all three observers, discriminated by background type (III and IV), and separated in panels according to the six adapted illuminations. 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 measured 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 of 3.2 ΔE^* ($\sigma=1.5$) and no differences were found according to background type. According to illuminants, it varies significantly ($F(5,102)=2.72$, $p<0.05$) but differences were only significant between the greenish (mean=3.7, $\sigma=0.8$) and unsaturated yellowish (mean=2.9, $\sigma=0.6$). According to colour categories, grey was the one with the lowest error (mean=2.54, $\sigma=0.44$) and purple with the largest (mean=4, $\sigma=1$). Also, there were small differences according to observers:

JR 2.66 ΔE* (σ=0.93), IR 2.83 ΔE* (σ=1.26) and JV 4.13 ΔE* (σ=1.65). Average time spend in each trial was 24.2 seconds (σ=5.5).

A series of sessions were included in the experiment to assess the ability of observers to reproduce the same colours over the experimental period. Appendix F contains a brief report about the repeatability of the chromatic settings following the same approach as described previously (see subsection 4.4.1). We conclude that observers were able to reproduce reliably the chromatic settings.
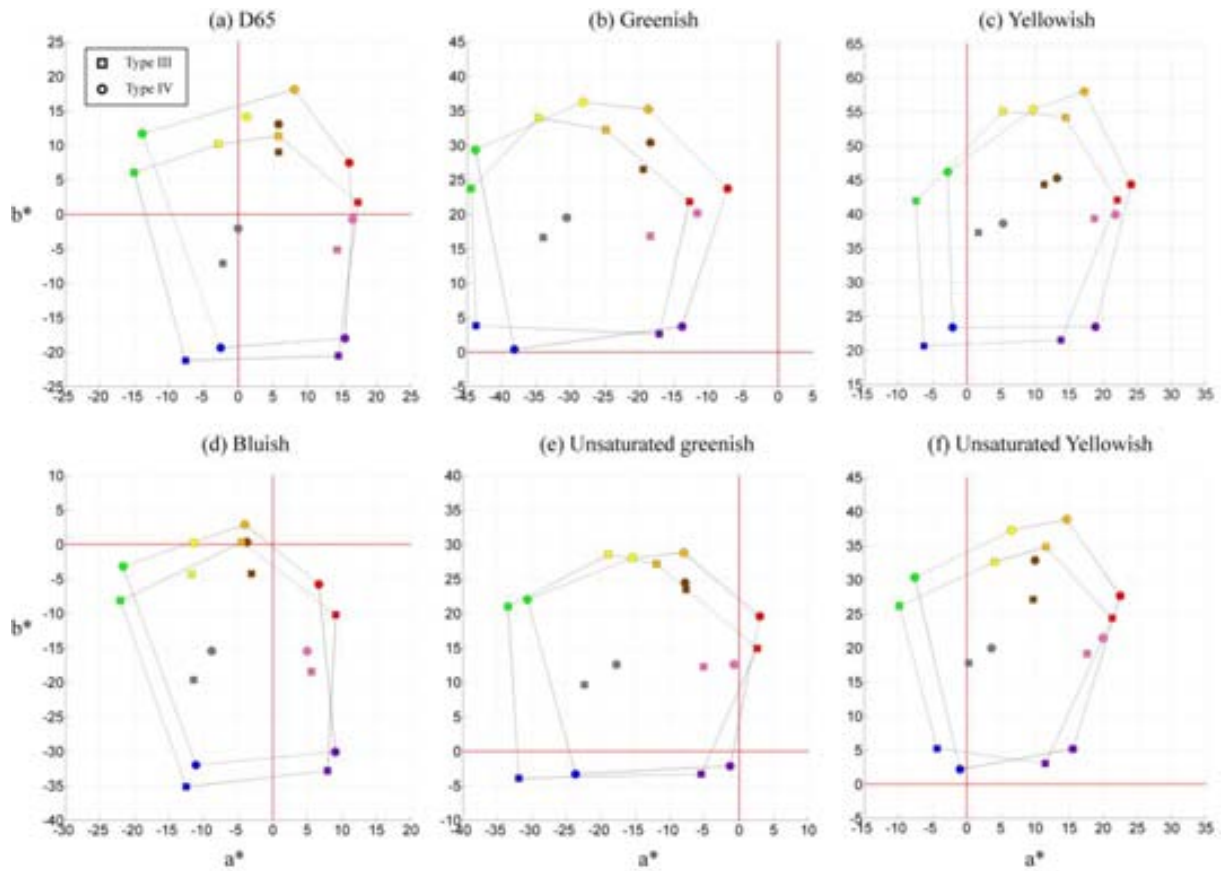
Figure 6.4 Chromatic settings averaged over all three observers for all six illuminants

. 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 global contrast effects, see text for details.

All six panels of Figure 6.4 show a consistent difference in the location of chromatic settings according to background type (linked by a black line). The tendency may be summarized as a broadly uniform shift, in magnitude and direction, for all categories between the same colour coded markers.

The shift was modelled for each colour category as the vector defined from the chromatic settings under background type III and IV (same colour coded square and circle). Its average magnitude was 7.2 ΔE^* ($\sigma=2.1$) and its direction was computed as the angular distance between the vector defined under the reference and test illuminants while keeping the same colour category. Average angular difference was 30 degrees ($\sigma=20$) which indicate an overall broad agreement. The existence of this shift is expected from global colour contrast adaptation, which adjusts for changes in the average colour across scenes (Webster 1996).

Along the experimental period we kept track of observers ability to reproduce the chromatic settings by means of two repeatability tests. These were conducted regularly at approximately three days' intervals and included a reference session where observers were asked to reproduce the original chromatic settings and a regular session with D65 illumination and Type 0 background. Figure 6.5 contains the chromatic settings for the reference sessions; columns correspond to observers and the first row the original chromatic settings and the other correspond to the repeatability test in temporal sequence. We quantified this ability by performing a series of t-tests to the same categories across different rows in each CIELab dimension. Our results show that the means of the results populations considering all observers were equal in 90% of the cases. The same approach was applied to the other kind of session and populations were equal in 70% of the cases (these results are reported in a Figure contained in Appendix F). As in Experiment II, this difference is likely to the absence of the Boundary Cylinder which increased uncertainty in the saturation dimension.
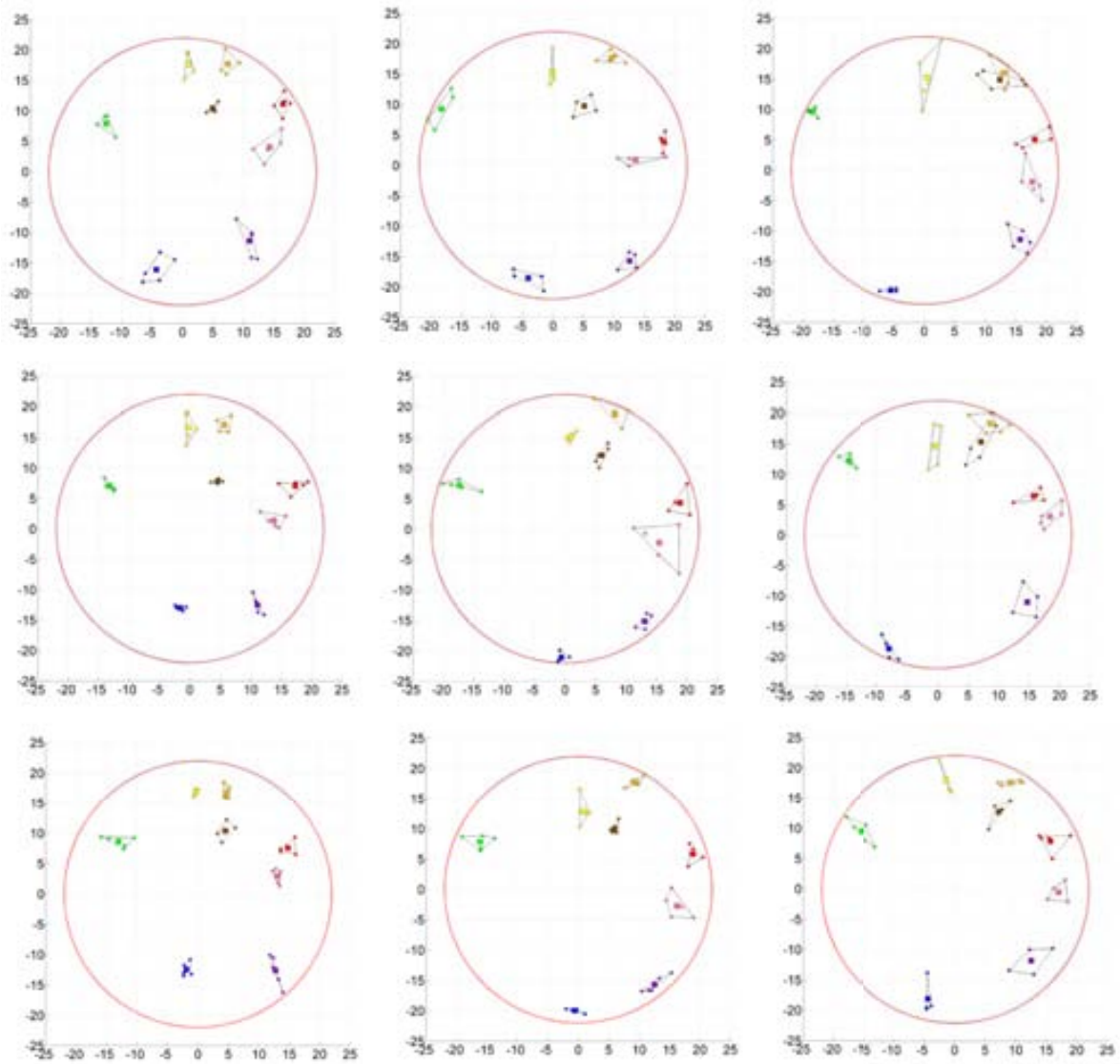
Figure 6.5 Chromatic settings in Experiment IV repeatability tests. First row shows the selected representatives chosen by the three subjects in the reference session. Following rows show the corresponding settings for the three subsequent repeatability tests. Square markers represent the average of individual trials (small dots joined by lines) and the large red circle corresponds to the Bounding Cylinder in a*b* chromaticity plane.

6.4.3 Hue disruption in chromatic settings due to illuminant changes

Each set of chromatic settings in Figure 6.5 seems to preserve somehow its inter-distances, something that was observer and quantified in Chapter 5 by means of the graph approach. We applied the same technique (see subsection 5.3) in order to find overall structural differences in the interrelations of chromatic settings motivated by illumination changes. Graphs were computed for each averaged set of chromatic settings in Figure 6.5 and distances computed between the reference and test illuminants while keeping the same background type. No differences were found according to background types.

Then averaged results for each illuminant were: 13% for greenish, 22% for yellowish, 10% for bluish and unsaturated greenish and 12% for unsaturated yellowish. Noticeably, this results are somehow in agreement with Chapter 4 findings which reported a higher degree of chromatic settings structural disruption under the yellowish illuminant motivated for different index values of the structural colour constancy index and linear models precision.

According to the graph approach, inter-distances among chromatic settings were highly stable (86.6% in average) and thus suggesting a high degree of stability in the categorical colour structure of colour space under changes of illuminant. However, the graph approach could hide cancellations among relatively small chromatic settings movements and also, results from our naming task (see next subsection) revealed the existence of consistent categorical changes. We further analysed our data by removing the saturation dimension from the analysis whilst focusing only in hue and luminance. The hue dimension was computed from the CIE 1976 u'v' coordinates of chromatic settings, i.e., for each chromatic setting we computed the angle between the vector defined as parallel to the u' axis from grey and the vector conformed by grey and the chromatic setting. The horizontal dimension of Panel a in Figure 6.6 corresponds these hue values, and the vertical dimension corresponds to the luminance. The hue and luminance coordinates of chromatic settings in Figure 6.6 are plotted according to backgrounds (panels) and illuminants (markers). Each panel in 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 in Experiment II). Notice that results in panel c were averaged after any significant differences were found among types 0, I and II.
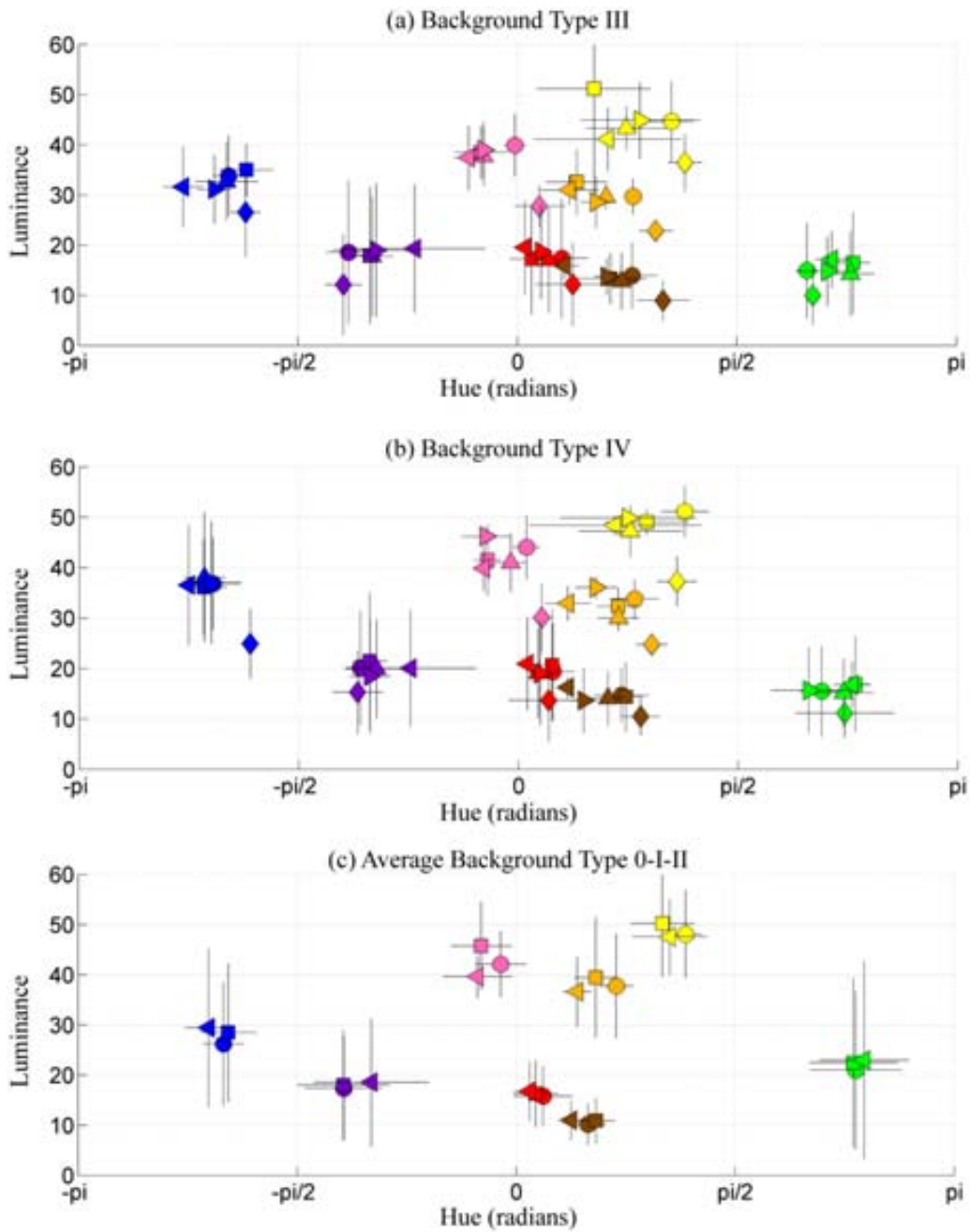
Figure 6.6 Hue and luminance differences in chromatic settings due to illuminant changes. Markers are colour coded according to the chromatic setting colour category and markers shape indicate the illuminant adapted. Each panel corresponds to results obtained under different backgrounds, see text for details. Black lines indicate standard deviation.

Figure 6.6 reveals three main facts. First, luminance information is not a distinctive factor despite that chromatic settings under bluish illumination (diamond in panels a and b) had lower luminance, something directly related to the lower intensity of bluish illuminant. Second, panels a and b has remarkable hue differences according to the adapted illuminants for nearly all categories. Third,

126

chromatic settings in panel c has no differences according to the adapted illuminants (D65, greenish and yellowish).

Figure 6.7 summarized the previous hue differences; it contains the difference (in degrees) between the corresponding colour *hues* under the reference and test illuminants (between the circle marker and the rest in Figure 6.6). Panel a of Figure 6.7 contains the results for background type III and results are grouped according to illuminants in colour coded bars for each chromatic setting. Notice the tendency for the greenish, yellowish, unsaturated greenish and unsaturated yellowish is that broadly exist two rotations with opposite directions at the same time: red, yellow, pink, orange and brown rotate to the left, and green and purple to the right (according to Figure 6.6 markers location). Panel b contains the results for background type IV and also contains these two opposite rotations, however less consistent. When results are compared over backgrounds there is a close similarity for the yellowish, bluish and unsaturated greenish illuminants. Notice that under the bluish illuminant hue disruption was minimal and also consistent to the same orientation in both background types.

The magnitude of hue disruption was averaged over backgrounds and categories and it was: 11.6 deg for greenish, 19.9 deg for yellowish, 6.6 deg for bluish, 9.1 deg for unsaturated greenish, and 10.7 deg for unsaturated yellowish. Interestingly these magnitudes were strongly correlated (r=0.91 and p=0.03) to the magnitude of the illuminant shift (computed as ΔE^* distances in a*b* plane from the origin (D65) to the illuminants coordinates: 37.4 for greenish, 43.9 for yellowish, 16.8 for bluish, 21.2 for unsaturated greenish and 25.2 for unsaturated yellowish). This correlations suggests that as more intense is the chromaticity saturation of the illuminant then higher distortions for chromatic settings are expected. However, the proportion of disruption was lower for the greenish illuminant.
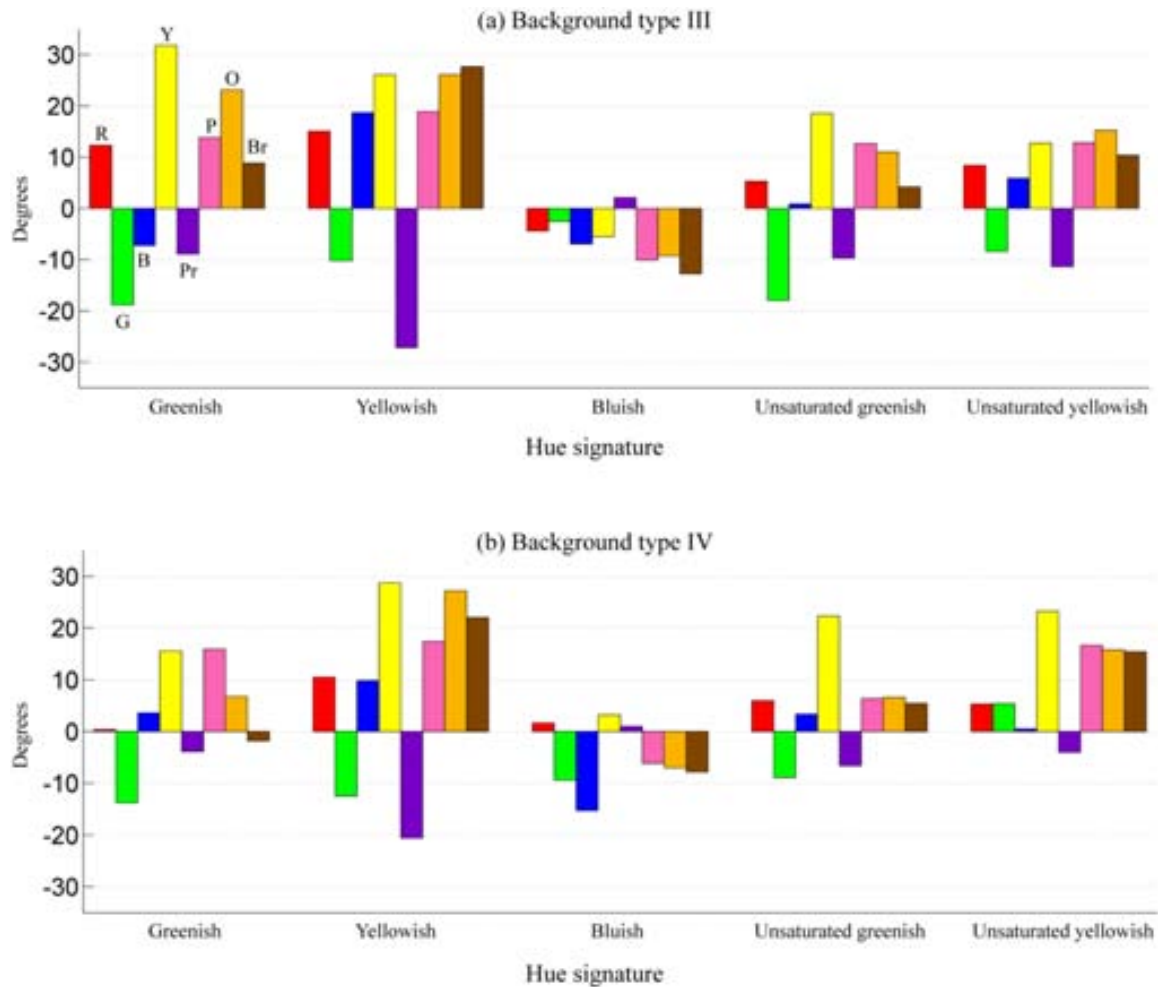
Figure 6.7 Hue distance for chromatic settingsunder the reference and test illuminants. Panel a corresponds to background type III results which are grouped by illuminants and colour coded according to chromatic setting categories (bars). Panel b corresponds to background type IV and follows the same description.

6.4.4 Extent of categorical colour constancy

Here we quantified the extent of colour constancy from two different approaches. First, we computed a traditional colour constancy index which quantified the magnitude of displacement between corresponding locations of chromatic settings in colour space under the test and reference illuminants. Second, we predicted the location of background colours in colour space once adapted to a test illuminant and compared to their locations under the reference illuminant. Both approaches contributed to describe the extent, in magnitude and quality, of categorical colour constancy.

Figure 6.8 contains the CIE 1976 uv coordinates of chromatic settings corresponding to the grey category for all tested observers and background/illumination conditions. As expected by the results of previous studies in successive colour constancy (Foster 2011) the location of this particular point lies

close to the joining line between illuminants coordinates. Notice that its close location to the coordinates of test illuminants suggests a high degree of colour constancy. Also, the small existent shift for measurements done under D65 confirms the global contrast effects described in subsection 6.4.2., consequence of the biased average chromaticity of background type IV.
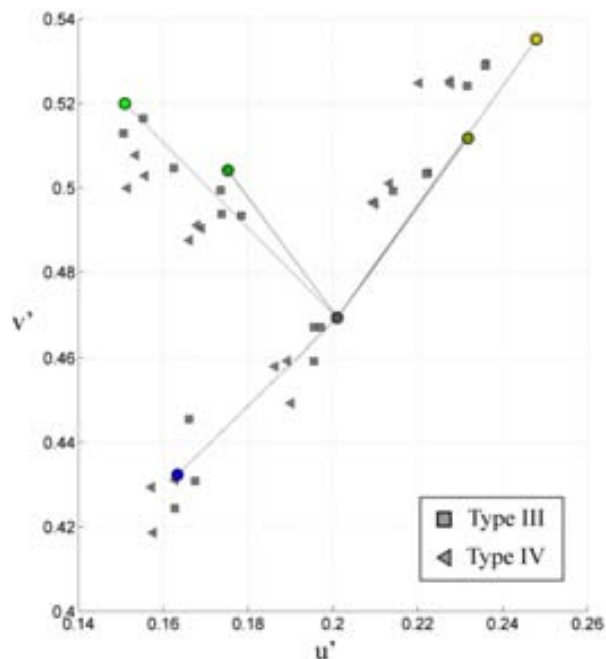
Figure 6.8 Coordinated of chromatic settings corresponding to "grey" in CIE 1976 uv colour space for all observers and background/illuminant conditions; squares and circles correspond to background type III and IV. Coloured circles indicate the illuminants location and are colour coded according to the illuminant *colour*.

We quantified the extent of colour constancy for each chromatic setting computing the Brunswick ratio between measurements done under the reference (D65) and test (the rest) illuminants. Computations followed the same approach as described in subsection 4.4.3 and Equation 4.1. Figure 6.9 shows these results according to colour categories (x-axis), background (panel) and illuminants (coded coloured lines). The overall average value was 0.83 (σ=0.17) and thus confirming a high degree of colour constancy. Differences in index values were studied according to backgrounds and no differences were found, also according to illuminants ($F_{(4,265)}$=2.99 and $p<0.05$) with the only significant difference between the greenish (mean=0.87, σ=0.19) and the yellowish (mean=0.77, σ=0.16) illuminants. To study differences according to colour categories we averaged index values over backgrounds and illuminants and significant differences existed ($F_{(8,81)}$=25.68, $p<0.001$), being

the highest for purple with a value of 1.07 (σ=0.11) and the lowest for yellow with a value of 0.67 (σ=0.05).
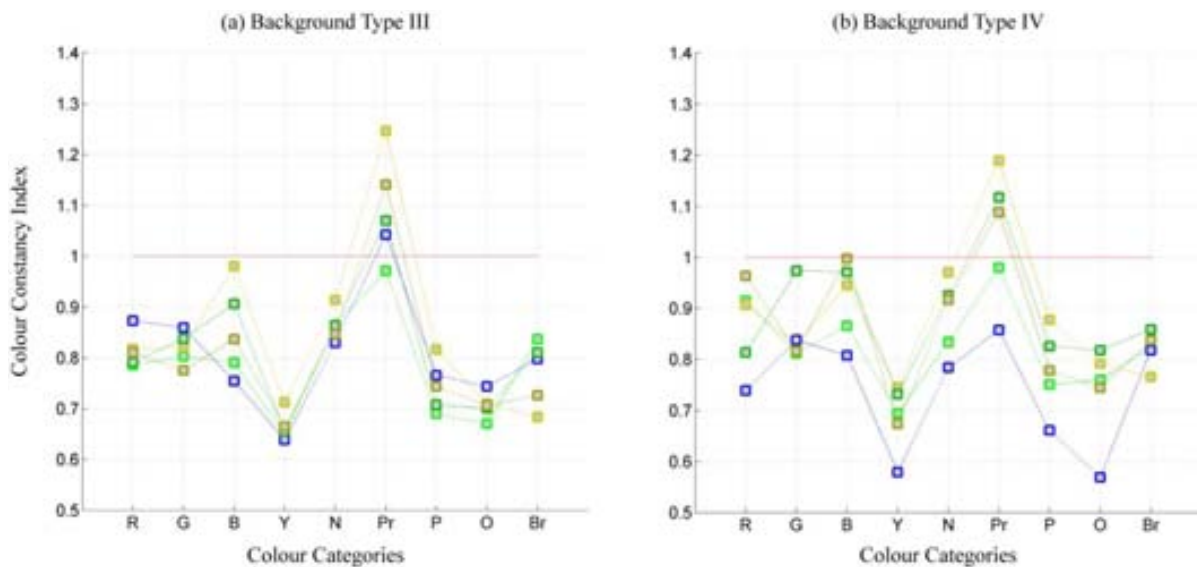
Figure 6.9 Brunswick ratio computed from chromatic settings of Experiment IV. Results are shown discriminated by colour categories and illuminants. Panels a and b correspond to background types III and IV respectively.

The Brunswick ratio values over colour categories were broadly uniform but not completely, the variation was still a 20% (coefficient of variation: 0.17/0.83=0.20), something that was in agreement with the hue distortions reported in subsection 6.4.5. Furthermore, it allowed the possibility to explain for the categorical results reported in subsection 6.4.4.

Next we quantified the extent of categorical colour constancy. We studied the relationship between background colours under the reference illumination and the "same" colours once adapted from a test illumination. To do so, we predicted the coordinates of background colours after adaptation using a linear model of colour constancy, as described in subsection 4.4.4. We used the chromatic settings between the test and reference illumination to fit the coefficients of an affine matrix, then it predicted the LMS coordinates of adapted colours under the reference illumination.

Figure 6.10 shows this approach where columns and rows correspond to observer and background type respectively. In each panel, the adapted colours are coded with a different symbol according to which were its test illuminant and, linked to the reference location with a black line. Coordinates shown are in CIELab a*b* plane with the reference illuminant as a reference white point. If adaptation was complete then different illuminant-symbols would be in the same location, but notice how the

distribution of adapted colours around the expected location under the reference illuminant (circle) is not uniform. We quantified the magnitude of this observation from ΔE^* distances between the reference and test illuminant locations. Table 6.8 contains its values observers averaged according to illuminants and backgrounds.

Illum / Colour	Type III			Type IV			Mean
	C_1	C_2	C_3	C_4	C_5	C_6	
Greenish	5.68	15.58	13.76	8.68	18.45	12.66	12.47
Yellowish	5.07	13.51	13.31	7.79	14.60	10.84	10.85
Bluish	5.12	5.47	8.47	3.42	4.29	8.98	5.96
Greenish unsaturated	5.77	7.94	8.04	4.02	7.63	6.75	6.36
Yellowish unsaturated	3.93	8.34	7.51	4.67	7.60	7.48	6.59
Mean	5.11	10.16	10.22	5.72	10.51	9.34	8.51

Table 6.7 Distance in ΔE^* between reference and adapted background colours. This table contains observer averaged distances observed in Figure 6.10 (black lines), see text for details.

Previously we observed the tendency that distortion grows among chromatic settings according to illuminant saturation, however Table 6.8 reveals another pattern; C1 and C4 colours keeps the same distance to the reference background colour (5.11 ΔE^* and 5.72 ΔE^*) despite the adapted illuminant. That fact suggests that colours C1 and C4, which corresponds to the green category, could have influenced the categorical perception under the adaptation process and thus inhibiting/constraining the existence of other colour categories.
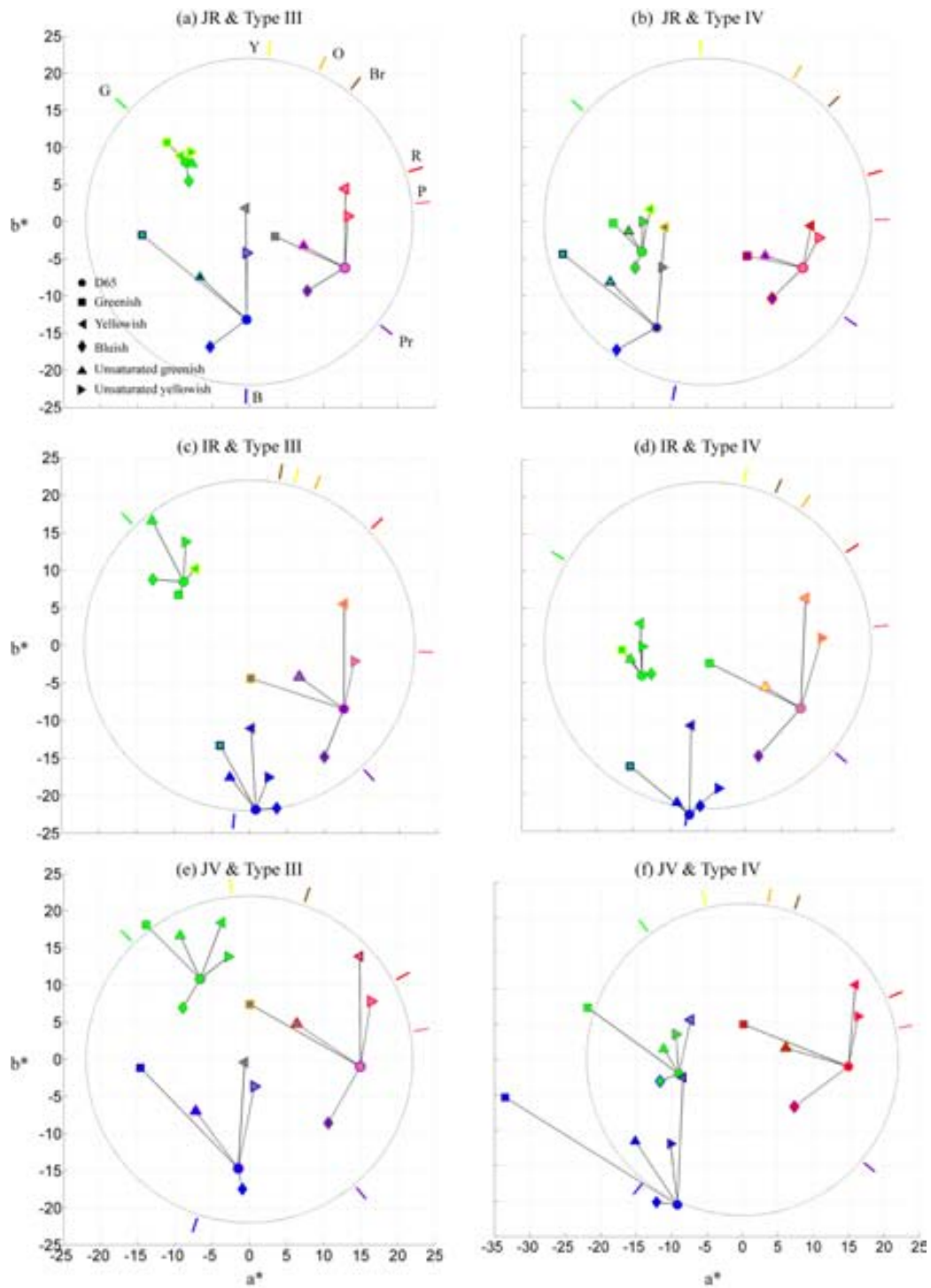
Figure 6.10 Prediction of adapted background colours under test illuminants . They are colour coded with its compound colour naming classification from the colour naming task. Each panel corresponds to one observer and background type. Illuminant are marker coded. Coloured lines outside the circle indicate hue of D65 chromatic settings. Solid squares indicate the SR location under D65 and Nameable background. The doubled coloured squares indicate the location of displayed colours under D65 and are colour

132

Each panel in Figure 6.10 has a set of coloured lines close to the circle, they indicate the hue of the chromatic settings under D65 accordingly to each observer in each panel. Visual inspection reveals the consistency of the chromatic settings with colour naming results, i.e., in general all hues are aligned with their categorical codifications. The validity to use the chromatic settings for a colour naming model is assessed in the subsection 6.5.1.

6.4.5 Simulating categorical colour perception on the achromatic axis

The performed colour naming task only provided categorical information for three points in colour space under each adaptation state and its results demonstrated that categorical colour constancy was not complete. We are interested into expand this categorical information to untested regions of colour space and to do so we used the categorical colour prediction model introduced in subsection 6.3.3. However, chromatic settings are the categorical parameters of the model and so the reliability of its predictions rely on their proximity to the chromatic settings. Thus we restricted the analysis to the achromatic axis with its lightness range limited to the chromatic settings one. Also only adaptations under background type III were studied.

Figure 6.11 shows the categorical colour predictions of the achromatic axis under the reference, test and adapted test illuminations. The achromatic axis was sampled by seven points according to each adaptation stated, i.e., points were selected to be equally spaced in the range defined by chromatic settings lightness. In each panel, the first column contains the categorical colour prediction of the achromatic points under the reference illumination (sorted according to lightness from lower-bottom to higher-top) and then follows five pairs of columns each corresponding to the categorical prediction of illuminated achromatic points and adapted achromatic points, both under the reference illumination. Thus, categorical colour perception was assigned according to the D65 one, i.e., the categorical colour prediction model (parameters: 0.25, 0.5 and 20) used the chromatic settings adjusted under D65 illumination. Notice that procedure was adjusted to each observer and its particular adapted state under a given illumination and background type. Colour codification used in Figure 6.11 follows the same rule than Figure 6.10.
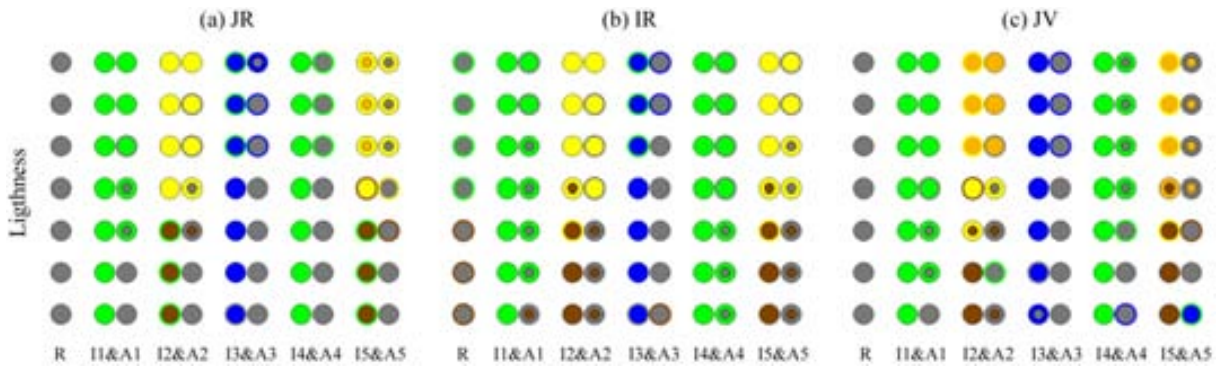
Figure 6.11 Prediction of categorical colour perception of the achromatic axis under reference, test and adapted test illuminations. Each panel corresponds to one observer using the adaptation measured under background type III. Each round marker is colour coded according to the compound basic terms classification: only one colour indicates one category, one colour with the border outlined indicates basic + basic adj and two concentric circles indicates the case basic + basic. Column indicated with R is colours seen under ref illum, then Ii corresponds to colour seen under i illum but categorized using d65 adapt and Ai are the categorization done on the adapted colours. See text for details.

We quantified how categorically constant are the predictions displayed in Figure 6.11 from the similarity index introduced in subsection 6.3.2. More precisely, we computed the similarity index between the prediction under the test illuminant and the reference illuminant for each pair of predictions corresponding to the same luminance points (that means between each lightness corresponding point of columns R and Ai). Table 6.8 contains these values averaged over all three observers and it reveals two tendencies. First, a lower similarity index for illuminants with higher saturation and second, a higher similarity index for lower lightness points. The latter probably due to the physics of illumination where darker colours are less influenced by illumination change and so they help constancy.

Colours / Illum	Greenish	Yellowish	Bluish	Greenish unsaturated	Yellowish unsaturated	Mean
Achromatic 1	0.25	0.00	0.79	0.79	0.54	**0.56**
Achromatic 2	0.25	0.08	0.83	0.79	0.54	**0.58**
Achromatic 3	0.33	0.12	0.87	0.79	0.66	**0.63**
Achromatic 4	0.58	0.54	0.96	0.83	0.71	**0.77**
Achromatic 5	0.66	0.79	0.96	0.79	0.87	**0.85**
Achromatic 6	0.75	0.92	0.96	0.83	0.96	**0.90**
Achromatic 7	0.96	0.87	1.00	0.79	0.62	**0.87**
Mean	**0.54**	**0.48**	**0.91**	**0.80**	**0.70**	**0.69**

Table 6.8 Similarity index computed between categorical predictions on reference and adapted achromatic colours

. Each row corresponds to one achromatic colour and are sorted from bottom to top into lower to higher lightness. Each column corresponds to the adaptation under the column header illuminant. Coefficients has been observer averaged. Notice that the similarity index acts as a categorical colour constancy index. The considerable number of repeated coefficients is due to the non continuous nature of the similarity index coefficients. See further details in text.

Interestingly the previous categorical predictions suggest large failures of categorical colour constancy for the most saturated illuminants, the greenish and the yellowish. For the other three illuminants the degree of categorical colour constancy is much higher, similar to the colour constancy indices reported before, however the categorical colour constancy is only completed under the bluish illuminant. The overall categorical changes observer are in agreement with the Helson-Judd effect (Derefeldt and Swartling).

Figure 6.11 is a visual approach to the hypothesis stated in the chapter introduction, i.e., which aim is priority either to recover reference colours or either to keep categorical colour perception of illuminated colours. In order to acquire a formal approach to this question we computed the same similarity index as in Table 6.8 but this time instead of using adapted colours we used the colours under the test illuminations. Table 6.9 reveals opposite tendencies to Table 6.8; saturated illuminants had a higher degree of categorical constancy than unsaturated ones and, colours with higher lightness had a higher degree of constancy.

Colours / Illum	Greenish	Yellowish	Bluish	Greenish unsaturated	Yellowish unsaturated	Mean
Achromatic 1	0.96	0.96	0.25	0.58	0.62	**0.67**
Achromatic 2	0.96	0.87	0.12	0.58	0.62	**0.63**
Achromatic 3	0.87	0.83	0.08	0.58	0.58	**0.59**
Achromatic 4	0.79	0.58	0.00	0.54	0.37	**0.46**
Achromatic 5	0.75	0.50	0.00	0.29	0.25	**0.36**
Achromatic 6	0.50	0.25	0.04	0.25	0.25	**0.26**
Achromatic 7	0.00	0.50	0.33	0.25	0.25	**0.27**
Mean	**0.69**	**0.64**	**0.12**	**0.44**	**0.42**	

Table 6.9 Similarity index computed between categorical predictions on reference and illuminated achromatic colours. The table description follows the same rationale as Table 6.8.

Notice how averaged results from Table 6.8 and 6.9 are inversely correlated, in broad terms. Take for instance similarity values under the green illuminant, 0.54 to the reference colours and 0.69 to the illuminated colours, these values indicate that adapted categorical colours were more similar to the illuminated ones than to the reference ones. From the global contrast property we can expect that adapted colours will lie close to the regions lead by the joining line between the reference and test illuminant coordinates, then previous results makes sense because somehow all categorical predictions

lie in this line and so values needs to be in overall inversely correlated (as it is). Something that also reinforces the validity of the approach of simulation on the achromatic axis. We will discuss further this point in the next subsection.

6.5 Discussion

6.5.1 Accuracy of the categorical colour prediction model

We studied the accuracy of the categorical colour prediction model (see subsection 6.3.3) according to its parameters: an achromatic threshold value, two categorical threshold values and two sets of chromatic settings measured under the test and reference illuminations. We used two tests, first we evaluated how well categorical data obtained from the naming task is predicted and second we compared categorical predictions between our model and fuzzy sets models of colour naming.

Our first test obtained the categorical colour prediction for each adapted colour (Markers of Figure 6.10) and computed the similarity index between this prediction and its corresponding colour naming result. We used the similarity index to address the quality of the prediction and it had an average (over all conditions) value of 0.79 (σ=0.16) when using the parameters values t_{achr}=12, t_1=0.25 and t_2=0.5. Next we studied the influence of t_1 and t_2 parameters to this quality index. We considered a sampling of the 0< t_1< t_2<1 parameter space in 0.01 steps and computed for each t_1, t_2 pair their related quality index. Panel b of Figure 6.12 shows the surface resulting from considering the function where for each pair of parameters obtains its quality index. Surface values range between 0.66 and 0.80 and its average value is 0.75 (σ=0.03) corresponding to (t_1,t_2) = (0.3,0.4). Also, notice the smooth properties of the surface, i.e., it is continuous and differentiable, and it has only one local maximum thus supporting the consistency of the approach. The high values of the similarity index demonstrated that the prediction method succeeded at least for the colour terms measured. However due to the small number and limited colour space location of tested colours we will not rely on threshold values for predictions of subsection 6.4.5 and use the more canonical values of 0.25 and 0.5.
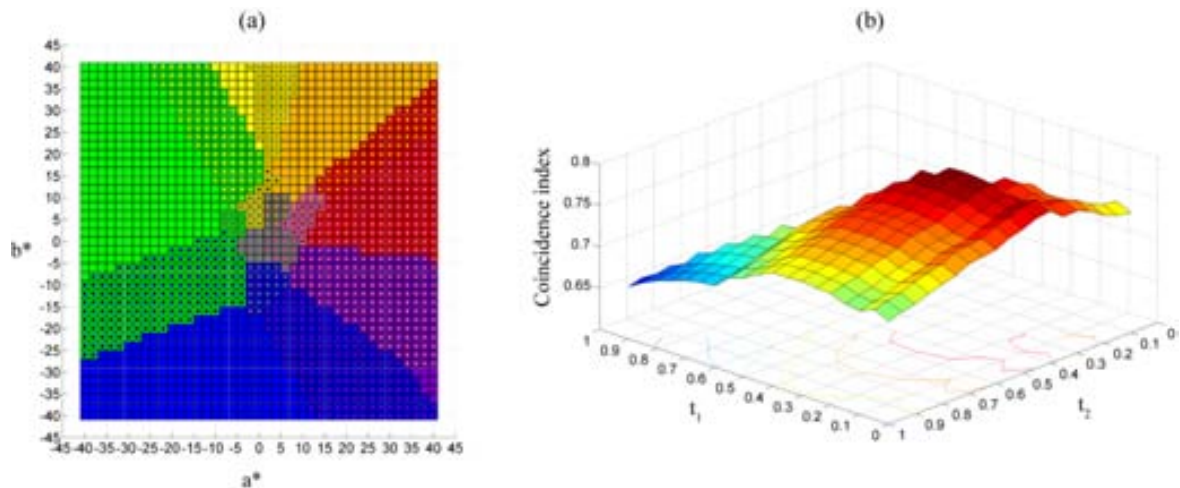
Figure 6.12 Accuracy of the categorical colour prediction model

according to its conforming parameters. Panel a: Section of CIELab colour space at L*=60 with its categorical model predictions colour coded as usual, for one observer. Notice how it reveals categorical transition in the plane as well as in the lightness dimension. Panel b: Surface relating threshold values to the coincidence index between adapted points predictions and its recorded colour naming results. Results are averaged over all conditions and observers.

Our second test studied the shape of categorical regions generated by the categorical model, according to its definition the shape is clearly radial from the adjusted grey. Due to the limited extension of chromatic settings in colour space only makes sense to study the inner part of colour space, this is the corresponding region limited by the chromatic settings lightness, saturation and hue ranges. Panel a of Figure 6.1 shows a cross section of categorized CIELab colour space at lightness L*=60 for one particular observer (IR). To render this figure we sampled the inner part of CIELab colour space in steps of 2 units (JND) for all three dimensions and categorized applying the categorical colour prediction model. Visual inspection of panel a confirms the consistency with previous approaches (Lammens 1995; Hansen, Walter et al. 2007; Benavente, Vanrell et al. 2008) containing the same basic features: large regions for basic colour terms, large border regions for green-blue, green-yellow and blue-purple, and limited round region for grey. As expected by model definition most of the category boundaries are represented by straight lines, something coincident to previous approaches. However this is clearly an oversimplification from what we have found in Chapter 3, an artefact created by the necessary simplicity of the model.

It is interesting to have recovered this particular shape with such a simple algorithm, something that challenges more complex approaches as the fuzzy sets (Lammens 1995; Hansen, Walter et al. 2007; Benavente, Vanrell et al. 2008).

137

6.5.2 Categorical inconstancy in successive colour constancy

Here we summarize previous results in terms of categorical changes occurred from changes of illumination. The number of colours is a visual cue for colour constancy mechanisms as demonstrated by previous works and seen from our own results, however in that case we do not study the amount of colour constancy but the categorical influences to them.
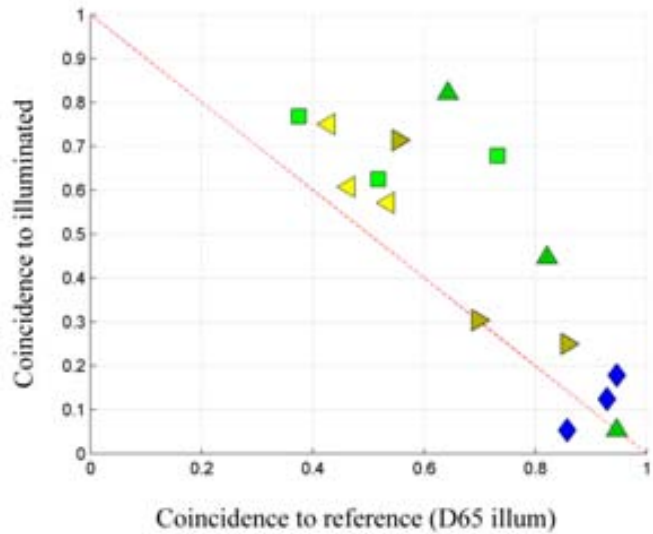
Figure 6.13 Coincidence to illuminated and reference colours.

Chapter 7 Conclusions

This chapter summarizes the main findings of previous chapters. In particular it summarizes the previous four experiments and their results and conclusions. Furthermore the results are linked from one chapter to the other and a the principal aim stated in the first chapter confronted. Also, as in the first chapter, the conclusions are stated in terms of possible contributions to the Computer Vision field. Finally future lines of research are proposed in visual perception as well as in computer vision.

7.1 Summary of findings

7.1.1 Chapter 3: Existence and particularities of colour categories

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.

7.1.2 Chapter 4: Colour categories under illumination changes

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.

7.1.3 Chapter 5: Structural constancy of colour categories

We have collected information on the perceptual interrelations of coloured surfaces under illuminant changes and modelled our measurements using graphs. Our results show that these interrelations remained 87% constant under an illumination shift, in contrast with the structural deformation undergone by the physical colours. This suggests that categorical perception may be used to guide colour constancy adaptation. This result is in accordance with previous studies, (Hansen, Walter et al. 2007; Olkkonen, Hansen et al. 2009), that reported categorical stability using colour naming techniques. Based on our previous results, we could state that categorical colour perception maintains a high degree of structural invariance under illuminant changes. It is not perfect. Remains some 20% of changes that still allow to border colours to change of categorical perception.

7.1.4 Chapter 6: Categorical changes under different illuminants

Experiment IV used the chromatic setting to measure differences in the categorical colour structure under different illuminants and non categorically balanced backgrounds. Results indicated a clear differences with those of Experiment II, i.e., two opposite rotations appeared to be operating in the chromatic settings hue. Thus the categorical information included in the background influenced the final categorical perception. In order to increase our knowledge about categorical changes we developed a colour naming model from the chromatic settings, the model allowed to predict categorical colours according to each particular observer and adaptation state. We validated the model using several complementary techniques and applied it to unveil categorical changes occurred for non measure colours for the constrained naming task. Our first approach was to model categorical changes in the achromatic axis and doing so we recovered the Helson-Judd effect. Results confronted the two initial hypothesis, is adaptation trying to keep categorical perception and so the categorical perception of initial colours influences the adaptation or is the possibility to recover the canonical colours under a reference illumination leading the adaptation? We have gathered support for both approaches according to the illumination used and thus indicating that the region of colour space where colours fall after illumination could constrain the process of global adaptation at certain noticeable degree.

7.2 Contribution to Computer Vision

Our results indicated the overall structural stability of categorical colour space, i.e., the stability of inter distances of chromatic settings under a change of illuminant. This structural stability was exploded by the computational colour constancy algorithm developed by Vazquez et al (Vazquez-Corral, Vanrell et al. 2011), which estimated the scene illumination according to its potential to map colours into parameterised regions of colour space, ideally regions corresponding to the basic colour terms.

However if to recover images with constant colours is the primary aim of this algorithms may be they should incorporate other approaches than to recover physical information. Recently neuroscience and in particular computational neuroscience introduced new approaches to solve the colour constancy problem. We know that colour constancy mechanism are divided into several levels in the HVS, cortical ones are still being delineated and still difficult to simulate (only functional attempts) but sensory levels are much further understood. Vanleeuwen et al (Vanleeuwen, Joselevitch et al. 2007) demonstrated feasible computational replications of the outer retina to model a large part of the colour constancy. This suggest clear path to solve the colour constancy just mimicking what the HVS does.

However, now it is still difficult to simulate the behaviour of neurons and their synaptic connexions in the large scale that such mechanisms would involve. As highlighted by Lindblad and Kinser (Lindblad and Kinser 2005) the behaviour of neurons is not linear so it is impossible to replicate its behaviour from only sums and products, then neurons should be modelled as dynamical systems (Izhikevich 2007) which allow a neuron to modify its behavioural pattern according to different inputs. However this approach is far but some interesting applications has been implemented in computer vision (Lindblad and Kinser 2005). Now it is not only possible to simulate this in terms of functional features of the mechanisms but also to simulate precisely from a close position to the actual physiology of neurons which has a behaviour of dynamical systems and it is far more complicated than computer vision traditional approaches. So a new field is claiming his own right to interact with the colour constancy problem, adding to the psychophysics.

7.3 Further work

Previous conclusions lead to the need to further work in order to clarify/increase our knowledge on colour constancy. Regarding visual perception the main lines of future research are:

- Get a reliable tool to quantify the magnitude of colour constancy, most probably by developing multidimensional indices to summarize the phenomenon.
- Get an empirical review of colour constancy; up to now there are only qualitative reviews of the phenomenon. To gather all empirical data, from different experiments and paradigms into on single colour space, then define a vector field that summarized this information and model it using differential equations. In this way the empirical behaviour of the phenomenon would be best summarized.
- To explore more contextual properties through the chromatic setting paradigm, it allows to measure precisely nine points in colour space as well as to get categorical information at the same time. In particular to study different geometrical distributions of colours in the background under a more extensive set of illuminants.

Also our conclusions allows to propose further work in the computer vision field:

- Most illuminant estimation approaches estimate the illuminant chromatic coordinates in a 3D colour space and so they use after a diagonal matrix to correct the image (Ebner 2007). From conclusions in chapter 4 we know that also other affine models can be used to increase the colour constancy modelling, also that our own non reported studies confirm.

- Colour constancy phenomenon has been divided into several mechanisms and neural levels in visual perception. Perhaps the computer vision approach should use this functional division to design two step (or more) colour constancy algorithms.

Finally highlight that neuroscience is increasing its influence and has become a new scientific field in its own among the older ones as psychology, Colorimetry, physics, computer vision, visual perception, etc. From its approaches we can attempt to simulate the precise behaviour of neuronal nets and in particular to simulate the behaviour of cones in the retina. Previous stated lines of future research should be contextualized in this emerging scientific field.

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 0.1 CIELab coordinates of light samples in Experiment I.

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 0.2 CIELab coordinates of surface samples in Experiment I.

Appendix B: Shape of ACNS regions in Experiment I
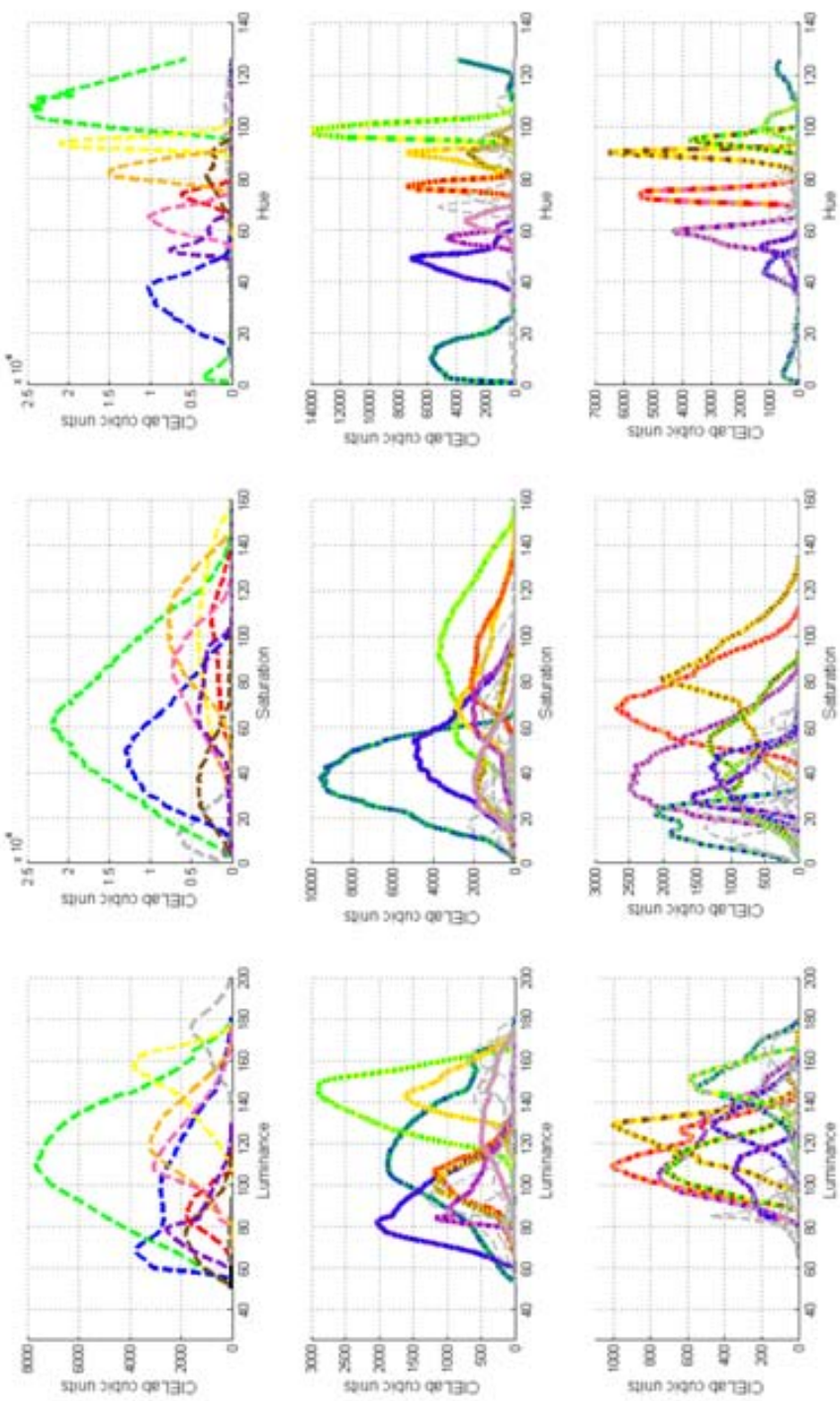
Appendix C: Chromatic setting training in Experiment II

Subjects were systematically trained in a two sessions loop procedure which lasted between one and two hours according to subject ability. Each loop consisted in two sessions: a reference session followed by a regular session with Type 0 background and D65 illumination. Notice that these sessions only differed regarding to the existence or not of the bounding cylinder. The conclusion that one particular subject was trained followed from two points: (a) Understanding of the chromatic setting methodology of first selecting and then reproducing colours and (b) Ability to reproduce the same colour at certain degree.

After the first loop, the first question was readily answered from the difference between corresponding categorical adjustments of both sessions: if the second session adjustments were at half distance more than the cylinder and the CRT gamut then the subject was enforced to do a second training loop. The second question was answered looking at the average distance between the adjustments and their mean for each colour category: if they were approximately uniform over categories and if they were similar to the previous tests performed by the authors (this is 4 DeltaE units).

All subjects undergone the training procedure except for the authors which were already trained from previous test experiments. Table X contains the training sessions details and demonstrates that the training procedures succeeded for all reported subjects. Explain in more detail its values. It shows the improvement of subject ability to set colours concluding that initial training sessions contributed to reduce the StDev by about 1.33 ΔE^* and the average time duration of a single run was reduced by 8.25 seconds. Also, subjects keep the trend to higher the precision as the experimental sessions went along, however this did not happen with time which stabilized readily after the training sessions.

Observer	Loop number	Precision (DeltaE*)			Time (seconds)		
		First session	Last session	Diff	First session	Last session	Diff
MS	1	3.27	2.28	0.99	31.76	20.76	11
XO	1	4.66	2.54	2.12	26.90	17.72	9.18
RB	1	8.79	4.19	4.6	37.68	15.05	22.63
LC	2	5.04	4.31	0.73	29.31	16.02	13.29
AB	2	5.21	3.68	1.53	22.78	16.34	6.44
RBV	2	3.52	4.19	-0.67	19.63	22.73	-3.1
JC	1	4.32	4.29	0.03	16.34	18	-1.66
Mean	-	4.97	3.64	1.33	26.34	18.08	8.25

Table C.0.1 Observers training results in Experiment II.

Appendix D: 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.
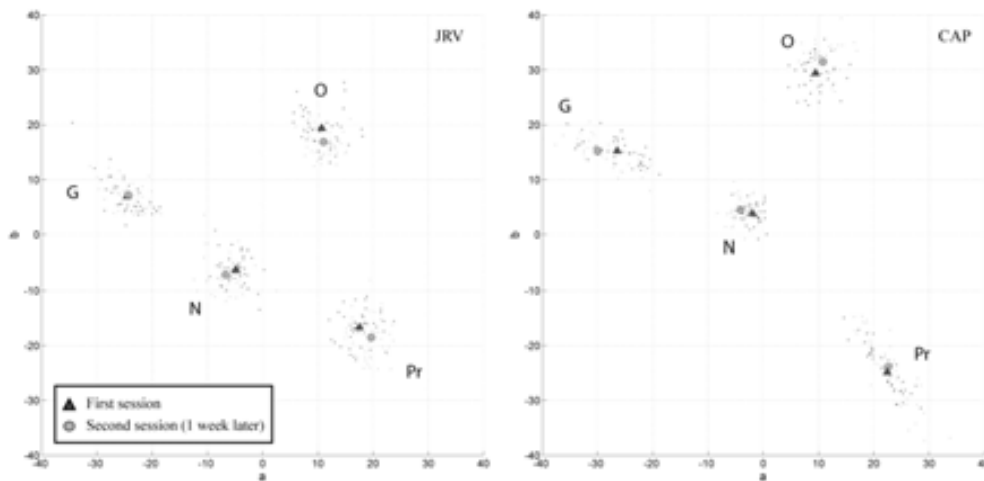
Figure D.0.1 Results of the long-term memory control experiment for two subjects. Four categories were tested (40 trials each). Dark and light dots were measured with a 7-day time difference. Averages are represented by triangles (first session) and squares (second session). The results clustered near the origin, are equivalent to those of a typical achromatic setting experiment.

Appendix E: 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).

Greenish	Yellowish

Model	3n	k	RSS	AICc	Δi	wi	RSS	AICc	Δi	wi
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 6. The 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: AIC_c, Δ_i and w_i. 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.

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.

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.
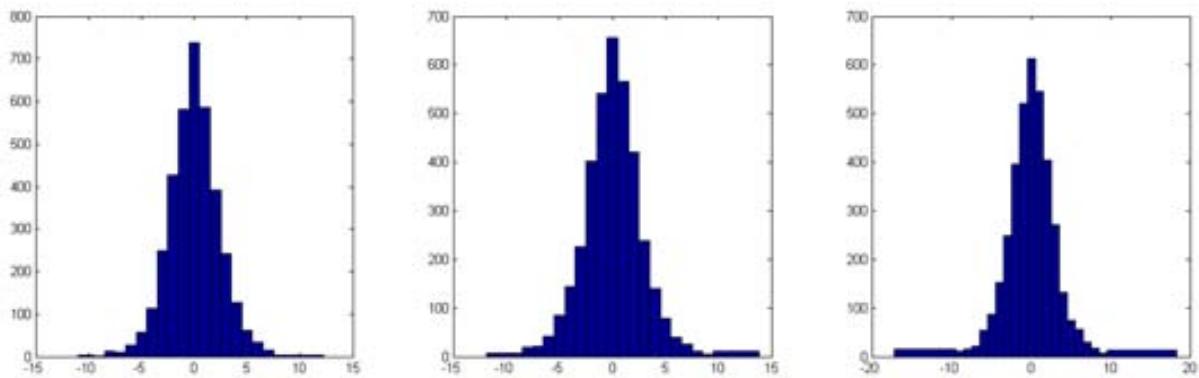
Figure E.0.1 Distribution of the Chromatic setting error measures of each CIElab dimension. Each histogram contains data from all results of Experiment II. This results demonstrates the hypothesis of AIC criterion. Graphs shows clearly how all follow a normal distribution.

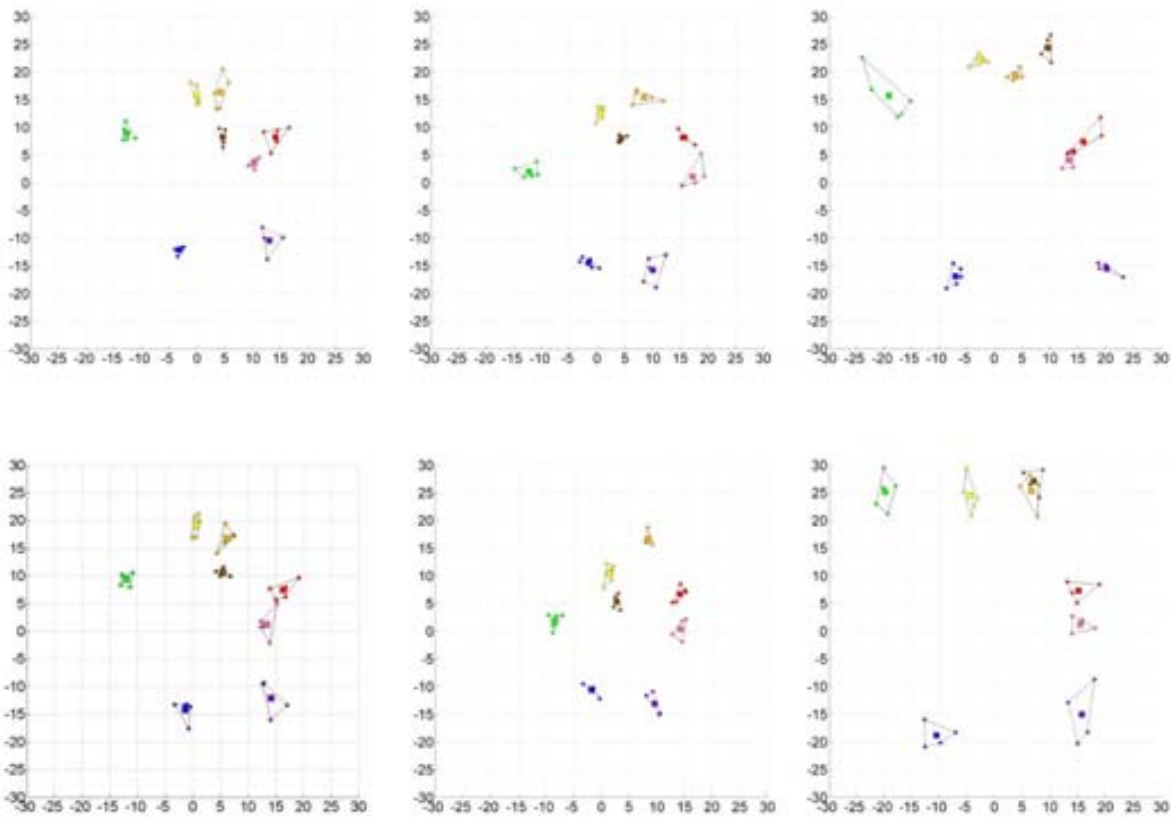# Appendix F: Repeatability of chromatic settings in Experiment IV

Figure F.0.1 Chromatic settings repeatability in the regular sessions under D65 illumination and background type x.

3.1 ΔE^* units (1.3 std).

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., C. A. Parraga, et al. (2009). "Colour categories boundaries are better defined in contextual conditions." Perception **38**(Suppl.): 36.

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. I. Asymmetric matches." Journal of the Optical Society of America A **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.

Derefeldt, G. and T. Swartling (1995). "Color Concept Retrieval by Free Color Naming - Identification of up to 30 Colors without Training." Displays **16**(2): 69-77.

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

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

Fairchild, M. D. (1998). Color appearance models. Reading, Mass. ; Harlow, Addison-Wesley.

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 TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE.

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.

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.

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.

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

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." 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.

Izhikevich, E. M. (2007). Dynamical systems in neuroscience : the geometry of excitability and bursting. Cambridge, Mass. ; London, MIT Press.

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." Annu Rev Psychol **40**: 1-22.

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 USA **96**: 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.

Lindblad, T. and J. M. Kinser (2005). Image processing using pulse-coupled neural networks. Berlin ; New York, Springer.

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.

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

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.

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.

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. 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." Philosophical Transactions of the Royal Society B-Biological Sciences **360**(1458): 1329-1346.

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.

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.

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.

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.

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.