Colour Communication
Within Different Languages

Dimitris Mylonas

Department of Computer Science
University College London

This Thesis is Submitted for the Degree of
Doctor of Philosophy

December 2019

Declaration

I, Dimitris MYLONAS, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Dimitris Mylonas

December, 2019

Abstract

For computational methods aiming to reproduce colour names that are meaningful to speakers of different languages, the mapping between perceptual and linguistic aspects of colour is a problem of central information processing. This thesis advances the field of computational colour communication within different languages in five main directions. First, we show that web-based experimental methodologies offer considerable advantages in obtaining a large number of colour naming responses in British and American English, Greek, Russian, Thai and Turkish. We continue with the application of machine learning methods to discover criteria in linguistic, behavioural and geometric features of colour names that distinguish classes of colours. We show that primary colour terms do not form a coherent class, whilst achromatic and basic classes do. We then propose and evaluate a computational model trained by human responses in the online experiment to automate the assignment of colour names in different languages across the full three-dimensional colour gamut. Fourth, we determine for the first time the location of colour names within a physiologically-based cone excitation space through an unconstrained colour naming experiment using a calibrated monitor under controlled viewing conditions. We show a good correspondence between online and offline datasets; and confirm the validity of both experimental methodologies for estimating colour naming functions in laboratory and real-world monitor settings. Finally, we present a novel information theoretic measure, called *dispensability*, for colour categories that predicts a gradual scale of basicness across languages from both web- and laboratory- based unconstrained colour naming datasets. As a result, this thesis contributes experimental and computational methodologies towards the development of multilingual colour communication schemes.

Impact statement

In today's global communication environments, understanding how people name colours is important for those wishing to develop effective digital image-related technologies. For example, there is growing interest in using colour naming data to improve applications in data visualisation, human-computer interaction, colour appearance modelling and e-commerce.

Since 2009, we have led an international collaborative project to collect unconstrained colour names with their corresponding colour ranges through an online experiment with thousands of observers in tens of languages (accessible at https://colournaming.org). This ongoing research is endorsed by the International Colour Association through its Study Group on the Language of Colour, and has attracted notable media attention, including articles in the Economist, New Scientist, United Press International and the Metro newspaper. We replicated the online experimental methodology in laboratory conditions to map, for the first-time, unconstrained colour names in the physiologically-based cone excitation space adopted recently by the Commission Internationale de l'Éclairage (CIE).

We developed robust computational tools trained by these multilingual datasets to automate the colour naming task across the full colour gamut. Through discussion with key companies in e-commerce, we have identified specific gaps in current capability with immediate applicability for our tools. We also successfully completed an EU-funded Short-Term-Scientific-Mission in Spain that made an advance towards the establishment of a procedure to improve the visualisation and accessibility of cultural heritage materials. Educational activities include the design of a colour card game, called *Colours of Babel*, which raised its funding goal in a crowdfunding campaign. The game has been played with university students as well as with the public in widening participation events in United Kingdom, Japan, South Korea and in United States. People can also engage on a day-to-day basis with our research through *Colournamer*, a web application where users can currently learn common colour names in seven languages.

In addition to practical applications of our research, the dataset we are constructing, together with the analysis tools, allows us to explore basic issues about how colour

names are arranged within different languages. Preliminary research outputs of our national and international collaborations have been presented in conferences, as book chapters and journal articles while more journal publications are in sight. Indirect academic outputs include two journal articles with Prof. Semir Zeki, FRS, on categorical colour constancy for which I was invited to give a talk at the Design Museum, London. Furthermore, a new international collaboration with researchers from Goldsmiths, University of London and University of Nimes (France) received a BA/Leverhulme Small Research Grant to adopt our experimental methodology for measuring colour naming distributions of indigenous people in Namibia in 2018. As a result, this thesis has application, as well as, theoretical-oriented impact.

Acknowledgments

Contents

Figures

Tables

Nomenclature

Roman Symbols

B	Number of trees
b	Tree index
c	Colour
i	Index
n	Name
P	Probability
r	Pearson correlation
T	Transpose
w	Word
$\tilde{x}$	Test colour sample
$\hat{y}(\tilde{x})$	Prediction of model given colour sample $\tilde{x}$

Greek Symbols

Δh	Hue angle difference
ΔE	Euclidean distance
ΔE_{00}	CIE ΔE 2000 Colour difference formula
λ	Wavelength of a sine wave
μ	Mean
φ	Nonlinear function

Acronyms / Abbreviations

2-D	2-Dimensional
3-D	3-Dimensional
AI	Artificial Intelligence
AIC	International Colour Association
Am	American
API	Application Programming Interface
AVA	Applied Vision Association
BA	British Academy
BC	Bhattacharyya
BCE	Before the Common Era
BCT	Basic Colour Terms
Br	British
C	Chroma
CAM	Colour Appearance Model
cd	Candela
CI	Confidence Interval
CIE	International Commission on Illumination
CNS	Colour Naming System
CRT	Cathode-Ray Tube

D	Daylight
DKL	Derrington-Krauskopf-Lennie space
En	English
EU	European Union
fMRI	Functional Magnetic Resonance Imaging
FRS	Fellow of Royal Society
Gr	Greek
H	Hue
HSL	Hue-Saturation-Lightness
IEC	International Electrotechnical Commission
IT	Inferior Temporal
LCD	Liquid-Crystal Display
LGN	Lateral Geniculate Nucleus
LMS	Long, Medium and Short
MAP	Maximum a Posteriori
MAP	Mean Average Precision
MB	Macleod-Boynton
MRD	Munsell Renotation Dataset
MSRD	Munsell Spectral Reflectance Data
NN	Nearest Neighbour
OSA	Optical Society of America
RF	Random Forest
RGB	Red-Green-Blue
RMS	Root Mean Square
RST	Rotated Split Trees
RT	Response time
Ru	Russian
SD	Standard Deviation
SQRT	Square Root
TFT	Thin-Film-Transistor
Th	Thai
Tu	Turkish
V	Value
V1, V2, V4	Visual areas 1, 2, 4
WCS	World colour survey

Chapter 1

Introduction

Colour naming describes the intriguing cognitive capacity to organise millions of discriminable colours into a smaller set of colour categories named, for example, as yellow, navy blue and dark olive green (Pointer & Attridge, 1998). Colours named in one language are given different colour names in other languages. Colour names vary across languages, lexically, in number and in range of reference. To augment colour communication within different languages, it is necessary to have a worldwide method for mapping perceptual to cognitive aspects of colour (Derefeldt, Swartling & Bodrogi, 2004).

The use of colour names in colour communication systems may seem inappropriate as they may convey different information to different speakers (Munsell, 1905). The advent of perceptual colour systems allows an unambiguous, numerical specification of colours, but the vast majority of people continue to use natural language in their everyday colour communication tasks, such as describing consistently the colour of a garment or a car in a continually changing visual world. People-centred design for scientific colour communication systems requires understanding of both the physical and the cognitive capabilities of the population it is addressed to (Bichard & Gheerawo, 2011). This need for better understanding calls for the development of new technologies through interdisciplinary research that will benefit each discipline and society as a whole. Colour naming has been the subject of many scientific disciplines including among others philosophy, psychology, physiology, linguistics, anthropology, and more recently computer science. Within computer science, existing computational colour naming methods essentially assume a universal system of a small number of colour categories but improved understanding of the differences between the large number of colour categories in each language of the world would improve their performance in interacting with their users (Evans & Levinson, 2009). Ultimately, to move towards the extension of human-artificial intelligence in the field of colour communication in multilingual environments, we will need to design data-driven colour naming systems without a priori theories of semantic universals that support the full complexity of colour languages across the world, including their relationship to physical, psychophysical and physiological aspects of colour. The work presented here extends previous research towards this direction.

In this thesis the field of colour communication within different languages is extended in five directions. Our work first focused on obtaining large colour lexicons in British and American English, Greek, Russian, Thai and Turkish and finding the distribution of colours for each name across the available colour gamut. We show that crowdsourced free colour naming methods offer a convenient and effective way to collect relevant data. Second, we compute linguistic, behavioural and geometric features for each common colour name and use a robust classifier to access the coherence of achromatic, primary and basic classes of colours based on these features. We show that members of achromatic and basic classes share coherent characteristics while this is not true for members of the primary classes. Third, we present a novel computational model that automates the colour naming task across the full three-dimensional gamut in different languages with excellent performance. Fourth, we locate unconstrained colour names in the physiologically-based cone excitation space adopted recently by the CIE through a calibrated laboratory-based experiment and show that both online and offline approaches produce consistent results. Fifth, we present a novel information theoretic measure that produces a gradual scale of basicness within different languages from both web- and lab- based unconstrained colour naming datasets. On the whole, this study aims to make the world a more colourful place by facilitating colour communication within different languages.

1.1.   Thesis Structure and Contributions

After the general introduction to this study presented here, in Chapter 2, we give a basic account of the underlying mechanisms and debates governing colour naming and the core components that this thesis is based on. This include a survey of colour spaces and colour difference measures, relevant colour naming experiments, an overview of existing colour naming models and information theoretic analysis in the context of colour language games.

In Chapter 3, we present an ongoing experimental methodology performed online for mapping multilingual colour names to colour coordinates for the first-time in British and American English (Mylonas, Purver, Sadrzadeh, MacDonald & Griffin, 2015; Griffin & Mylonas, 2016; MacDonald & Mylonas, 2016; Mylonas & MacDonald, 2017; Mylonas, MacDonald & Griffin, 2017; Griffin & Mylonas, 2019), Greek, Russian (Griber & Mylonas, 2015; Griber, Paramei & Mylonas, 2017; Paramei, Griber & Mylonas, 2018; Griber, Mylonas & Paramei, 2018), Thai (Katemake, Mylonas, MacDonald & Prasithrathsint, 2015) and Turkish (Ulusoy, Griffin & Mylonas, 2017). In the collection of our behavioural

colour naming data, we extend earlier cross-cultural studies which used only the most saturated colour samples (n=330) on the surface of the Munsell system (Berlin & Kay, 1969/1991; Kay, Berlin, Maffi, Merrifield & Cook, 2010; Lindsey & Brown, 2014), by also sampling (n=600) the interior of the colour solid. A further methodological improvement includes the departure from usual methods which would use a small number of observers and/or the use of only a restricted set of monolexemic terms (Berlin & Kay, 1969/1991; Boynton & Olson, 1987; Sturges & Whitfield, 1995). Instead, thousands of volunteers from linguistically and demographically diverse populations named freely a large number of colours online (Moroney, 2003; Mylonas & MacDonald, 2010; Munroe, 2010). We argue that participating in an online experiment in your own familiar environment, with your own equipment, and without the physical attendance of the examiner would give more ecological validity to the underlying categories responsible for colour naming (Reips, 2000). We also depart from previous research by taking the different colour names given by the observer in our online task to reflect a categorical distinction important to the observer; and in our data analysis, we will not use statistical procedures to look for similarities between given terms to summarise them into smaller groups (Lindsey & Brown, 2014).

In Chapter 3, we also consider families of linguistic, behavioural and geometric features. Previous studies have shown that as the length of a word increases, the frequency of its use decreases, due to communicative pressures (Zipf, 1935; Piantadosi, Tily & Gibson, 2011). Hence, our first linguistic measurement is the length of a colour name determined by the number of letters of each response in our online colour naming experiment (Brown & Lenneberg, 1954; Berlin & Kay, 1969/1991; Boynton & Olson, 1987). We will also count the number of derivative forms of each colour name, such as, 'greenish, 'greener, or 'pale green' (Berlin & Kay, 1969/1991; Kerttula, 2007). Our third linguistic measurement is the frequency of colour names in a million of Twitter messages as a measure of psychological salience (Hays, Margolis, Naroll & Perkins, 1978, Corbett & Davies, 1997; Mylonas et al., 2015). In our second set of features, we consider the behavioural measurements of frequency of occurrence, response latency and agreement (consensus) across observers in our colour naming experiment that provide an indication of easy and hard to name colours (Brown & Lenneberg, 1954; Boynton & Olson, 1987; Sturges & Whitfield, 1995). The third set of measurements considers the geometrical characteristics of the regions of colour space corresponding to each colour name. Here, we measure the mean location – called a centroid (Boynton & Olson, 1987) – of each colour category in terms of their perceptual attributes: lightness, chroma and hue, their size / volume (Mylonas & MacDonald, 2016) and their shape-sphericity (Basser & Pierpaoli, 1996; Gärdenfors, 2000).

In Chapter 4, we make the claim that if a subset of colours has a foundational role in the system of colour naming, then that will leave measurable marks in the properties of those colours when compared to other colours. We employ machine learning methods to discover criteria in the above families of features of colour names and access the coherence of proposed special classes of colours, including achromatic, primary and basic classes. For the primary class we report the coherence of the Hering primary colours (*black, white, red, green, blue* and *yellow*) and for the basic class we report the coherence of the Berlin & Kay's basic colour terms (Hering's primaries plus *purple, orange, pink, grey* and *brown*). For the achromatic class we consider only three terms (*black, grey* and *white*) to check whether smaller classes are necessarily less coherent because they have fewer examples from which to determine a membership criterion. Our findings provide evidence to substantiate the coherence of basic and achromatic classes, but we found no support for the coherence of any version of a primary class. The examination of the contribution of the families of features in the assessment of the best performing basic class showed that none of the three families is as good as all three families together. These results provide evidence against primaries playing a fundamental role in the development of colour categories and challenge explanations based on this claim (Kuehni, 2005; Philipona & O'Regan, 2006; Regier, Kay & Khetarpal, 2007).

In Chapter 5, we focus on algorithmic components needed for applications. For example, given the numerical coordinates of a sample in some colour space, what is the best name to describe it in multiple languages? We present the evaluation of a range of computational models trained by human observers of the colour naming experiment to automate the assignment of colour names across the full three-dimensional (3D) colour gamut of different colour spaces (Mylonas, Andrews & Griffin, 2016). Each method is assessed by cross validation and scored by root-mean-square (RMS) of Bhattacharya distances between observed and interpolated histograms of colour naming responses. A Rotated Split Trees (RST) approach performs best in automating the colour naming task. An evaluation of RST in several colour spaces (linear RGB, sRGB, CIEXYZ 1931, CIELAB, CIELUV and CIECAM02-UCS) showed that overall the algorithm performed best in CIELUV, in agreement with the reports of a recent study on colour clustering (Douven, 2017). Using these tools, we infer histograms of naming responses for any colour, and compute their entropy as a measure of naming variability. Our analysis revealed structure of easy and hard to name regions on the interior as well as on the surface of colour space that can be used to guide decisions for easy and hard to identify colour palettes in colour design (Mylonas & MacDonald, 2017). A key contribution of this

chapter is the training of our computational colour naming model with multilingual datasets in British and American English, Greek, Russian, Thai and Turkish to assign colour names to colours across the full colour gamut rather than using only a single language of existing methods (Lammens, 1994; Seaborn, 2005; Mojsilovic, 2005; Benavente, Vanrell & Baldrich, 2008; Weijer, Schmid & Verbeek, 2007, Heer & Stone, 2012; Parrage & Akbarinia, 2016). Our findings suggest that each language of the world should be approached in its own terms when automating the colour naming task, especially in the blue and green regions where some languages acquired prominent *turquoise*, *sky blue* and or *lime green* categories while others not.

In Chapter 6, we determine for the first time the location of unconstrained colour names within the physiologically-based cone excitation space (Stockman & Sharpe, 2000; CIE 170-1: 2006; CIE 170-2: 2015) through a laboratory-based experiment using a calibrated cathode-ray tube (CRT) monitor. The landmark colour names usually associated with unique hues (Boynton & Olson, 1987), *red* and *green* were not colinear with *white* but *yellow* was colinear through *white* with *blue*. *Red* was nearly complementary with *turquoise* and *green* with *magenta*. We also show that the loci of the basic colour terms obtained in the offline experiment are consistent with the loci of these names in the online experiment. Our findings support the validity of both, online and offline methods in estimating colour naming functions in laboratory and real-world monitor settings.

In Chapter 7, we focus on which colour names are shared and well comprehended among speakers in each language. We extend upon previous cross-cultural research, which used multiple questionable criteria for the identification of basic colour terms (BCTs; Berlin & Kay, 1969/1991; but see Crawford, 1982; Saunders & van Brakel, 1997; Levinson, 2000; Biggam, 2012), by contributing a simple, language-independent measure – called *dispensability* – that produces a graded scale of basicness from both web- and lab- based unconstrained colour naming data in different languages (Mylonas, Stockman & Griffin, 2018). We show that in all three datasets in English (British, American and calibrated) the 11 BCTs of Berlin & Kay (1969/1991) had lower dispensability scores than all non-BCTs. Our measure was also able to capture the indispensability of the proposed second *blue* basic term in Greek, Russian, Thai and to a lesser degree in Turkish (Prasithrathsint, 1988; Sturges & Whitfield, 1995; Özgen & Davies, 1998; Androulaki, Gômez-Pestañ, Mitsakis, Jover, Coventry, Davies, 2006; Mylonas & MacDonald, 2016; Paramei et al. 2018). Our results support growing evidence that communication efficiency provides a better framework to understand colour naming than opponent theory (Jameson & D' Andrade, 1997; Lindsey, Brown, Brainard, & Apicella, 2015; Regier, Kemp & Kay, 2015; Abbot, Griffiths & Regier, 2016; Gibson,

Futrell, Jara-Ettinger, Mahowald, Bergen, Ratnasingam, Gibson, Piantadosi & Conway, 2017).

In Chapter 8, we give a general discussion of the findings and limitations of this study while the thesis concludes in Chapter 9 with suggestions for future developments.

1.2. Published work

Parts of this thesis have been published in the following papers:

1.2.1. Direct

1. Mylonas, D., Purver, M., Sadrzadeh, M., MacDonald, L., Griffin, L. D. (2015). The Use of English Colour Terms in Big Data. In *AIC 2015 Midterm Meeting*. Tokyo, Japan.

2. Katemake, P., Mylonas, D. MacDonald, L., Prasithrathsint, A. (2015). Comparison Among Three Methods for Thai Colour Naming. In *AIC 2015 Midterm Meeting*. Tokyo, Japan.

3. Грибер Юлия Александровна & Милонас Димитрис (2015) Картография цвета: эмпирический анализ цветонаименований русского языка, Человек и культура. (In Russian, trans.: Yulia Griber & Mylonas Dimitris (2015) Color mapping: empirical analysis of color terms in the Russian language).

4. MacDonald, L., Mylonas, D. (2016). Colour Naming: Linking Vision and Speech. In *3rd International Conference 'Colour, Culture, Science'*. Kraków, Poland.

5. Mylonas, D., Andrews, J., Griffin L.D. (2016). Variability in free colour naming across the full colour gamut. In *AVA Christmas Meeting 2016*. Queen Mary University of London, London, UK.

6. Griffin L.D., Mylonas, D. (2016). Why so much talk about "red"? In *AVA Christmas Meeting 2016*. Queen Mary University of London, London, UK.

7. Mylonas, D., MacDonald, L. (2017). *Colour Naming for Colour Design*. In Colour Design, ed. Best, J., Elsevier

8. Mylonas, D., MacDonald L., Griffin, L.D. (2017). Differences in Color Naming between British and American English Speakers. In *13th AIC Congress*, Jeju, South Korea.

9. Griber Y., Paramei G.V., Mylonas, D. (2017). Gender Differences in Russian Colour Naming. In *13th AIC Congress*, Jeju, South Korea.

10. Ulusoy B., Griffin L.D., Mylonas, D. (2017). Turkish Colour Naming on the Net. In *13th AIC Congress*, Jeju, South Korea.

11. Paramei, G. V., Griber, Y. A., & Mylonas, D. (2018). An Online Color Naming Experiment in Russian Using Munsell Color Samples. *Color Research & Application*, *43*(3), 358–374. https://doi.org/10.1002/col.22190

12. Griffin, L. D., & Mylonas, D. (2019). Categorical colour geometry. *PLOS ONE*, *14*(5), e0216296. https://doi.org/10.1371/journal.pone.0216296

13. Mylonas D., Stockman A., Griffin L.D. (2018) Basic Colour Terms are Indispensable. In *AVA Christmas Meeting 2018*, Birkbeck University of London, London, UK.

14. Mylonas D., Griffin L.D., Stockman A., (2019) Mapping Colour Names in Cone Excitation Space. In *25th Symposium of ICVS*, Riga, Latvia.

15. Mylonas D. (2019) A Perpetual Ride to Knowhere, In Colour & Poetry, ed. Volley, J., The Small Press.

1.2.2. Indirect

1. Zeki S., Cheadle S., Pepper J., Mylonas D. (2017). The constancy of the colored after-images. *Frontiers in Human Neuroscience, 11*.

2. Zeki S., Javier A., Mylonas D. (2019). The Biological Basis of the Experience and Categorization of Colour. *European Journal of Neuroscience*.

1.2.3. In preparation

1. Coherence of Achromatic, Primary and Basic Classes of Colour
2. The indispensability of basic colour terms across languages.
3. Locating unconstrained colour names in cone excitation space.
4. Colour naming in American and British English.
5. A multilingual computational colour naming model.
6. Online and offline colour naming experiments.
7. Colour naming in remote societies

1.2.4. Invited Talks & Workshops

1. Mylonas, D. (2015) The role of psychological primaries in cognitive colour spaces: a cross-cultural colour naming study. Talk in the *Annual Colour Vision Meeting*, Colour Group Great Britain, Inst. Ophthalmology, UCL, London, UK.

2. Mylonas D. (2017) Where our brain meets the universe. Talk in the *Chromophobia Colour in Architecture*. Design Museum, London, UK.

3. Mylonas D. (2017) Features of Green colour categories across languages. Talk in the *Colour Green*, Colour Group Great Britain, Newcastle, UK.

4. Mylonas, D. (2018) Colour brings people together. Workshop in the *Munsell Centennial Color*, ISCC-AIC, Boston, US.

5. Mylonas, D. (2018) Colour naming within and across languages. Workshop in the *Munsell Centennial Color*, ISCC-AIC, Boston, US.

6. Mylonas, D. (2019) Mapping colour names in LMS cone excitation space. Talk in the *Annual Colour Vision Meeting*, Colour Group Great Britain. City University, London, UK.

7. Mylonas, D. (2019) Colour naming: a perpetual ride to knowhere. Talk in the *Colour & Poetry: A Symposium*, Slade Research Centre, University College London, UK.

1.2.5. Awards

1. Mylonas D. (2015). Characterisation and visualization of cultural heritage materials through colour naming. COST-STSM-TD1201-041015-067757, European Cooperation in Science and Technology (Brussels, Belgium).

2. Mylonas D. (2017). Differences in color naming between British and American English speakers. WD Wright Award, Colour Group Great Britain, UK.

3. Mylonas D. (2019). Mapping colour names in cone excitation space. WD Wright Award, Colour Group Great Britain, UK.

4. Mylonas D. (2019). Mapping colour names in cone excitation space. Applied Vision Association (AVA) Travel Award, UK.

Chapter 2

Background

This chapter gives a basic account of the underlying mechanisms governing colour naming. Its purpose is to give a general context of the basis and development of lexical colour categories. We then describe the core components that this thesis is based on, which include a survey of colour spaces and colour difference measures, relevant colour naming experiments, an overview of existing colour naming models and information theoretic analysis in the context of language games.

2.1. Colour naming and colour vision

Colour naming is the process of organizing millions of discriminable colours (Pointer & Attridge, 1998) into a smaller set of categories named, for example, as red, orange and purple. In a short treatise, Aristotle (350 B.C.E.) was one of the first to theorize on the underlying mechanisms of colour categorisation. He suggested that five pure colours - crimson, green, cyan, purple and possibly yellow (Sorabji, 1972) - arise from the mixture of white (light) and black (darkness) and from these all the other impure or irregular colours are generated. Aristotle justified this reduction into seven rational categories to simple numerical ratios – similar to other senses – and offered an analogy with music concords:

"...we may regard all these colours as analogous to the sounds that enter into music, and suppose that those involving simple numerical ratios, like the concords in music, may be those generally regarded as most agreeable; as, for example, purple, crimson, and some few such colours, their fewness being due to the same causes which render the concords few."

(Aristotle 350 B.C.E. / trans. by Beare, 2000)

This analogy between colour and the octave division of pitch influenced Isaac Newton (1730) in his prominent work to name initially five principal colours - red, yellow, green, blue and violet in the spectrum - and consecutively seven, by adding orange and indigo (Topper, 1990), arranged in a circular representation (see Figure 2.1). Neutral or achromatic colour names have no specific wavelengths and, as such, black was considered as the absence of colours and white as the presence of all colours.
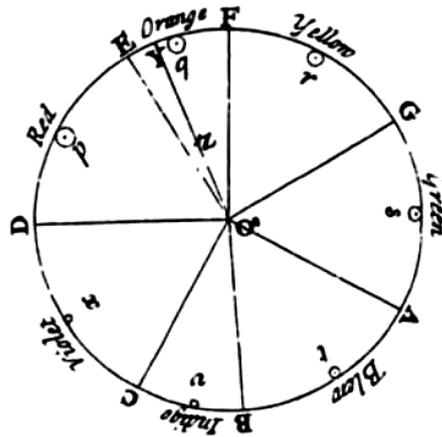
Figure 2.1 Seven spectral colour names of Newton arranged in a circular representation (reproduced from Newton, 1730).

Based on Newton's findings and his wave theory of light, Thomas Young (1802) suggested that it is necessary to assume only a limited number of three principal colours - initially labelled red, yellow and blue, but later switched to red, green and violet - as it would be impossible for every receptor in the retina to be sensitive to all different wavelenths of light. Helmholtz (1911) extended the trichromatic hypothesis by suggesting three different types of receptors with overlapping sensitivities. The Young-Helmholtz theory of trichromatic colour vision was supported by the colour matching studies of Grassmann (1853) and Maxwell (1872) which further advanced our scientific understanding of colour vision and led to the development of the Red, Green and Blue (RGB) additive colour models, which are widely used in input and output devices today. Trichromacy can predict with reasonable precision whether two coloured lights with different spectral power distributions match perceptually, but it cannot account for colour appearance phenomena. For example, why mixing yellow with blue lights would produce a perfect white. This limitation led Ewald Hering (1878/1964) to propose an 'opponent colours' theory, which assumes four chromatic processes arranged in opponent pairs (*red* versus *green* and *yellow* versus *blue*) and two achromatic processes (*white* versus *black*) to account for colour appearance, and, for the contestable observation that these colours are perceived to be unique while others are not (Hurvich & Jameson, 1957; but see also Malkoc, Kay & Webster, 2005; Bosten & Boehm, 2014).

The two theories can be unified to provide a comprehensive explanation for colour vision by assuming that colour appearance mechanisms operate upon the output of the cone mechanisms. This is better explained in a mechanistic framework shown in Figure 2.2, where the first stage consists of the trichromatic properties of cones, the cone-opponent mechanisms subsist the second and the third stage consists of the colour-opponent

mechanisms of colour appearance (Judd, 1949; De Valois & De Valois, 1993; Stockman & Brainard, 2010).

In the first stage, the three types of Long-, Medium- and Short- wavelength sensitive cones in the retina produce univariant responses relative to their absorbed amount of light (Rushton, 1972). The outputs of the three cones (L, M and S) are combined in the second stage by three postreceptoral mechanisms: L+M, L-M and S-(L+M). The first luminance mechanism adds the inputs from the L and M cones. The second antagonistic mechanism encodes signals in the range between *turquoise* and *red* by computing the difference between the L and M cones inputs; and consists effectively of a single main axis with only a modest – if any – contribution from the S cones. The output of this antagonistic stage is then modulated by the S cone in the third mechanism where the sum of L- and M- cone signals is differenced by the S-cone signal to split the main axis into two separate axes and encode signals varying from *lime* to *purple*. In the third conjectured stage, the outputs of the second stage mechanisms are summed to produce the four colour opponent mechanisms - Red/Green, Green/Red, Blue/Yellow and Yellow/Blue. These three stages illustrate the colour processing that takes place in the visual pathway from the cones of the retina to the visual cortex through the axons of the retinal ganglion cells and the lateral geniculate nucleus in the thalamus (Stockman & Brainard, 2010).

The mechanisms of the second stage correspond well with the three physiologically and anatomically defined channels in the lateral geniculate nucleus (LGN; De Valois, Abramov & Jacobs, 1966; Derrington, Krauskopf & Lennie, 1984) and the three cardinal dimensions proposed by Krauskopf and his colleagues (1982). The later authors carried out psychophysical experiments to define three independent channels in colour vision, called cardinal directions. The first direction refers to luminance and the second and third ones to colour-opponent directions. Notably, these cardinal dimensions do not coincide with the labelled dimensions of the colour opponent mechanisms of the third stage proposed by Hering (1878/1964; Abramov & Gordon, 1994; Valberg, 2001; Wuerger, Atkinson & Cropper, 2005). Therefore, other higher-order colour mechanisms are required to account for colour appearance, but the nature of their involvement remains an open question in colour vision (Gegenfurtner & Kiper, 2003; Stockman & Brainard, 2010). Subsequent studies showed evidence of activity of larger number of chromatic channels (Krauskopf et al., 1986; Gegenfurtner & Kiper 1992). Hansen & Gegenfurtner (2006) suggested multiple broadly tuned mechanisms, while Eskew and his associates suggested the existence of six unipolar postreceptoral colour mechanisms, rather than three bipolar mechanisms (Eskew, 2009; Shepard et al. 2016; 2017).
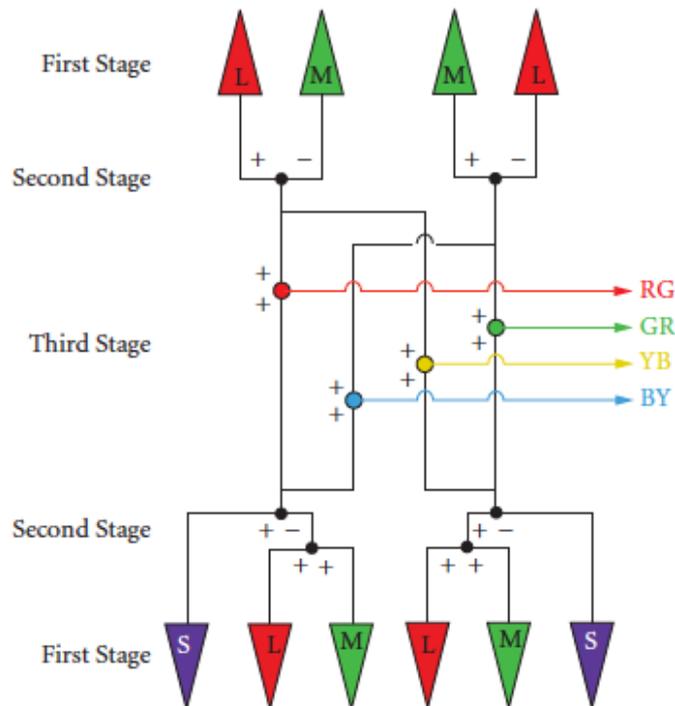
Figure 2.2. Three stage colour mechanisms. The first stage (top & bottom) involves L-, M-, S-wavelength sensitive photoreceptors. The second stage consists of the L-M and M-L cone opponent (top) and S-(L+M) cone opponent (bottom) mechanisms. The third stage includes the summation of the cone-opponent second stage mechanisms to produce the labelled colour opponent mechanisms (reproduced from Stockman & Brainard, 2010).

The higher stages of colour processing take place in the visual cortex, where information from the retino-geniculate channels is combined to construct colour perception. Our understanding of the colour processing in the cortex is less clear. Livingstone & Hubel (1988) suggested a specialized system for colour processing based on their observations of segregated cells in layers of the primary visual cortex (V1) analyzing either luminance or colour information. This was not supported by subsequent studies (Lennie, Krauskopf & Sclar, 1990, Gegenfurtner, Kiper & Fenstemaker, 1996). A key point with regards to the hue preference of cells in V1 (Lennie et al., 1990), in secondary visual cortex V2 (Kiper, Fenstemaker & Gegenfurtner, 1997) and cells in LGN (Derrington et al., 1984) is that cortical cells do not respond to hues associated with the unique hues of Hering, but prefer instead intermediate hues such as *limes* and *oranges*. This tuning of cortical cells in V1 and V2 might play a role in the construction of colour categories, but their responses correlate predominantly with the wavelength composition of light and only with local colour contrasts (Lennie et al., 1990; Wachtler, Sejnowski, & Albright, 2003) rather than with the perceived constant colours of more naturalistic large fields of view (Zeki, 1973, 1980, 1983).

Colour constancy is the ability of the colour vision system to discount changes in illumination conditions and assign constant colours to objects or surfaces in scenes. Mechanisms that mediate colour constancy can be grouped into two broad categories: a) chromatic adaptation at receptoral level (von Kries, 1905) and b) spatial ratio-taking operations at cortical level (Land & McCann, 1971, Zeki, 1980). The basic idea of chromatic adaptation models is that the three cone types in the retina selectively adapt to changes of illumination by dividing their Long, Medium and Short spectral sensitivity to light reflected from a coloured surface by their corresponding sensitivity to an estimated reflectance of a white surface under the same illumination. In contrast, Edwin Land (1974) proposed a computational theory ('Retinex') to account for colour constancy based on a comparison of three lightness measurements of the wavelength composition reflected from a surface and from its surrounds. Retinex hypothesise that the three lightness records correspond to the sensitivity of the three L-, M-, S- cones in the retina and the spatial comparisons take place in the cortex. Nevertheless, neither chromatic adaptation nor spatial computational methods can fully account for human colour constancy (Brill & West, 1986; Kraft & Brainard, 1999; Golz & MacLeod, 2002). By colour constancy, we should do not mean that the colour of an object preserves its exact shade across illuminations. The shade of a coloured surface will naturally change with changes in the wavelength-energy composition of the light in which it is viewed. Instead, objects maintain their perceived colour category to stabilise their appearance in different viewing conditions. Therefore, a better description would be 'constant colour categories', a definition than brings colour constancy closer to colour categorisation (Jameson, 1983; Olkkonen, Hansen & Gegenfurtner, 2009; Zeki et al., 2017; Zeki et al., 2019).

The actual cortical site at which constant colour categories are generated attracted significant scientific interest. Zeki suggested that the extrastriate visual area V4 plays a critical role in colour constancy in both monkeys (Zeki, 1973. 1980, 1983) and humans (McKeefry & Zeki, 1997). The peak sensitivity of narrow-band cells in V4 is distributed through the spectrum - including the extra-spectral purple. This area can account for the narrow ranges in colour space of colour categories (Zeki, 1980). Subsequent studies challenged (Schein, Marrocco, & de Monasterio, 1982; Schein & Desimone, 1990; Shapley & Hawken, 2011 for a review) but also supported (Conway & Tsao, 2006; Wade et al., 2008; Tanigawa et al., 2010; Brouwer & Heeger, 2013) the existence of a colour processing centre in the brain. Another area suggested to be involved in the elaboration of colour categories is the inferior temporal (IT) cortex implicated also in object vision. Komatsu et al. (1992) found a uniform distribution of cells with colour preferences, even down to a very narrow range of colours such as pale pink and pale red but found no

activations for blue and cyan in this area. The involvement of the IT cortex in colour processing has also been confirmed by fMRI studies which show multiple colour-biased regions (Lafer-Sousa & Conway, 2013). These conflicting views suggest that V4 may not act in isolation, but in cooperation with multiple cortical areas for the construction of constant colour categories.

Colour naming offers a direct and natural method of measuring colour constancy (Uchikawa, Uchikawa & Boynton, 1989; Troost & Weert, 1991; Foster, 2011, Zeki et al., 2017). Moreover, the strong correlation between naming consistency across illuminants and across observers suggests a close link between categorical colour constancy and consistent colour communication (Olkkonen et al., 2009, 2010; Zeki et al. 2019). Colour naming is considered one of the last stages in colour processing and it is likely to involve both visual and language areas of the brain, even though further research is needed to draw firm conclusions. Rare cases of cortical damage where patients lost only the ability to name colours of objects (Damasio, McKee & Damasio, 1979; Davidoff, 1996) indicate the existence of a specialised area of the brain for colour naming processing. In a positron emission tomography study with healthy human subjects, the production of colour words activated the ventral temporal lobe – an anterior region to V4 associated with object knowledge (Martin, Haxby, Lalonde, Wiggs & Ungerleider, 1995). In recent neiroimaging studies, Brouwer & Heeger (2013) located activations in human V4 and VO1 during colour naming tasks while Bird and his colleagues (2014) found no activations in any of the traditional visual and language areas of the brain for categorical perception of colours. The elusive relationship between the psychophysical mechanistic stages, colour constancy and colour naming remains one of the unsolved mysteries in colour vision.

2.2.   Primary colours in colour naming

The primary colours are widely considered to be *red, green, yellow, blue, black* and *white* (Kaiser & Boynton; 1996). These colours have been also proposed as a universal basis for colour naming across languages (Berlin & Kay, 1969/1991; Kay & McDaniel, 1978; Regier, Kay & Cook, 2005). A range of explanations has been suggested about their fundamental role in colour categorization based on different stages of the physical-optical-physiological causal sequence that underlies colour sensation.

An explanation of the physical type is that the primary colours have special status in visual ecology (Shepard, 1992; Mollon, 2006) and in the subtractive mixing of colours as

occurs with pigments or in the additive mixing of colour lights. Further along the causal chain of colour sensation is the idea that the primaries are special because they correspond to sensations that have a less ambiguous relation to their underlying reflectance than other colours (Philipona & O'Regan, 2006). There are other theories that purely consider the range of possible cone responses; for example, that the primaries are maximally spaced within the 3D sub-volume of cone response space corresponding to possible surface colours. A widely cited account of this type by Regier, Kay and Khetarpal, (2007) based on the suggestion by Jameson & D'Andrade (1997) argues that colour categories are determined by optimising the division of an irregular perceptual colour space to maximize similarity within a category and minimise similarity across categories. Furthest along the causal physiological chain is the idea that the primaries align with postulated postreceptoral (opponent) channels (Kay & McDaniel, 1978). As discussed earlier, these primary colour categories contain examples that are unique in that they are perceived to contain no other colour; and are considered an important physiological component in the formation of colour categories (Kuehni, 2005).

Contrary to these theories is the view that primary colours play no special role in the formation of colour categories. The cultural view for the origin of all colour names holds that it is the need to communicate about the surface properties of objects that generates colour names (Brown & Lenneberg, 1954; Lucy & Shweder, 1979; Davidoff, Davies & Roberson, 1999; Levinson, 2000; Steels & Belpaeme, 2005; Davidoff, 2015; Gibson et al., 2017). Neurobiological findings do not support either the concept of primary colours in cortical regions (Hubel, 1988), as the wavelengths of peak sensitivity of neurons are distributed across the spectrum; while some neurons are sensitive to desaturated and extra-spectral colours that do not correspond to any single wavelength of light (Zeki, 1980; Komatsu, 1992; but see Stoughton & Conway, 2008 and Mollon, 2009 for a reply). The idea that primary colours are associated with the opponent-process cells in early vision (Hering, 1878/1964; Hurvich & Jameson, 1957; De Valois et al., 1966) has also been disputed (Abramov & Gordon, 1994; Valberg, 2001; Wuerger, et al., 2005).

There are at least three reasons why the special status of primary colours has been maintained despite the inability to match them to opponent-process physiology. First, it is still widely held that primary colours are necessary and sufficient to describe all colours (Hardin, 2005). Secondly, primary colours show higher frequency as qualifying adjectives in texts (Corbett & Davies, 1997). Third, many tasks have used only a constrained set of colour names, allowing authors to claim (see Ocelák, 2014 for arguments that the claim is spurious) that primary colour names are the most important for colour matching and naming (Berlin & Kay, 1969/1991; Boynton & Olson, 1987;

Sturges & Whitfield, 1995; Kay et al., 2010; Panorgias, Kulikowski, Perry, McKeefry & Murray, 2010). Nevertheless, the importance of primary colours has been disputed.

The primary colours being the basis of colour naming systems has been questioned on conceptual grounds (van Brakel, 1993; Jameson & D'Andrade, 1997; Ocelák, 2014) and in a re-analysis of the World Color Survey (Kay et al., 2010) by Jameson (2010); but these concerns have not been met with widespread acceptance. Doubts have also been raised about the superiority of primary colour terms over non-primaries with respect to yielding consensus for naming (Boynton & Olson, 1987; Lindsey & Brown, 2006, 2009; Uchikawa & Boynton, 1987), for visual search (Wool, Komban, Kremkow, Jansen, Li, Alonso & Zaidi, 2015) and for hue cancelation (Malkoc et al., 2005; Bosten & Boehm, 2014). Regarding hue cancelation, both studies above found no differences between unique-hue judgments of binary hues (i.e., *orange*, *purple*) and those of their corresponding primaries.

2.3. Basic colour terms in colour naming

Berlin and Kay (1969/1991) suggested that after languages acquire terms for primary colours, they then acquire terms for *grey, purple, brown, orange* and *pink*, which, together with the primary terms make up the 'universal basic colour terms'. These undergo a seven-stage evolution in the development of colour vocabulary with the following order of emergence:

Stage I: Black and white
Stage II: Red
Stage III: Either green or yellow
Stage IV: Both green and yellow
Stage V: Blue
Stage VI: Brown
Stage VII: Purple, pink, orange, grey

Berlin and Kay defined BCTs by a combination of linguistic and psychological criteria:

1. A BCT should consist only of a single word.
2. Its scope should not overlap with any other colour term.
3. It should not be restricted to a limited class of objects.
4. It should be psychologically salient for speakers of the language in question.
5. Its meaning is not divisible or determined by its parts.
6. It is not the name of an object.
7. It is not a foreign loan word.
8. Its morphology is not complex.

While these criteria might sound logical, they have been strongly criticised as being not equally applicable across languages (Lucy & Sweder, 1979; Saunders and van Brakel, 1997; Biggam, 2012). For example, orange, which is considered a basic term in English, derives from the name of an object (fruit); and it is a foreign loan word from old French, therefore violating the sixth and the seventh criteria.

Subsequent studies found that BCTs are responded to quicker and with greater consensus within and across observers than non-basic colour terms (Boynton & Olson, 1987; Uchikawa & Boynton, 1987; Sturges & Whitfield, 1995; Regier et al. 2005; Lindsey & Brown, 2009). In machine vision, Griffin (2006) found that basic colour terms are performing better than any other set of colour terms in object classification tasks and proposed a pressure-to-optimality explanation for their basis. A symmetry analysis of the eleven BCTs by Griffin (2001) revealed similarities between the psychological structure of the basic colours and the physical structure of colour space (Figure 2.3).
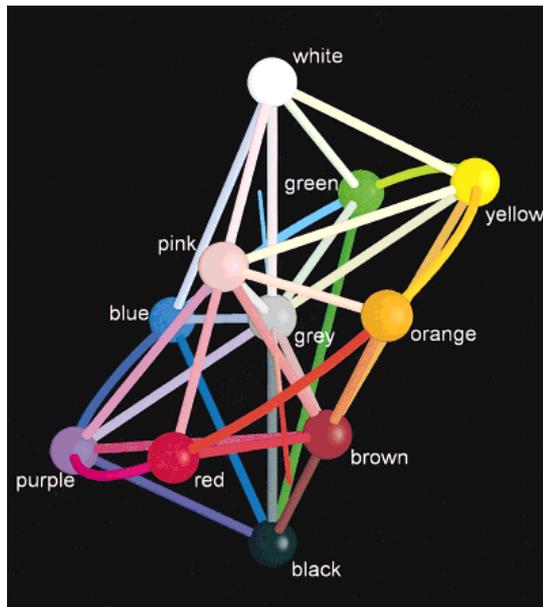
Figure 2.3 The psychological structure of the Basic Colour Terms. The spheres correspond to the 11 BCTs and the lines their adjacencies (reproduced from Griffin, 2001).

There has been considerable debate over what makes these eleven colour terms basic (see also Levinson, 2000; Lyons, 1995). Most importantly, Kay and his colleagues did not regard all basic terms as equivalent, even if subsequent investigations often did (Boynton & Olson, 1987; Sturges & Whitfield, 1995; Griffin, 2001, 2006; Yendrikhovskij, 2001; Lindsey & Brown, 2014; Lindsey, Brown, Brainard & Apicella, 2015). In its current form, the Universalist hypothesis suggests that a biological explanation for the origin of basic colour terms may be true only for the six landmark colours corresponding to the opponent primaries of Hering, while the other five may be formed under the influence of higher cognitive mechanisms. Logically, this opens the way for languages to acquire more than eleven basic colour terms, and for secondary terms to be considered as a potential group out of which new basic colour terms can arise (Hardin & Maffi, 1997).

2.4.   Basic colour terms in different languages

In this section we summarise the identification of BCTs by previous studies in the American and British English, Greek, Russian, Thai and Turkish colour lexicons that will be used in this study.

In American English, Berlin & Kay (1969/1991) reported eleven BCTs: *white, black, red, yellow, green, blue, brown, purple, pink, orange and grey* and classified English in the highest stage (VII) of evolution in the development of colour vocabularies. Boynton and Olson (1987) confirmed that American English speakers use the eleven BCTs more

consistently, with greater consensus and more quickly than non-BCTs in laboratory settings. Lindsey and Brown (2014) confirmed the importance of the eleven BCTs in the American colour lexicon but also reported nine additional statistically significant non-basic colour categories featuring *peach, teal, lavender* and *maroon* as high-consensus terms.

Sturges and Whitfield (1995) showed that BCTs have shorter response times and also higher consistency and consensus than non-BCTs in British English. *Cream* was suggested as a strong candidate for a missing twelfth BCT, but with a subtle lower consistency from the other eleven BCTs. Recently, Mylonas & MacDonald (2016) suggested the extension of the English inventory from 11 basic colour terms to 13 terms, with the addition of *lilac* and *turquoise*. The authors analysed colour naming responses from an online colour naming experiment and computed the mean of the ranks for each colour term across six different measures (frequency, consensus, response time, consistency, volume and inter-experimental agreement) to obtain a gradual index of basicness. In terms of index differences, separation from the lowest ranked BCTs, of *white, red* and *orange*, to *lilac* and *turquoise* was moderate; but there was a considerable jump in index value to the following non-basic term, *tan* (14th). Both terms appear to reduce the uncertainty of colour naming from using only the 11 BCTs, as *lilac* partitions the large colour category of *purple* in light and dark segments, while *turquoise* appears at the border between *green* and *blue*.

The Greek colour language was studied through the literature of Homeric Greek and was categorized as Stage IIIb by Berlin & Kay (1969/1991). A recent study (Androulaki et al. 2006) of Modern Greek colour terminology reported twelve BCTs including two blues: *aspro/white, mavro/black, kokkino/red, kitrino/yellow, prasino/green, ble/blue, galazio/light blue, kafe/brown, gri/grey, mov/purple, roz/pink, and portokali/orange.* Further evidence for the existence of the two basic Greek blues provided by two studies on Greek-English bilingualism (Athanasopoulos, 2009) and on categorical perception of the blue region between Greek, German and Russian speakers (Maier & Rahman, 2018).

Berlin and Kay (1969/1991) noted that Russian speakers may have 12 BCTs in their colour lexicons, including two *blues*: *belyj/white, čërnyj/black, krasnyj/red, žëltyj/yellow, zelënyj/green, sinij/blue, koričnevyj/brown, fioletovyj/purple, rozovyj/pink, oranževyj/orange, seryj/grey and goluboj/sky blue.* This exemption to the universal inventory of the eleven BCTs, triggered an abundance of studies confirming the second basic *blue* in Russian (Morgan & Corbett, 1989; Moss, 1998; Paramei, 2005). In a recent

study, we also verified that the two Russian blues divided the *blue* unitary area in English along the lightness dimension, and that their centroids deviated from the centroid of English *blue* (Paramei et al., 2018)*.*

Berlin & Kay (1969/1991) classified the Thai colour language as Stage VII with 10 BCTs, excluding *grey* but recent studies identified twelve BCTs including two *blues*: *white /khaw/, black /dam/, red /dang/, yellow /leaung/, green /khiaw/, light blue /fa/, blue /namngen/, brown /namtan/, grey /thaw/, purple /muang/, pink /chompu/ and orange /som* (Prasithrathsint, 1988; Engchuan, 2003). In a recent study, we compared the frequency and location of the twelve Thai basic colours terms in three experimental methodologies - two were conducted in controlled viewing conditions and one over the Internet (Katemake et al., 2015). Although the frequencies of colour names across the methods differed, they produced ranks within each method that were similar to the ranks obtained in the other two methods. We found good correspondence amongst the three methods in terms of the location of basic colour terms in terms of hue and lightness but large differences in chroma dimension.

The Turkish colour lexicon was not investigated by Berlin & Kay (1969/1991), but consequent studies reported the existence of twelve BCTs including possibly two *blues*: *beyaz/white, siyah/black, kirmizi/red, sari/yellow, yeşil/green, mavi/blue, kahverengi/brown, mor/purple, pembe/pink, turuncu/orange, gri/grey and lacivert/dark blue* (Özgen & Davies, 1998; Ekici, Yener & Camgöz, 2006)*.* The low consensus reported for the second *blue* term *lacivert* (Rätsep, 2011), diminished the claim for its basicness but in a recent analysis of the Turkish data of this study we found that *lacivert* was the fourth term with the highest consensus (Ulusoy et al., 2017).

Overall, these results suggest that basicness is a continuous rather than a binary characteristic of lexical colour categories (Lindsey et al., 2014; Mylonas & MacDonald, 2016; Gibson et al., 2017; Witzel, 2018; for a review). Furthermore, the existence of two basic *blue* terms dividing the unitary English *blue* area is far more common than previously thought and implies the existence of a Stage VIII in the development of basic colour terms.

2.5. Achromatic colours in colour naming

The most common achromatic colour names include *black*, *white*, *grey* and their modifiers. *White* and *black* are considered pure categories according to the Aristotelian view (350 B.C.E.) as well as the opponent process theory (Hering, 1878/1964). Hering

differentiated the white - black opponent processes from the mutually exclusive chromatic opponent processes because *grey* appears simultaneously whitish and blackish. This view was challenged by previous studies, which suggested that *grey* is also a primary category as it shares similar characteristics with the other elemental categories (Dimmick 1925; Boring, 1949). Pure *grey*, they suggest, can be seen when *white* and *black* or other chromatic opponent axes are in a state of perfect equilibrium, but as such has no complementary colour. Quinn and his colleagues (1982) supported Hering's view that *grey* is a composite colour, but their results were based on a small number of subjects (n=3) and a constrained experimental methodology.

2.6. Colour specification systems

The mapping of perceptual and linguistic aspects of colour is essential in order to understand their relationship. Colour systems are usually three-dimensional geometric spaces because the human colour visual system is comprised of three types of receptors sensitive to long, medium and short wavelength of light that allow the description of colours with numerical coordinates as points (Kuehni, 2003). The representation of perceived colours in such space is often based on colour matching experiments without the use of language. Researchers can specify the referents of colour names in terms of coordinates in such colour systems. The grid of the colour system can be subdivided into regions/categories where the same colour name is used by a large number of observers. Using the same colour system in cross-cultural research allows measurements of differences and similarities of colour categories across languages. In the following sections, we briefly review colour systems which will be covered in this work to specify the referents of colour names.

2.6.1. Munsell colour system

The Munsell colour system was originally designed as an educational tool to help art students to describe colours and their relations (Munsell, 1905), but it was further revised with extensive visual experiments (Newhall et al., 1943) to standardise colour specification in virtually any area of colour application. The system consists of three perceptual dimensions - *hue*, *chroma* and *value* - in a cylindrical colour space (Figure 2.4). The Munsell *hue* refers to the quality by which we distinguish one colour from the other, for example a *blue* from a *green*, and it is measured by degrees around horizontal meridians. The *hue* dimension is divided into five principal hues of *Red (5R), Yellow (5Y), Green (5G), Blue (5B) and Purple (5P)* along with 5 intermediate hues which are further subdivided into 10 steps. This produces 100 hue steps in the full cycle. *Value* refers to the relative colour quality of lightness or darkness within a colour and it is measured

vertically from black being set to 0 and white to 10. *Chroma* refers to the purity or intensity of a colour and it is measured radially from neutral axis towards the surface of the colour solid. Each *hue* is extended to its maximum *chroma* at each *value* forming an irregular spheroid.
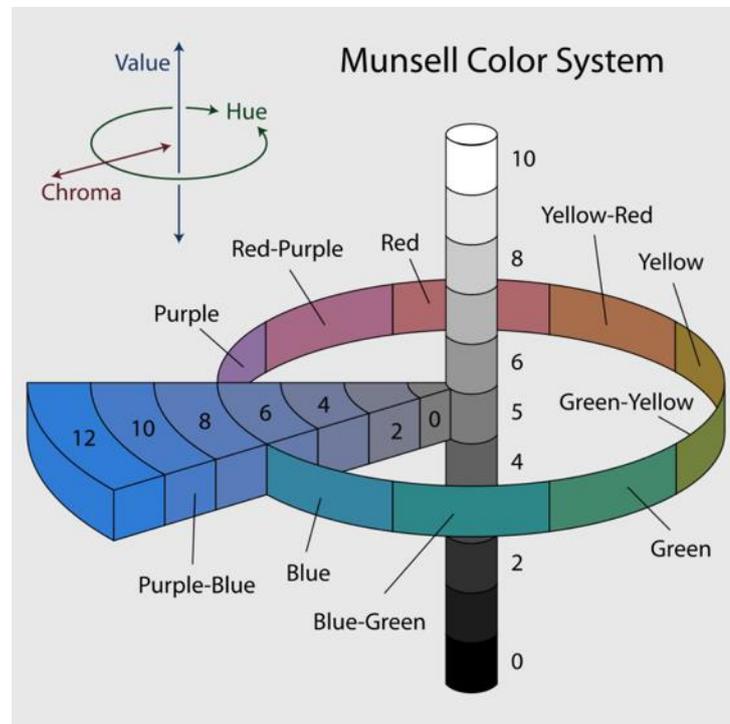
Figure 2.4 Munsell Colour system, reproduced by Jacob Rus, distributed under a CC BY-SA 3.0.

A major drawback of the Munsell system is that while each sample is specified colourimetrically, there are no mathematical equations to relate its coordinates to the physically measurable values of colourimetric coordinates. Another limitation is the unspecified colour appearance of the samples under different viewing condition (Fairchild, 2005). Finally, given the cylindrical shape of the solid, the uniformity of the sampling varies at different chroma levels.

The positive characteristics and the accuracy of the colour reproduction of the Munsell books of colours influenced a large number of researchers to use this system extensively in colour naming studies (Brown & Lenneberg, 1954; Berlin & Kay, 1969/1991; Sturges & Whitfield, 1995; Olkkonen et al., 2009; Kay et al., 2010; Gibson et al., 2017). In the online colour naming experiment of our ongoing study (Mylonas & MacDonald, 2010), we employ 600 simulated chips spaced approximately uniformly in the Munsell colour system (see section 3.1.).

2.6.2. OSA-Uniform Colour Scale

As discussed above, the Munsell system suffers from poor uniformity because of its radial sampling. The Optical Society of America developed a colour order system based on a cuboctahedron structure, which results in the colour system consisting of a uniform spacing in all three dimensions (Nickerson, 1981). The O.S.A Uniform Colour Scales (UCS) consists of 424 colour samples on a regular 2-unit grid, where 12 samples surround each sample with equal distances in *L, j* and *g* notation (Figure 2.5). Despite its great uniform features, the limitation of the system to sample colour in constant hue and high chroma has reduced its practical applications and popularity (Berns, 2000).
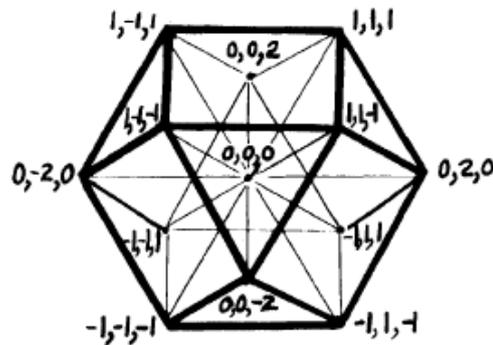
Figure 2.5 Cuboctahedral shell of OSA-UCS with 12 points at unit distance from central point where L,j,g = 0,0,0 (reproduced from Nickerson, 1981).

2.6.3. The colourimetric system

Colourimetry is the science of colour measurement where the objective is to specify a physically defined visual stimulus with numbers (Wyszecki & Stiles, 1982). In the CIE colourimetric system, this quantitative description of colour is expressed with a triplet of values *X, Y, Z* calculated in practice from spectra as:

$$X = \sum_i \bar{x}_i P_i \delta\lambda \tag{2.1}$$

$$Y = \sum_i \bar{y}_i P_i \delta\lambda \tag{2.2}$$

$$Z = \sum_i \bar{z}_i P_i \delta\lambda \tag{2.3}$$

where $P_i \delta\lambda$ is the spectral radiant power distribution of the colour stimulus at the interval i of wavelength $\lambda \in [360, 830]$ and $\bar{x}, \bar{y}, \bar{z}$ are the colour matching functions of the standard colourimetric observer for stimuli at an angular subtense of 2° as defined by the CIE in 1931. For stimuli with an angular subtense size greater than 4°, CIE defined in 1964 the set of colour matching functions $\bar{x}_{10}, \bar{y}_{10}, \bar{z}_{10}$ of the supplementary standard colourimetric observer. CIE colour spaces are often referred to as device independent colour spaces, as they do not depend on any particular device or medium.

These tristimulus values form a perceptual 3D space in which two stimuli sharing the same values under the same viewing conditions will, when viewed by an observer with normal trichromatic vision, match in colour. CIE complements the previous formalisations with the addition of a set of standard illuminants A, B, C and more recently the illuminant D series for natural daylights (Wyszeki & Stiles, 1982). The CIE chromaticity diagram (Figure 2.6) is a plot for visualising colour stimuli of the CIE XYZ 1931 space in a chromatic plane defined by the coordinates *x* and *y*:

$$x = \frac{X}{X + Y + Z} \tag{2.4}$$

$$y = \frac{Y}{X + Y + Z} \tag{2.5}$$

The CIE *XYZ* 1931 colour space is the basis for all CIE defined colour spaces where we can measure differences between any colour but lacks perceptual uniformity (MacAdam, 1942); this limits its application in measuring the magnitude of the perceived differences between mismatched colour stimuli.
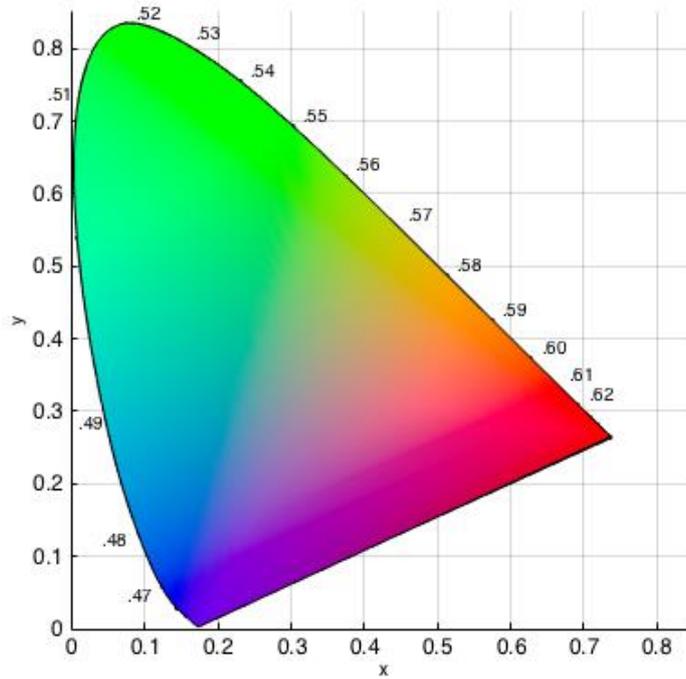
Figure 2.6 CIE 1931 chromaticity diagram.

2.6.4. CIE 1976 *L*u*v** (CIELUV)

In 1976 the CIE proposed an approximately perceptual uniform colour space defined by the following transformation of CIEXYZ 1931 coordinates to *L*u*v** coordinates in a rectangular grid (Wyszecki & Stiles, 1982):

$$L^* = 116(Y/Y_n)^{1/3} - 16 \qquad (2.6)$$

$$u^* = 13L^*(u' - u'_n) \qquad (2.7)$$

$$v^* = 13L^*(v' - v'_n) \qquad (2.8)$$

for $Y/Y_n > 0.01$. For $Y/Y_n \geq 0.008856$, a modified L_m^* is defined as follows:

$$L_m^* = 903.3\, Y/Y_n \qquad (2.9)$$

The quantities *u', v' and u_n', v_n'* can be computed by:

$$u' = 4X/(X + 15Y + 3Z) \quad and \quad v' = 9Y/(X + 15Y + 3Z) \tag{2.10}$$

$$u'_n = 4X_n/(X_n + 15Y_n + 3Z_n) \quad and \quad v'_n = 9Y_n/(X_n + 15Y_n + 3Z_n) \tag{2.11}$$

where X_n, Y_n, Z_n are the tristimulus values of the illuminant, an example of which is the standard daylight illuminant D65 [95.047, 100.00, 108.883] which will be used extensively in this thesis. The *u_n', v_n'* correspond to the chromaticity coordinates of the white point. The *u', v'* coordinates can be used to plot the UCS chromaticity diagram (Figure 2.7), and *L** to predict the lightness of stimuli. CIELUV is also associated with the perceptual attributes of hue, chroma and saturation which can be calculated by:

$$h_{uv} = arctan(u^*/v^*) \tag{2.12}$$

$$C_{uv} = \sqrt{(u^*)^2 + (v^*)^2} \tag{2.13}$$

$$S_{uv} = C^*_{uv} / L^* \tag{2.14}$$
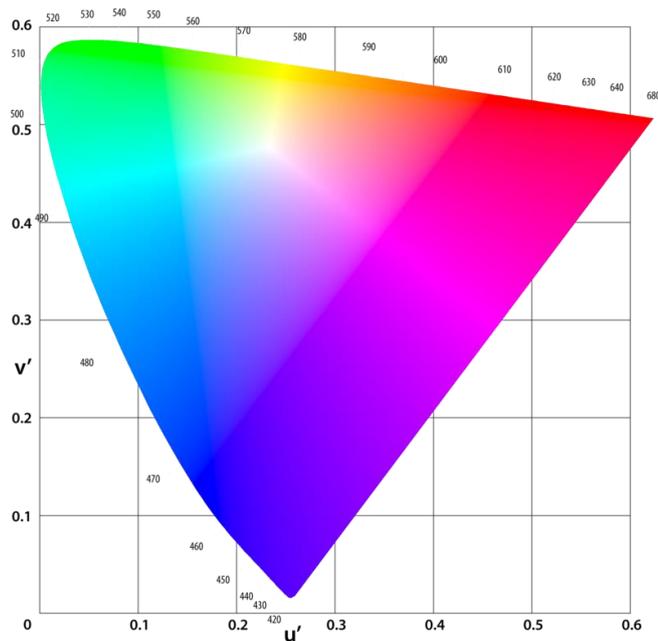
Figure 2.7 CIE 1976 UCS (u' v') chromaticity diagram.

41

The CIELUV colour space is used widely for coloured lighting applications, including display monitors.

2.6.5. CIE 1976 *L* a* b** (CIELAB)

The CIELAB colour space was the second approximately perceptual uniform colour space proposed by CIE in 1976 and it is more often used in the printing industry. *The L* a* b** coordinates are defined as follows:

$$L^* = 116(Y/Y_n)^{1/3} - 16 \text{ if } Y/Y_n > 0.008856 \qquad (2.15)$$

$$L^* = 903.3\left(\frac{Y}{Y_n}\right) \text{ if } Y/Y_n \leq 0.008856 \qquad (2.16)$$

$$a^* = 500\left[f\left(\frac{X}{X_n}\right)^{1/3} - f\left(\frac{Y}{Y_n}\right)^{1/3}\right] \qquad (2.17)$$

$$b^* = 200\left[f\left(\frac{Y}{Y_n}\right)^{1/3} - f\left(\frac{Z}{Z_n}\right)^{1/3}\right] \qquad (2.18)$$

where X_n, Y_n, Z_n, represents the chosen reference white point, and if $f(N/N_n)$>0.00856 then $f(N/N_n)=(N/N_n)^{1/3}$, otherwise $fN/N_n=7.787(N/N_n)+16/116$. The perceptual attributes of hue and chroma can be calculated using Equations (2.12 and (2.13 by replacing *u** with *a** and *v** with *b**. As the above ratios of the tristimulus values are not linear there is no chromaticity diagram for CIELAB, and hence no saturation.

2.6.6. Colour differences

The colour differences between stimuli in CIELUV can be calculated using 3D Euclidean distances:

$$\Delta E_{uv} = [(\Delta L^*)^2 + (\Delta u^*)^2 + (\Delta v^*)^2]^{1/2} \qquad (2.19)$$

where *ΔL, Δu** and *Δv** denote the arithmetic differences between the coordinates of the stimuli in CIELUV. The same formulae can be used to measure colour differences in CIELAB by substituting *a** for *u** and *b** for *v** (Wyszecki & Stiles, 1982). This UCS 1976

formula is widely used but performs poorly for saturated colour stimuli, poorly in the sense that the ΔE_{uv} values do not correlate well with perceived colour differences. To improve the uniformity of colour differences, CIE recommends the use of CIE ΔE2000 formula in CIELAB that can be calculated from:

$$\Delta E_{00} = \sqrt{\left(\frac{\Delta L'}{k_L S_L}\right)^2 + \left(\frac{\Delta C'}{k_C S_C}\right)^2 + \left(\frac{\Delta H'}{k_H S_H}\right)^2 + R_T \left(\frac{\Delta C'}{k_C S_C}\right)\left(\frac{\Delta H'}{k_H S_H}\right)} \qquad (2.20)$$

where k_L, k_C, k_H in this thesis is set to unity. For details about the calculation of $\Delta L'$, $\Delta C'$, $\Delta H'$, S_L, S_C, S_H, R_T, see Green (2002). Both formulae will be used for measuring differences between colour names in this study.

2.6.7. Colour Appearance Models

Colour appearance models (CAM) are developed with the aim of extending basic colourimetry to allow prediction of how colours appear in different viewing conditions. In the most common form, colour appearance models involve a chromatic adaptation transform, a dynamic response function and a transformation to a uniform colour space (Fairchild, 2005). CAM02 is the current recommended colour appearance model of CIE. As input data the model accepts the CIE XYZ 1931 tristimulus values of the colour sample, the tristimulus values of the white point, the adapting luminance, the relative luminance of the surround and a decision whether to apply the process of discounting the illuminant or not. The model can predict a wide range of colour appearance dimensions such as lightness, brightness, chroma, colourfulness, saturation and hue. It can also be used to predict the influence on colour appearance of different states of adaptation, surround and luminance levels. CAM02 Uniform Colour Space (Luo, Cui & Li, 2006) is an extension of the original model to improve its performance in predicting colour discrimination data while a new dataset of unique hues has been proposed to improve its hue uniformity and chromatic adaptation under mixed illumination conditions (Xiao, Wuerger, Fu & Karatzas, 2011; Xiao, Fu, Mylonas, Karatzas & Wuerger 2011; Xiao, Mylonas, Fu, Karatzas & Wuerger, 2011). In this thesis we will make little, if any, use of this type of models, and the description of their rather large number of equations – which result in computational failures in certain cases – is beyond its scope. The reported mathematical problems are addressed in the new CAM16 model (Li, Li, Wang, Xu, Luo, Cui, Melgosa, Brill, Pointer, 2017).

2.6.8. Default Internet RGB colour space sRGB

The sRGB colour space was proposed by Microsoft and Hewlett-Packard as a default colour space for monitors, cameras, printers and the Internet (IEC, 1999). The sRGB space is based on the *ITU-R BT.709-5* primaries, a typical overall transfer gamma function of *2.2* and specified viewing conditions that allow a straight-forward transformation to the CIE *XYZ* 1931 colour space from:

$$RGB_{linear} = RGB_{sRGB} / 12.92, \text{ if } RGB_{sRGB} \leq 0.04045$$

$$RGB_{linear} = (RGB_{sRGB} + 0.055 / 1+0.055)^{2.4}, \text{ if } RGB_{sRGB} > 0.04045$$

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9505 \end{bmatrix} \begin{bmatrix} R_{linear} \\ G_{linear} \\ B_{linear} \end{bmatrix} \tag{2.21}$$

Currently sRGB is the default colour space for virtually all monitor manufacturers but some differences between implementations are to be expected. In an earlier industry related research project, we assessed the error of colour reproduction in a wide range of mobile display devices (Mylonas, Karatzas & Wuerger, 2010) and uncorrected LCD/TFT desktop monitor displays (Xiao & Mylonas, 2010), against an sRGB calibrated monitor (Apple Cinema). For each display the colour reproduction of a set of 18 × 18 × 18 samples evenly distributed in the RGB cube were evaluated against a reference sRGB display using a Piecewise Linear Chromaticity Constancy (PLCC) characterization model and 21 spectroradiometric measurements. In total, we evaluated 11 desktop and laptop monitors (Apple MacBook, Asus, Dell, Hewlett Packard, Samsung, and Sony) and 4 popular mobile displays (iPhone, Samsung, HTC and Nokia), under four different lighting conditions (daylight-D65, home-incandescent, office-CWF and dark). Overall, the mean colour reproduction error for the desktop displays was larger (ΔE_{00}=8.06; STD=4.49) than for mobile displays (ΔE_{00}=6.02; STD=1.75) while the mobiles were fairly consistent under the four lighting conditions (max STD=0.46; min STD=0.14).

2.6.9. Physiologically relevant colour matching functions.

Despite the practical success of CIE *XYZ* 1931, there is strong evidence that can be significant errors in the specifications of the 2° Standard Colourimetric Observer, as the colour matching functions are too insensitive at the shorter wavelengths of the visible spectrum (Stockman & Sharpe, 1999; Stockman, 2006). The precise specification of the cone spectral sensitivities is essential not only for modelling colour vision, but also for practical applications of colour matching, colour measurements and colour naming. CIE proposed (CIE 170-1:2006) a new set of physiologically-relevant cone fundamentals

based on data from experimental studies with real observers, aiming to ground colourimetry on physiology (Stockman & Sharpe, 2000). The cone fundamentals for 2° (see Figure 2.8) and 10° fields of Stockman & Sharpe (2000) were determined from spectral sensitivity measurements under chromatic adaptation of dichromatic and normal observers, and analysis of the 10° colour matching functions of Stiles and Burch (1958). These functions can be found online at www.cvrl.org. The tristimulus coordinates of *L, M* and *S* for a colour stimulus $\varphi_\lambda(\lambda)$ can be then obtained by:

$$L = k_L \int \varphi_\lambda(\lambda) \cdot \bar{l}(\lambda) \cdot d\lambda \qquad (2.22)$$

$$M = k_M \int \varphi_\lambda(\lambda) \cdot \bar{m}(\lambda) \cdot d\lambda \qquad (2.23)$$

$$S = k_s \int \varphi_\lambda(\lambda) \cdot \bar{s}(\lambda) \cdot d\lambda \qquad (2.24)$$

where $\bar{l}(\lambda), \bar{m}(\lambda), \bar{s}(\lambda)$ are the cone fundamentals for 2 degrees visual dimeter colour stimuli normalised to unity peak and *k* are normalizing constants (CIE 170-2: 2015).
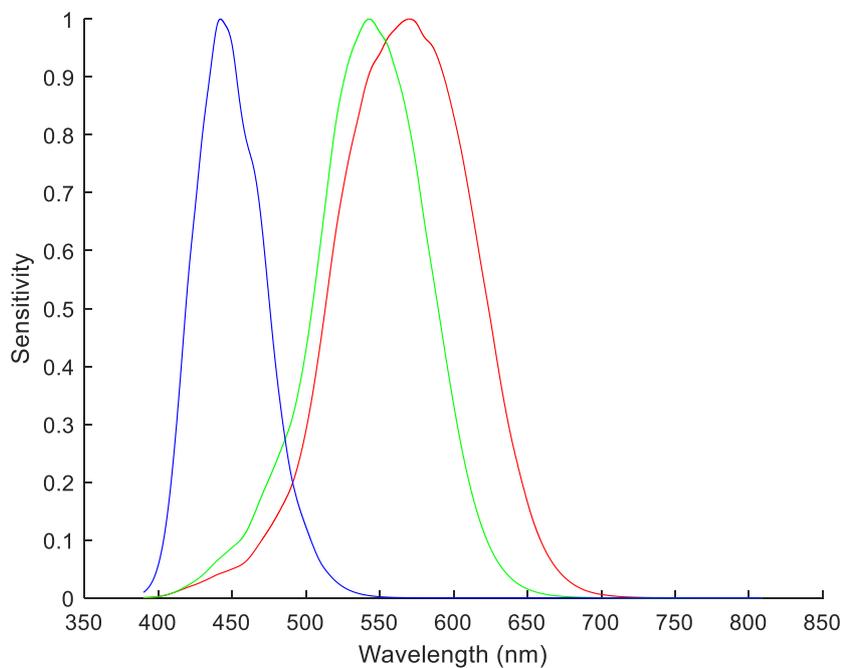
Figure 2.8 Spectral sensitivities of L-, M-, S-cones for 2° (Stockman & Sharpe, 2000)

In 2015, CIE adopted a new set of colour matching functions $\bar{x}_F(\lambda), \bar{y}_F(\lambda), \bar{z}_F(\lambda)$ which are linear transformations of the $\bar{l}(\lambda), \bar{m}(\lambda), \bar{s}(\lambda)$ cone fundamentals (Stockman & Sharpe, 2000; CIE 170-1: 2006; CIE 170-2: 2015). The functions that determine the cone fundamental based tristimulus values for 2 degrees field size can be also obtained by a matrix equation:

$$\begin{pmatrix} \bar{x}_F(\lambda) \\ \bar{y}_F(\lambda) \\ \bar{z}_F(\lambda) \end{pmatrix} = \begin{pmatrix} 1,947\,354\,69 & -1,414\,451\,23 & 0,364\,763\,27 \\ 0,689\,902\,72 & 0,348\,321\,89 & 0 \\ 0 & 0 & 1,934\,853\,43 \end{pmatrix} \begin{pmatrix} \bar{l}(\lambda) \\ \bar{m}(\lambda) \\ \bar{s}(\lambda) \end{pmatrix} \qquad (2.25)$$

where λ=390nm to 830nm in 1 resolution steps (CIE 170-2: 2015). The new CIE X_F, Y_F, Z_F cone fundamental based tristimulus values can be then calculate by substituting $\bar{x}$, $\bar{y}$, $\bar{z}$ in Equations 2.1-2.3 with $\bar{x}_F(\lambda), \bar{y}_F(\lambda), \bar{z}_F(\lambda)$. The cone fundamental based chromaticity diagram can be computed by substituting X, Y, and Z with X_F, Y_F, Z_F in Equations 2.4 and 2.5.

2.6.10. Cone chromaticity diagram

The CIE chromaticity diagrams do not provide a satisfactory visual relationship between the colour representations and the underlying cone-opponent mechanisms. Therefore, CIE recommends the use of the MacLeod & Boynton (MB; 1979) cone chromaticity diagram based on the cone fundamentals of Stockman and Sharpe (2000). In LMS, the *MB* diagram (Figure 2.9) is parallel to S-axis and the S dimension is scaled for convenience in the [0, 1] range. The total contributions of L- and M- cones remain constant within the plane to determine its orientation. *MB* chromaticity coordinates can be obtained for 2 degrees field size by the equations:

$$l_{MB}(\lambda) = 0,68990272\,\bar{l}(\lambda)/(0,68990272\,\bar{l}(\lambda) + 0,348\,321\,89\,\bar{m}(\lambda)) \qquad (2.26)$$

$$m_{MB}(\lambda) = 0,348\,321\,89\,\bar{m}(\lambda)/(0,68990272\,\bar{l}(\lambda) + 0,348\,321\,89\,\bar{m}(\lambda)) \qquad (2.27)$$

$$s_{MB}(\lambda) = 0,037\,159\,71\,\bar{s}(\lambda)/(0,689\,902\,72\,\bar{l}(\lambda) + 0,348\,321\,89\,\bar{m}(\lambda)) \qquad (2.28)$$

where $\bar{l}(\lambda), \bar{m}(\lambda), \bar{s}(\lambda)$ are the cone fundamentals (Stockman & Sharpe, 2000; CIE 170-1: 2006).
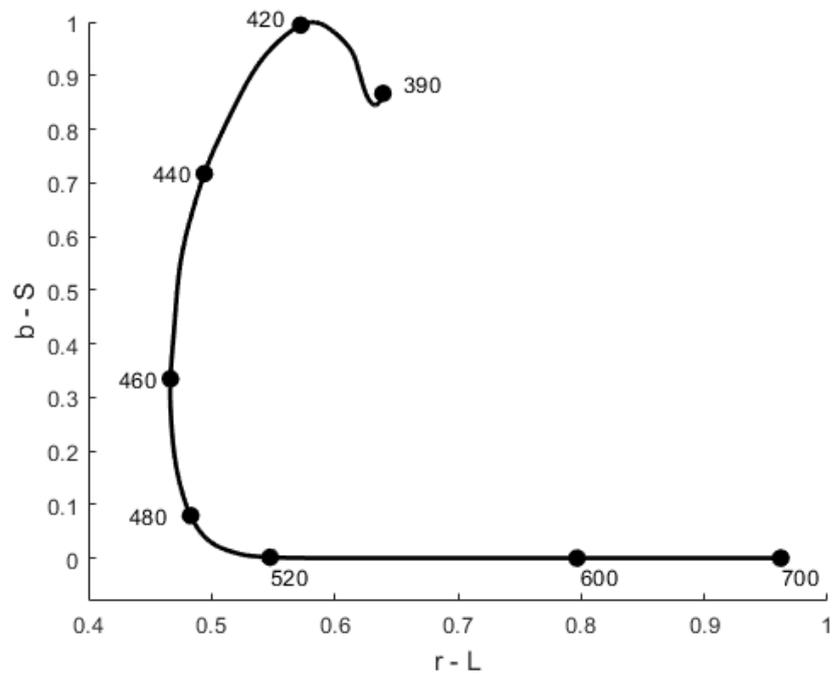
Figure 2.9 MacLeod-Boynton diagram of the spectral locus with s vs l coordinates based on

L-, M-, S-cones for 2° (MacLeod & Boynton, 1979; Stockman & Sharpe, 2000).

2.6.11. Derrington-Krauskopf-Lennie (DKL) colour space

Chromaticity diagrams represent the initial encoding of light by the cones in the first stage of colour processing in the retina, but their dependence on the adapting white of the scene limits their applications for evaluating colour categories across different adaptation states (Krauskopf & Gegenfurtner, 1992). The Derrington-Krauskopf-Lennie (DKL; 1984) opponent modulation colour space represents the second stage of colour processing in which differential cone signals are combined into three postreceptoral mechanisms with respect to the adapting background (Brainard, 1996; in Kaiser & Boynton, 1996). The axes of DKL represent the proposed cardinal directions (Krauskopf et al., 1982) of a luminance: (L+M) and two, colour opponent mechanisms: L-M and S-(L+M). Note, however, that the loci of unique hues do not align with the cardinal axes of the DKL space nor do the boundaries of their corresponding categories. (Malkoc et al., 2005; Bosten & Boehm, 2014; Witzel & Gegenfurtner, 2018).

2.7. Colour naming experiments

In a seminal study, Brown & Lenneberg (1954) performed colour-naming experiments using the Munsell colour system to examine behavioural consequences of naming on recognition, known as linguistic codability. In their experimental procedure, the authors first mounted 240 colour samples of the most saturated Munsell colours on cards in a systematic arrangement and asked 5 subjects to pick the best example for eight colour terms in wide cultural use in English (*red, orange, yellow, green, blue, purple, pink* and *brown*) from these 240 chips. In a second task, 24 subjects were asked to name 24 colour chips covering the colour space approximately uniformly; while including the 8 chips that were most frequently picked for each colour name. The codability index was then measured in five ways:

a) The mean number of syllables of colour names produced to each colour
b) The mean number of words of colour names produced to each colour
c) The mean reaction time for every colour sample
d) Interpersonal agreement
e) Intrapersonal agreement

A colour name with high codability index would be a shorter word that is identified faster and with greater degree of agreement about the referent colour across observers. The second part of the experiment involved a recognition/memory task where subjects were exposed to four colours simultaneously. After the colours were removed, they were asked to identify them by pointing at them within a larger array of 120 colours. The authors found a high degree of correlation between codability and recognition.

Berlin & Kay (1969/1991) used an elicitation method to identify the most common colour terms in different languages. They employed a stimulus palette with 320 of the most saturated colours for each *value* level and 10 achromatic tonal values from the Munsell system (Figure 2.10) and asked subjects to identify the best colour examples for each colour term in each language under unspecified viewing conditions.
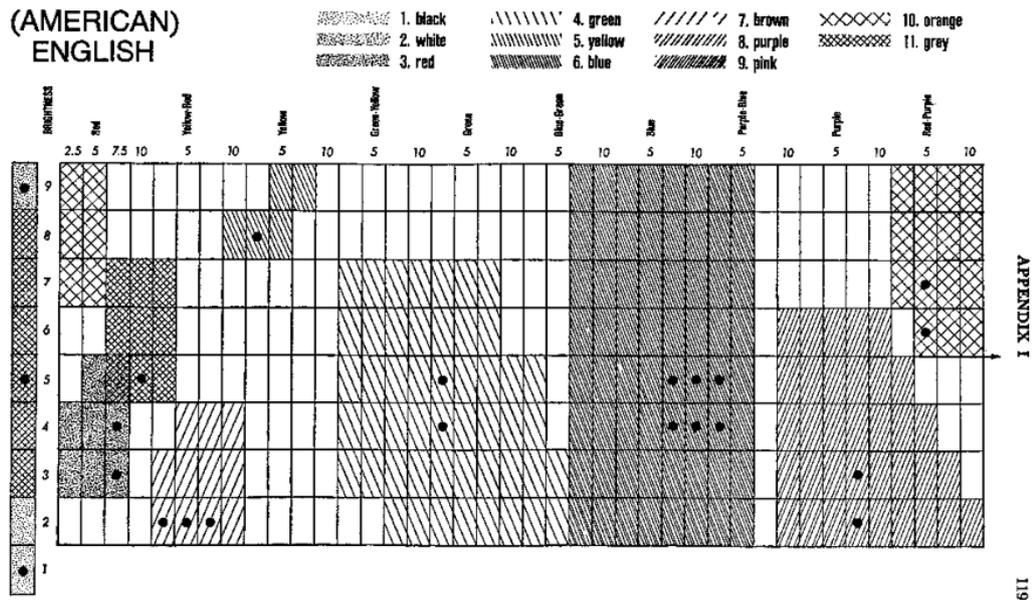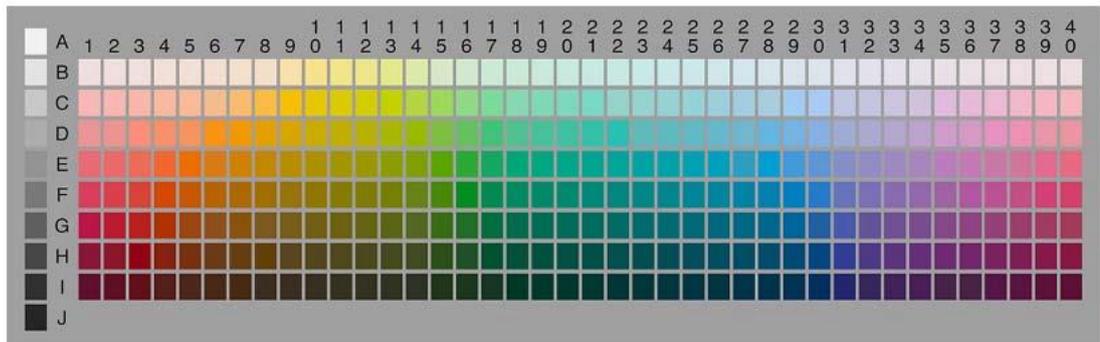
Figure 2.10. Colour Array of Munsell value vs hue (top). Basic Colour Terms in American English mapped on the Munsell array (bottom; Berlin and Kay, 1969/1991)

The World Color Survey (WCS) was initiated in the late 1970's to test the hypotheses advanced by Berlin and Kay regarding: (1) the existence of universal constraints on cross-language colour naming; and (2) the existence of a partially fixed evolutionary progression according to which languages gain colour terms over time. Colour naming data was collected in 110 unwritten languages using the same stimulus palette (Figure 2.10) for a constrained naming task and a focus (best example) task. Kay and his colleagues (Berlin & Kay, 1969/1991; Kay & Regier, 2003; Kay et al., 2010) demonstrated that different languages tend to classify the surface of the Munsell system in a similar way, though with some differences. One of the most intriguing finding was that these inter-language differences were smaller than intra-language differences among individuals (Berlin & Kay, 1969/1991; Webster & Kay, 2007). The analysis of the data supported the conclusion that the colour space is partitioned under universal constraints and detected a hierarchical order in the lexical partitions (Kay et al., 2010).

Boynton and Olson (1987) conducted a colour naming experiment in American English to locate the referents of the BCTs in the OSA space. The experiment involved 424 uniformly spaced colour samples, presented against a neutral grey background of 20% reflectance under a photoflood lamp of 3,200K. Response times (RTs) were measured from the onset of the stimulus to the start of the subject's vocalisation. Six observers were asked to use solely monolexemic colour terms and a seventh was instructed to restrict his responses only to BCTs. Their study showed that the 11 BCTs were used more frequently, more consistently, with greater consensus and more quickly than non-BCTs. The differences between individual were large but there was no overlap between the loci of the BCTs. The authors also suggested an emergent twelfth BCT in the region between white, yellow, orange, pink and brown. The word most frequently used for that region was peach, but it was not qualified as a BCT.

In a follow-up study, Sturges and Whitfield (1995) located colour terms in the Munsell system for British English (Figure 2.11). The experiment involved 446 colour samples presented randomly against a neutral grey background of Munsell N7 (matte) under a CIE D65 simulator. Their results confirmed that BCTs have shorter response times and higher consistency and consensus than non-BCTs. An interesting finding was that purple ranked third in terms of consistency and frequency, along with short response times, and appeared to cover a larger area of the Munsell than the OSA space. Cream was suggested as a candidate for a twelfth BCT, as it was used frequently and consistently but with a clear differentiation from the other 11 BCTs. The different sampling, as well as the different illuminant under which the experiments were carried out, may explain the relatively large mean colour difference (ΔE_{ab}=13.92) between the location of the chromatic basic colour terms of this study against the findings of Boynton & Olson (1987).
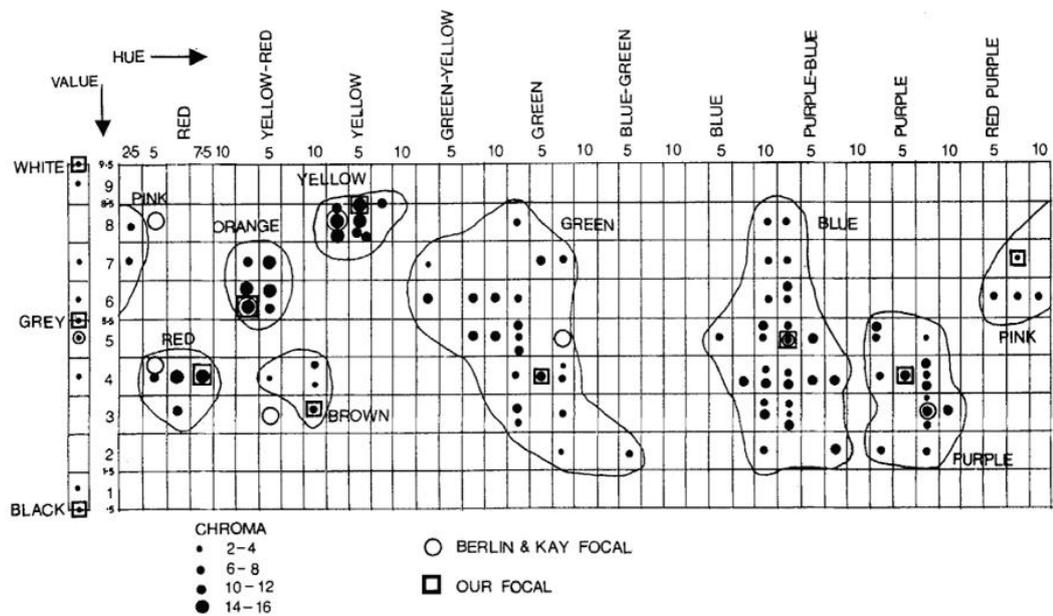
Figure 2.11 Location of focal samples of Basic Colour Terms on Munsell array reported by Berlin & Kay (1969/1991; discs) and Sturges & Whitfield (1995; squares).

Davies & Corbett (1995) proposed a faster method of identifying BCTs. The procedure included two tasks: firstly, an elicited list task, and secondly, a process of mapping the names onto a set of 65 colour tiles to measure the frequency of colour names and the consensus across respondents respectively. For consensus, the authors adopted the dominance index and the specificity ratio described in Moss et al. (1990). Dominance index represents the number of colours for which each colour name was used by more than 50%, while the specificity ratio is calculated by dividing the total number of dominant responses of each name by their frequency to produce a stability scale. The authors estimated saliency of a colour term as a combined index of both the frequency of the term in the listing task and the consensus in the mapping task.

Moroney (2003) used 'distributed psychophysics', to collect a small number of colour names from a large number of observers over the web. Participants were asked to give the best names for seven patches of colours selected randomly from a 6 × 6 × 6 non-perceptually uniform grid sampling of the RGB cube, viewed on a display against a white background. Results of the online experiment were validated against the results of Boynton and Olson (1987) and Sturges and Whitfield (1995), both obtained under controlled laboratory conditions, and showed a high degree of correlation with the chromatic basic colours terms, expressed as hue angles in CIELAB.

In 2009, we launched an ongoing colour naming experiment to collect broad sets of colour names within different languages with their corresponding colour ranges in sRGB and Munsell specifications over the Internet (Mylonas & MacDonald, 2010). The colour

naming responses are associated with metadata regarding the cultural background, colour deficiency, hardware/software components and viewing conditions of the observers. The experiment was initially translated in three languages, English, Greek and Spanish. In earlier studies, we presented the results of the English data collected in the period 2009-2010 (Mylonas & MacDonald, 2010; Mylonas, MacDonald & Wuerger, 2010; Mylonas & MacDonald, 2016). From 2009 to 2018, the experiment was translated into twenty-two languages (English, Greek, Spanish, German, Catalan, Italian, traditional and simplified Chinese, Korean, French, Danish, Lithuanian, Thai, Portuguese, Swedish, Russian, Japanese, Turkish, Vietnamese, Dutch, Norwegian and Polish) and has gathered colour naming responses from many thousands of observers (until February 2018, *n*=7,000). In this thesis, we consider – for the first-time – responses in American English, British English, Greek, Russian, Thai and Turkish for which up to 2018 we collected 5,000+ responses in each language. In Chapter 3, we provide details regarding the experimental procedure and the produced datasets. In 2018, the interface of the ongoing experiment was redesigned to run on all new devices and minimize security threats (accessible at: https://colournaming.org). It is currently collecting colour naming responses in 11 languages but the data from the new interface will be considered in future studies.

In a parallel online colour naming experiment, the author of the web comic XKCD, Randall Munroe (2010) collected a dataset of 3.4 million unconstrained responses mainly in English. This very large dataset is associated with metadata regarding the sex, language skills, colour-blindness, and information about the monitor settings. Observers were free to name as many sets of colour swatches as they liked presented against a white background. Each colour swatch was uniformly sampled from the full RGB cube. The dataset was made available to the public and was considered for training colour naming models by a number of subsequent studies (Heer & Stone, 2012; Lindner et al., 2012).

2.8. Colour naming models

A wide variety of colour naming models have been proposed to facilitate colour communication. In 1955, the National Bureau of Standards (NBS) published the ISCC-NBS dictionary of colour names, based on the recommendations of the Inter-Society Color Council (ISCC), in order to facilitate colour communication between different colour vocabularies in the fields of art, science and industry (Kelly & Judd, 1955). This dictionary consists of 7,500 English colour names that represent 267 regions of the Munsell colour system. The dictionary is limited by the lack of systematic syntax of the system and its

specialised vocabulary. Nevertheless, this method of designating colours is an important documentation of colour naming and has inspired new colour naming systems to this day.

The Colour Naming System (Berk, 1982) was designed with the intention to simplify the syntax of ISCC-NBS system. Hence it was formed by the same lightness and saturation values as the ISCC-NBS, except for the combined modifiers of *brilliant*, *pale* and *deep*. Additionally, the hue terms have been considerably simplified down to the basic colour terms of Berlin & Kay (1969/1991), with the exception of the pink term. This modification provides the CNS with a formal syntax so that the intermediate hues can be defined by the combination of all adjacent hue terms to generate a total of twenty-four hue names. The CNS system encodes in total 627 colour names quantised in *HSL* colour space, of which only 480 were actually realised, because of its cylindrical shape. Tominaga (1985) described a colour naming method for predicting the colour name of digitised colour samples. The colour-naming system was structured systematically with basic colour terms and modifiers to different levels of accuracy (see Figure 2.12). At level 1, the Munsell colour solid is assigned to 16 colour terms, at level 2 to 25 colour terms, at level 3 to 92 colour names and lastly at level 4 to 236 colour names. The reported reliability of the system was satisfactory up to level 3.
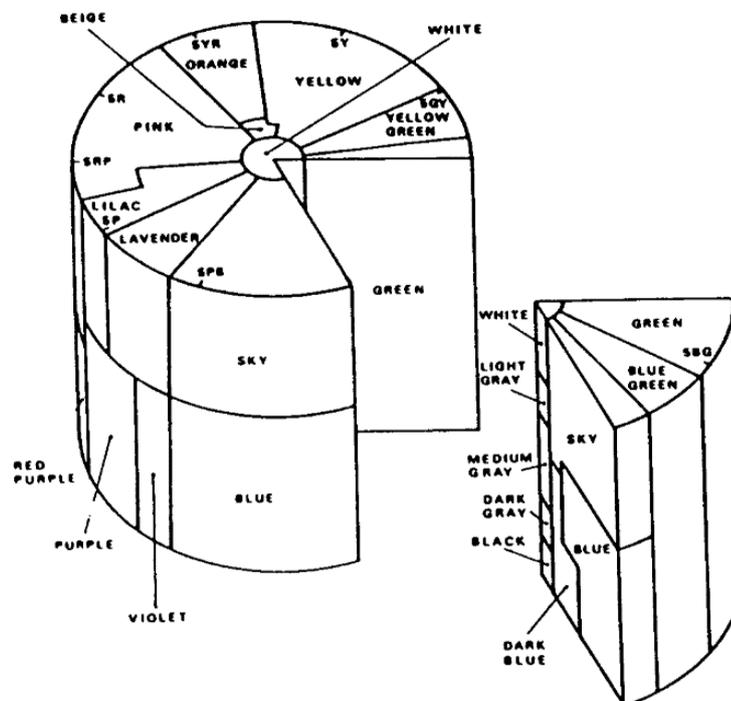
Figure 2.12 Colour naming block at Level 2 (Tominaga, 1985).

The computational colour naming model of Lammens (1994) was influenced by the fuzzy set theory of Kay & McDaniel (1978). Every colour was assigned to one of the eleven focal colour locations of basic colour terms (Berlin & Kay, 1969/1991) with a variant of a Gaussian distribution model:

$$G_n(x) = e^{-\frac{1}{2}\left(\frac{\sqrt{\sum_{i=1}^{N}(x_i-\mu_i)^2}}{\sigma}\right)^2} \tag{2.29}$$

where x-μ is the Euclidean distance of a point x to the mean μ of a colour category.

A categorical colour mapping method was proposed by Motomura (1997) for cross-media colour reproduction. The most interesting characteristic of this approach is the maintenance of an identical colour name in both source and destination mediums, while preserving the relative relationship of the colours of each categorical cluster. To determine the categorical classification of a colour to one of the eight chromatic basic colour terms plus an achromatic category, the author utilized a set of *Mahalanobis distances:*

$$D_i = \sqrt{(X - \mu_\iota)^t \sum_i^{-1} (X - \mu_\iota)} \tag{2.30}$$

where X is a test colour, μ_ι is an average vector of colour categories and Σ_i the covariance matrix of their distribution.

Lin, Luo, MacDonald, Tarrant (2001a, 2001b) focused on the boundaries of each colour category instead of the location of the prototypes of each colour category. The distribution of the eleven basic colour terms was determined by a combination of unconstrained and constrained experimental data of English and Chinese subjects, according to which the borders of each category were confined by crisp thresholds. Seaborn et al., (2005) proposed a nonparametric model based on fuzzy k-means algorithm to measure similarity and dissimilarity between colours based on Sturges & Whitfiled (1995) data. Menegaz and her colleagues (2006) proposed a fuzzy partitioning of the colour space into the eleven categories of BCTs based on linear interpolation of membership functions obtained in psychophysical experiments. The test stimuli consisted of 424 samples from OSA-UCS set, each of which is associated to a vertex of a three-dimensional tetrahedron via a three-dimensional Delaunay triangulation of the CIELAB. Benavente et al. (2006; 2008) extended the computational model of Lammens

(1994) with a combination of Gaussian-Sigmoid distribution functions. Similarly, Parraga and Akbarinia (2016) proposed a physiologically inspired model of colour categorisation by concentrating on the boundaries of the basic colour terms based on Neural Isoresponsive Colour Ellipsoids (NICE) in a cone contrast space.

Figure 2.13 shows the currently state-of-art performance of NICE classifying the surface colours of the Munsell system against psychophysical data (Berlin & Kay, 1969/1991; Sturges & Whitfield, 1995). An alternative, and considerably quicker approach for training colour naming models was described by Weijer et al., (2007). The authors estimated the colour distribution of the eleven basic colour terms from image statistics instead of colour naming experiments using Probabilistic Latent Semantic Analysis (PLSA):

$$PLSA - bg = P(z|w) \propto P(z)P(w|z) \tag{2.31}$$

where $P(z|w)$ describes the likely image pixels z that word w is referring to, and $P(w|z)$ describes the distribution where image pixels z may be labelled by word w for classifying single pixels to colour names. The images for each colour term were retrieved using the Google Images API.
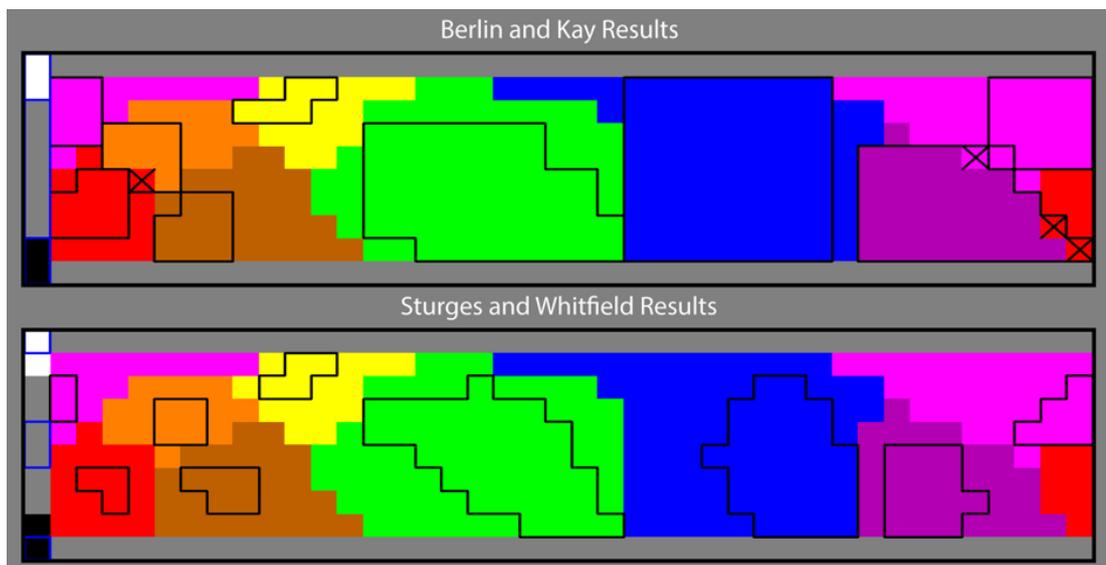
Figure 2.13. Surface colours of the Munsell system categorised into eleven basic terms by NICE (Parraga & Akbarinia, 2016) and Berlin & Kay's (1969; top) and Sturges & Whitfields's (1995; bottom) psychophysical data shown by black lines. Cross marks denote differences in naming by the NICE model and Berlin & Kay's results.

Mojsilovic (2005) proposed a computational method for categorizing, naming, and extracting colour compositions of images. The author adopted the 267 colour regions of the ISCC-NBS system and employed a nearest neighbour classifier based on perceptual

distances in CIELAB. Moroney (2008) proposed a lexical classification method in which non-parametric histograms are used to represent the colour range of each colour name. The lexical processing algorithm functions on the *n* most frequently used colour names obtained from an ongoing web-based colour-naming experiment with thousands of participants. In 2008, Chuang et al. proposed a non-parametric probabilistic method to model the categorical association between colours:

$$P(C|c) = \sum_w P(C|w)P(w|c) \tag{2.32}$$

where *P(W|c)* is the conditional probability for each word *w* that have be chosen to describe colour *c*, and *P(C|w)* the conditional probability of colour *c* being the cause of a word *w*. This model considers all possible words for describing a colour as a probability distribution, and as such is robust against the noise in the colour naming data set of unconstrained colour naming experiments. Heer & Stone (2012) extended this probabilistic model of colour naming using data from an online survey (Munroe, 2010) to develop a colour dictionary and colour selection and editing applications. The authors measured the degree that a colour is unique named by the entropy of the conditional probability *P(W|c)* defined as saliency:

$$Saliency(c) = -H\big(P(W|c)\big) = \sum_w p(w|c) \log p(w|c) \tag{2.33}$$

Mylonas et al. (2010) proposed a probabilistic interpretation of Mahalanobis distances to automate the assignment of colours to a large number of common colour names (n=47) based on a Maximum a Posteriori estimator (Figure 2.14). For each colour name y from a set of colour names $y_1,....,y_T$ offered by the observers for colour samples viewed against a neutral grey background, they calculated the empirical mean m_y and covariance matrix S_y of test colours $x_1,....,x_n$. The probability density function was then estimated by:

$$\hat{f}_{norm}(x|y) = const_y exp\left(-\frac{1}{2}(x - \mu_y)^T \sum_y^{-1}(x - \mu_y)\right), x \in \{x_1, ..., x_n\} \tag{2.34}$$

where x is the test colour specified by the triplet $x = (x_{(L)}, x_{(a)}, x_{(b)})^T$ and $const_y$ is a normalizing factor that depends on m_y, S_y and $x_1,....,x_n$ and ensures that the sum of the

probability distribution is equal to 1. Using the Bayes' theorem, the MAP estimator was then defined as:

$$\hat{y}_{MAP}(x) = \underset{y \in \{y_1, \dots, y_T\}}{\mathrm{argmax}} \left(\frac{\hat{f}_{norm}(x|y)\hat{f}(y)}{P(X = x)} \right)$$

(2.35)

The MAP estimator favours colour names with high probability to maintain congruence between observed and predicted data. This means that frequent and consistent colour categories tend to subsume less common and inconsistent neighbour categories.

Figure 2.14. Segmentation of synthetic image to colour categories by MAP estimator with training set of 47 colour names (Mylonas et al. 2010).

A similar approach to Weijer et al., (2007), was used by Lindner et al. (2012) to develop a multilingual colour thesaurus of 9,000 colour names. The authors translated manually, via dictionaries and single native speakers, 900 English colour names from an online colour survey (Munroe, 2010) to ten different languages and used a statistical framework to determine their colour distribution via the Google Images API. Recently, a convolutional neural network approach trained on thousands of hand-labelled images, was proposed for identifying pedestrians in public spaces, by classifying the colour of their clothes to the eleven BCTs (Cheng, Li & Loy, 2016). In the same domain of person re-identification, Yang et al., (2014) obtained satisfactory results using 16 terms while Yu et al., (2018) showed that considering 39 colour names in total outperformed earlier colour name descriptors on this task.

2.9. Information theory and colour language games

Information theoretic analysis (Shannon, 1948) has also been used to shed light on the origin of colour lexicons in the context of language games (Wittgenstein, 2009/1953).

Language games were first introduced in colour naming studies by Lantz & Stefflre (1964) and were revived by recent computational approaches (Steels & Belpaeme, 2005; Loreto et al., 2012; Regier, Kemp & Kay, 2015; Lindsey et al., 2015; Gibson et al., 2017).

In communication theory (Shannon, 1948), the entropy equation provides the total amount of information content in an entire probability distribution of an event *x* based on its frequency:

$$H(X) = -\sum_{i=1}^{n} P(x_i) log_2 P(x_i)$$

$$(2.36)$$

where log_2 produces units of entropy in bits. Lantz & Stefflre (1964) measured communication accuracy using a language game of encoders and decoders where subjects asked to assign a colour name to a colour from the Farnsworth-Munsell array in such a way that another person will be able to pick it up. The authors found that communication accuracy performed better than naming agreement to predict memory colours. Its high correlation with codability (Brown & Lenneberg, 1954) provided further evidence for the influence of language on non-linguistic behaviours. Steels & Belpaeme (2005) considered three philosophical propositions - nativism, empiricism and culturalism - as formal artificial intelligence (AI) models to explore the mechanisms of communication of a set of perceptually grounded colour categories within a population of autonomous agents. The authors showed that model-based colour categories between agents converge to allow effective communication. In other words, colour categories are achieved collectively through the communication process; therefore, supporting the influence of language in the formation of colour concepts.

Partial evidence for the evolutional hierarchy of BCTs proposed by Berlin & Kay (1969/1991) was provided by Loreto and his associates (2012) using simulated language games between multiple agents. The authors found that agents achieved first consensus for colour names based on just-noticeable differences of specific regions of the hue dimension. In other words, they concluded that the perceptual structure of colour space can partly explain the hierarchy of colour categories. Agreement between pair of agents was measured by:

$$\frac{2 \sum_{i=1}^{N} \sum_{j=i+1}^{N} match(i,j)}{N(N-1)}$$

$$(2.37)$$

Regier et al., (2015) showed that the six primary colour categories (Regier et al., 2007) may be formed from communication pressures between simplicity (the number of colour terms) and informativeness (the precision of colour terms) using simulated language games based on multilingual data from WCS (Kay et al., 2010) and Kullback-Leibler Divergence to measure the information lost in communication:

$$D(s \parallel l) = \sum_{i \in u} s(i) log \left(\frac{s(i)}{l(i)} \right)$$

(2.38)

where *s* is the actual colour naming distribution and *l* is an approximation to *s* for all colours *u.*

Lindsey and colleagues (2015) used information theoretic analysis and simulated languages games based on colour naming data in Hazda, Somali and American English to show that even the smaller colour lexicon of a hunter-gatherer population aligns with other world languages. The authors measure the communication cost by mutual information:

$$GMI(C_S; C_R) = \sum_{s,r} p_N(s,r) log_2 \left(\frac{p_N(s,r)}{p_N(s)p_N(r)} \right)$$

(2.39)

where C_R are the test samples in the language game, C_S the utterances by the speaker for the test samples, p_N(s,r) is a matrix of a joint distribution of the random variables C_S and C_R and p_N(s), p_N(r) are the marginal distributions on C_S and C_R. Gibson et al., (2017) showed that surprisal, an information theoretic measure of communication cost, tends to be higher for cooler than warmer colours reflecting according to the authors the usefulness of a colour. Surprisal is computed by:

$$S(c) = \sum_{w} P(w|c) log \frac{1}{P(c|w)}$$

(2.40)

where the communication cost for each colour *c* is measured by summing the cost for each word *w* that might have be chosen to describe colour *c* multiplied by the log of the probability that colour *c* was the cause of a word *w*.

2.10. Discussion

In this chapter, we reviewed the main components of this thesis including colour vision mechanisms, their associated classes of colours, colour specification systems, colour naming experiments and models and information theoretic approaches in the context of colour language games.

We have seen that the primary focus in colour vision research is the psychophysical specification of different mechanistic stages of colour processing in the visual pathway. The first stage defined by the spectral sensitivities of the three types of cones in the retina is now reasonably well understood but the relationship between the second stage of colour discrimination/detection mechanisms with the hypothetical third stage of colour appearance and the higher-order mechanisms of colour constancy and colour naming remains elusive. Doubts have also been raised about the fundamental status of the primary colours associated with the colour opponent mechanisms in the development of colour naming systems whilst the quest for a cross-culturally legitimate approach to identify basic colour categories within different languages remains unsettled.

The Munsell system has been the gold standard in colour naming research but there is no straightforward way to map its coordinates to physical properties. The OSA space provides an exceptional geometry of equidistant colours but the ability to sample colours in high chroma is limited. In colourimetric spaces, we can represent all visible colours and because they are based on psychophysical measurements, they can be reasonably mapped on different stages of colour vision processing. For example, the new CIE XYZ 2015 offers a chromaticity diagram with physiological axes as it is a linear transformation of the LMS cone fundamentals. The DKL space can be used to represent the second stage of colour processing and it is closely related to the CIE uniform colour spaces that are widely used to measure colour differences. Colour appearance models extent basic colorimetry and allow – to some extend – the prediction of how observers match colours under different viewing conditions. Finally, the practical success of sRGB in all areas of colour reproduction allows researchers to conduct colour naming research over the Internet.

We described how earlier colour naming studies used these colour specification systems in the field and in laboratory settings and how information communication technologies have enabled new methods for collecting unconstrained colour naming data by crowdsourcing over the Internet. Despite the usefulness of experiments with carefully calibrated displays in the field of vision science, experimenters in colour

naming studies face several limitations. First, the controlled viewing conditions, while advantageous for accurate colorimetric specification of stimuli, limit the ecological validity of the predicted colour naming functions in real-world settings. Second, the pool of available observers is often constrained to a small number of college students or to the authors of the study that makes it difficult to generalize the results to heterogenous groups of the population. The large individual differences reported in colour naming studies makes the generalization of the experimental findings even more problematic (Webster & Kay, 2007; Lindsey & Brown, 2009). Furthermore, the constrained colour naming method is able to capture only a small fraction of the richness of colour languages of the world.

In contrast, web-based colour naming experiments provide greater ecological validity than traditional approaches by allowing simultaneous participation of observers in their own familiar space, in their own time, with their own equipment and without the physical attendance of the examiner (Reips, 2000; Moroney, 2003, Mylonas & MacDonald, 2010; Munroe, 2010). A further methodological improvement includes the departure from usual methods which would use a small number of observers and/or the use of only a restricted set of monolexemic terms. Instead, thousands of observers from linguistically and demographically diverse populations name freely a large number of colours online and produce larger colour lexicons that improve the precision of colour names in colour space. Online methods also depart from previous research by distributing the colour naming task to minimize the influence of single individuals in estimating colour naming functions at a population level. In the collection of multilingual colour naming data, online experiments also extend earlier cross-cultural studies which used only the most saturated colour samples on the surface of the Munsell system (Berlin & Kay, 1969/1991; Roberson et al., 2005; Kay et al., 2010; Lindsey & Brown, 2014; Gibson et al., 2017), by also sampling the interior of the colour solid. Yet, online experimental methodologies often receive criticism as not meeting the exacting standards demanded for rigorous vision research because of the uncalibrated colour reproduction and viewing conditions. These unknown confounding factors contribute to a lack of complete quality control and the full potentials of crowdsourcing in colour research and applications remain in an infant stage. To respond to these criticisms, a direct comparison between web- and laboratory- based experimental methodologies in estimating colour naming functions in calibrated and uncalibrated settings respectively is essential.

Computational methods are also becoming important in reproducing colour names that are meaningful to human observers (Harnard, 1987). The majority of earlier efforts constrained the focus of their research towards a small number of basic colour terms in

a single language rather than towards the development of colour naming systems that support subtler colour identifications in multiple languages described in this thesis. The idea that information theory can provide a better framework to advance our understanding in colour naming is gaining traction in the field. We described the use of information theoretic analysis to shed light on the basis and development of colour lexicons in the context of language games but – to our knowledge – there are no earlier reports of the inverse game that we describe in the last chapter of this thesis where the guess is about the name rather than the colour.

In order to move towards the extension of human-artificial intelligence in the field of colour communication within different languages for basic and applied science, we will need to design data-driven colour naming systems that support the full complexity of colour languages across the world, including their relationship to physical, psychophysical and physiological aspects of colour. The work presented in this thesis extends earlier work in this direction.

Chapter 3
Online colour naming experiment

People use a large number of colour names to communicate about colours. Each name may consist of any number of words, such as *yellow*, *salmon pink* and *light periwinkle blue.* As we have seen in the literature review, multilingual data about unconstrained colour names and their colour referents across the full colour gamut is limited. In 2009, we designed an ongoing colour naming experiment to collect unconstrained colour names in English, Greek and Spanish with their corresponding regions across the full colour space over the Internet. In this thesis, we consider much larger colour naming datasets in American and British English, Greek, Russian, Thai and Turkish collected up to 2018. A description of the methods used in this study and its participants, will be followed by the data analysis techniques to obtain lexical, behavioural and geometric features of colour names from raw responses. In closing, we compare the location of BCTs in the above six language-datasets.

3.1. Materials and procedure

An online colour naming experiment was designed to collect a small number of unconstrained colour naming responses from a large number of participants. Participation was voluntary and anonymous, and the experimental sessions were conducted after obtaining online informed consent (Varnhagen et al., 2005). In earlier studies, we presented the results of the English data collected in the period 2009-2010 (Mylonas & MacDonald, 2010; Mylonas et al., 2010; Mylonas & MacDonald, 2016). In this thesis, we consider for the first-time responses in American English, British English, Greek, Russian, Thai and Turkish for which up to 2018 we collected 5,000+ responses in each language. The colour naming responses are associated with metadata regarding the cultural background, colour deficiency, hardware/software components and viewing conditions of the observers.

Six hundred colour samples (see Figure 3.1) were selected from the Munsell Renotation Data set (Newhall et al., 1943), including eleven additional achromatic (greyscale) samples. The colour samples were specified in the sRGB standard colour space for the Internet, and out of gamut colours were removed. To achieve approximately uniform sampling, we followed the suggestions of Billmeyer in Sturges and Whitfield (1995); see

also Mylonas and MacDonald (2010). The colour stimuli were presented against a neutral grey background with a black outline of 1 pixel. Stimulus size (width by height) on the display was 147 by 94 pixels, which for a display resolution of 3.3 pixels per mm (83 pixels per inch) would be 45 by 30 mm, subtending an angle of approximately 5 by 3.4 degrees at a viewing distance of 50 cm.
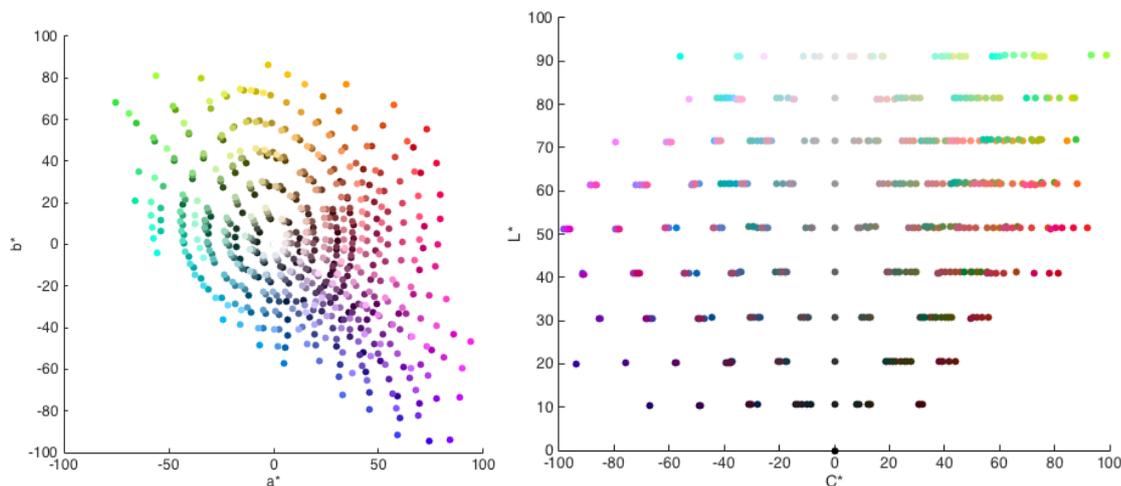
Figure 3.1. Colour stimulus set (n=600) of online colour naming experiment in CIELAB space, a* vs. b* plane (left) and L* vs. C* plane (right).

The experimental procedure consisted of six steps (see Figure 3.2). Depending on their expertise in colour technology, observers were first asked to set their display to sRGB settings either through an advanced instrument-based calibration procedure, or by a given basic set of instructions to adjust the manufacturer settings and the brightness of their monitor, so that all twenty- one steps of a grey scale ramp were visible. In the second step, participants answered questions related to their lighting conditions, their environment and properties of their display. In the third step, we screened our participants for possible colour deficiencies with a web-based Dynamic Colour Vision Test developed at the City University London (Barbur, Harlow & Plant, 1994).

The fourth step involved an unconstrained colour-naming task: each observer was presented with a sequence of 20 randomly selected colours from the 600 total samples. Hence, our randomised order of single colour patches (Roberson, Davies & Davidoff, 2000) avoided bias introduced by constant large target arrays (Berlin & Kay, 1969/1991; Kay et al., 2010) or by making a small constant selection for all observers (Lindsey et al., 2015). The task instructions were always visible at the top of the screen. Observers were asked to type in the name of each colour patch with the most representative colour name that they could remember. To estimate naming consistency, one colour sample was

presented twice with more than ten stimuli between the repetitions but without notifying participants about the recurrence of the stimulus. Response times were measured from the onset of the stimulus to the observer's first keystroke of the typed colour name. The next colour sample appeared after pressing return or clicking on a submission icon button. Observers were informed that the response time would be recorded. The web interface also included two questionnaires to collect information about the viewing conditions, display properties and cultural background of each participant.

In the fifth step, we collected information about participant's country of residence, nationality, language proficiency, educational level, age, gender and experience in colour applications. Finally, in the last step, participants are provided with a summary of their responses and an optional communication form for comments. The communication form is detached from the experimental data to ensure the anonymity of the participants. Screenshots of the interface are available in Figure A.1.
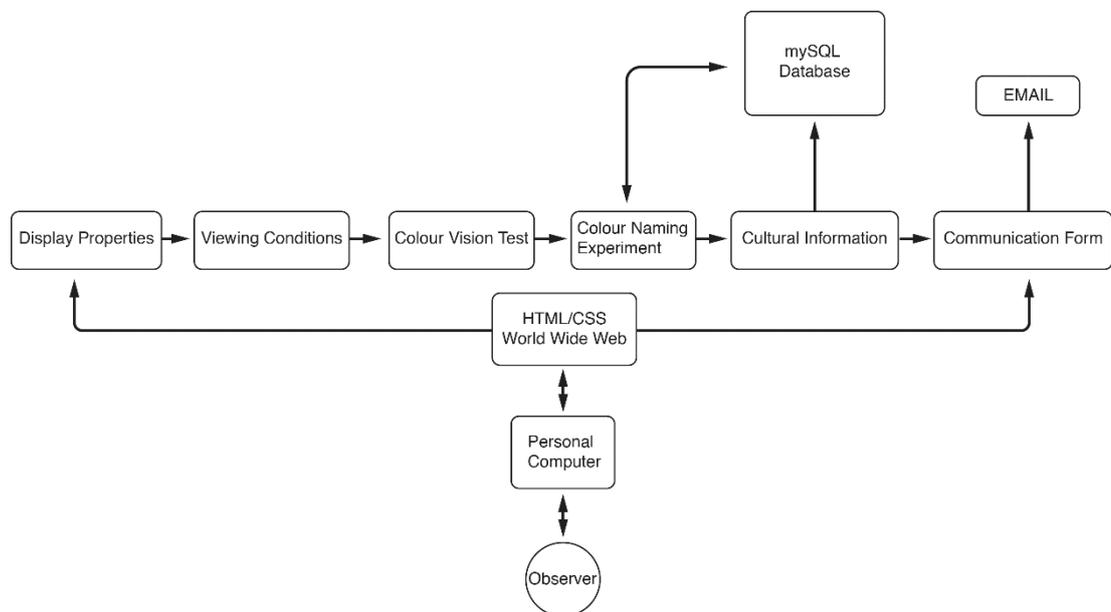
Figure 3.2. Schematic diagram of the web-based colour-naming experiment.

3.2. Participants

In this chapter, we present the colour naming responses in British English. The presentation of the American English, Greek, Russian, Thai and Turkish data can be found in Appendix A. We retrieved 10,000 raw responses from 500 British English observers of the online colour naming experiment. We excluded disruptive observations, for example numerical responses or responses in languages other than the language of the instructions (1%), and for further analysis only responses from participants with no self-reported colour deficiencies (90.3%) were considered. This filtering resulted in a

dataset for 447 respondents. Their mean age was 33 years (SD = 13 years). Females provided 63% of the responses while males provided 37%.

3.3. Data Cleaning

In the raw responses, typographic conventions and leading/trailing spaces were removed. Hyphenated, comma separated and words in parenthesis were treated as multiword colour expressions. Different word orders (i.e. orange-red or red-orange) were considered as different names. Incomplete, numerical and responses written with characters of languages other than the language of the instructions of the experiment were excluded from the analysis. All capital letters were converted to lower case. Finally, a supervised semi-automatic spell-checking procedure was performed by a native speaker before the data analysis. We scrutinised only distinct colour names (n=478) given by two or more observers resulting in 7,405 responses. Unique responses from single observers were excluded because we could not be confident that other observers will understand the colour name used and therefore these responses were considered idiosyncratic.

3.4. Number of Words

The occurrence of colour descriptors with varying word number for British English speakers was: monolexemic BCT 30%; monolexemic non-BCT 24%; two-word colour names 40% and colour descriptors containing ≥ 3 words 5% (Figure 3.3).
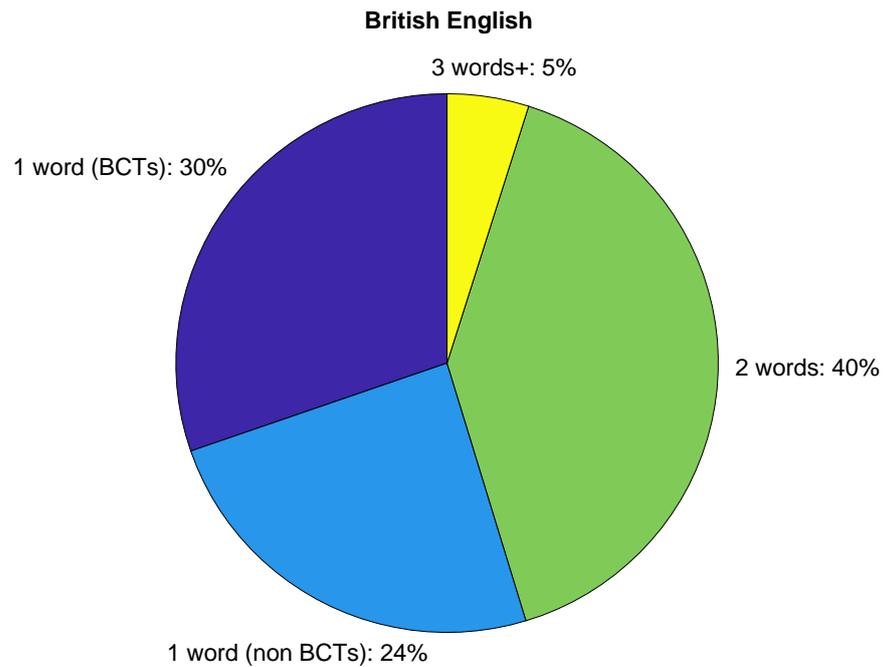
British English

Figure 3.3 Number of words in colour descriptors for British English speakers.

3.5. Linguistic features

In this section we present linguistic measurements of word length and derivative forms of colour names from the online colour naming experiments but also their frequency in linguistic data obtained from Twitter.

3.5.1. Words length

Colour name length, qualified as number of letters in all words of the name has been previously coincided with phonetic length and negatively correlated with frequency of usage and information content (Zipf, 1935; Piantadosi et al. 2011). The basic *red* and the non-basic *tan* were the colour names with the shortest length (Figure 3.4). *Purple*, *yellow* and *orange* were the basic terms with the longest length and were not ranked in the top 30 positions.
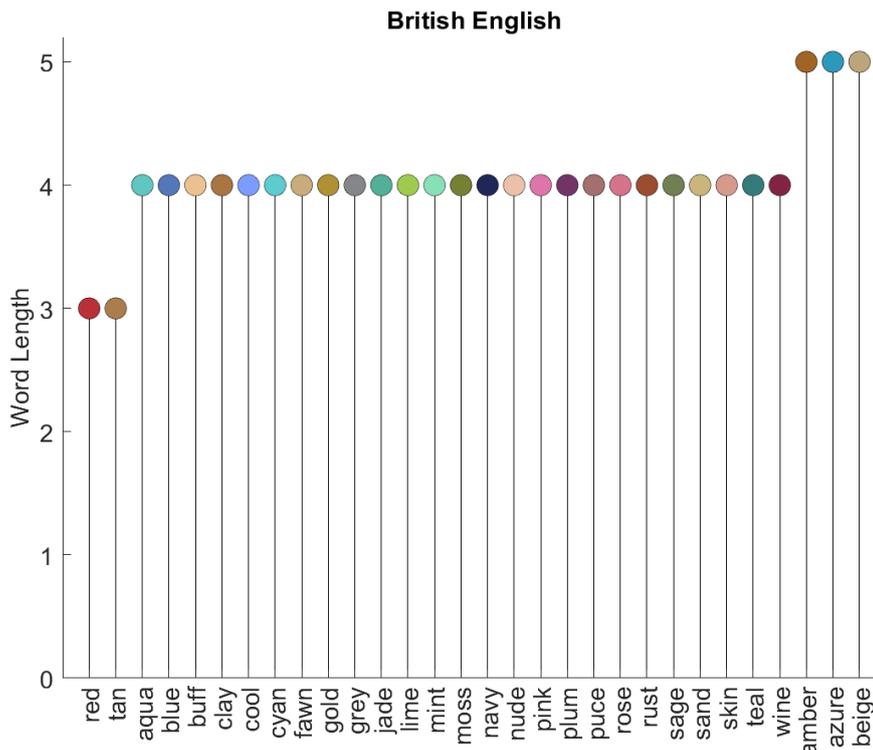
Figure 3.4 Top 30 colour names with shorter word length for British English speakers. The coordinates of their centroids were used to colour each colour name. Colour reproduction may vary depending on the medium.

3.5.2. Number of derivative forms

Derivative production is a measure of the number of derivative types of a colour name in colour naming responses (Corbett & Davies, 1997; Kerttula, 2007). This includes, for example, the suffix –ish (e.g. greenish) or –er (e.g. greener) and compound colour words (e.g. light green or sea green). *Green* was found with the largest number of derivative forms followed by *blue* and *pink* (Figure 3.5). *Turquoise* and *lilac* were the non-basic terms with largest derivative production in the 10th and 12th position. *Black* (21st) was the basic colour term with the smallest number of derivative forms. The high rank of *bluish* can be explained by the common use of this term as modifier of blue tinted categories by the observers.
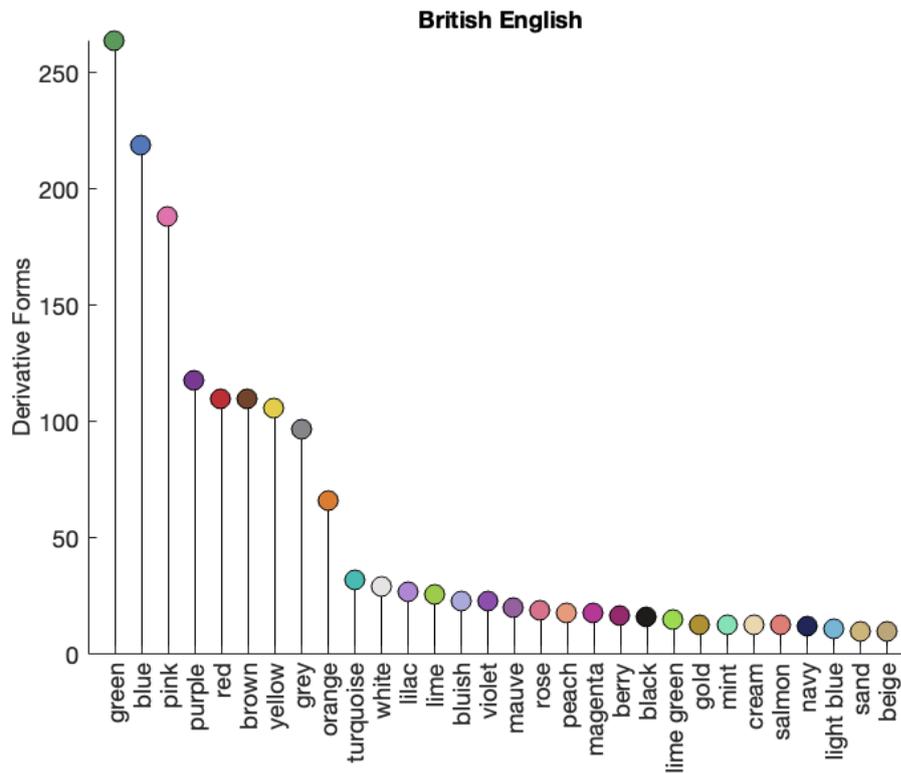
Figure 3.5 Top 30 colour names with highest number of derivative forms for British English speakers.

3.5.3. Linguistic frequency

Linguistic frequency is a measure of the usage of a colour name in a literary language (Hays et al., 1972). To examine the frequency of British English colour names in everyday online conversations, we measured their probability of occurrence in 1,036,103 random tweets from the Twitter API. Similarly, to the online colour naming experiment, messages in Twitter are given voluntarily and provide greater volume and variability than other sources (Corbett & Davies, 1997). We consider this dataset as more representative of ordinary language use. We filtered Twitter's public stream with the geo-location coordinates [-5.4, 50.1, 1.7, 55.8] that correspond to a rectangle with its edges approximately at the edges of Britain. We excluded tweets in other languages than English {'lang':'en'}. Each tweet was tokenised using the Natural Language Toolkit (Bird, Klein & Loper, 2009). *Black* followed by *white* and *red* were the most frequent colour names in Twitter (Figure 3.6). The 11 basic colour terms were found in the top 12 positions. The basic term *orange* ranked in the 7th and the non-basic term *cream* ranked in the 4th position but they can also be used in a non-colour sense.
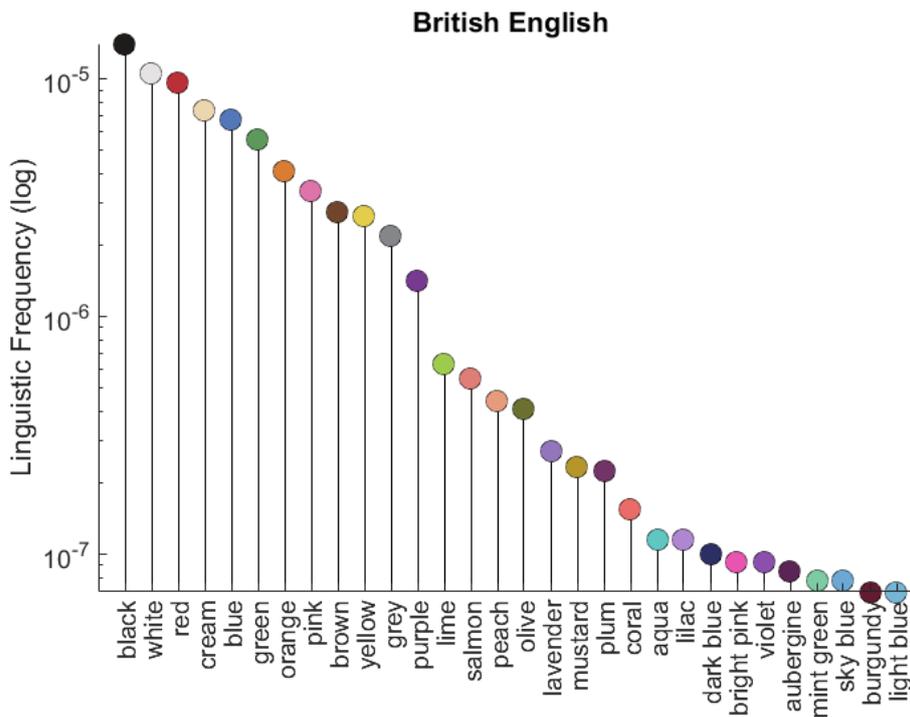
Figure 3.6 Top 30 most frequent colour names in Twitter for British English speakers.

3.6. Behavioural features

In this section we present the behavioural measurements of frequency of occurrence, response time and consensus of colour names in the dataset obtained from the online colour naming experiment.

3.6.1. Frequency of use in colour naming experiments

Frequency in colour naming experiments quantifies the total number of times that each colour name was used to describe any colour stimuli by all observers (Boynton & Olson, 1987; Sturges & Whitfield, 1995). This is determined by the extent of the category on colour space, and by the rate at which colours within this extent are named as such. *Purple* was the most frequent colour name followed by *pink*, *blue* and *green* (Figure 3.7). *Green* was nearly twice as frequent as the next most frequent colour *brown*. The least frequent basic term was *white*, found in the 22nd position and *red* in the 15th. The non-basic terms *lilac* and *turquoise* were found in the 6th and 7th positions respectively while *violet* and *light blue* were in 11th and 12th positions.
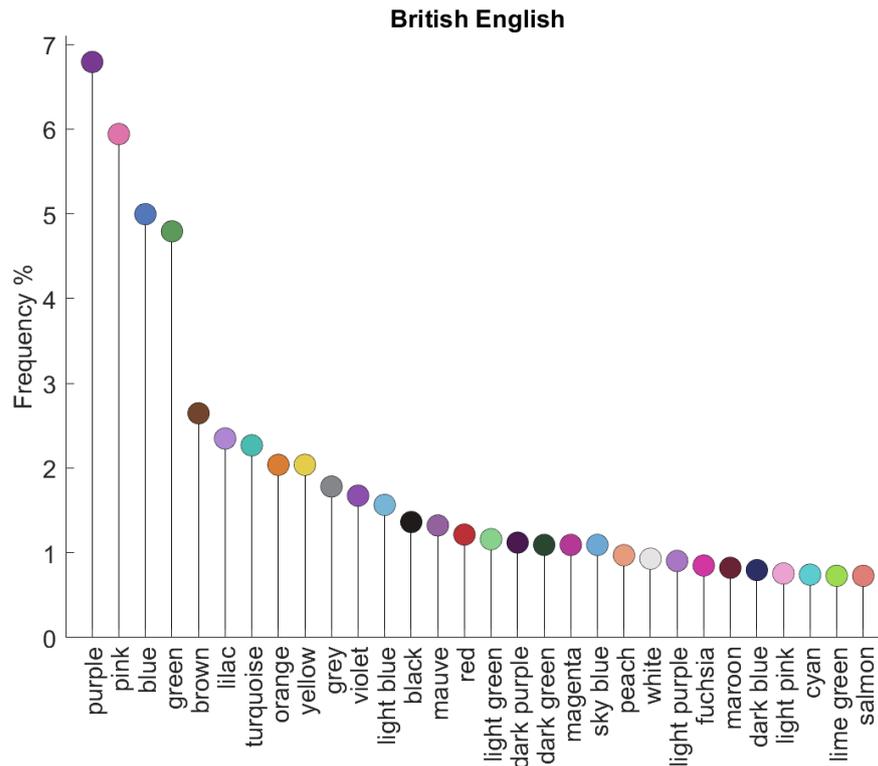
Figure 3.7. Frequency of top 30 colour names in online experiment for British English speakers.

3.6.2. Consensus

Consensus describes the agreement among observers in naming colour samples (Brown & Lenneberg, 1954; Boynton & Olson, 1987; 1990; Davies & Corbett, 1994; Sturges & Whitfield, 1995). Previous studies have defined consensus against a threshold. To provide a consensus measure for all names, we compute it as the peak of each naming distribution over colour samples by:

$$consensus(n) = \max_{c \in C} P(n|c) \tag{3.1}$$

where *P(n|c)* is the conditional probability that name *n* will be assigned to a colour *c* given the *c*=600 colour stimuli of the experiment and the *n*=478 distinct colour names offered by the observers. In Figure 3.8, we show the 30 colour names with the highest consensus. *Red* was the highest followed by *blue* and *yellow*. The top nine ranked colour names were all basic colour terms. *Lilac* was again found earlier than *white* and *green* while *lime green* and *royal blue* are equal to *green*.
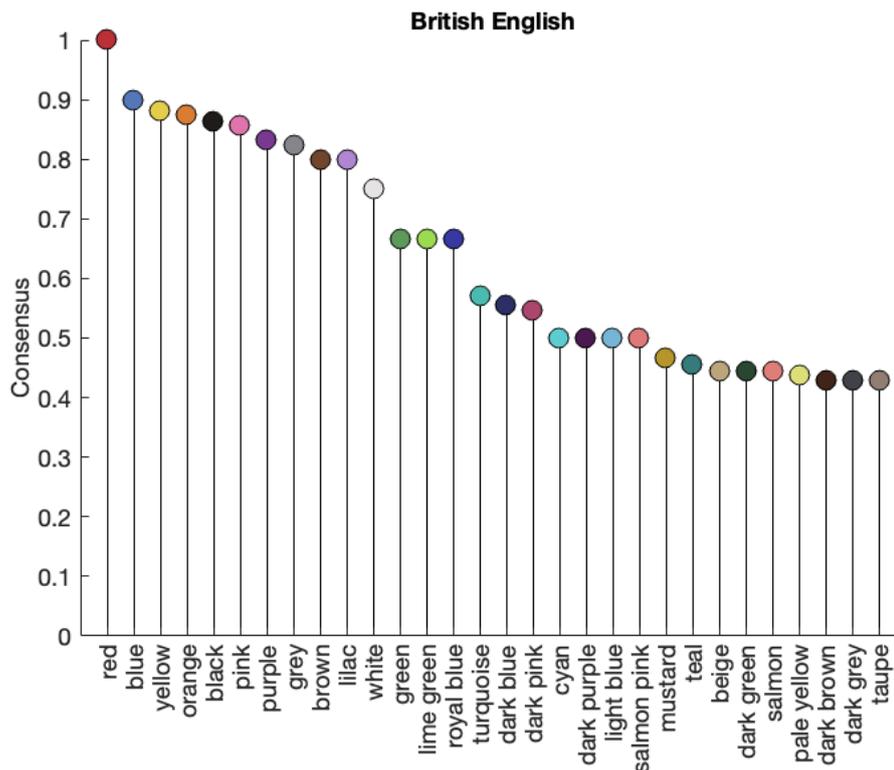
Figure 3.8 Top 30 colour names named with consensus across samples and observers for British English speakers.

3.6.3. Reaction time

The reaction time, also called latency, is the time required by each observer to complete a colour naming task (Brown & Lenneberg, 1954; Boynton & Olson, 1987; Sturges & Whitfield, 1995). In the online colour naming experiment, latencies were measured from the onset of the stimulus to the observer's first keystroke of the typed colour name. Response time distributions are rarely Gaussian as their shape rises rapidly on the left followed by a long tail on the right. Therefore, we report the median range and 95% confidence interval of response latency for each colour name (Whelan, 2008). *White* and *red* were the fastest to name colours followed by *pink* and *grey* (Figure 3.9). Not all eleven basic colour terms ranked in the top positions. *Pale khaki* and *lighter purple* were found in the 9th and 10th position while *purple* in the 15th.
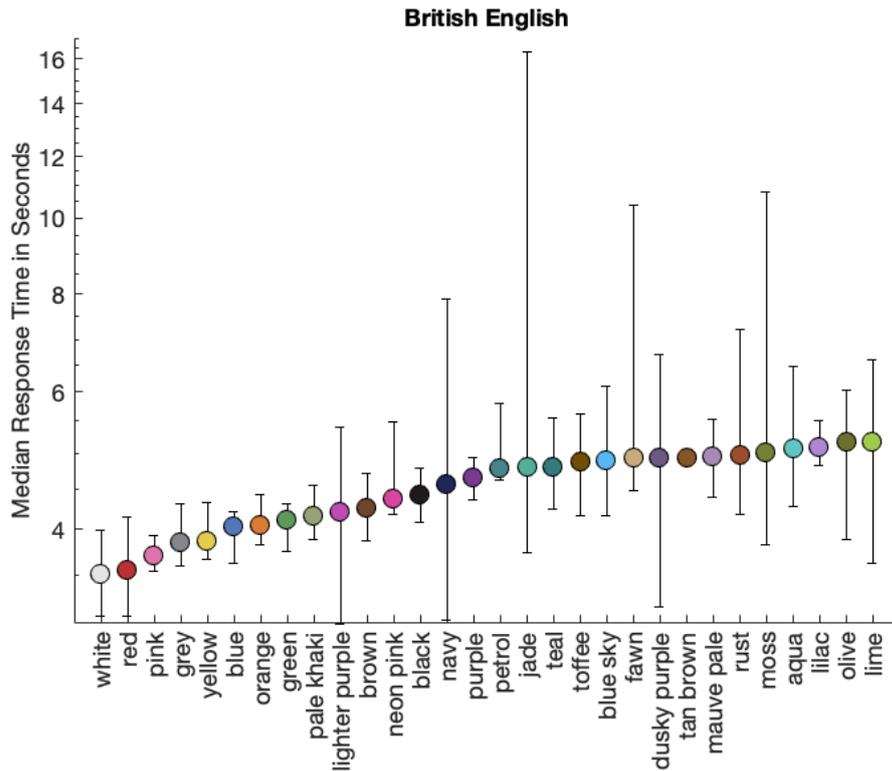
Figure 3.9 Median response time of top 30 fastest responding colour names for British English speakers. Error bars denote 95% confidence interval.

3.7. Geometric features

In this section we present the geometric measurements of size, shape and location of colour names in colour space from the dataset obtained in the online colour naming experiment.

3.7.1. Size

The size of colour categories was measured by their volume in colour space. To approximate the volume of each lexical colour category in CIELAB, we first described the dispersion of their distribution by their covariance matrix. Volume was then measured as the square root of the determinant of the uncertainty ellipsoids. To avoid possible redundancies from the sampling used in the experiment that could in principle produce near to zero volumes for ellipsoids thin in one direction despite having substantial spread in other directions, we regularised the covariance matrix by adding an identity matrix multiplied by the mean colour difference of the four nearest neighbours across stimuli (mean $\Delta E_{ab} = 7.14$). The category with the label *unknown* that summarises all responses where participants replied 'I don't know' covered almost the entire of colour space and was the largest category followed by *puce* for which it seems that participants didn't know the referent colours for this name. The third largest category was *violet* and then *green*.

Only five basic terms were ranked in the top positions of the largest colour categories scale (Figure 3.10). There is a weak Pearson positive correlation between frequency and volume, $r = 0.21$, $p < 0.0025$ (Figure 3.11) that confirms the differences between the two measurements.
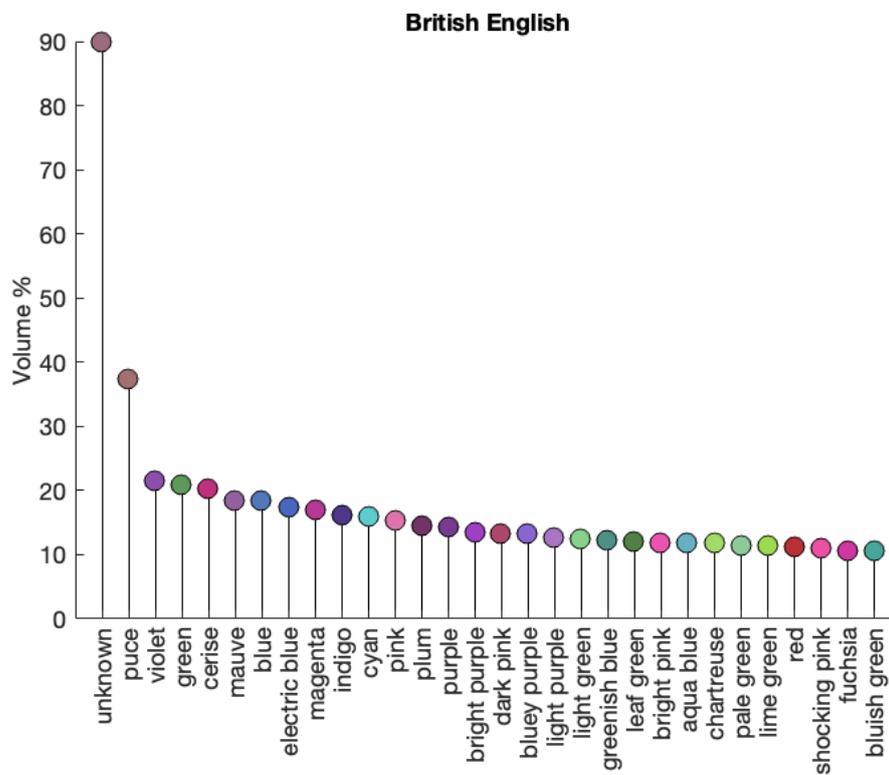
Figure 3.10. Volume of top 30 largest lexical colour categories in colour space for British English speakers.
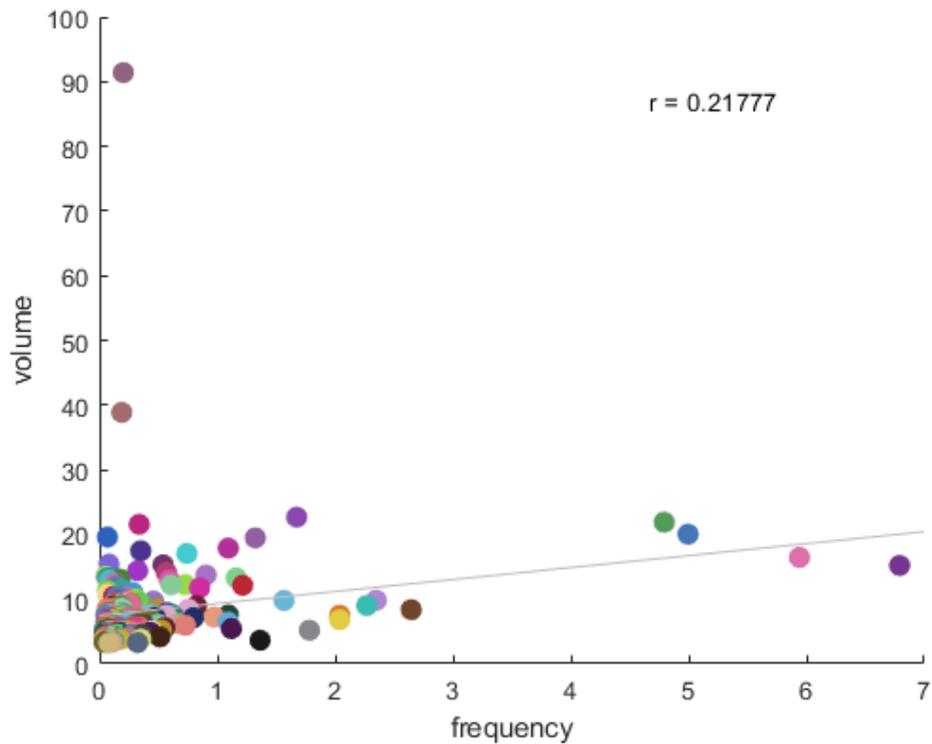
*Figure 3.11 Correlation between frequency and v*olume of colour names in online colour naming experiment.

3.7.2. Shape

The shape of colour names in colour space was measured by their convexity. To assess convexity, the dispersion of each distribution of a colour name across samples was described by its covariance matrix. Sphericity was then defined as the fractional anisotropy of the covariance matrix. Fractional anisotropy (Basser & Pierpaoli, 1996) is a size invariant, pure-shape measure that ranges from 0 for isotropic spherically distributed 3-D data, up to unity for data which is constrained to a line, hence maximally anisotropic. Intermediate values indicated degrees of anisotropy. Hence, colour categories that are near spherical, whether large or small, will have low anisotropy scores; elongated categories will have high scores; and flattened categories will have intermediate scores (Figure 3.12). *Dark grey* was the most spherical colour category followed by *terracotta* and *light brown* (Figure 3.13). *Brown* and *yellow* were the only basic colour terms found in the top 30 positions.
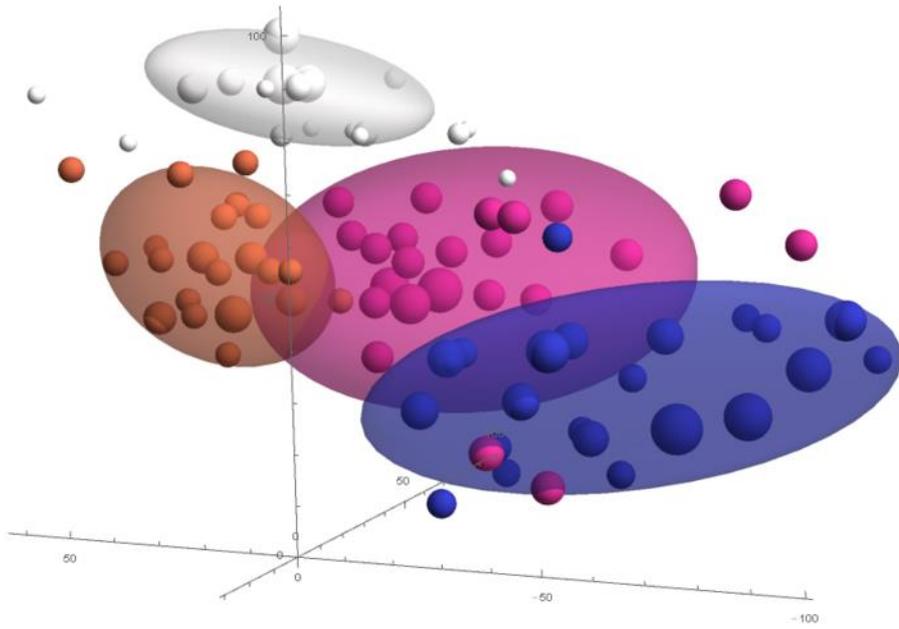
Figure 3.12 Covariance ellipsoids of terracotta, cerise, white and royal blue. The fractional anisotropy of these examples is 0.33, 0.62, 0.86 and 0.90 respectively. The spheres indicate the location of colour samples producing the response, with volume proportion.
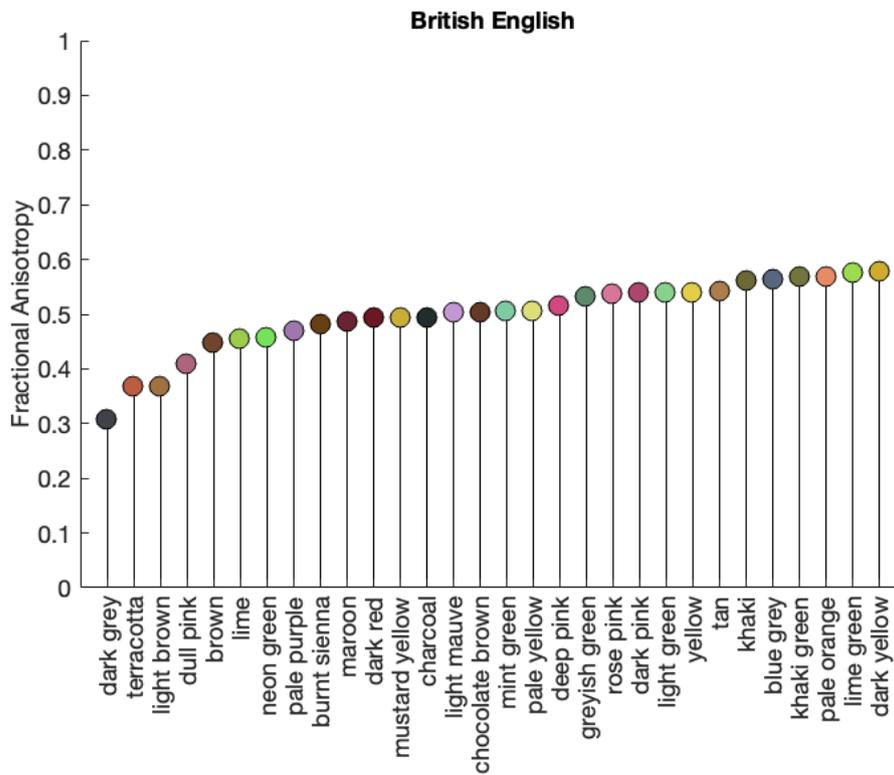
Figure 3.13 Fractional anisotropy of top 30 most spherical colour names for British English speakers.

3.7.3. Location

Centroids are a measure of the centre of mass of the location of colour categories in colour space. For each colour name, we determined its distribution over the colour stimuli in CIELAB and then computed the mean location of its domain for each of the three coordinates L^*, a^*, b^*. The Cartesian coordinates a^* and b^* (Figure 3.14, top) were then converted to the perceptual coordinate of *Chroma* (C^*) and shown against lightness L^* (Figure 3.14, bottom). *Lime green* and *fuchsia* were the colour names with the highest Chroma, while the highest lightness was found for *white* and the lowest for *black*.
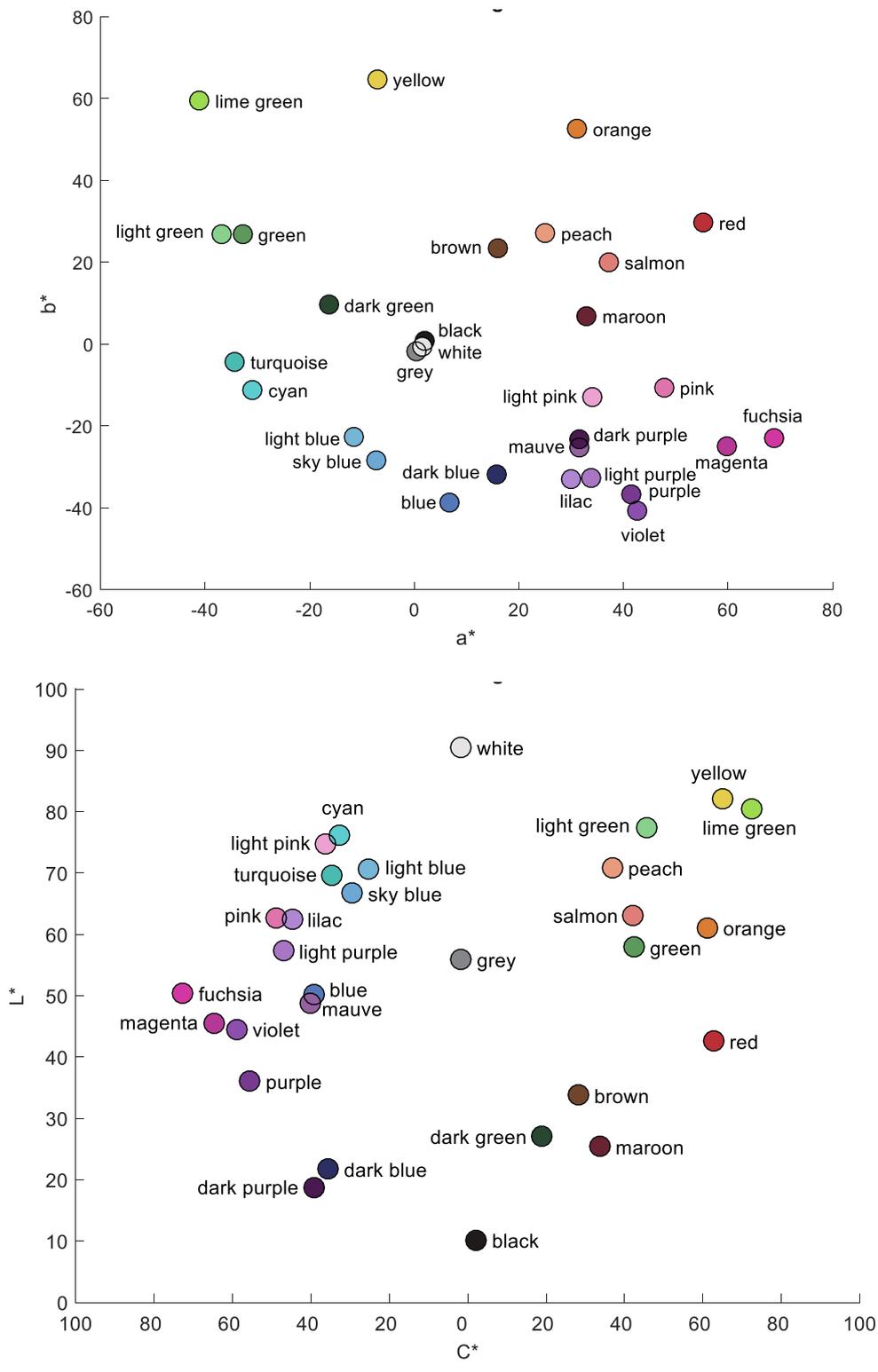
Figure 3.14 Centroids of 30 common colour names in a*b* plane (top) and L*C* plane (bottom) of CIELAB for British English speakers.

3.8.   Centroids of Basic Colour Terms within different languages

In order to explore the agreement between the BCTs across languages, in Figure 3.15 we show their centroids in British English, American English, Greek, Russian, Thai and Turkish. Overall, except for *blue*, there is a very good correspondence between BCTs across languages in terms of hue, Chroma and Lightness. For the *blue* term, there is a good agreement between British, American and Turkish (*mavi*), but there are large differences against Greek (*ble*), Russian (*sinij*) and Thai (*fa*). The Thai main BCT for *blue* (*fa*) coincides more with the *sky blue* in Greek (*galazio*) and Russian (*goluboj*); while the main Greek *blue* term (*ble*) coincide well with the dark blues in Russian (*sinij*) and in Thai (*namngen*). The navy blue term in Turkish (*laçikvert*) appears to differ from the *blue* term (*mavi*) mainly in the lightness dimension with not much differences in the hue dimension. Furthermore, its location is not in good agreement with the *dark blue* BCTs in Greek, Russian and Thai.

Table 3.1 shows the colour differences in terms of the CIE ΔE2000 formula between the BCTs in British English and all other above languages that confirms the good correspondence of the visual inspection. The second basic terms in the blue region proposed in Russian, Greek, Thai and Turkish are compared against the same *blue* centroid in British English. The best agreement is found between British and American English with a mean ΔE_{00}=1.49, and the worst agreement is against Thai with a mean ΔE_{00}=5.01. Except for *blue*, the comparison between British English and American English, Greek, Thai, and Turkish for the loci of the 10 BCTs results in mean ΔE_{00} of 1.83, 2.13, 2.39 and 2.23 respectively. In Table A.1, we report the colour differences in terms of Euclidean distances ΔE_{ab} and the pairwise Euclidean distances for each name. The largest colour difference was found in average >20ΔE_{ab} between the *blue* BCTs while the term with the second largest colour differences was *pink* with a mean ΔE_{ab}=4.52. The BCT with smallest mean colour difference across languages was *purple* ΔE_{ab}=2.35.
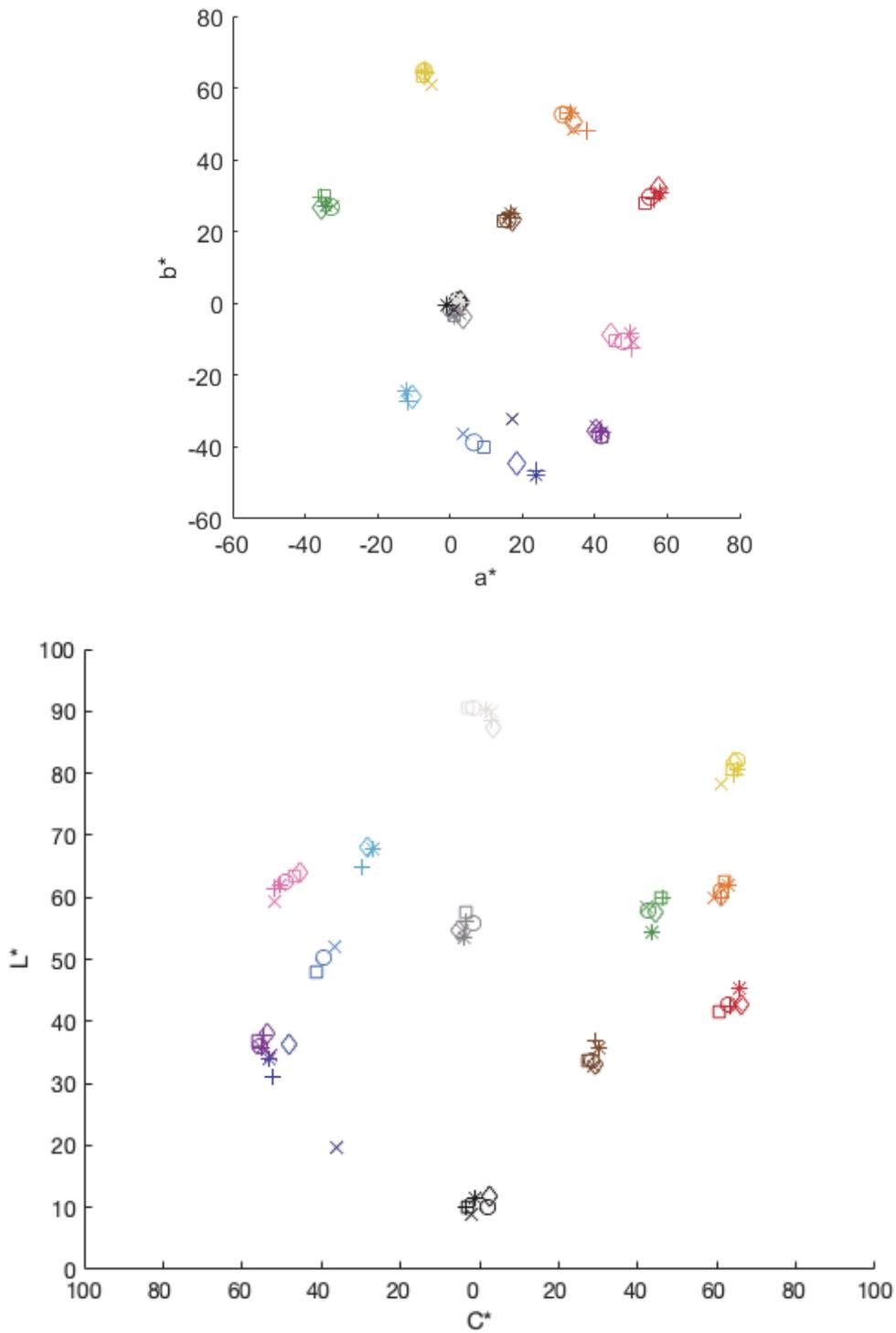
Figure 3.15 Centroids of 11 BCTs in British English (circle), American English (square) and 12 BCTs in Greek (diamond), Russian (asterisk), Thai (plus) and Turkish (cross) in a*b* plane (top) and L*C* plane (bottom) in CIELAB. Centroids with a hue angle >180° are shown on the left side and with a hue angle of <180° on the right side of neutral axis.

Table 3.1 Colour differences CIE ΔE2000 between centroids of BCTs in British English (Br) and American English (Am), Greek (Gr), Russian (Ru), Thai (Th) and Turkish (Tu).

	Br vs Am	Br vs Gr	Br vs Ru	Br vs Th	Br vs Tu
white	1.55	3.02	1.49	2.36	2.58
black	1.51	1.40	4.51	3.59	3.05
red	1.08	0.96	2.59	0.59	0.56
yellow	1.10	0.32	1.15	1.62	3.10
green	2.16	1.21	3.31	2.30	0.67
blue	2.70	14.31	16.98	17.24	2.26
brown	0.72	0.84	1.89	2.68	0.93
purple	0.75	1.65	0.55	1.51	1.51
pink	0.97	1.85	1.41	1.56	3.00
orange	1.39	2.51	1.43	5.21	3.13
grey	2.43	4.58	2.95	2.44	3.77
blue (2)	NA	18.84	19.26	19.00	26.88
mean	1.49	4.29	4.79	5.01	4.29

3.9. Discussion

In this chapter, we presented an ongoing online colour naming experiment designed to collect a small number of unconstrained colour naming responses (n=20) from each of a large number of participants (n=500). In the collection of our behavioural data, we extended previous cross-cultural studies which used only the most saturated colour samples (n=330) on the surface of the Munsell system (Berlin & Kay, 1969/1991; Kay et al., 2010) by sampling (n=600) also the interior of the colour solid. A further methodological improvement includes the departure from usual methods which would use a small number of observers and/or the use of only a restricted set of monolexemic terms (Berlin & Kay, 1969/1991; Boynton & Olson, 1987; Sturges & Whitfield, 1995; Benavente et al., 2006; Lindsey & Brown, 2014; Parraga & Akbarinia, 2016). Instead, thousands of volunteers from linguistically and demographically diverse populations freely named a large number of colours online (Moroney, 2003; Mylonas & MacDonald, 2010; Munroe, 2010). We argued, that participating in an online experiment in your own familiar environment, with your own equipment, and without the physical attendance of the examiner, would give more ecological validity to the underlying categories

responsible for colour naming. We also depart from previous research by taking the different colour names given by the observer in our online task to reflect a categorical distinction important to the observer. So, in the analysis of our data, we did not use statistical procedures to look for similarities between given terms to summarise them into smaller groups (Lindsey & Brown, 2014).

We presented sets of lexical, behavioural, and geometric features of colour names obtained from the analysis of the behavioural responses in British English but also from linguistic data obtained from Twitter. None of the features alone were sufficient to demarcate the BCTs from non-basic colour names in agreement with earlier studies that required multiple criteria for this task (Berlin & Kay, 1969/1991; Corbett & Davies, 1997; Mylonas & MacDonald, 2016). Nevertheless, the obtained features will allow us to examine questions regarding the origin and development of colour categories in Chapter 4.

Except for *blue*, comparison of the centroid location of the BCTs showed a good correspondence (<=5 ΔE_{00}) between languages despite the online experimental methodology and the linguistic diversity of the observers. Overall, these results demonstrate that different languages tend to categorise colours into BCTs similarly. However, the mean colour difference between British and American English was about 3 times smaller than the mean colour differences between British English and the other 4 test languages indicating that speakers of similar linguistic groups agree more on the location of BCTs than speakers of different languages. The large differences for *blue* can be explained by the proposed second basic blue in the languages other than English that splits the unitary blue category. In Russian, Greek and Thai these two basic *blues* correspond well but this was not the case in Turkish contrary to an earlier report (Özgen & Davies, 1998). The range of the basic *blue* term (*mavi*) in Turkish overlapped with the *blue* term in British English while the *navy blue* term (*laçikvert*) appear to differ from the *blue* term (*mavi*) mainly in the lightness dimension with no many differences in the hue dimension. These findings raise doubts for the basicness of the proposed second *blue* in Turkish. We will revisit the important question of basicness of colour names within different languages in Chapter 7.

Chapter 4

Coherence of Classes of Colour

A range of explanations have been advanced for the systems of colour names found in different languages. Some explanations give special, fundamental status to a subset of colours from which all other categories derive. We argue that a subset of colours, if fundamental, will be coherent; meaning that there exists a non-trivial criterion which distinguishes them from the other colours. In Chapter 3, we presented a set of measurements for capturing features of colour names obtained in an online experiment and in linguistic usage data. Here, we make use of these features to test the coherence of subsets of achromatic, primary and basic colours. We apply a supervised machine learning method to discover criteria which distinguish subsets of colours, and so assess their coherence. We find that achromatic and basic colours are coherent subsets but not primaries. These results reinforce the ongoing argument against the special role of primaries in the formation of colour categories.

4.1. Coherent classes

We make the claim that if a subset of colours has a foundational role in the system of colour naming then that will leave a trace in the properties of those colours, in comparison to other colours. We formalise this idea as a subset of colours forming a *coherent* class, defined by a generalisable membership criterion. We define a criterion to be generalisable if it can be reliably identified from a subset of members of the class. This rules out trivial list-membership style criteria. If we are able to show that some subset of colours cannot be distinguished by a generalisable criterion, and hence do not form a coherent class, we suggest that this presents a challenge to any explanation for colour naming that gives that subset a fundamental role, as no trace of that role exists. The class membership criteria that we will consider are based on sets of behavioural, linguistic and geometric features of colour names, presented in the previous chapter. We restrict our analysis to 73 colour names in wide cultural use which were produced at least 20 times in our data to give us confidence in their measures. This accounts for 62% of the responses. In Appendix B, we report features for all 73 common colour names.

In order to discover criteria that can distinguish class membership based on features we use methods from machine learning. These machine learning methods represent the

state-of-the-art in practical classification problems and subsume many previous classical theories such as 'necessary and sufficient conditions' (Berlin & Kay, 1969/1991), 'similarity to prototypes' (Rosch Heider, 1972), or 'networks of family resemblances' (Rosch & Mervis, 1975).

Specifically, we use membership criteria expressed by an ensemble (forest) of decision trees (Figure 4.1). Each decision tree expresses membership with a binary tree B of criteria such as 'C is in X if $f_1 \leq t_1$, or $f_1 > t_1$ and $f_2 \leq t_2$, otherwise not' – where C is a colour, X is a class, f_i are feature values and t_i are thresholds. The trees in an ensemble are purposely constructed to be different, so some trees may deem C to be a member of X, and others not. The proportion of votes for membership, across the ensemble, is considered a class membership confidence. So, while the decision boundary in feature space of an individual tree is necessarily piecewise-linear, axes-aligned, and sharp, the decision boundary of an ensemble of trees can be curvilinear, and sharp in some parts while fuzzier in others.
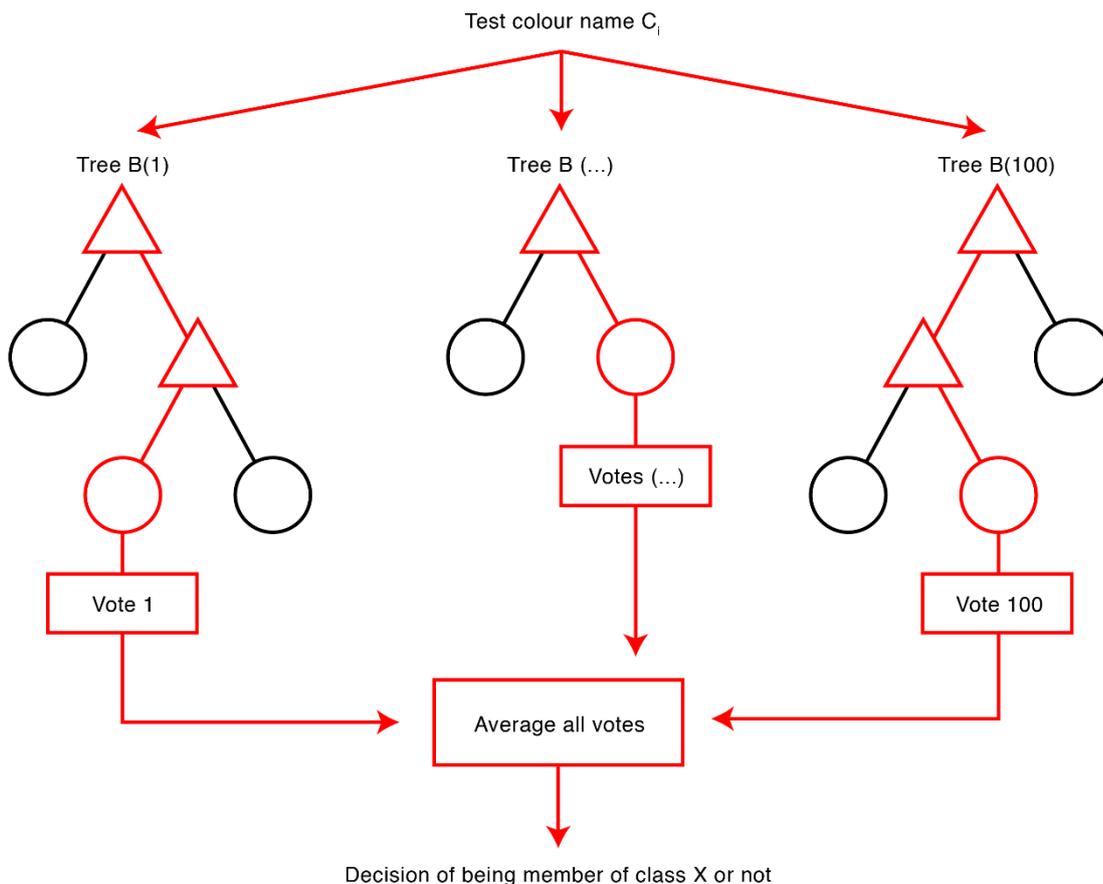
Figure 4.1 Graphical example of a decision tree ensemble for determing membership of a colour C to class X. Triangles denote nodes, circles binary decisions, red colour routes True decisions.

84

Effective algorithms ('Random Forests') for construction of ensembles of decision trees based on training examples, ensure that the trees are varied by constructing each on a different random subset of the training data; and by choosing the splitting feature at each branch of the binary tree not from all possible features, but from a different random subset. Random Forests have been shown to be highly effective for many diverse classification problems (Breiman, 2001; Gislason et al., 2006; Cutler et al., 2007). An advantage of them, useful for our application, is that they do not assume commensurate feature dimensions, or normally-distributed features values.

To assess the coherence of a class of colours, we measure how well it is defined by a generalisable criterion. We enforce generalisability by using a leave-one-out evaluation: for each colour (in class or out) we build a random forest classifier using all *other* colours, together with their labels as in-class or out; and then evaluate the class membership confidence of the left-out colour using that classifier. Finally, we evaluate whether the membership confidences of in-class terms are higher than those of out-of-class terms.

We report the coherence of the Hering primary class (*black, white, red, green, blue* and *yellow*) and the basic class (Hering's primaries plus *purple, orange, pink, grey* and *brown*). Additionally, we report the coherence of an achromatic class (*black, grey* and *white*) to check whether smaller classes are necessarily less coherent because they have fewer examples from which to determine a membership criterion. In the Appendix, we report results for other plausible sets of primary and basic colours.

4.2.   Families of features

For each of the 73 common colours, three sets of behavioural, geometric and linguistic features were computed. For linguistic features, we use name length measured in letters; the number of derivative forms (e.g. greener, greenish, and sea green are all consider derivatives of green) in the online experiment; and usage frequency based on counts in the social media dataset. The behavioural features, computed from the online response dataset, are: frequency of occurrence, response latency and inter-subject consensus. The geometric features, computed from the online response dataset, are: the mean colour space location of the distribution of samples that generate that response, and the size and shape of that distribution.

4.3. Classifier

We constructed criteria for demarcating classes of colours using the Random Forests algorithm (Breiman, 2001). As an input, the algorithm receives a training dataset of colours, each described by a vector of the above feature values, and associated with a binary label indicating whether it is in-class or out-class. Based on this input, the algorithm creates an ensemble (forest) of 100 independently-generated decision-trees. We have confirmed that a larger forest does not change the results.

Each tree is grown using a separate dataset created by bootstrap sampling-with-replacement from the training data. Trees are grown down from a root node at which all training data arrives. At each node a feature dimension is chosen to be the basis for a splitting rule. The choice of dimension is made from a subset of all feature dimensions, chosen randomly for that node. Following the standard recommendation, if there are n feature dimensions, then the subset size is (rounded) square root of n; so, in our trees, at each node, a subset of three feature dimensions were considered out of the full eleven. Given the feature subset, the particular feature and threshold value that best segregates the data arriving at the node according to its labels is identified. The arriving data is then sent to left and right sub-nodes according to this criterion. Sub-nodes are iteratively constructed below nodes until leaf nodes are reached that receive only a single training data sample. After tree construction, the unique dataset generating the tree is discarded but the structure of the tree, the splitting dimension and threshold at each node, and the label of the datum in each leaf node is retained.

After construction of the forest a new datum is classified by passing it through the structure of each tree, directing it to sub-nodes according to its feature values, and recording the label of the leaf node at which it finally arrives. The proportion of trees of the forest that classify it as in-class is the overall in-class classification confidence of the forest.

4.4. Colour classes

We assess the class coherence of three subsets of colours:

Achromatic (n_{in}=3): *white, black* and *grey.*
Primary (n_{in}=6): *white, black, red, green, yellow* and *blue.*
Basic (n_{in} =11): *white, grey, black, red, orange, yellow, green, blue, purple, pink* and *brown.*

For each class, the other colours of the common set (n_{out}=73-n_{in}) were considered out-of-class.

4.5. Evaluation of classification

Evaluating classifiers on data on which they were trained is generally misleading. To avoid this, and to ensure that the computed class criteria are generalisable, we employ a leave-one-out cross-validation strategy. For each class that we assess, we build 73 separate classifiers. Each is trained on 72 colours, with a different colour left out. The in-class confidence of each colour is then computed by the classifier which was trained with it left out. To assess the coherence of a class we quantify the extent to which the class confidences of the in-class colours are higher than those of the out-of-class. For this quantification we use a measure based on rank precision. Precision is the fraction of correct positive classifications to a test class over all positive classifications. MAP is the mean average precision of the ranks at the top k positions, where k is the size of the test class. MAP will be 1.00 if all in-class confidences are higher than all out-of-class; 0.00 if all in-class confidences are lower than all out-of-class; and intermediate if the range of in-class confidences overlaps the range of out-of-class.

To examine the importance of features for the coherence of each class, we repeat the full leave-one-out assessment and MAP computation, but with classifiers trained with only a subset of features. The subsets we assessed were: all features except one, two out of three families of features, and single families of features. The importance of features, or families of features, for each class of colours is quantified by how much the MAP score decreases compared to using all features. In the next section we examine the cohesion of achromatic, primary and basic classes, and determine the contribution of different features to that coherence.

4.6. Coherence of Achromatic class

In our first assessment, we examined the coherence of an achromatic class consisting of *black*, *white* and *grey*. The Random Forests classifier gave all three in-class colours higher confidences than all non-class colours, giving a maximum possible MAP score of 1.00. In Figure 4.2, we present the confidence for each colour to belong to the achromatic class. The in-class confidence of each colour is assessed by a classifier that is trained on all colours apart from it. *White* was the colour with the highest confidence, followed closely by *black*. *Grey* was found in the third position but with lower confidence. *Light grey* was the out-of-class colour with the highest in-class confidence.
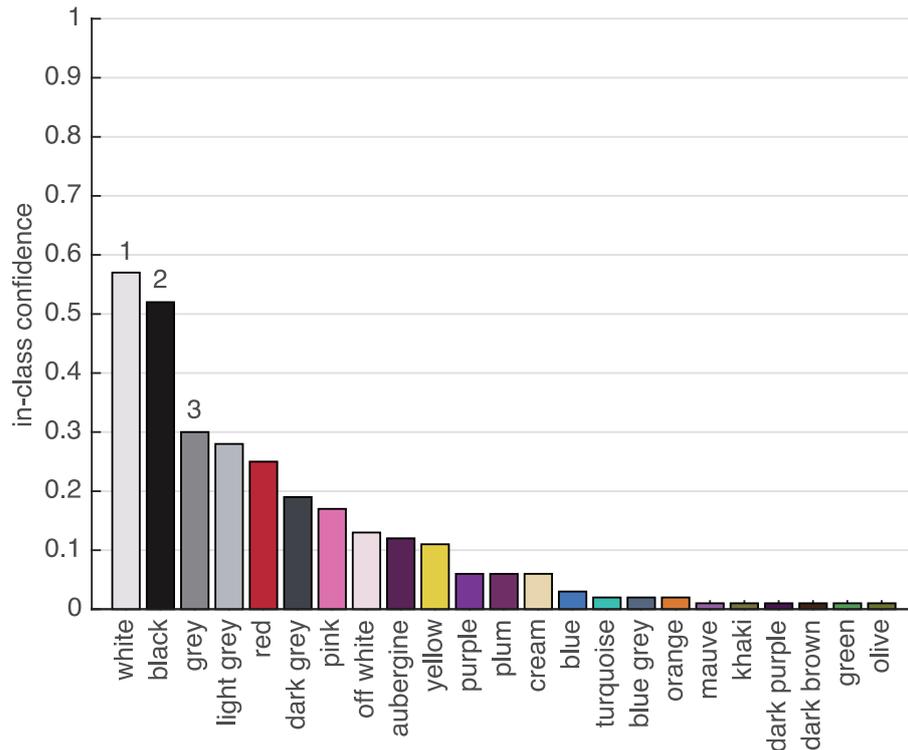
Figure 4.2 Class confidences for the achromatic class. Only colours with an in-class confidence above 0 are shown for clarity. Numbered bars indicate class members. The display colour of each bar is the sRGB centroid of the distribution of samples generating that colour name response.

4.7.   Coherence of Primary class

As a primary class we took the six opponent Hering primaries: *white*, *black*, *red*, *green, yellow* and *blue*. The classifier produced a MAP score of 0.50. Examination of the confidences for individual colours (Figure 4.3) showed that this low coherence score was due to failure of the class criteria to generalise to all in-class members (especially *yellow*), and erroneous generalisation to non-class members (especially *pink*, *grey* and *brown*).
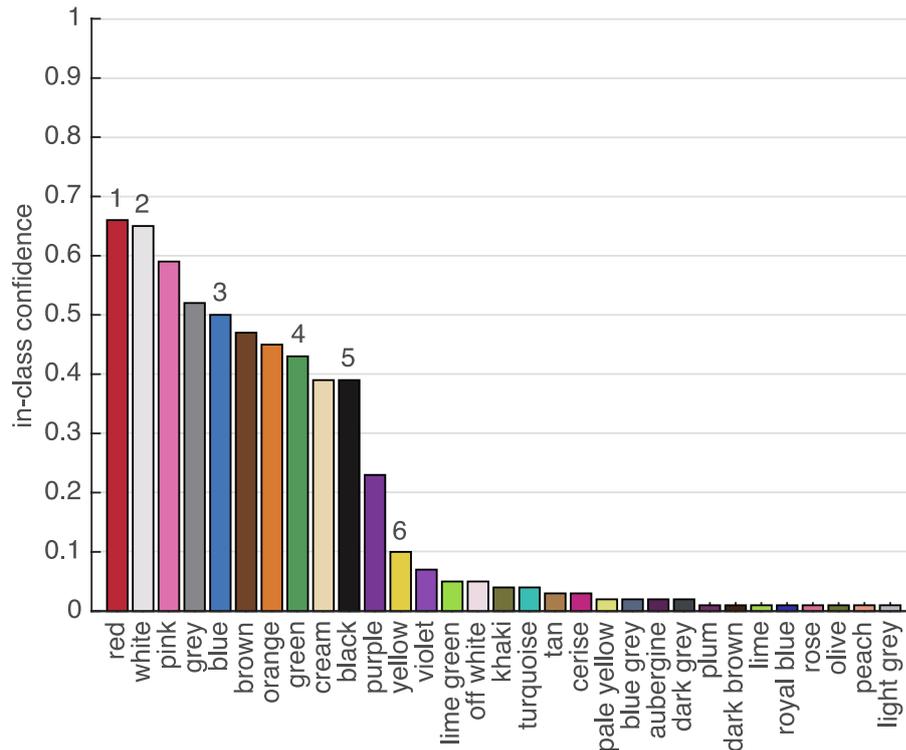
Figure 4.3 Class confidences for the primary class. See Figure 4.1 for additional notes.

4.8. Coherence of Basic class

For the assessment of the basic class, we considered the 11 basic colour terms of Berlin & Kay (1969/1991), *white, black, grey, red, orange, yellow, green, blue, purple, brown* and *pink*. All basic colours were given higher confidences than all non-basic, resulting in a maximum possible MAP score of 1.00. Amongst the basics, *blue, pink* and *brown* were given the highest confidences and *purple* the lowest (Figure 4.4). Amongst the non-basics, *olive* was given the highest confidence.

A summary of all evaluations is given in Table 4.1.

Table 4.1 MAP scores, expressing class cohesion, for achromatic, primary and basic classes. A score of 1 is perfect cohesion according to our assessment.

Class	MAP
Achromatic class (n=3)	1.00
Primary class by Hering (n=6)	0.50
Basic class by Berlin & Kay (n=11)	1.00
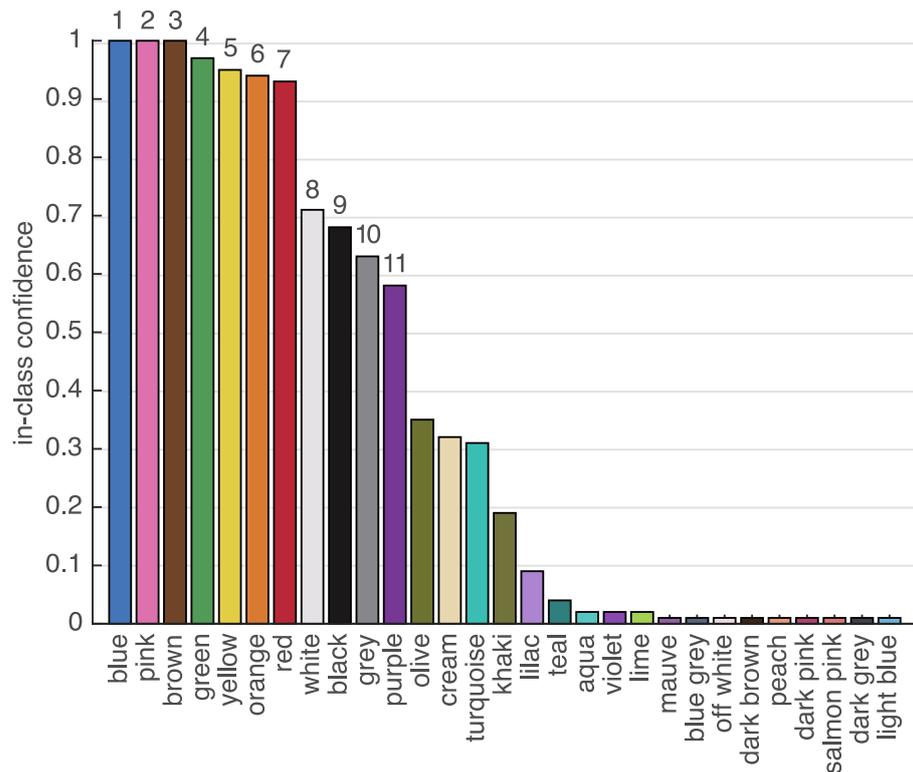
Figure 4.4 Class confidences for the basic class. See Figure 4.1 for additional notes.

4.9. Feature Contribution

To examine the importance of each feature and each family of features we assessed class coherences using different feature subsets, specifically:

a) All features (n=11)
b) All features bar one (n=10), eleven variants
c) Behavioural plus Geometric features (n=8)
d) Geometric plus Linguistic features (n=8)
e) Behavioural plus Linguistic features (n=6)
f) Geometric features (n=5)
g) Behavioural features (n=3)
h) Linguistic features (n=3)

MAP

| | 0 | 0.2 | 0.4 | 0.6 | 0.8 | 1 |

All Features

Frequency [-]
Response Time [-]
Consensus [-]
Size [-]
Shape [-]
Lightness [-]
Chroma [-]
Hue [-]
Linguistic Frequency [-]
Word Length [-]
Derivative Forms [-]

■ Achromatic   ■ Primary   ■ Basic
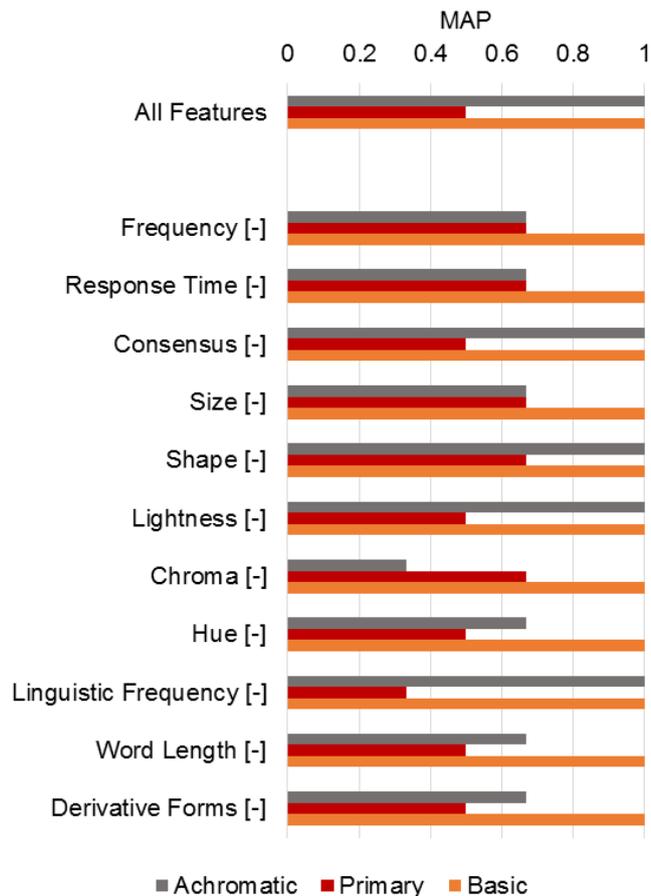
Figure 4.5 Class coherence scores when a single feature is excluded [-], compared to when all features are used (top).

When excluding individual features, the most important feature for the achromatic class, fittingly, is the Chroma feature as the MAP score was reduced from 1.00 to 0.33 when it is excluded (Figure 4.5). Exclusion of consensus, shape, lightness and linguistic frequency had no effect. For the primary class the most important feature was Linguistic Frequency, which reduced the MAP to 0.33 from 0.50 when excluded. Excluding frequency, response time, size, shape or chroma improved the MAP score. This is presumably because these features are useful to demarcate some of the class but generalise inconsistently. The greatest improvement was when Response Time was excluded, raising the MAP score from 0.50 to 0.67. In this case, *pink* and *cream* remain as false positives at ranks 5 and 6, with class confidences higher than *green* and *yellow*. For the basic class of colours none of the excluded features reduced the MAP score below 1.00.
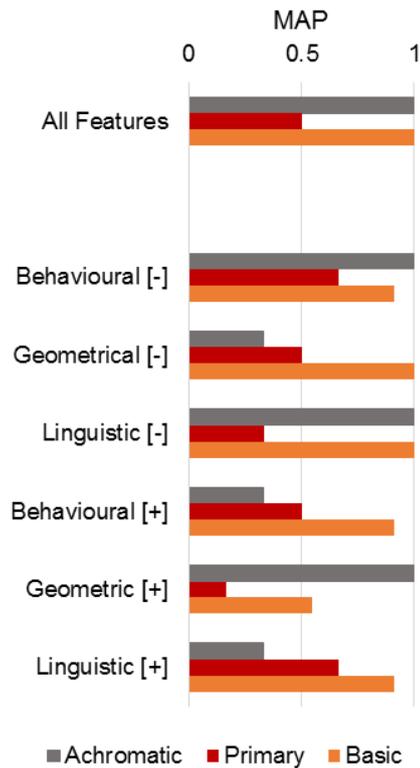
Figure 4.6 Class coherence scores when one family of features is excluded [-], and when only one family of features is retained [+], compared to all features (top).

Considering exclusion of single families of features: for achromatic, so long as geometric is retained the MAP score is 1.00, otherwise it is 0.33 (Figure 4.6). For primary, the exclusion of linguistic produced the lowest MAP score of 0.33 and the exclusion of behavioural the highest MAP score of 0.66. For basic, the exclusion of geometric and linguistic did not influence the coherence of the class with a MAP score of 1, but excluding the behavioural reduced the MAP score to 0.90 because *cream* was then given higher confidence than *white* and *black*.

The assessment of retaining single families of features resulted in a MAP score 0.33 for the achromatic class when either behavioural or linguistic were retained, and a MAP score of 1 when geometric was retained. For the primary class, keeping only geometric features produced a MAP score of 0.17 while retaining the linguistic resulted in a maximum MAP score of 0.66. The retainment of either behavioural or linguistic alone produced a MAP score of 0.90 for the basic class, but geometric alone gave a MAP score of 0.54.

4.10. Discussion

A point of contention that frequently arises regarding the basis of colour categorisation is whether there are subsets of colours with a special fundamental status, from which all other categories derive. Different subsets have been suggested as fundamental, and no consistent assessment of each of their claims has been previously been made. Here, we argue that a fundamental subset of colours should form a coherent class, with a generalisable membership demarcating it. To test this, we analysed a large dataset of colour naming responses from an online colour naming experiment and public social media posts to examine the class coherence of achromatic, primary and basic colours. Our findings provide evidence to substantiate the coherence of basic and achromatic classes, but we found little support for the primary class. Indeed, the best generalisable criteria for demarcating the primaries consistently also capture secondary colours. These results suggest that the primary class of colours does not play a foundational role in colour categorisation. A summary of all assessments is given in Table 4.2 including the results for other plausible sets of primary and basic colours reported in Appendix B.

In our assessment of the primary class, we considered the primaries of Hering's opponent process theory because of a widely held view that these colours are the basis of colour naming systems across languages (Kuehni, 2005; Regier et al., 2005; Philipona & O'Regan, 2006). Still, the number and the members of the primary class vary in the literature (Aristotle, 350 B.C.E.; Newton, 1730; Maxwell, 1872; Hering, 1878/1964; Eskew, 2009; Skelton, Catchpole, Abbott, Bosten & Franklin, 2017). In Appendix B (see Table 4.2 for a summary), we tested primary classes with different proposed members than those of Hering but again we found no evidence to substantiate the coherence of any primary class. The coherence of primary classes proposed by Eskew (0.63), Aristotle (0.57) and Newton (0.57) was higher than Hering's primary class, but those proposed by Maxwell (0.30) and Skelton et al. (0.40) were lower. All these primary classes are smaller ($3 \leq n \geq 7$) than the basic class ($n = 11$) but this does not explain their low MAP scores, since the even smaller achromatic class ($n = 3$) had perfect coherence (MAP = 1.00) because its members have distinctive, common characteristics.

Table 4.2 MAP scores, expressing class cohesion, for achromatic, primary and basic classes and all variants. A score of 1.00 is perfect cohesion according to our assessment.

Class	MAP
Achromatic class (*n=3*)	1.00
Primary class by Maxwell (*n=3*)	0.33
Primary class by Hering (*n=6*)	0.50
Primary class by Eskew (*n=6*)	0.63
Primary class by Skelton (*n=5*)	0.40
Primary class by Aristotle (*n=7*)	0.57
Primary class by Newton (*n=7*)	0.57
Random classes (n=6)	μ=0.13
Secondary basics + 1 of Hering's primaries classes (*n=6*)	μ=0.53
Basic class by Berlin & Kay (*n=11*)	1.00
Basic class by Berlin & Kay (*n=11 + olive*)	0.92
Basic class by Berlin & Kay (*n=11 + cream*)	1.00

Random classes with equal number of randomly selected colours ($n = 6$) had an average MAP score of 0.13 (Figure A.19), while an equally sized class of secondary basics colours (*brown, orange, purple, pink* and *grey* plus one of Hering's primaries) had an average MAP score of 0.53 (Figure A.20). This indicates that primaries are not a completely haphazard class but are not more coherent than classes of secondary colours; consistent with previous studies in adults (Boynton & Olson, 1985) and in infants (Franklin, Pitchford, Hart, Davies, Clausse & Jennings, 2008).

Considering why the class coherence was low for all systems of primary evaluated, we note that *yellow* (considered primary in all schemes) was consistently given low class-confidence. The particular characteristics of *yellow* that might explain these results is its narrower distribution (see Figure A.9) and higher lightness (see Figure 3.14) than other chromatic members of the primary class. Interestingly, *yellow* was absent in Aristotle's original text where he named only six out of the seven pure categories; and also missing from the wavelength sensitivities of cells in V4 reported by Zeki (1980). A second reason for the universally low coherence scores for primary classes were the consistently high in-class confidences given to *pink* and *brown* (non-primary in all schemes). *Pink* and *brown*, similarly to *green* and *blue*, were responded to very frequently, in a very short period of time and with very good agreement between subjects. It is also interesting to note that *pink* and *brown* appear as a symmetrically related pair within the cognitive

structure of the basic colours determined through analysis of similarity, relative lightness and adjacency (Griffin, 2001), suggesting that the salience of these two categories may have a shared explanation.

In contrast to the poor coherence of the primaries, the 11 basic colours (Berlin & Kay, 1969/1991) had a perfect MAP score of 1.00. The coherence of the basic class was also apparent when the classifier was trained with reduced features: behavioural or linguistic features alone gave a score of 0.90, together 1.00. Geometric features contributed little. Coherence of the basic class is unsurprising given that they were originally identified according to a criterion based on features similar to the ones we use. Our results are a confirmation that the Berlin and Kay's basic colours can be distinguished from other colours by such a criterion in English.

Discussion of the basic colours is frequently concerned with why these particular colours satisfy this criterion, rather than some other colours. Different candidate answers have been advanced, placing different emphasis on the role of physiology or natural world properties. On the one hand, Griffin (2001) has shown that the cognitive similarity structure of the 11 basic colours has a symmetry which corresponds to a symmetry of the cone response functions. At the other end of the spectrum of explanations, is grounding in the statistical regularities found in natural images (Yendrikhovskij, 2001), or optimal performance at tasks where semantics must be inferred from appearance (Griffin, 2006). Any explanation wherever it lies in the spectrum, must account for the variation in the number of basic colours across languages; and some authors have questioned whether the same set of basic colours is coherent in all cultures, dependent on the communication needs of semantic categories that are locally most important (Davidoff et al., 1999; Gibson et al., 2017). A cross-language extension of the current methodology could shed light on this.

The examination for a possible *additional* 12th basic colour term in the supplementary section showed a slight deterioration of the coherence of the class, except when *cream* was added which also produced a perfect MAP score of 1.00. The reversal of the confidence ranking of *cream* and *olive*, when *olive* or *cream* is added to the basic class (compare Figure A.21 and Figure A.22) is surprising but explicable. Consider the category of *flying birds*. What animal is the closest to being in-class by generalising from the class? Possibly penguins, with emus further behind. But when penguins are grouped with flying birds, then the criterion which demarcates the class from all other animals would change substantially (promoting the importance of feathers perhaps), and emus could become more in-class than penguins. *Cream* was also suggested as a candidate

for a 12th basic colour terms in a previous study (Sturges & Whifield, 1995) but similarly to our findings with much lower scores than the other eleven basic terms. This indicates that the upper limit of the basic class has some fuzziness and new basic terms may arise (Hardin, 1997; Mylonas & MacDonald, 2016).

Could our results be influenced by our online experimental methodology, the quality of features and absent features? Regarding the uncontrolled colour reproduction of the web-based colour naming experiment, the comparisons against results of previous studies conducted in laboratory conditions produced similar centroids for the primary colour names in English and in different languages (Sturges & Whitfield, 1995; Mylonas & MacDonald, 2016; Paramei et al., 2018). The agreement for the location of the basic terms, including achromatic and primaries, between British and American English in the online experiment (ΔE_{00}=1) (Mylonas, MacDonald & Griffin, 2017) was better than what has previously reported between laboratory-based studies (ΔE_{00}=7) (Boynton & Olson, 1987; Sturges & Whitfield, 1995). Furthermore, the response times reported here, albeit longer than latencies recorded in laboratory settings, replicate the advantage of the basic terms and the equality of primary and secondary basic terms reported in previous studies (Boynton & Olson, 1987, Corbett & Davies, 1997).

With respect to different computational approaches for determining the features of each colour, we recognise that there are alternative reasonable ways to compute some of these. For example, replacing the reported *median* response time with the *mean* as used in previous studies (Boynton and Olson, 1987) or replacing the probabilistic calculation of consensus of this study with a more information-based computation (Gibson et al., 2017). We have not found that variants of computations for either response time or consensus substantially alters our results.

A possible missing feature could be the purity of each colour and a hue cancelation task could provide a better measure than our naming task to determine the coherence of primary classes. Nevertheless, previous studies (Malkoc, et al., 2005; Bosten & Boehm, 2014) found no differences between unique-hue judgments of non-primary (i.e., orange, purple) and primary hues (i.e. red, yellow, green and blue), suggesting that inclusion of such a feature would not be sufficient to make the primaries coherent. A different type of missing feature would be relational features, such as the good configuration of colours in a class (Jameson & D'Andrade, 1997; Regier et al., 2007). Our class coherence approach is unable to accommodate relational features since they belong jointly to a class, not separately for each colour. Indeed, we consider that the most compelling

justification for most systems of primaries is not their fundamental role in colour categorisation but their practical success in subtractive or additive colour mixing.

In conclusion, we show that primary colours do not form a coherent class, whilst achromatic and basic classes do. These results provide evidence against primaries playing a fundamental role in the development of colour categories and challenge explanations based on this claim.

Chapter 5

Computational colour naming models

In this chapter, we present a range of computational colour naming models to automate the assignment of colour names to colours across the full three-dimensional colour gamut. We evaluate the performance of four supervised nonparametric algorithms (1-NN, kNN, Random Forests and Rotated Split Trees) trained by responses from human observers in the online colour naming experiment. The best performing algorithm, Rotated Split Trees (RST), is also evaluated in several colour spaces (linear RGB, sRGB, CIEXYZ 1931, CIELAB, CIELUV and CIECAM02-UCS) where it achieved the best performance in CIELUV. Using this method, we infer histograms of naming responses for any colour, and compute their entropy as a measure of naming variability in colour space. We also compare the classification of the most saturated colours of the Munsell Array into colour terms by RST against previous colour naming models based on monolexemic psychophysical data. We then show the performance of RST in segmenting a synthetic colour wheel when trained by British and American English, Greek, Russian, Thai and Turkish speakers.

5.1. Learning from colour naming data

Machine learning is a subfield of Computer Science that provides computers the capacity to automate a learning task without being explicitly programmed. It is being used in many applications across fields such as search engines, computer vision, natural language processing, medical imaging and computational finance. In colour naming studies, machine learning methods have been used to map perceptual and linguistic aspects of colour with an aim to reproduce colour names that are meaningful to human observers.

To automate the colour naming task, the majority of studies make use of supervised models trained by responses of human observers (Lammens, 1994; Lin et al., 2001b; Mojsilovic, 2005; Seaborn, 2006; Benavente, 2008; Chuang et al., 2008; Mylonas et al., 2010; Parraga & Akbarinia, 2016) or annotated images (Weijer et al., 2007; Lindner et al., 2012). Interestingly, an unsupervised colour categorization model based on the minimum perceptual distance criterion for natural image statistics produced a system of basic colour categories that resemble those of human observers (Yendrikhovskij, 2001; Berlin & Kay, 1969/1991). Similarly, the use of reinforcement learning resulted in

distributions similar to universal colour categories based on the communication between multiple agents for uniform and natural distributions of colours (Belpaeme & Bleys, 2005). Given that the performance of the latter two approaches in order to reproduce human colour categories have been demonstrated only for a small number of basic categories, we will focus here on supervised models trained by the large number of unconstrained colour names in different languages of the online colour naming experiment. In *supervised learning*, the computer is given a training set with a known property (e.g. an RGB triplet with a label such as red) as an input and the task is to predict this property for new instances (e.g. the colour name for an unlabelled RGB triplet). For classification problems the task is to predict whether or not a name is applicable to a test sample, while for regression problems the task is to predict the distribution of names for a test sample.

Computational colour naming approaches can be also grouped into *parametric* and *non-parametric* models based on their assumption for the form of the underlying mapping function from colours to names. On one hand, parametric models are easier to use as the training data is characterised with a set of predefined tuneable parameters that determine the shape of each colour category (Lammens, 1994; Benavente, 2008; Parraga & Akbarinia, 2016; Griffin & Mylonas, 2019). On the other hand, non-parametric models make fewer strong assumptions about the mapping function and the shape of the categories is determined by the training data (Mojsilovic, 2005; Seaborn, 2006; Weijer et al., 2007). In this chapter, we present the performance of four supervised nonparametric algorithms, which make fewer assumptions than parametric approaches and, as a result, provide wider applicability and increased robustness.

5.2. Methods

The training set for our models consists of $n = 600$ approximately uniformly distributed colours x and $T = 1544$ distinct colour names y in British English offered by 500 observers in the online colour naming experiment. To generalize our observations from the training labelled points $\{(x_i, y_i)\}_{i=1}^{n}$, with $x = (x_{(L)}, x_{(a)}, x_{(b)})^T$ and $y = (y_1, \dots, y_T)$ to the entirety of colour space in CIELAB, a regression-based machine learning model $M = f(x)$, can be applied to predict the histogram of naming responses for any colour. The assignment of the most likely colour name $y \in Y$ to a test colour $\tilde{x}$ specified by the triplet $\tilde{x} = (\tilde{x}_{(L)}, \tilde{x}_{(a)}, \tilde{x}_{(b)})^T$ can be then expressed as:

$$\hat{y}(\tilde{x}) = \underset{i}{\text{argmax}} \, f(\tilde{x}_i) \tag{5.1}$$

In the following sections, we describe a set of supervised nonparametric models with an aim to infer a function $f(\tilde{x}_i)$.

5.2.1. Nearest neighbour and k-Nearest Neighbours

A nearest-neighbour (NN) approach is a simple algorithm that works well for basic machine learning problems. It only requires a colour metric to be chosen to determine the distances between test colour $\tilde{x}$ and training colours x. In CIELAB, an appropriate distance metric could be the CIE ΔE_{ab} (Euclidean distance) or the more recently proposed CIE ΔE_{00} colour difference formulae. A k-Nearest Neighbours algorithm allows you to specify the k closest training samples and here the optimal k that minimises the error between observed and predicted histograms is determined using two modes of cross validation described in the following section. Given our training set of colour points $X = x_1, ..., x_n$ with their naming responses $Y = y_1, ..., y_T$, and an optimal k, a kNN algorithm aggregates the histogram of colour names among the k closest neighbours of test colour $\tilde{x}$ to predict its histogram of naming responses by:

$$f(\tilde{x}) = \frac{1}{K} \sum_{x_i \in N_{k(\tilde{x})}} y_i \qquad (5.2)$$

where $N_{k(\tilde{x})}$ are the k closest training observations to query colour $\tilde{x}$. For $k = 1$ the 1-nearest neighbour algorithm is expected to produce good results for individual colours near the centre of categories, but the predictions would be sensitive to noise near the boundaries of categories. Larger values for k will be more robust to noise and would produce smoother predicted lexical colour categories but it might produce larger errors for individual colours. A very large k would result in over-smoothed boundaries between categories. Changing the value of k can change the assigned colour name for each test colour, and different values for k perform differently for different test colour sampling that make the prior selection of the most appropriate k problematic.

5.3. Random Forests

Random Forests (RF) can be viewed as a method for choosing an adaptive k for each test colour $\tilde{x}$ in a kNN framework (Lin & Jeon, 2012). RF is a popular ensemble decision trees algorithm that is being used here in regression mode (Breiman, 2001). A decision tree is a collection of nodes and branches in a treelike structure, where the internal nodes split for each decision. Directed branches represent possible decisions leading to leaf nodes where the final decision is being made. In addition, the non-parametric nature of

regression trees makes the decision models more robust as they rely on fewer assumptions - at the cost of having less power than parametric models – and this makes them suitable for the large colour lexicons in many languages. However, when decision trees are fully grown, they can lead to overfitting on the training set with poor generalisation performance. Ensemble learning methods, like RF, reduce the variance of the expected generalization error by ensembling randomly constructed decision trees – a process known as bagging.

Given our training set of colour points $X = x_1,..., x_n$ with their naming responses $Y = y_1, ..., y_T$, and a free parameter $B = 100$ trees, bagging constructs a forest of B uncorrelated regression trees $f_b = \{T_1,..., T_B\}$ by sampling for each tree T a random subset of the training set with replacement. At each node, the feature to split is selected as the *best* amongst a randomly chosen subset of features of the labelled set of training samples. The splitting process continues for each node until the Gini impurity index (Breiman, Friedman, Stone & Olshen, 1984) cannot be further decreased. The trees in an ensemble are purposely constructed to be independent, so some trees may predict a distribution of Y for X, and others not. The decision boundary in feature space of RF can be curvilinear, and sharp in some parts, while fuzzier in others. After training, the histogram of naming responses for test colour $\tilde{x}$ is estimated by averaging the predictions of all trees on $\tilde{x}$ by:

$$f(\tilde{x}) = \frac{1}{B} \sum_{b=1}^{B} f_b(y_b)$$

$$(5.3)$$

where B is the number of trees, b the tree index, and y_b is the histogram of naming responses on the training points x_b computed as their arithmetic average.

5.4.   Rotated Split Trees

Similarly, to Random Forests, the Extra-Trees algorithm constructs a set of random binary decision trees but each tree is growing using the whole training dataset selected without replacement and splits nodes at random (Geurts, Ernst & Wehenkel, 2006). The produced extreme randomised split trees overcome the perturbations caused by the search for the optimal split during tree growing (Breiman, 2001; Cutler & Zhao, 2001) with competitive performance in terms of accuracy and computational efficiency. Considering that tree-based ensembles are essentially a set of hyper-rectangles that can be sensitive to rotations when partitioning the decision space; prior the construction of each random split tree, we also randomly rotate the representation space to further induce diversity within the constructed forest and as a result to improve the accuracy of

the algorithm at determining the form of colour categories in a three-dimensional space (Blaser & Fryzlewicz, 2016; Andrews, Jaccard, Rogers & Griffin, 2017). In this work, this Rotated Split Trees (RST) approach is being used to predict the histogram of colour names for test samples in regression mode.

Given our training set of colour points $X = x_1, ..., x_n$ with their naming responses $Y = y_1, ..., y_T$, and free parameter $B = 100$ trees, RST ensembles B random-split tress, $f_b = \{T_{1,...,}T_B\}$ by using for each tree T the full training set. Prior to any splitting, a proper rotation matrix R is generated using Householder QR decomposition (Householder, 1958; Blaser & Fryzlewicz, 2016). This rotation matrix is orthogonal with positive diagonal elements, and the attribute space, for each tree, can be then rotated to give a unique coordinate system and increase the diversity in the construction of regression trees. As opposed to un-rotated trees, the rotated trees have different orientation and vastly dissimilar data partition and are capable of producing smoother non-axis parallel decision boundaries. For growing a tree, RST splits the training data at each node independently of the target variable at random, unlike the optimum criterion of RF. Top-down binary recursion continues until no further splits are possible, that is, until all samples have been partitioned into their own leaf node. The predictions of each tree are then aggregated to predict the distribution of colour names for a test colour sample $\tilde{x}$ by:

$$f(\tilde{x}) = \frac{1}{B} \sum_{b=1}^{B} y_b$$

(5.4)

where $B=100$ is the number trees, b is the tree index, and y_b is the histogram of naming responses on the training points x_b computed as their arithmetic average.

5.5.   Evaluation Metric

To assess the performance of each colour naming model M, we measure how well it is defined by generalisable criteria. We enforce generalisability by using a Leave-One-Out and Leave-Planes-Out cross-validation. In the Leave-One-Out mode, we exclude a test chip from the training data, predict its histogram of colour names from the trained model, and score the difference between the predicted and observed histograms using Bhattacharyya (1943) distances. We aggregate scores for each chip in turn left out using the RMS Bhattacharyya distances. In the Leave-Planes-Out mode, we exclude the test chip and all the chips with the same, chroma, or lightness or hue dimensions. And again, we score each interpolation method by RMS of Bhattacharyya distances:

$$RMSE = arccosBC(p, q) = \sum_{x \in X} \sqrt{p} \ \sqrt{q} \qquad (5.5)$$

where *p* and *q* are the predicted and observed colour naming distributions for colour *x* respectively. The range of the Bhattacharyya angle is between 0 for perfect overlap and 1.5708 (π/2) for complete separateness.

5.6. Interpolation of colour naming responses across the full colour gamut

In Table 5.1, we compare the performance of the different computational models. The first step in our procedure was to quantify the minimum RMSE that could be achieved given the sparseness of data. We resampled the distribution of colour names across samples 100 times and we measured the overlap between observed and bootstrapped distributions with an RMSE = 0.51. In the second step, we measured as a baseline performance the RMSE between the observed distributions and the mean distribution of all other test samples using the leave-one-out with an RMSE = 1.35 and leave-planes-out with RMSE = 1.36 cross validations.

The 1-Nearest Neighbour interpolation approach based on ΔE_{ab} and ΔE_{00} colour difference formulas produced $RMSE_{\Delta Eab}$ = 1.11 and $RMSE_{\Delta E00}$ = 1.10 for the Leave-One-Out and $RMSE_{\Delta Eab}$ = 1.19 and $RMSE_{\Delta E00}$ = 1.16 for the Leave-Planes-Out cross validations respectively. For kNN, RF and RST, we also present the results of applying a square root (SQRT) transformation to the histograms that empirically improved their performance. In this variant, we interpolated the square root of the histograms, and then squared and normalised them to unit sum after estimating at a test colour. The k-Nearest Neighbours method produced the minimum RMSE using the ΔE_{00} colour difference formula, but it required tuning the number of *k* nearest neighbours for different cross validation modes. Random Forest with 100 trees, performed slightly worse in the Leave-One-Out validation, but it is sensitive to rotation of the feature axes and its performance deteriorated in the Leave-Planes-Out validation. The RST with 100 trees performed best in terms of accuracy and simplicity of parameter tuning. We limited the number of trees to 100 because there was no need to further increase the processing time in terms of RMSE (Figure 5.1) while increasing the number of trees to 500 would, in practice, increase the error again (Probst & Boulesteix, 2017). In Figure 5.2, we show an example of observed and predicted distributions for a single colour sample using RST and in Figure 5.3 the total histogram of RMSE for RST.

Table 5.1 Comparison of interpolation methods for automating the colour naming task

	Leave-One-Out	Leave-Planes-Out
Resampling	0.51	
Baseline	1.35	1.36
1-NN (ΔE_{ab})	1.11	1.19
1-NN (ΔE_{00})	1.10	1.16
k-NN (ΔE_{ab})	1.00, k=9	1.06, k=8
k-NN (ΔE_{00})	1.00, k=9	1.05, k=7
RF	1.08	1.19
RST	1.03	1.10
k-NN (ΔE_{ab}-SQRT)	0.90 k=15	0.96 k=11
k-NN (ΔE_{00}-SQRT)	0.89 k=18	0.96 k=9
RF (SQRT)	0.92	1.04
RST (SQRT)	0.91	0.96
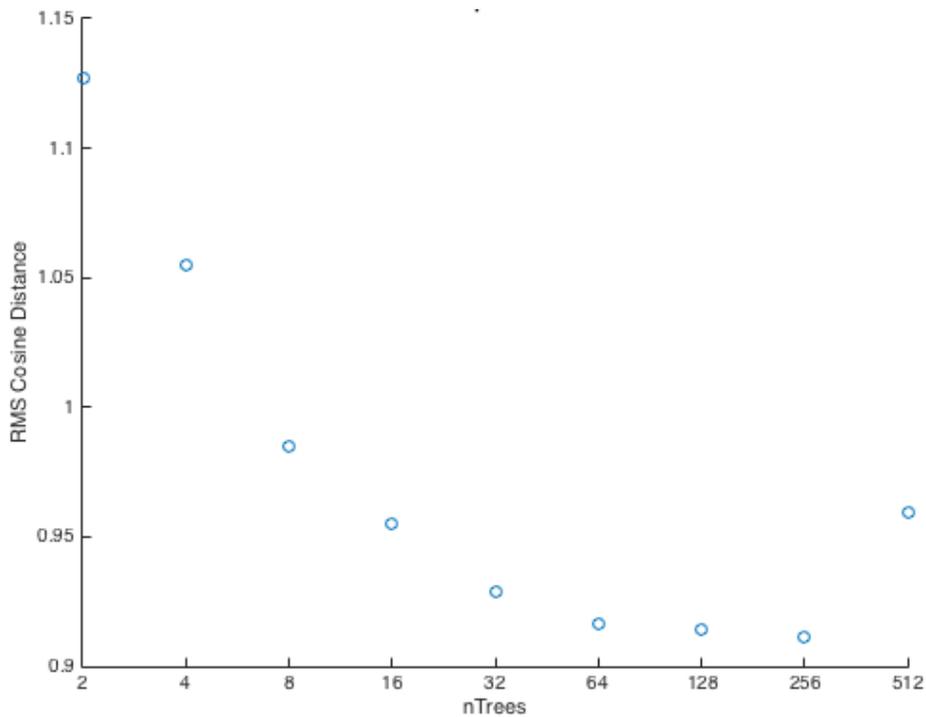
Figure 5.1. Root mean square error between observed and predicted colour naming distribution per number of trees for RST.
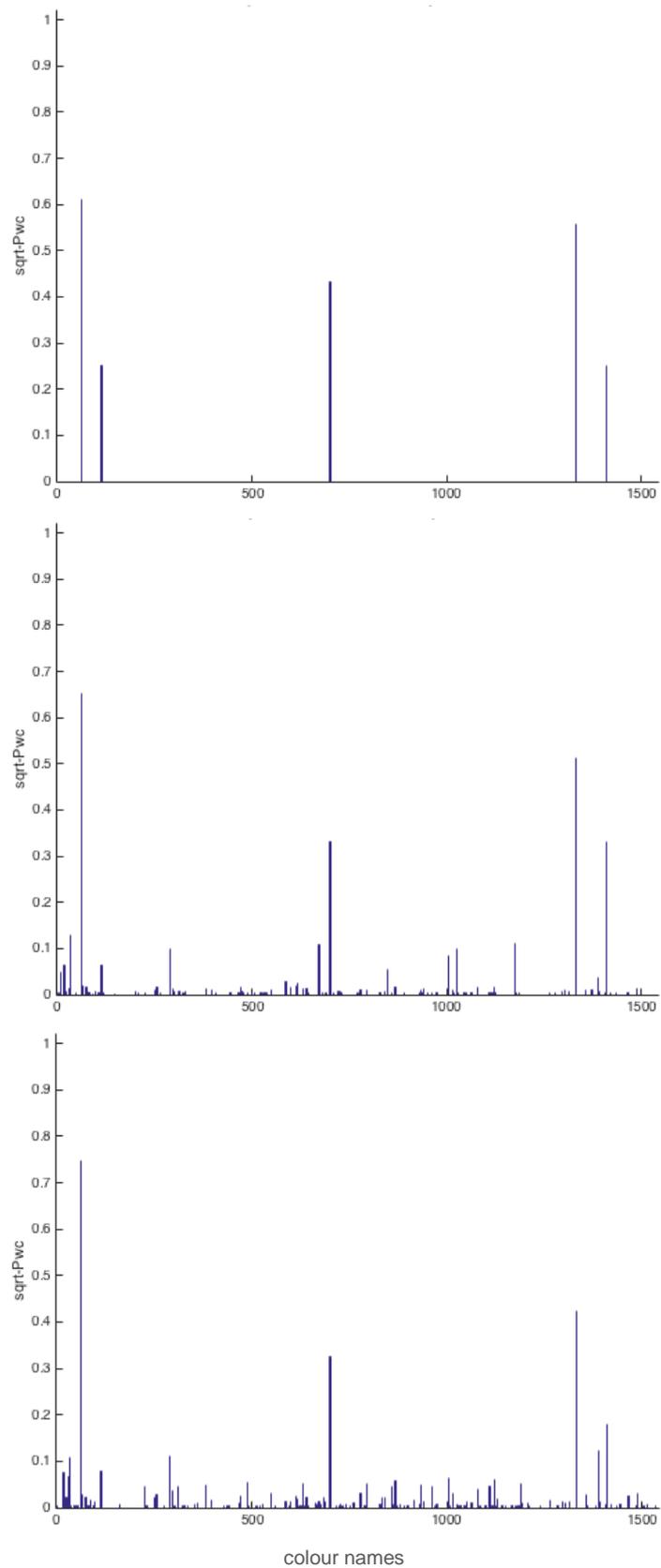
Figure 5.2. Observed (top) and predicted colour naming distributions for a test colour (id=321) using RST in leave-one-out (middle) and leave-planes-out (bottom) cross validation modes.
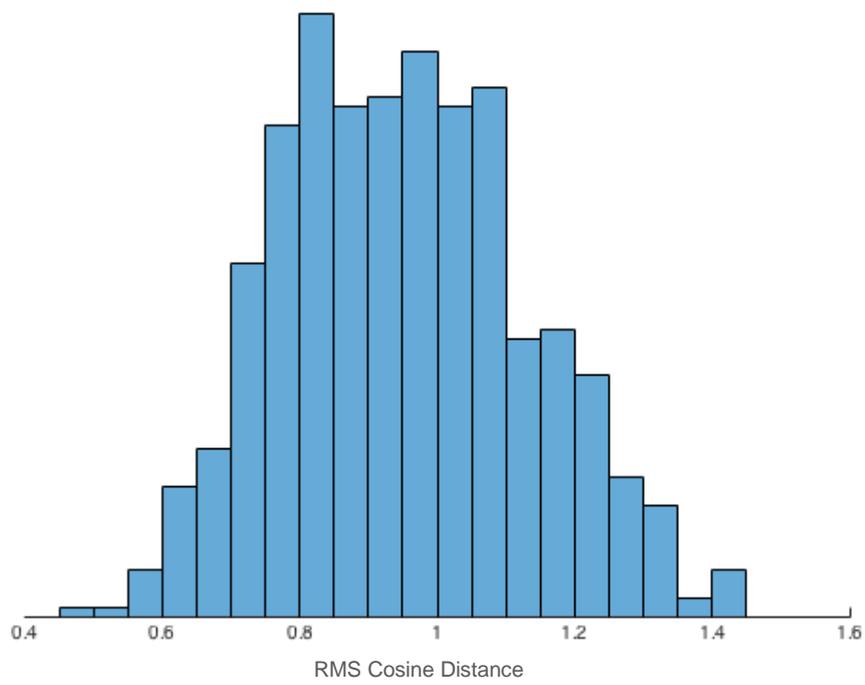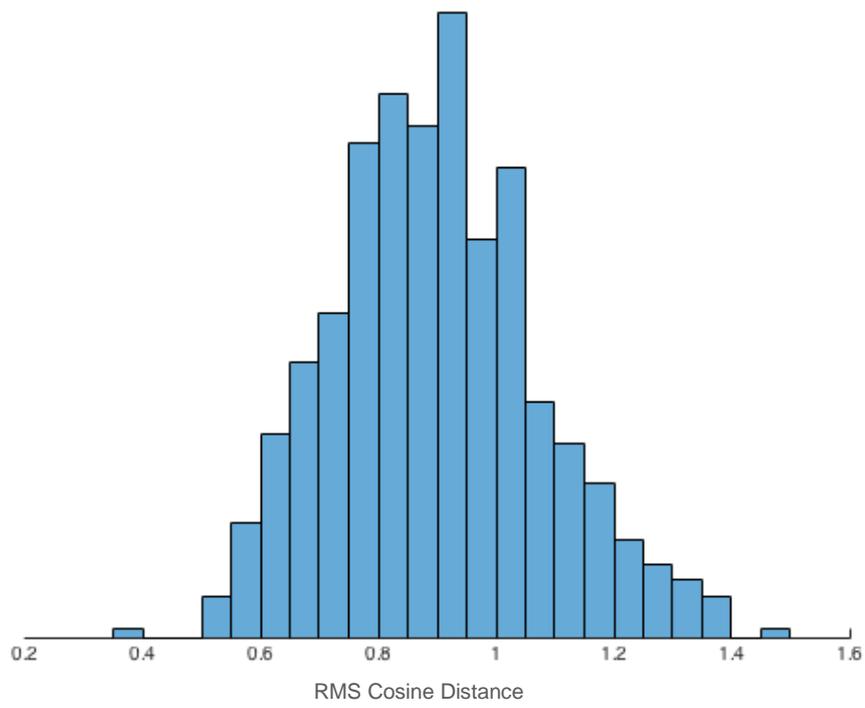
Figure 5.3. RMSE histogram between observed and predicted distributions using RST in leave-one-out (top) and leave-planes-out (bottom) cross validation modes.

5.7. Variability in colour naming

As an example of RST in operation, in this section we infer histograms of naming responses for any colour and compute their entropy as a measure of naming variability. First, we classified cross-sections of the sRGB gamut to the maximum predicted colour names in CIELAB (Figure 5.4). In L*<50, purple and brown were the largest categories followed by green, dark green, and pink and black. In L*>50, colour names with the largest partitions were green and pink followed by lilac, grey, yellow and turquoise.
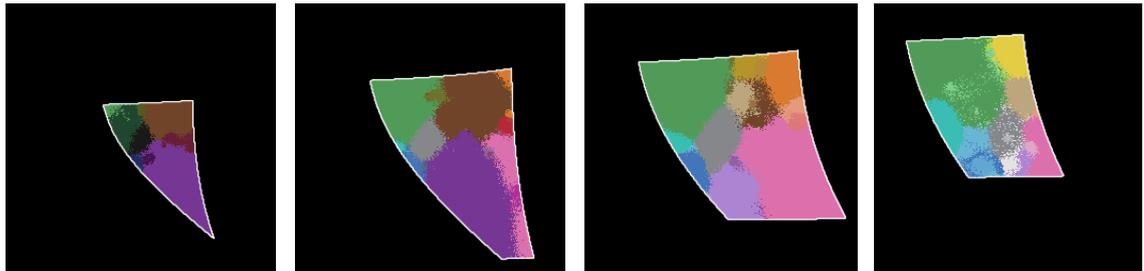
Figure 5.4. Most likely colour names at each colour estimated by RST on planes of sRGB gamut in CIELAB for lightness levels of L*=20, 40, 60 and 80 (from left to right). Coordinates of their centroids were used to colour each name category.

Second, in Figure 5.5 we present the variability in colour naming distributions measured by their entropy. The entropy of colour name distributions was lower at the corners of the gamut and around the neutral axis. These low entropy regions are associated to the eleven basic colour terms but also to lilac, turquoise and light blue.
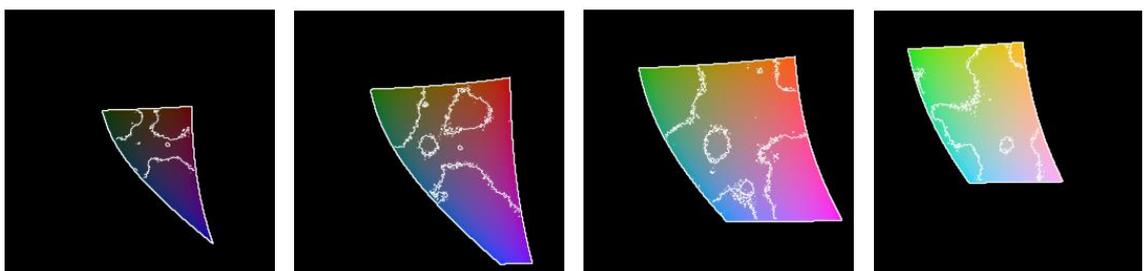
Figure 5.5. Contour of entropy of colour naming distribution overlaid on planes of sRGB gamut in CIELAB for lightness levels of L*=20, 40, 60 and 80. Low entropy in areas enclosed by contour.

We also compared the variability of free colour naming at basic colour foci in English (Berlin & Kay, 1969/1991), with the variability at non-focal locations on the exterior of the Munsell colour space (Figure 5.6). At foci, the mean and standard deviation were 2.13 and 0.53 respectively; at non-foci 2.93 and 0.65. All foci of basic colour terms fall in regions of low entropy, except red. Foci red were most frequently named as red but were

also named as orange, brick red, terracotta and maroon. Low variability was also observed for light blue, turquoise and lilac.
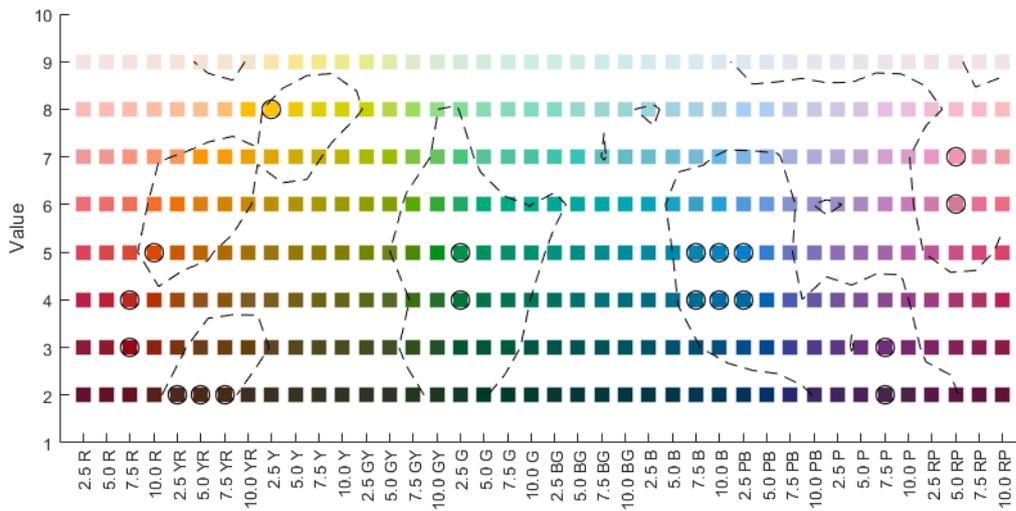
Figure 5.6. Contour of entropy of colour naming distribution overlaid on Mercator of the Munsell Colour System. Low entropy in areas enclosed by contour. Foci of Berlin & Kay (1969/1991) are shown in black circles.

5.8. Evaluation of colour spaces

To investigate whether the choice of colour space influences the results of the classification model, we compared the performance of RST_{sqrt} using Leave-One-Out and Leave-Planes-Out cross-validations in several colour spaces (RGB-linear, CIE XYZ 1931, CIELAB, CIELUV and CIECAM02), assuming the sRGB viewing conditions (IEC, 1999). As in our evaluation above, each colour space was scored by RMS of Bhattacharyya distances shown in Table 5.2.

Table 5.2 Comparison of colour spaces for interpolation of colour naming distributions using RST$_{sqrt}$.

	Leave-One-Out	Leave-Planes-Out
RGB (linear)	0.99	1.03
sRGB	0.99	1.03
CIEXYZ$_{1931}$	0.97	1.14
CAM02UCS$_{sRGB}$	0.95	1.02
CIELAB$_{D65}$	0.92	0.97
CIELUV$_{D65}$	0.91	0.96

Overall, the predictions of the RST$_{sqrt}$ algorithm were better in the approximately perceptually uniform colour spaces (CIELAB, CIELUV, CAM02-UCS) than in the non-uniform (RGB, sRGB, CIE XYZ 1931) spaces. The best colour space in terms of accuracy of predictions in both cross-validation modes was by a margin, CIELUV.

5.9. Comparison to earlier colour naming models

To compare the performance of RST against previous colour naming models based on the monolexemic psychophysical data of Berlin & Kay in American English (1969/1991) and of Sturges and Whitfield (1995) in British English, we first followed the approach described by Guest & Laar (2000) and restricted the responses to their last word resulting in 320 distinct colour terms instead of just the eleven terms of previous colour naming models. We trained the RST with these monolexemic responses and inferred their histograms for the 330 patches of the simulated Munsell array in CIELUV.

RST assigned the 330 chips to 16 colour terms (11 BCTs plus turquoise, lilac, maroon, peach, mauve and teal) by taking the peak of the estimated distribution for each chip. In Figure 5.7 we show the segmentation of the colour chart by RST compared to the classification reported by Berlin & Kay (1969/1991), shown with black boxes. RST classified all foci of BCTs to their corresponding categories but misclassified 27 patches at the borders of the categories which were often classified to a non-BCT.

In Figure 5.8, we show the segmentation of the simulated Munsell array against Sturges & Whitfield's colour naming results drawn with black boxes. Despite the five additional non-BCTs identified by RST, all the patches 100% of the BCTs categories were assigned to the correct term.

In Table 5.3 we show the comparison of the performance of the RST model against previous colour naming models of Lammens's Gaussian model (LGM; 1994); MacLaury's English Speaker (MES; 1992); Benavente and Vanrell's Triple Sigmoid model (TSM; 2004); Seaborn's fuzzy k-means model (SFKM; 2005); Benavente et al's Triple Sigmoid- Eliptic Sigmoid model (TSMES; 2008); van de Weijer et al's Probabilistic Latent Semantic Analysis (PLSA; 2007); Parrage & Akbarinia's Neural Isoresponsive Colour Ellipsoids model (NICE; 2016); and Mylonas et al's Maximum a Posteriori (MAP; 2010).

We trained our earlier colour naming model (MAP; Mylonas et al., 2010) using the same dataset of 320 monolexemic terms similarly to the RST model as described above. MAP identified 17 colour terms on the Munsell array and its performance against the results of both Berlin & Kay and Sturges & Whitfield can be found in Appendix C. MAP was able to classify all foci of BCTs to their corresponding categories but misclassified 34 patches at the borders of the categories against the results of Berlin & Kay. Against the results of Sturges & Whitfield, MAP misclassified only 1 out of 111 chips at the *lime green* region between *green* and *yellow*.

RST performed equally well (100%) with other state-of-the-art colour naming models (SFKM, TSMES and NICE) on Sturges & Whitfield's results while RST identified five additional terms on the Munsell array. This means that the estimated distribution of BCTs using the RST model is tighter than those of previous models that constrained their responses only to the eleven BCTs. On the other hand, the performance of RST against Berlin and Kay's results, as to be expected given the additional categories, was poorer than other models constrained to the 11 BCTs.
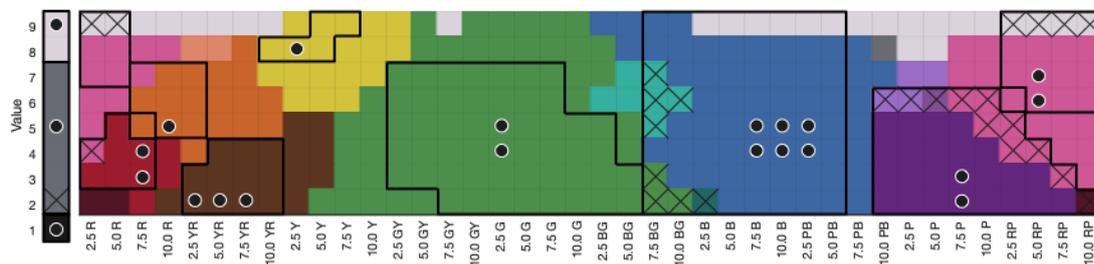
Figure 5.7 Segmentation of simulated Munsell array into 16 monolexemic colour terms by RST model. Coordinates of their centroids were used to colour each name category. Berlin and Kay's foci of BCTs in American English are drawn with dots and their distribution with black boxes. Cross marks denote differences in naming by the RST model and Berlin & Kay's results.
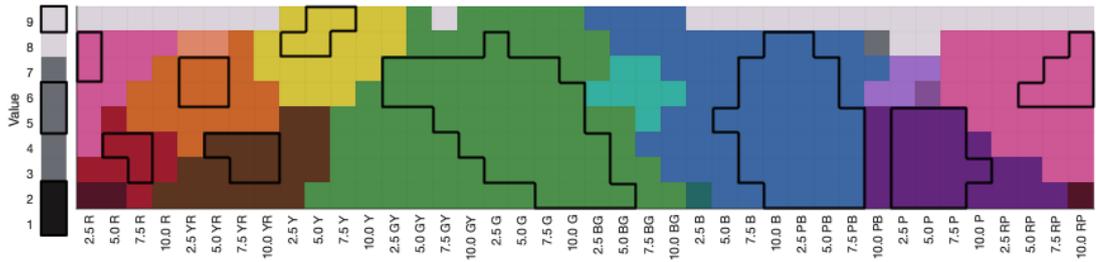
Figure 5.8 Segmentation of simulated Munsell array into 16 monolexemic colour terms by RST model. Coordinates of their centroids were used to colour each name category. Sturges & Whitfield's mapping of BCTs in British English are drawn with black boxes.

Table 5.3 Comparison of colour naming models on the Munsell array (n=330 chips) against Berlin & Kay (1969/1991) and Sturges & Whitfield (1995) results. The data for LGM, MES, TSM, SFKM, TSEM, PLSA and NICE was obtained from Table 4 in Parrage & Akbarinia (2016).

	Berlin & Kay results			Sturges & Whitfield results		
	Coincidences	Errors	%	Coincidences	Errors	%
LGM	161	49	23	92	19	17
MES	182	28	13	107	4	4
TSM	185	25	12	108	3	3
SFKM	193	17	8	111	0	0
TSEM	193	17	8	111	0	0
PLSA	187	23	12	109	2	2
NICE	206	4	2	111	0	0
MAP	176	34	16	110	1	1
RST	184	27	13	111	0	0

5.10. Computational colour naming within different languages

For automating the colour naming task in different languages, we trained the RST model using the multilingual datasets in British and American English, Greek, Russian, Thai and Turkish, in order to segment the colour space into lexical colour categories. For clarity, we visualise the categories on a two-dimensional (2D) plane. We created a synthetic image by taking a cross section of a conic representation of the HSL colour solid where the additive and secondary primaries (red, green, blue and yellow, cyan,

magenta) are arranged around the outside edge of the solid at maximum Saturation of 1 and Lightness of 0.5. The HSL coordinates were then converted to CIELUV via sRGB. The synthetic image is not isoluminant in CIELUV but has hills and valleys (Figure 5.9).
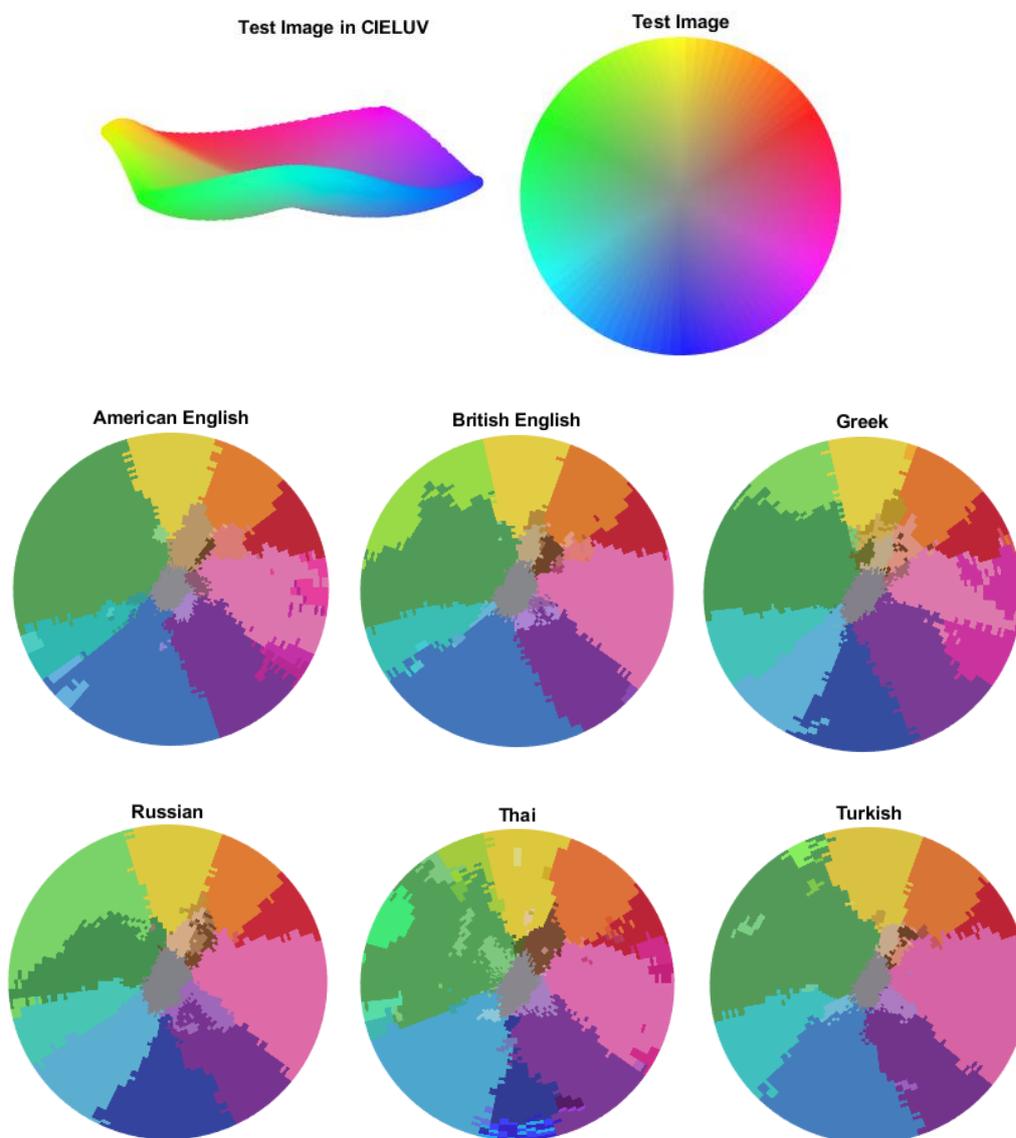
Figure 5.9 Segmentation of synthetic test image. Test image in CIELUV (1st row – left), test image in HSL (1st row – right). Segmentation of test image by American English (2nd row – left), by British English (2nd row – middle), by Greek (2nd row – right), by Russian (3rd row – left), by Thai (3rd row – middle) and by Turkish (3rd row – right) colour names. Coordinates of their centroids were used to colour each name category. Some small categories cover <1% of the synthetic image.

Learning from British English speakers, the RST algorithm assigned the colour coordinates of the synthetic image into 30 colour names. The seven largest categories were BCTs including *green, blue, grey, pink, purple, yellow* and *orange*. *Red* and *brown* were the 10th and 13th largest categories. *Black* and *white* were not assigned to any

coordinate as these regions were not sampled in the synthetic image. *Turquoise* (8th) and *lime green* (9th) were the non-basics with the largest coverage in the test image with *lilac* (11th) and *beige* (12th) found also to cover regions larger than *brown*. *Red* was restricted to the most saturated colours with *salmon*, *peach*, *pink* and *orange* covering the pale region of the same hue angles. *Turquoise* was assigned to pixels all the way from the neutral axis to the limit of the gamut while *lilac* was restricted to the pale regions of *purple*.

Trained by American speakers, RST identified 26 colour names in the synthetic image. Similar to the segmentation by British English, the seven largest categories were BCTs – with *red* and *brown* being the 10th and 17th largest categories respectively. *Turquoise* was found again as the 8th largest category followed by *tan* (9th). *Lavender*, instead of the British *lilac*, was the 11th largest category. *Salmon* was assigned to the 13th largest category while *magenta* and *teal* to the 15th and 16th largest categories respectively. No pixels were assigned to *lime green* or to *chartreuse*.

Learning from the Greek dataset, RST identified 28 lexical colour categories in the test image. The five largest categories were the BCTs *green* (*prasino*), *purple* (*mov*), *grey* (*gri*), *blue* (*ble*) and *pink* (*roz*). *Yellow* (*kitrino*) was the 8th, *orange* (*portokali*) was the 10th, *red* (*kokkino*) was the 12th and *brown* (*kafe*) was the 19th largest categories. *Sky blue* (*galazio*) the proposed second blue basic category in Greek was the 6th largest category covering regions from the neutral axis to the limits of the gamut. Similarly, turquoise (*tirkuaz*) was the 7th most common category. *Lime green* (*lahani*) and *fuchsia* (*fouxia*) were also very popular categories followed by *beige* (*bez*), *salmon* (*somon*) and *olive* (*ladi*). *Lilac* (*lila*) was assigned to only 3 pixels.

In the synthetic image segmented by RST using the Russian dataset, 23 categories were identified. The largest categories were the BCTs *grey* (*seryj*), *pink* (*rozovyj*), and *green* (*zelënyj*). *Blue* (*sinij*), *purple* (*fioletovyj*) and *yellow* (*žëltyj*) were the 6-8th largest categories while *orange* (*oranževyj*) was the 10th; with *red* (*krasnyj*) the 12th and *brown* (*koričnevyj*) the 14th most popular categories. The second basic *blue* (*goluboj*) was the 4th while the non-basics *lime green* (*salatovyj*), *turquoise* (*birûzovyj*) and *lilac* (*sirenevyj*) were the 5th, 9th and 11th largest categories respectively.

Learning colour names from Thai speakers, the RST algorithm identified 46 colour names in the synthetic image. Again, the seven largest categories were the BCTs *green* (*khiaw*), *grey* (*thaw*), *sky blue* (*fa*), *pink* (*chompu*), *purple* (*muang*), *yellow* (*leaung*) and *orange* (*som*). *Brown* (*namtan*) and *red* (*dang*) were the 10th and 12th largest categories.

The proposed second basic blue (*namngen*) was the 9th most common category. The largest non-basics were *light green* (*khiawon*) and *light purple* (*muangon*). Compared to all other languages, the *turquoise* category in Thai (*faomkhiaw*) was assigned to a much smaller number of pixels (>1%).

The RST algorithm identified 26 colour names when was trained by Turkish speakers. The six largest categories were the BCTs, *green* (*yesil*), *pink* (*pembe*), *blue* (*mavi*), *grey* (*gri*), *purple* (*mor*) and *yellow* (*sari*). *Orange* (*turuncu*) was the 9th, *red* (*kirmizi*) the 11th and *brown* (*kahverengi*) the 15th largest categories. The proposed second basic *blue* (*lacivert*) was not found in the synthetic image. *Turquoise* (*turkuaz*) and *lilac* (*lila*) were the 7th and 9th largest categories.

5.11. Discussion

In this chapter, we evaluated several supervised nonparametric colour naming models using Leave-One-Out and Leave-Planes-Out cross-validation. Each method was scored by RMS of Bhattacharya distances between observed and interpolated histograms of colour naming responses. A Rotated Split Trees (RST) approach demonstrated the best performance. RST is a supervised nonparametric model that chooses each attribute and split at random rather than by selecting the attribute that best splits labelled sets of training samples (Geurts et al., 2006; Breiman, 2001). In addition, the random rotation of its decision space infuse diversity within the constructed forest and improves its accuracy at determining the form of colour categories in a three-dimensional space (Blaser & Fryzlewicz, 2016; Andrews, Jaccard, Rogers & Griffin, 2017). Nonparametric models make fewer assumptions about the shape of colour categories than parametric approaches and as a result provide wider applicability and increased robustness.

An evaluation of the predictions of RST in several colour spaces (linear RGB, sRGB, CIEXYZ 1931, CIELAB, CIELUV and CIECAM02-UCS) showed that overall the algorithm performed best in approximately perceptually uniform colour spaces (CIELAB, CIELUV, CAM02-UCS) than in the non-uniform (RGB, sRGB, CIE XYZ 1931) spaces. The best colour space in terms of accuracy of predictions was CIELUV in agreement with a recent study that suggested CIELUV as the best space for performing colour clustering algorithms (Douven, 2017). These results reinforce the CIE recommendation for using the CIELUV colour space for colours displayed on display monitor that was the presentation mode of our online colour naming experiment.

Using RST, we inferred histograms of naming responses for any colour, and computed their entropy as a measure of naming variability. All foci of basic colour terms fall in regions of low entropy, except *red*. Foci *red* were most frequently named as *red* but were also named as *orange*, *brick red*, *terracotta* and *maroon*. Low variability was also observed for *light blue*, *turquoise* and *lilac*.

We also compared the classification of the most saturated colours of the Munsell Array into colour terms by RST against previous colour naming models (Lammen, 1994; MacLaury; 1992; Benavente & Vanrell, 2004; Seaborn, 2005; Benavente et al, 2008; Weijer et al., 2007; Mylonas et al., 2010; Parrage & Akbarinia, 2016). Despite the identification of 16 (11 BCTs + *turquoise, teal, lilac, mauve* and *maroon*) distinct colour terms rather than only the 11 BCTs considered by previous models, RST achieved the same level of state-of-the-art performance for the psychophysical results of Sturges & Whitfield (1995). The performance of RST against Berlin and Kay's results in American English was poorer than the performance of other models constrained to the 11 BCTs at the boundaries of colour categories. However, the early data of Berlin & Kay's study (1969) was not generated from a colour naming task but from a best example task, and as such they are not appropriate to evaluate category boundaries. The assumption that BCTs can name all colours on the surface of the colour space is not supported by empirical findings (Boynton & Olson, 1987; Sturges & Whitfield, 1995; Mylonas & MacDonald, 2016).

Furthermore, we trained RST with multilingual datasets in British and American English, Greek, Russian, Thai and Turkish from the online colour naming experiment to segment a synthetic colour wheel that includes the most saturated regions of the RGB cube. Overall our nonparametric model produced linear boundaries between categories except for categories falling in the interior of the colour space like the achromatic grey and skin colours. Except for *white* and *black* which were not sampled in the synthetic image, the other 9 BCTs were assigned relatively to large categories in all languages. *Turquoise* occupied also a large well-defined region of the colour space between *blue* and *green* extending from the neutral axis to the limits of the gamut in all languages except in Thai. Despite the relatively low sampling of the *light purple* region in the synthetic image, *lilac* was also very popular in all languages except in Greek. These results support the suggestion of adding *turquoise* and *lilac* to the set of BCTs not only in English but also in other languages (Mylonas & MacDonald, 2016). Our results also support the addition of the proposed second basic blue terms – *galazio* in Greek (Androulaki et al., 2006; Athanasopoulos, 2009), *goluboj* in Russian (Morgan & Corbett, 1989; Paramei, 2005; Paramei et al. 2018) and *namngen* in Thai (Prasithrathsint, 1988; Engchuan, 2003) – to

the set of BCTs, as the these categories covered large regions in these languages but not in the other test languages. Our findings do not offer support to the proposed second basic *blue* term, *lacivert*, in Turkish (Özgen & Davies, 1998; Ekici et al., 2006) as it was not assigned to any colour of the synthetic image. A specific analysis assessing the basicness of colour categories is presented in Chapter 7.

In conclusion, we presented a supervised nonparametric computational colour naming model in order to automate colour naming within different languages. Our tools and data allowed the analysis of colour names across the full 3D colour gamut and revealed structure in the interior of colour space. Our model performs best in CIELUV and achieves the same level of state-of-the-art performance as earlier models while it identifies 5 additional colour categories on the surface of the Munsell system. The performance of the model in different languages is also supported by the empirical findings of earlier studies in colour naming of these languages.

Chapter 6

Offline colour naming experiment

Online colour naming experiments are often criticised for the uncalibrated colour reproduction of different displays and viewing conditions of the participants. On the other hand, laboratory- based (offline) experiments are also judged as being unable to predict colour names in real world monitor settings. To respond to these criticisms, a comparison between web- and laboratory- based experimental methodologies in estimating colour naming functions is needed. Furthermore, an accurate measurement of colour names represented on a fundamental scale with physiological axes is of great importance. The recent establishment of the physiologically-based colour matching functions (Stockman & Sharpe, 2000; CIE 170-1:2006; CIE 170-2: 2015) provides a satisfactory determination of the cone excitation space, but its relationship to higher-order cognitive processes of colour appearance remains uncertain. In this chapter, we determine the location of colour names within the new cone excitation space through an unconstrained colour naming experiment of 600 simulated samples of the Munsell system on a calibrated CRT monitor and compare the findings against the results of the online experiment.

6.1.   Internet- and laboratory- based colour experiments

Online experimental methodologies often receive criticism as not meeting the exacting standards demanded for rigorous colour research. Web-based colour naming experiments provide greater ecological validity than traditional approaches by allowing a large number of participants to name colours freely in their own environment, in their own time, with their own equipment and without the physical attendance of the examiner (Reips, 2000; Moroney, 2003). Laboratory-based experiments on the other hand, offer the opportunity to calibrate and characterise the colour reproduction device, usually a cathode-ray tube (CRT) monitor, using a colorimetric specification and controling the viewing conditions. This linear relationship between the RGB monitor and, for example, the new CIE colour matching functions is an important requirement for the accurate determination of the location of colour names within a cone excitation space. However, predicting colour naming functions based on responses from only a small number of observers in a laboratory test environment also has limitations. Here, we compare the results of the online colour naming experiment against an offline colour naming

experiment and at the same time investigate the performance of the latter in estimating colour naming functions measured under uncalibrated conditions.

6.2. Cone excitation space and colour naming

In Chapter 2, we described the physiologically-relevant cone fundamentals (Stockman & Sharpe, 2000) and their linear transformations into a new set of colour matching functions recently adopted by the CIE (CIE 170-1: 2006; CIE 170-2: 2015). This precise specification of the L, M, S cones spectral sensitivities provides the basis to represent colour in a cone excitation space, but its relationship with colour appearance mechanisms is still unknown. Earlier attempts to link colour names to cone excitations (Cao, Pokorny & Smith, 2005; Parrage & Akbarinia, 2016) focused only on a small number of 11 basic colour terms (Berlin & Kay, 1969/1991) and utilised the cone fundamentals based on the revised by Judd and Vos, CIE 1931 colour matching functions (Smith & Pokorny, 1975; Wyszecki & Stiles, 1982). Smithson and her colleagues, utilised differences of response time in a reverse Stroop task to map the new cone fundamentals to five focal colours of red, orange, yellow, green and blue (Smithson, Khan, Sharpe & Stockman, 2006). In the experiment described in this chapter, we set out to measure directly, for the first-time, unconstrained colour naming distributions in the new CIE adopted physiologically-based cone excitation space (Stockman & Sharpe, 2000; CIE 170-1:2006; CIE 170-2: 2015).

6.3. Materials and procedure

Observers were seated one metre away from a CRT monitor (22-inch Mitsubishi Diamond Pro, 2070SB) in an otherwise dark room with neutral grey painted walls. The CRT monitor was calibrated using a ColorCal CRS (Cambridge Research Systems) colorimeter and characterised using a RadOMA spectroradiometer (Gamma Scientific, San Diego, California) positioned at 1 metre distance and aiming straight at the centre of the screen. The measured CIE 1931 chromaticity coordinates of the white point of the monitor were $x = 0.3126$, $y = 0.3296$ with a correlated colour temperature of 6507K and a luminance of 80.17cd/m^2. The stimulus presentation and response timing were controlled by PsychoPy-version 1.84.2 software (Pierce, 2007) and by a DATAPixx display driver (Vpixx Technologies Inc.) with 16 bits per RGB gun resolution. The spectral power distribution for each R, G and B gun of the CRT monitor and a linearity evaluation can be found in Appendix C. The task of the observers was to name out loud the colour of the stimulus, so that others will know to which colour they were referring. Observers were free to use as broad or narrow names as they liked. A head-mounted microphone

was used to record vocal responses. Response times are measured using a detection limit of 4 standard deviations of the blank (silence) to avoid false positive responses due to noise.

6.4.   Observers

Ten international English speakers from Britain, Unites States and Australia (3 males, 7 females, mean age 46.5 years; SD = 14.3) living in London for more than 10 years participated in the experiment. All observers were screened for colour vision deficiencies using the City University Test (Fletcher, 1978). Participants provided written informed consent and were compensated with cash for their time. The study was approved by the Research Ethics Committee at the University College London (3387/001).

6.5.   Stimuli

Test stimuli were uniformly coloured 2 degrees of visual diameter discs with a black outline of 1 pixel against a D65 neutral grey with luminance of $40cd/m^2$. Similar to the sampling of the online colour naming experiment, the stimuli consisted of 589 simulated samples from the Munsell colour order system plus 11 achromatic samples. The 600 colour samples were presented one at a time in random order for each observer.

6.5.1. Correction of spectral reflectance of Munsell chips

The Munsell colour system (Munsell, 1905) was designed with the objective of representing perceptually uniform visual spacing of Hue, Chroma (saturation) and Value (lightness) dimensions. The system was revised and specified colorimetrically for CIE illuminant C and the 1931 standard observer (Newhall et al., 1943). The resulting Munsell Renotation Dataset (MRD) is available at the website of the Munsell Color Science Laboratory. By constrast, the Munsell Book of Colour is a physical reproduction of the system available in glossy or matte finish editions. The Measured Spectral Reflectance Data (MSRD) of the matt Munsell chips are accessible at the website of the Spectral Color Research Group at the University of East Finland. A number of studies (e.g. D'Andrade & Romney, 2003; Olkkonen et al., 2009; Vazquez-Corral, O'Regan, Vanrell & Finlayson, 2012; Skelton et al., 2017) assumed that the colourimetry of MRD and MSRD data match, but in Figure 6.1, we show the differences between them (n=1021) in CIELAB for consistency with earlier reports by Derhak and Berns (2012) and Li & Lee (2014). The main differences are found in the Chroma and Lightness dimensions with a mean colour difference of CIE ΔE_{00} = 3.2652; SD = 0.9238.
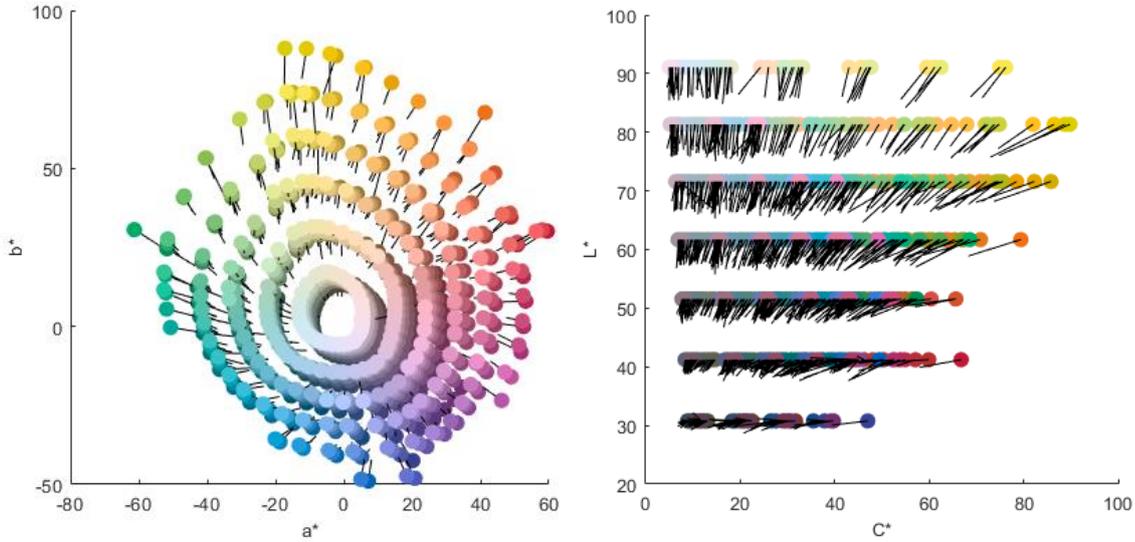
Figure 6.1 Errors between intersection (n=1021) of Munsell Renotation Dataset (markers) and Munsell Spectral reflectance Dataset, Matte edition (tails) in a* vs. b* (left) and L*(lightness) vs. C*(chroma; right) planes of CIELAB.

We followed the method described by Derhak and Berns (2012) to correct the spectral reflectance of MSRD from 390nm to 780nm at 1nm wavelength resolution. For each of the 1021 common samples of MRD and MSRD, we estimated a transformation matrix E (391 × 3) computed by:

$$E = RC^{-1} \tag{6.1}$$

where R is a matrix of spectral reflectance of MSRD (1021 × 391) and C^{-1} is the pseudo-inverse matrix (1021 × 3) of the corresponding tristimulus values of MRD. The corrected spectral reflectance R_{corr} can be then computed by:

$$R_{corr} = R + E(C_{ref} - C) \tag{6.2}$$

where C_{ref} is the tristimulus values of the MRD and C the tristimulus values of the MSRD. Hereinafter, we use the corrected reflectance R_{corr} for the Munsell chips.

6.5.2. Transformation of CIE XYZ 1931 tristimulus coordinates of the Munsell Renotation Data to CIE XYZ 2015

To specify the colour samples of MRD used in the online experiment (*n*=589) – some of which are not included in the MSRD – in a cone excitation space, the CIE*xy* 1931 chromaticities and the luminance of MRD should be linearly transformed to the L-, M- and S- cone fundamentals (Stockman & Sharpe, 2000; CIE 170-1: 2006). However, the

V(λ) used in the CIE XYZ 1931 standard observer is profoundly insensitive at the short wavelengths of the spectrum and is not linearly related to CIE LMS 2006 (Stockman & Sharpe, 1999). Although there is a linear model for mapping CIE LMS 2006 cone fundamentals to CIE XYZ 2015 chromaticities (CIE 170-1: 2006; CIE 170-2: 2015), there is no linear relationship between CIE LMS 2006 and CIE XYZ 1931 (Golz & MacLeod, 2003). Instead, we utilise the *fmincon* function of Matlab to implement a constrained nonlinear multivariable optimisation and find the best transformation matrix shown in Equation 6.3 that minimises the RMSE between the intersection of CIE XYZ 1931 of MRD, and of CIE XYZ 2015 derived from the corrected MSRD.

$$\begin{pmatrix} X_{15} \\ Y_{15} \\ Z_{15} \end{pmatrix} = \begin{pmatrix} 0.9708 & 0 & 0 \\ 0.0130 & 0.9669 & 0.0175 \\ 0 & 0 & 0.9102 \end{pmatrix} \begin{pmatrix} X_{31} \\ Y_{31} \\ Z_{31} \end{pmatrix} \tag{6.3}$$

In Equation 6.4, we give the inverse of the transformation matrix for changing the CIE XYZ 2015 coordinates of MRD back to CIE XYZ 1931.

$$\begin{pmatrix} X_{31} \\ Y_{31} \\ Z_{31} \end{pmatrix} = \begin{pmatrix} 1.0301 & 0 & 0 \\ -0.0138 & 1.0342 & -0.0199 \\ 0 & 0 & 1.0987 \end{pmatrix} \begin{pmatrix} X_{15} \\ Y_{15} \\ Z_{15} \end{pmatrix} \tag{6.4}$$

In Figure 6.2, we show the RMSE before and after the optimisation routine. The main errors were found in the *z* dimension (RMSE = 3.7848) that is related to the short wavelength sensitive cone (Stockman & Sharpe, 1999). This error was reduced substantially (RMSE = 0.3760) after applying the optimised transformation matrix of Equation 6.3. Then, we checked whether any of the colour stimuli were out of gamut using a round-trip clipping method and no stimuli were found above the 3 RMSE (or 5*ΔE$_{ab}$*) threshold (Kang, 2006). Figure 6.3 shows the 589 colour samples of the MRD used in the online experiment in the new chromaticity diagram (CIE 170-2: 2015).
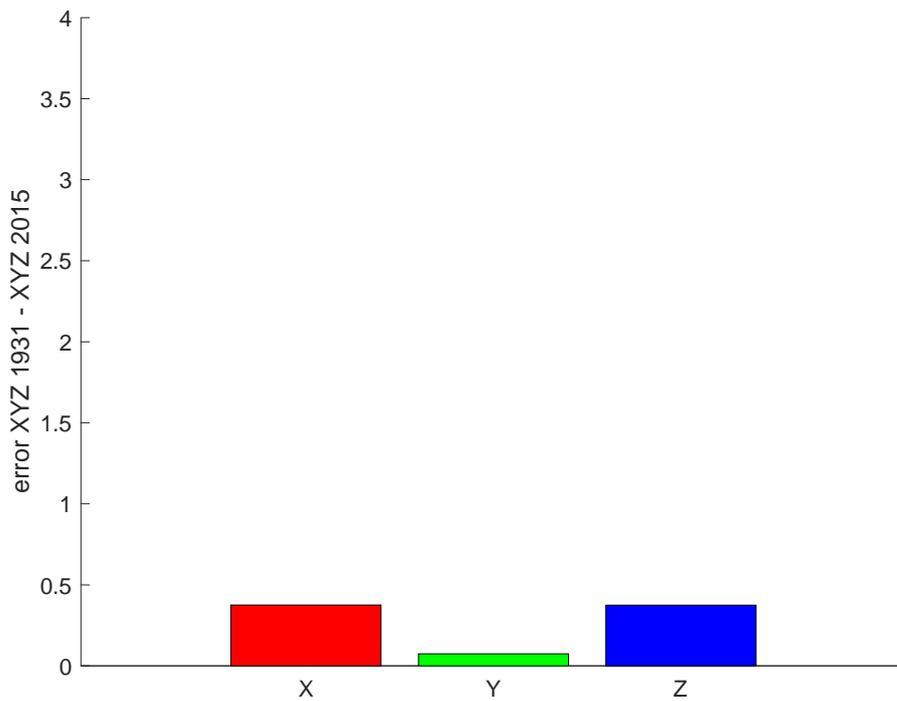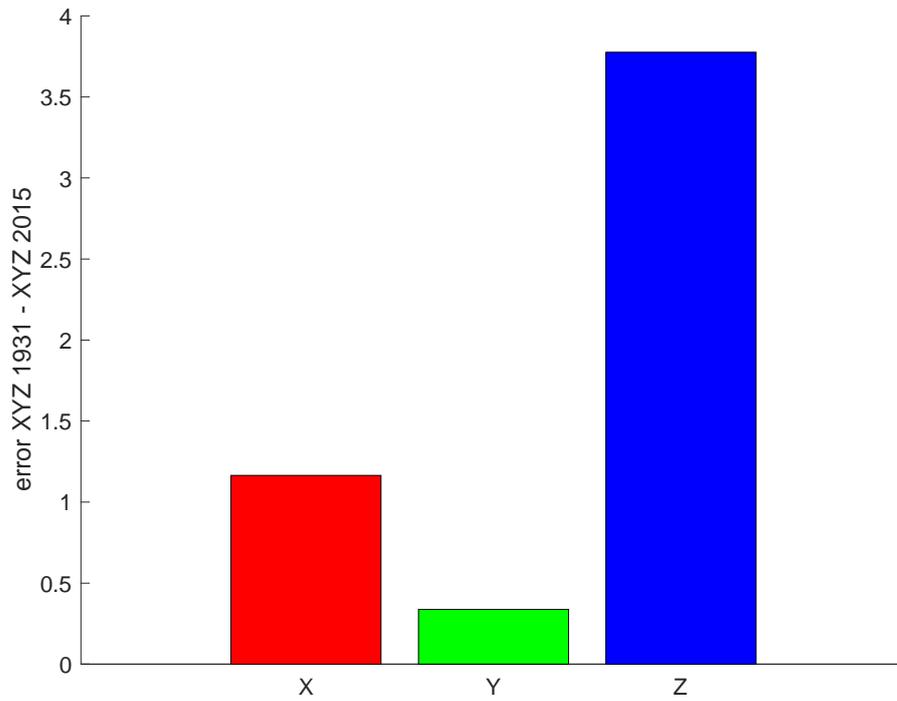
Figure 6.2 Root-Mean-Square Errors (RMSE) between tristimulus values X, Y and Z of Munsell Renotation Data in CIE XYZ 1931 and of corrected Spectral reflectance Munsell Data in CIE XYZ 2015 (top). RMSE between tristimulus values X, Y and Z of transformed Munsell Renotation Data to CIE XYZ 2015 (Eq 6.3) and of corrected spectral reflectance Munsell Data in CIE XYZ 2015 (bottom).
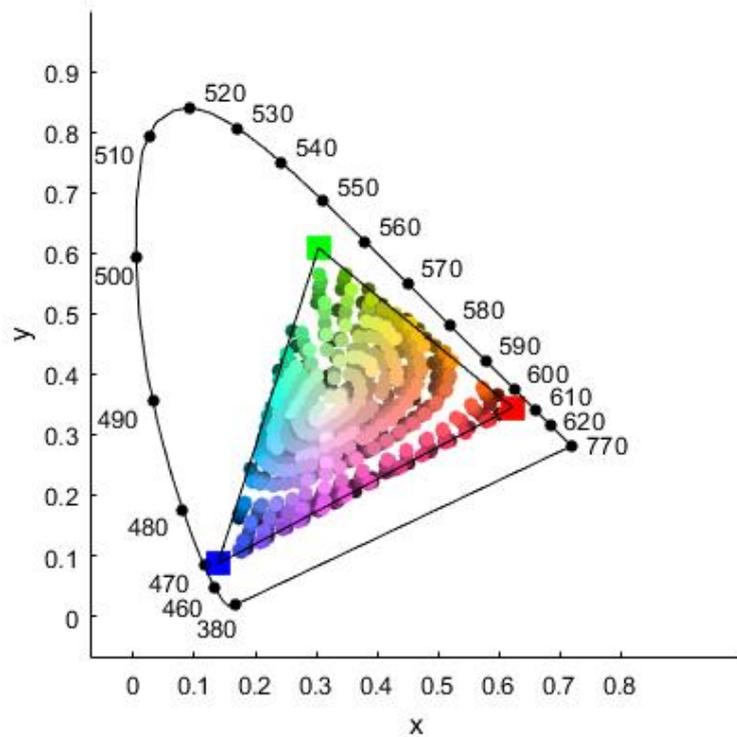
Figure 6.3 Simulated Munsell chips of different luminance levels (n=589; discs) and Red, Green, Blue phosphors (squares) of CRT monitor in CIE 2015 chromaticity diagram. Outline draws the gamut of the CRT monitor.

In the next step of our sampling procedure, we added 11 achromatic colour samples which are missing in the original Munsell Renotation Data. The nine samples specified at the mean plane for each level of Munsell Value and two samples, a white and a black, at the extremes of the RGB cube [R, G, B = 1, 1, 1] and [R, G, B = 0, 0, 0]. This resulted in a total of 600 colour stimuli. The final step involved a transformation of the CIE XYZ 2015 coordinates of the 600 stimuli to LMS cone excitation space shown in Figure 6.4 via the linearized CRT monitor and the cone fundamentals of Sharpe & Stockman (2000).
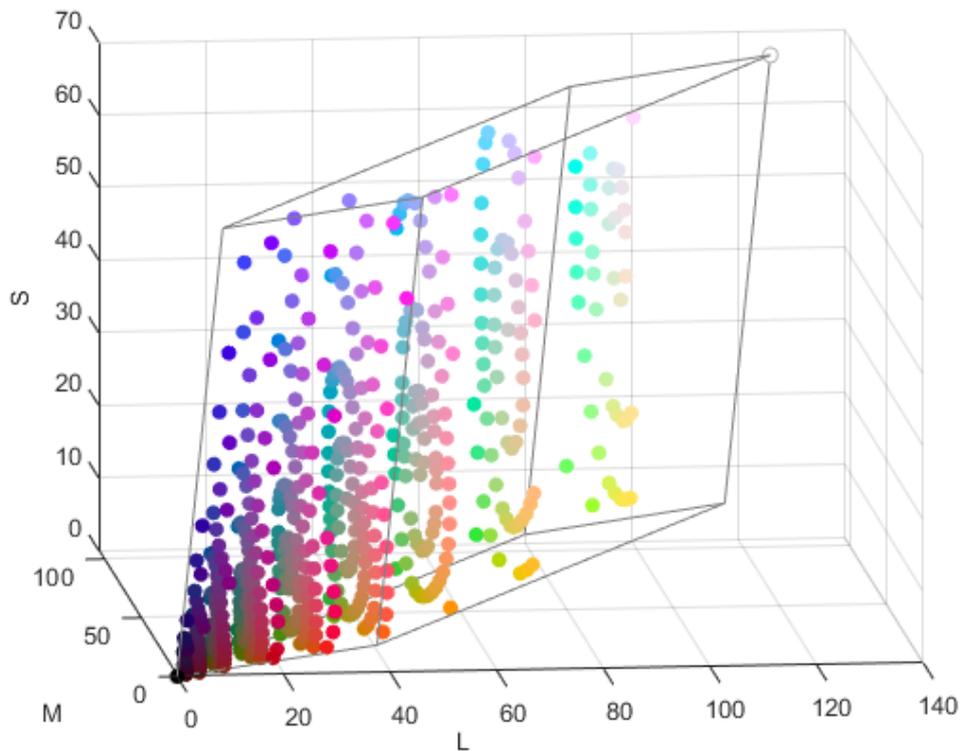
Figure 6.4 Colour stimuli (n=600) of colour naming experiment in CRT gamut in LMS cone excitation space. Outline draws the gamut of the CRT monitor.

6.6. Colour naming dataset

The vocal responses were transcribed and entered in a table for further analysis and the audio files were archived. The colour naming dataset includes in total 7,400 naming responses for 600 colour samples from ten observers. Half of the observers (n=5) repeated the entire experiment a second time to measure intra-observer consistency of the responses. The data from one session of one observer was discarded because the transcription of the responses was not feasible due to clipped audio files. From the 6,000 responses offered in the first session from all observers, here we consider only 247 distinct colour names including a *not-known* category, given by two or more observers resulting in 4,812 responses. Unique responses from single observers were excluded because we could not be confident that other observers will understand the colour name used and therefore these responses were considered idiosyncratic.

6.7. Analysis per name

In the naming responses, 50% of the data included single terms, 42% two word- and 8% three- word descriptions. The 11 basic colour terms proposed by Berlin & Kay (1969/1991) in English accounted for 35% of the responses and other non-basic single terms in 15% (Figure 6.5).
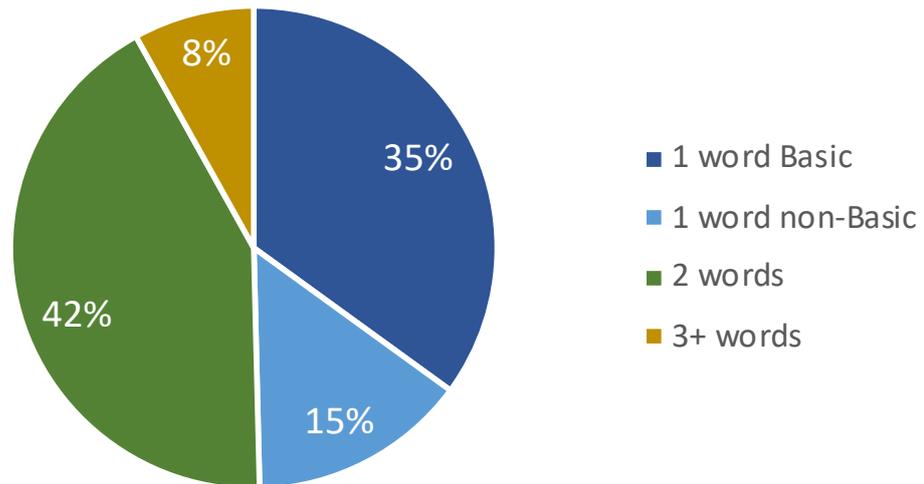
Figure 6.5 Number of words in colour naming dataset of lab-based colour naming experiment.

In Figure 6.6, we show the top 30 most frequent colour names offered by the observers in the offline experiment. These names account for 63% of responses. We chose 30 for clarity of the visualisation, but also because non-expert observers are able to identify 30 colour names in their native language without training (Derefeldt & Swartling, 1995). *Green* was the most frequent term, followed by *pink*, *purple* and *blue*. The *not-known* category that summarizes all empty responses was found in the 8th position. *Lilac*, a non-basic colour term, was more frequent than the basic *grey, yellow* and *white. White* was the least frequent basic term followed by the non-basic terms, *turquoise* and *magenta*.
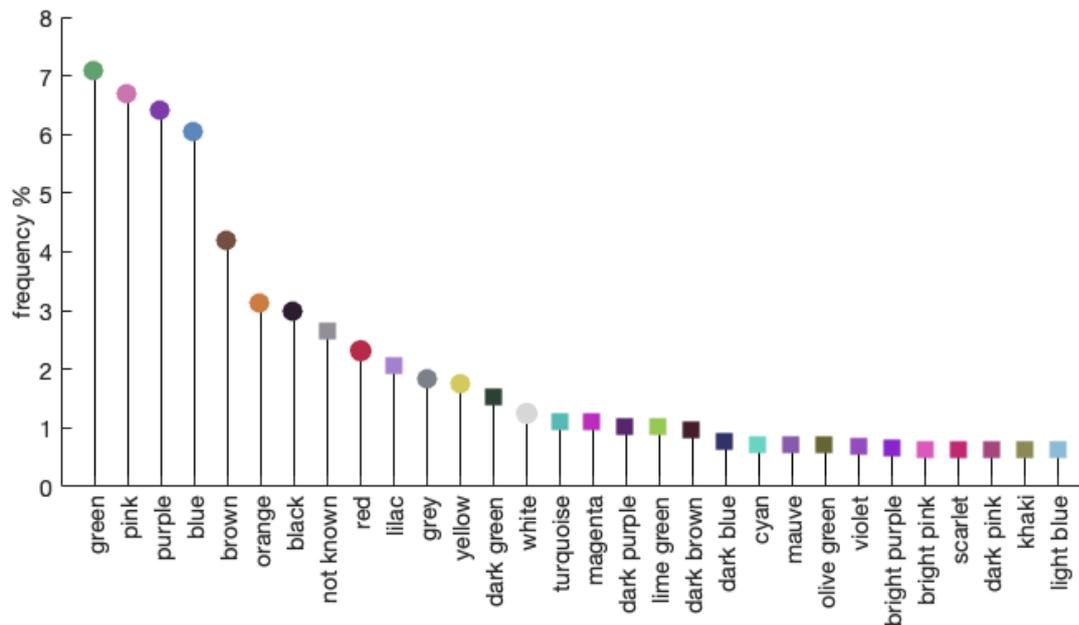
Figure 6.6 Top 30 most frequent colour names in offline experiment. Basic Colour Terms are drawn with discs and non-basics with squares.

In Figure 6.7, we present response times for the fastest named colours. Response time distributions are rarely Gaussian as their shape rises rapidly on the left followed by a long tail on the right. Therefore, we report the median and its 95% confidence interval of response latency for each colour name (Whelan, 2008). *Black, white* and *pink* were the fastest named categories; followed, unexpectedly, by the non-basic term, *olive*. *Navy*, a non-basic term, was also the 7th most quickly named colour. The slowest to name basic term was *grey*, ranked in the 19th position.
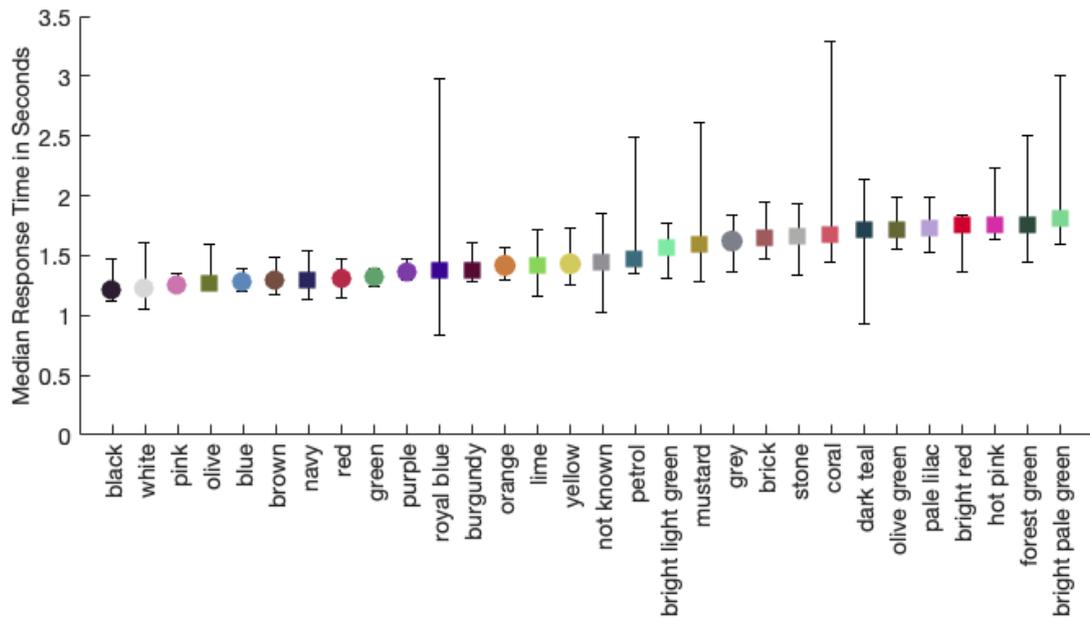
Figure 6.7 Top 30 fastest responding colour names in offline experiment. Basic Colour Terms are drawn with discs and non-basics with squares. Error bars denote 95% confidence interval.

Consistency measures the agreement between two responses for the same sample from observers who repeated the experiment twice. In total, 62% of the repeat responses were consistent between the first and second presentation. Colour names varied in how often they used consistently. In Figure 6.8, we show a rank of the 39 most consistent colour names from the 246 distinct colour names. Here, all basic terms were ranked in the top 11 positions but *lilac*, a non-basic term, was also equally consistent with *black* and *white*. *Green* had the largest number of samples named consistently, followed by *pink*, *purple* and *blue*. The top six position resemble the rank of the most frequent names.
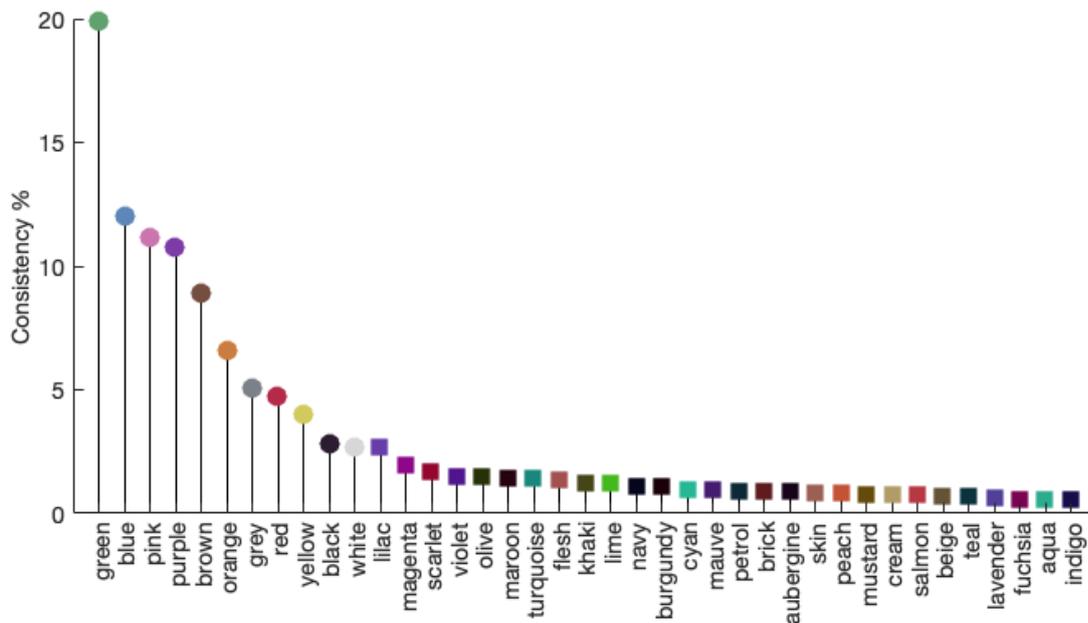
Figure 6.8 Consistent colour names in offline colour naming experiment. Basic Colour Terms are drawn with discs and non-basics with squares.

6.8. Location of Basic Colour Terms based on the CIE XYZ 1931

In this section, we use the inverse transformation matrix (Eq. 6.4) to represent our results and measure distances between the location of colour terms in the approximately uniform colour spaces of CIELAB and CIELUV which are based on the CIE XYZ 1931. The reason for using the CIE XYZ 1931 coordinates instead of the CIE XYZ 2015 is two-fold. First, at this stage of development there is no uniform colour space associated with the new CIE XYZ 2015 where we can measure meaningful distances between colour terms. Second, the relationship between the CIE XYZ 1931-based results of our online and of previous studies is not linear to the new CIE LMS 2006 and CIE XYZ 2015. We will present our results in the new cone excitation space in the next section (6.9.

In Figure 6.9, we compare centroid location for the 11 BCTs estimated in the offline study with the centroids obtained in our online experiment and the earlier laboratory-based studies of Boynton & Olson (1987), and Sturges & Whitfield (1995). There is a good correspondence between the 4 sets of data that can be also accessed in Table 6.1 in terms of *CIE ΔE 2000* colour differences. The smallest mean colour differences were observed against the online data (ΔE_{00} = 4.58), despite the uncalibrated colour reproduction of the web-based experiment. Yellow and orange had the smallest differences and black the largest. The largest discrepancies were observed against Boynton & Olson's study (ΔE_{00} = 9.67) because the authors located the basic terms in the OSA space that has a different geometry and is less saturated than the Munsell

system. In Table 6.2, we report the coordinates of the centroids of the BCTs of the offline and online study in CIELAB (D65).
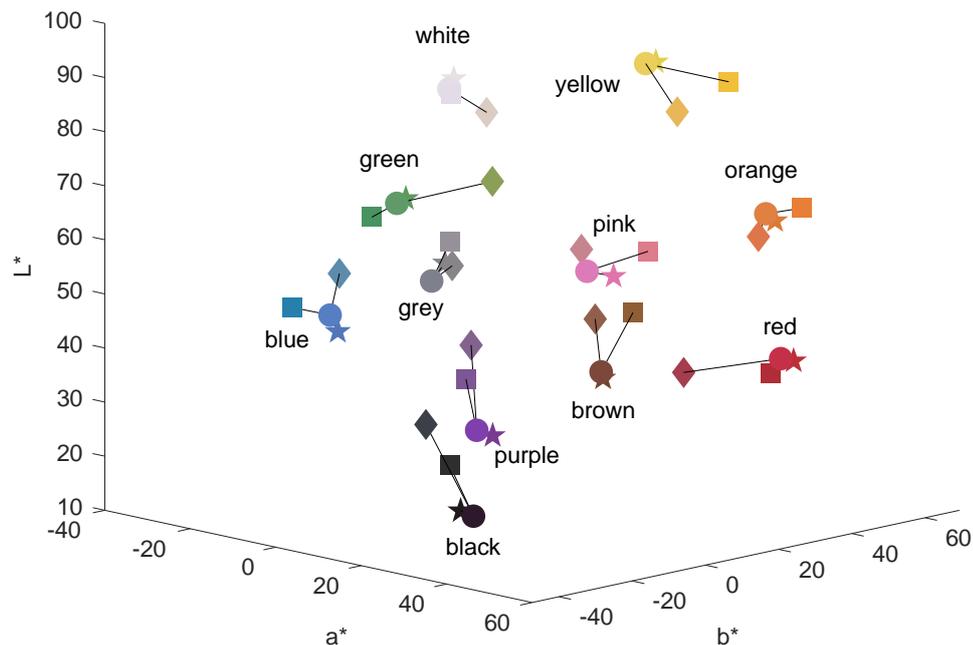
Figure 6.9 Comparison of centroids of Basic Colour Terms obtained in the offline (discs), online (stars), Boynton and Olson (1987; diamonds), and Sturges and Whitfield (1995; squares) studies in CIELAB (D65) based on XYZ 1931.

Table 6.1 Colour differences using the CIE ΔE 2000 formula between centroids of Basic Colour Terms in CIELAB (D65) based on XYZ 1931.

	Offline vs Online	Offline vs B&O	Offline vs S&W	S&W vs B&O
black	11.99	17.65	12.75	7.36
blue	3.03	7.25	9.90	6.80
brown	6.91	8.51	12.64	7.74
green	2.91	10.46	3.99	13.56
grey	5.07	5.19	7.44	5.21
orange	2.47	5.97	1.95	7.17
pink	2.82	12.72	12.96	5.21
purple	4.58	14.00	8.42	6.00
red	5.01	6.44	5.73	6.56
white	2.98	9.35	2.12	10.33
yellow	2.58	8.81	7.06	5.12
Mean	4.58	9.67	7.72	7.37

Table 6.2 Location of centroids of Offline and Online colour naming experiments in CIELAB (D65) based on XYZ 1931.

	Offline			Online		
	L*	*a**	*b**	*L**	*a**	*b**
black	12.32	13.17	-8.69	10.17	2.43	0.35
blue	52.81	6.09	-39.60	49.91	7.33	-38.93
brown	36.52	20.81	17.25	33.60	15.81	23.64
green	58.56	-29.76	20.50	57.97	-33.06	26.81
grey	53.66	2.01	-7.08	56.02	0.69	-1.89
orange	63.06	32.20	48.97	61.11	31.31	52.52
pink	64.15	45.63	-15.56	62.83	47.60	-10.60
purple	39.63	47.02	-47.32	36.24	41.81	-36.88
red	44.83	59.04	21.69	42.70	55.49	29.37
white	88.84	3.06	-3.59	90.14	1.82	-0.81
yellow	83.39	-3.62	57.96	82.19	-6.84	64.50

6.8.1. Location of centroids and perceptual structure

To explore whether perceptual structure – embedded in the 600 simulated Munsell samples named in both online and offline experiments – can explain the consistent location of their BCT centroids, we used the k-means algorithm to construct a set of imaginary colour naming systems based on Euclidean distances between the colour samples in CIELAB without any colour naming observation (Zaslavsky, Kemp, Regier & Tishby, 2018).

The k-means algorithm clustered the 600 samples in k categories, where k was set to be equal to the number of distinct colour names in the offline study ($k = 247$) offered by at least two observers. Then, we constructed a distance matrix between the centroids of the hypothetical k categories, and the centroids of the observed categories obtained in the offline experiment and then we assigned optimally the first to the latter categories using the Munkres assignment algorithm, also known as the Hungarian method (Kuhn, 1955). We repeated this process 50 times and computed the mean Euclidean colour differences in CIELAB between the observed and hypothetical centroids to stabilise the procedure.

Despite the relatively large k value for the number of hypothetical colour categories that should minimise the error by definition, the mean colour difference between offline and hypothetical BCT centroids was ΔE_{ab} = 14.61, double than the mean ΔE_{ab} = 7.30 between the observed BCT centroids of the offline and online experiment. The mean colour differences between hypothetical and online BCT centroids was even larger ΔE_{ab} = 16.54. The mean agreement between each imaginary colour naming system was ΔE_{ab} = 12.59. Setting k = 11 equal to the number of BCTs in English would produce larger colour differences between hypothetical and observed BCTs of the offline ΔE_{ab} = 22.02 and online ΔE_{ab} = 24.16 experiments. Therefore, perceptual structure embedded in the approximately uniformly distributed colour sampling used in both online and offline experiments cannot explain the good agreement between their BCT centroids. In Table 6.3, we report the Euclidean colour differences (ΔE_{ab}) between centroids of BCTs obtained in the offline and online colour naming experiments and an indicative k'-means imaginary colour naming system. Larger differences between hypothetical and observed BCTs were found for purple, followed by red and orange; the smallest was found for grey.

Table 6.3 Euclidean colour differences ΔE_{ab} (D65) based on XYZ 1931 between centroids of BCTs in online, offline and an indicative imaginary colour naming system based on k-means.

	Offline vs. Online	Offline vs. k'Means	Online vs. k'Means
black	14.20	9.61	9.83
blue	3.22	15.84	14.42
brown	8.62	19.31	14.14
green	7.15	17.55	10.68
grey	5.85	2.32	6.41
orange	4.15	22.08	18.79
pink	5.50	7.37	10.60
purple	12.15	26.16	38.05
red	8.72	21.72	27.84
white	3.31	12.12	10.06
yellow	7.39	1.96	7.48
Mean	7.30	14.19	15.30

6.8.2. Individual differences on the location of Basic Colour Terms

Centroids of BCTs were computed for each of the ten individual observers of the offline experiment to quantify their differences in colour naming (Figure 6.10). Although the

centroids for each BCT are clustered together for all individuals; there is considerable variation between individual observers with a mean colour difference between their centroids of $\Delta E_{00} = 7.91$. The smallest differences were found for *red* ($\Delta E_{00} = 4.77$) and the largest for *blue* ($\Delta E_{00} = 15.48$). These inter-individual differences in BCT centroid location are larger than inter-experimental differences between the online and offline population mean ($\Delta E_{00} = 4.58$) reported in the previous section. The inter-individual differences are also larger than the inter-language differences ($\Delta E_{00} = 5.01$) reported in Chapter 3.
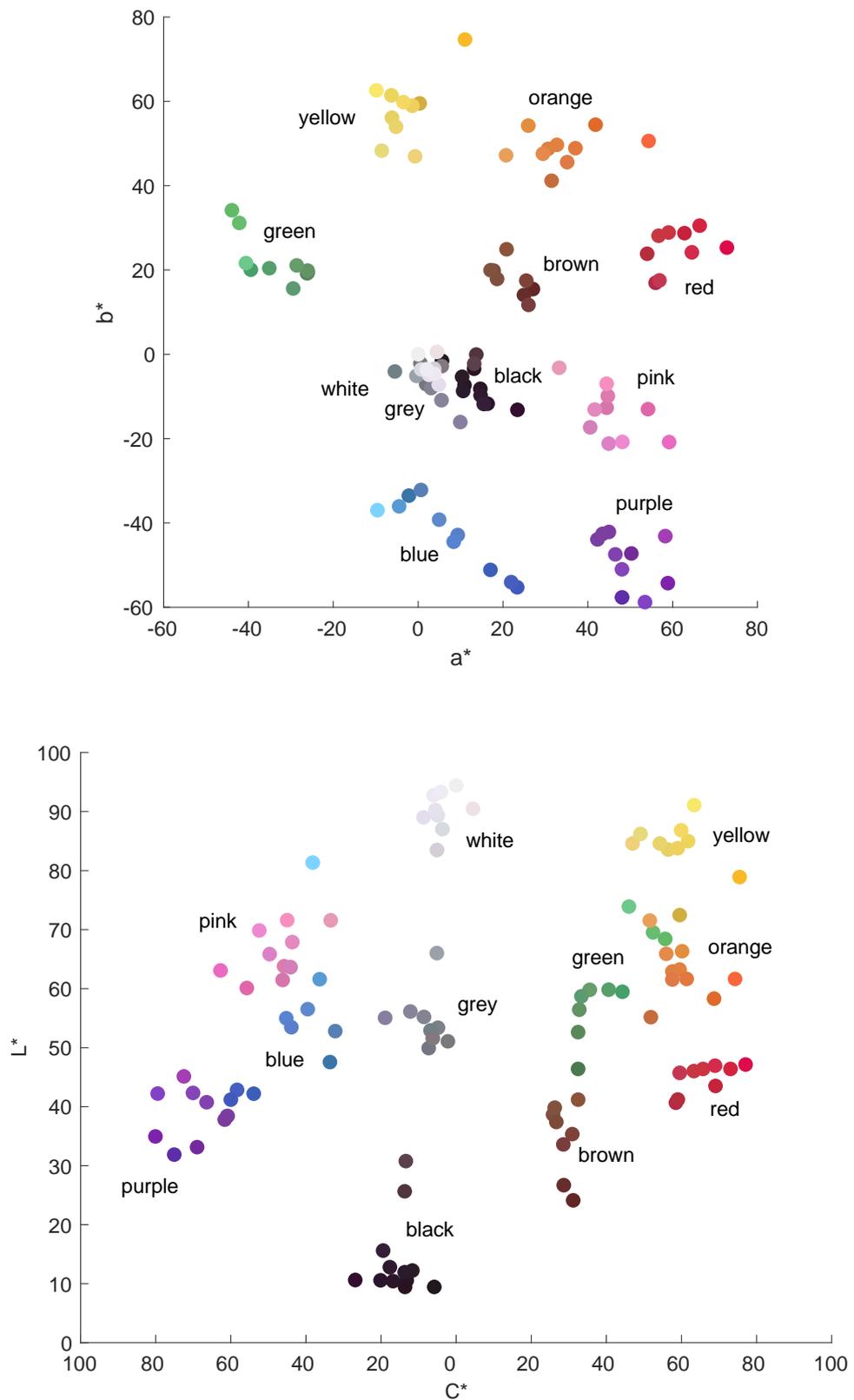
Figure 6.10 Location of BCT centroids for each individual observer in a* b* (top) and L* C* (bottom) planes of CIELAB (D65) based on XYZ 1931. Centroids with a hue angle >180° are shown on the left side and with a hue angle of <180° on the right side of neutral axis.

6.8.3. Colour names and unique hues

Earlier studies reported a good correspondence between unique hue settings and the corresponding colour names *red, green, blue* and *yellow* (Kuehni, 2005). To explore the relationship between colour names and unique hue settings, in Figure 6.11 we compare the centroids of the lab-based colour naming responses against the unique settings that we reported in an earlier study in CIELUV (Xiao, Fu, Mylonas, Karatzas & Wuerger, 2011). Except for *blue*, there was a good correspondence of the unique hue settings and the corresponding colour terms. The smallest hue difference between mean unique hue settings and centroids of colour terms was found for *yellow Δh* = 1.8, followed by *green* (*Δh* = 6.46) and *red* (*Δh* = 6.65). The largest hue difference was found for *blue* with *Δh* = 14.65. The centroids of the colours names usually associated with unique hues, *red* and *green* were not colinear with *white* but *yellow* was colinear through white with *blue*. *Red* was nearly colinear with *turquoise* and *green* with *magenta*. In the next section we locate colour terms in the physiologically-based cone excitation space to examine further the relationship between colour naming and colour vision mechanisms.
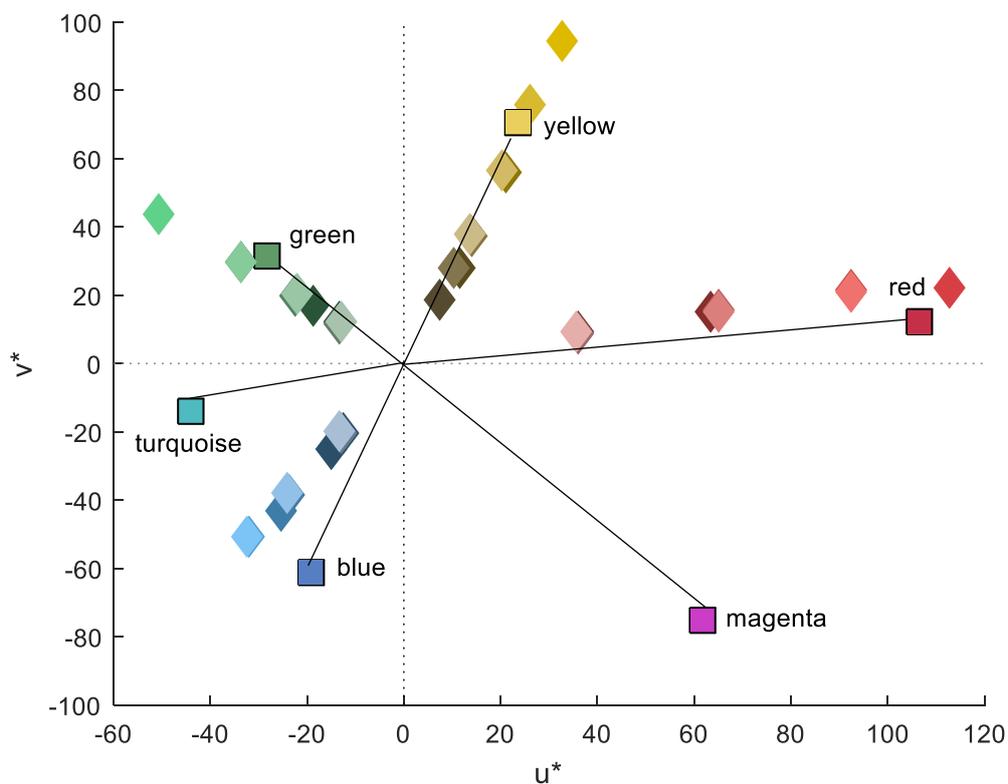
Figure 6.11 Centroids (squares) of red, green, blue, yellow, turquoise and magenta and unique hue settings (diamonds) reported by Xiao et al., (2011) in u*v* plane of CIELUV (D65) based on XYZ 1931.

6.9.   Mapping colour names in cone excitation space

A precise measurement of colour names in a cone excitation space with physiological axis is of great importance for basic and applied colour research. Here, we report a) centroids of BCTs b) consensus of naming each sample, c) response times required to name each sample, d) the foci and centroids of dominant colour names, e) the connectedness between dominant names and f) a colour naming model for dominant colour names, in the LMS cone excitation space (Stockman & Sharpe, 2000; CIE 170-1:2006).

In Figure 6.12, we show the centroids of the 11 BCTs as squares and the 600 colour samples of the lab-based experiment as discs in CIE LMS 2006 where the size of each disk is related to the proportion of the most frequent name for each sample. The larger the disk, the higher the consensus for naming this sample. Inspecting the figure reveals clusters of colour samples named with high consensus corresponding roughly to the 11 BCTs of Berlin and Kay (1969/1991) with some unidentified high consensual regions.
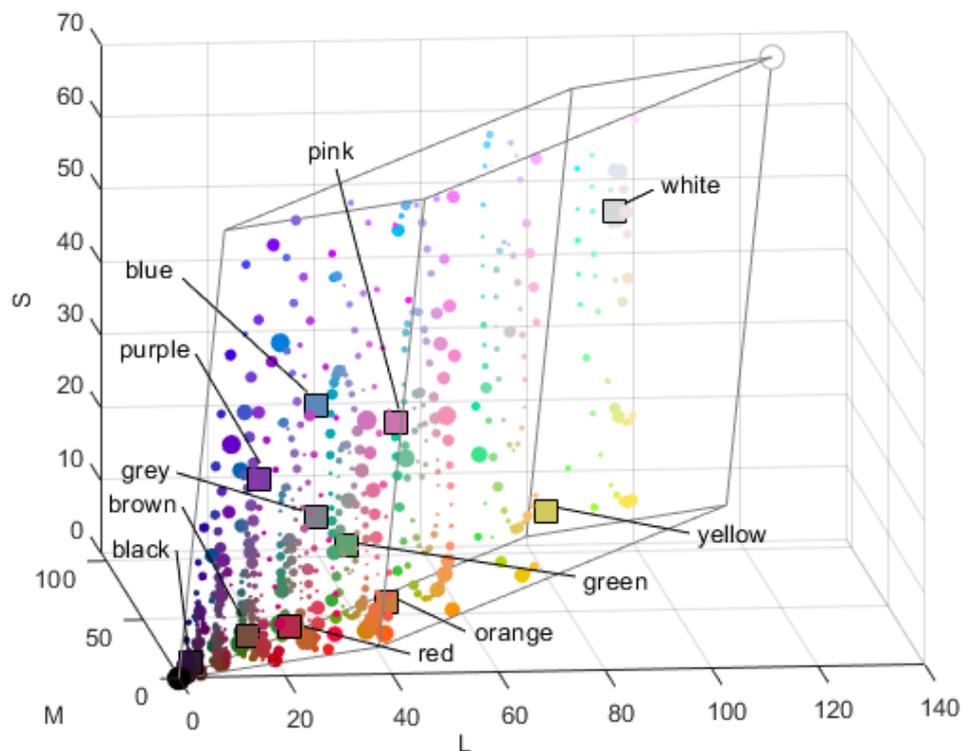
Figure 6.12 Naming consensus for each colour sample (discs) and centroids of BCTs

(squares) in LMS cone excitation space. The larger the disc, the higher the consensus.

Outline draws the gamut of the CRT monitor.

In Figure 6.13, we show again the centroids of BCTs and the mean response latencies required for naming each sample. There is a moderate Pearson negative correlation between response time and consensus, $r = -0.39$, $p < 0.001$ (Figure 6.14). Hence, longer response times were observed at the borders between colour categories where there is an overlap of colour naming distributions than for regions closer to the limits of the colour gamut named with higher consensus. Both measures, consensus and response time, will be used to identify the best examples (foci) of dominant colour categories in the next section.
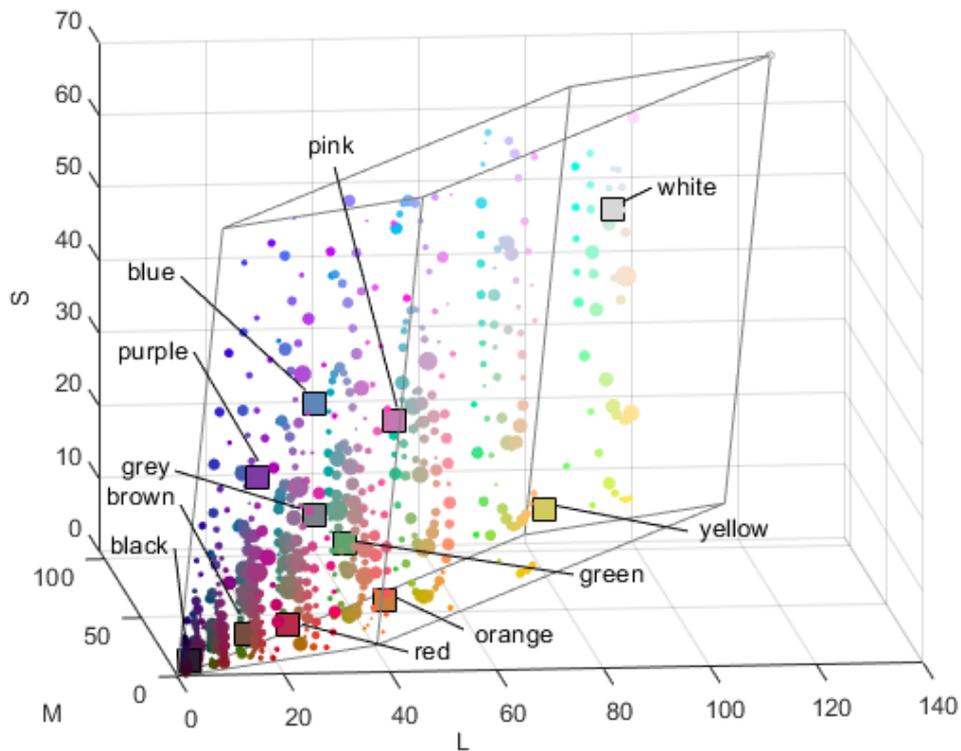
Figure 6.13 Mean response time required to name each colour sample and centroids of BCTs (squares) in LMS cone excitation space. The larger the disc, the longer the average response time required to name each sample. Outline draws the gamut of the CRT monitor.
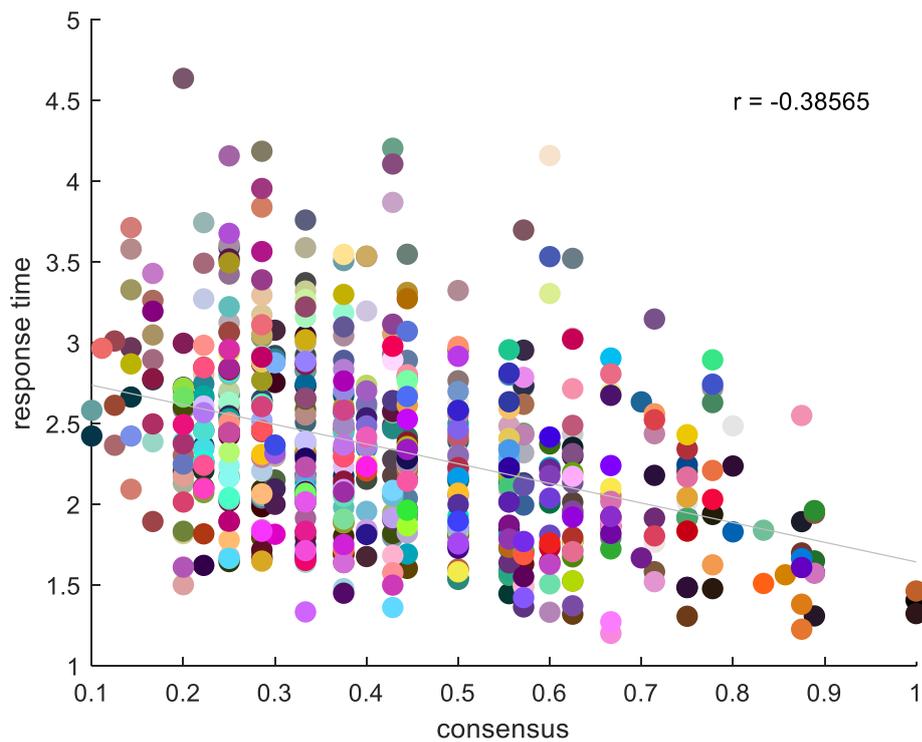
Figure 6.14 Correlation between consensus of naming and responses time required to name each sample in CIE LMS 2006.

6.9.1. Foci and centroids of dominant colour names

In this section, we report the best examples, called focal colours, of colour names. Focal colours are determined by the shortest mean response time of colours within categories named with consensus across observers (Boynton & Olson, 1987). We give foci for all dominant colour names over other names (>=50%) used to describe each colour sample together with their corresponding centroids in CIE LMS 2006 to show the direction of the most consensual regions within each category (Figure 6.15). In Figure 6.16, we present the results in the 2D Macleod-Boynton (1979) cone-opponent chromaticity diagram based on Stockman & Sharpe's (2000) fundamentals, where the axes correspond to the L/(L+M) and S/(L+M) retino-geniculate main pathways (Krauskopf et al., 1982). In Table 6.4, we give the coordinates of the focal colours in the Munsell system with their consensus level and response time.
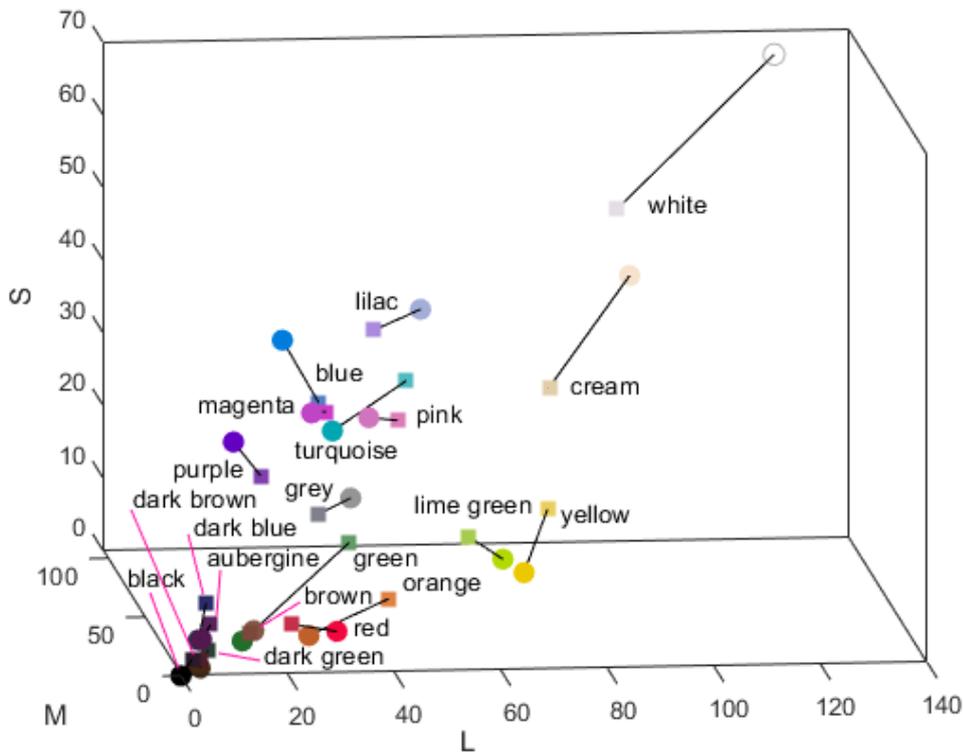
Figure 6.15 Focal colours (discs) of dominant colour names and their centroids (squares) connected with black lines in CIE LMS 2006.
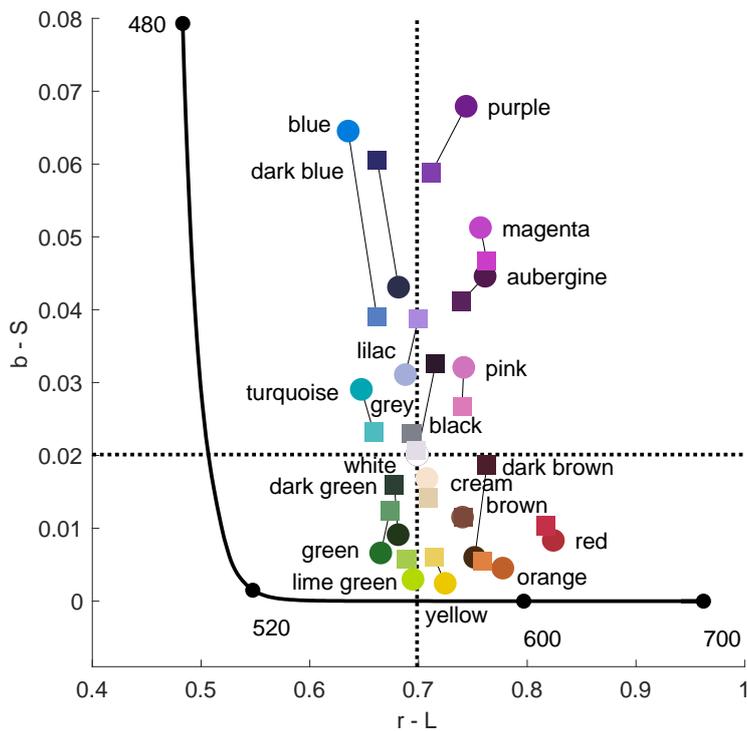
Figure 6.16 Focal colours (discs) of colour names with dominance of >=50% and their centroids (squares) connected with black lines in the Macleod & Boynton diagram. Dotted horizontal and vertical lines correspond to the adapting background chromaticities. Spectrum locus is shown with a thick black line and the selected wavelengths are given in nm.

Names	hue	Value	Chroma	RT	Consensus
Aubergine	10P	2	8	2.20	0.63
Black	NA	0	0	1.13	1.00
Blue	5PB	5	14	1.51	0.88
Brown	2.5YR	4	4	1.75	0.89
Cream	7.5YR	9	2	2.86	0.75
Dark blue	7.5PB	2	4	2.11	0.57
Dark brown	7.5YR	2	4	1.85	0.63
Dark green	7.5GY	2	4	1.55	0.67
Green	10GY	4	8	1.42	0.89
Grey	NA	6	0	1.90	0.88
Lilac	7.5PB	7	6	1.91	0.51
Lime green	5GY	8	12	1.44	0.63
Magenta	7.5P	5	18	1.46	0.60
Orange	2.5YR	5	10	1.31	1.00
Pink	10P	6	12	1.19	0.89
Purple	5P	3	14	1.43	0.88
Red	5R	4	12	1.52	0.86
Turquoise	10BG	6	8	1.77	0.63
White	NA	10	0	0.92	1.00
Yellow	5Y	8	12	1.47	0.75

Only 3 samples named as *black*, *white* and *orange* were named with 100% consensus. The foci and centroid of *white* overlapped with the foci of *black* but the centroid of *black* was shifted towards the purple region because most of the dark purplish tones at Munsell Value 1 were named as *black*. At 75% consensus, we found all eleven BCTs plus *cream* describing 52 samples. At >50% consensus, we found in addition to the 12 names of the previous level, 8 more colour names including *turquoise, lilac, magenta aubergine, lime green,* and the *dark* modifiers of *blue, green* and *brown*. In total, 20 dominant colour names were assigned to 243 samples.

Overall focal colours were closer to the spectrum locus than centroids. Except for the achromatic colours all other 17 pairs of centroids and foci of colour names lie within the axes of colour discrimination relative to the adapting background in the MB diagram

(Skelton et al., 2017). But, the location of some centroids (*turquoise, lilac, purple, yellow, lime green, cream* and *dark brown*) near these lines indicate that some categories might cross over the axes and requires further investigation (see section 6.10. for modelling the full extent of colour categories).

6.9.2. Linked colour categories in LMS

Lexical colour categories share their boundaries with each other in colour space. To reveal overlapping colour names in a cone excitation space, we measured their connectedness. Two lexical categories are connected when the same colour is identified with both names (Boynton & Olson, 1987). Figure 6.17 shows the connectedness of all dominant colour names.

The 11 BCTs tend to share fewer common samples between them than non-basic terms. *Red* and *pink* shared the most common samples between BCTs. *Purple* and *green* are the most connected categories, followed by *grey*, *brown* and *blue. Lime green, yellow, cream, turquoise* and *magenta* exhibit the smallest number of connections with other names. *Green* is neither connected to *red* nor to *purple. Blue* is not connected to *yellow* or to *orange*. Unexpectedly, *yellow* is not connected to *white* and *cream* seem to be a crucial node connected with *yellow*, *orange*, *white*, *pink* and *green*. Non-BCTs tend to share a larger number of common samples with other categories. *Turquoise* is strongly connected to *blue* and less so to *green*, while *lilac* is strongly linked to *purple* and *pink. Magenta* overlaps strongly with *pink* and *purple. Lime green* is more connected to *green* than to *yellow*.
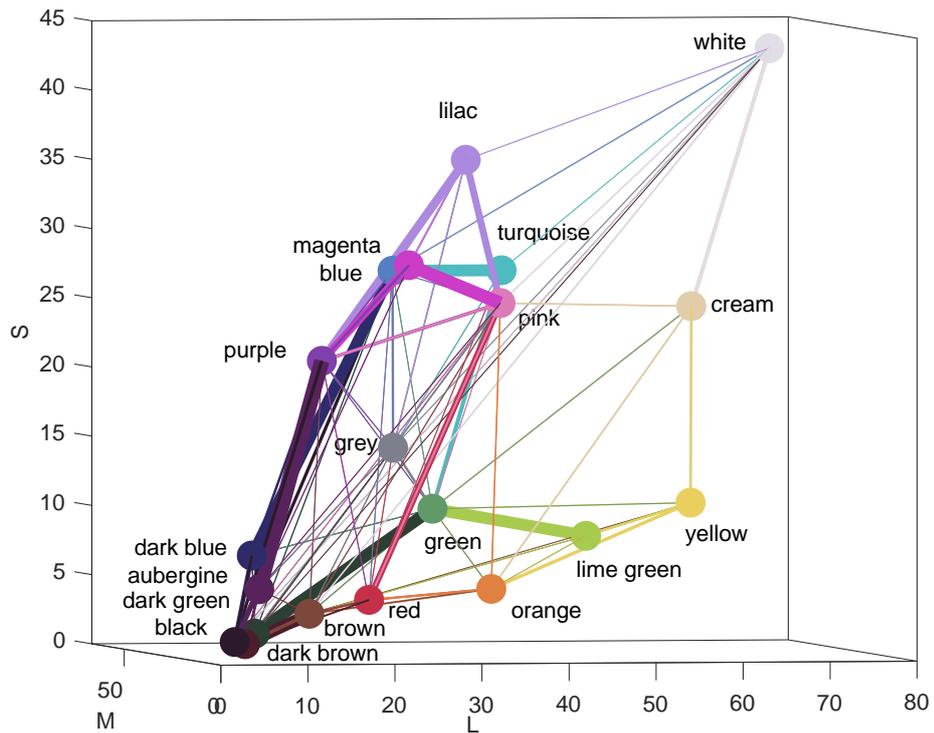
Figure 6.17 Connectedness of dominant colour names in CIE LMS 2006. The colour and location of the discs corresponds to the coordinates of the centroids of the colour categories. The lines between the circles link colour categories that share common colour samples. The width of the lines indicates the number of these common samples.

6.10. Segmentation of cone excitation space

To generalise our observation from the 600 points used in the experiment to the entire 3D colour gamut of the CRT RGB monitor in cone excitation space, we used a probabilistic colour naming model based on Maximum a Posteriori (MAP) described in Mylonas et al. (2010). We used the MAP method rather than the RST approach because the performance of the latter deteriorates in non-uniform colour spaces (see Table 5.2).

The algorithm was trained by all dominant names identified with >=50% of consensus to segment in LMS a grid of 2^{15} = 32,768 points uniformly distributed in the RGB cube of the CRT into 20 lexical colour categories (Figure 6.18). Coordinates of the centroids of the dominant colour names were used to colour each category.
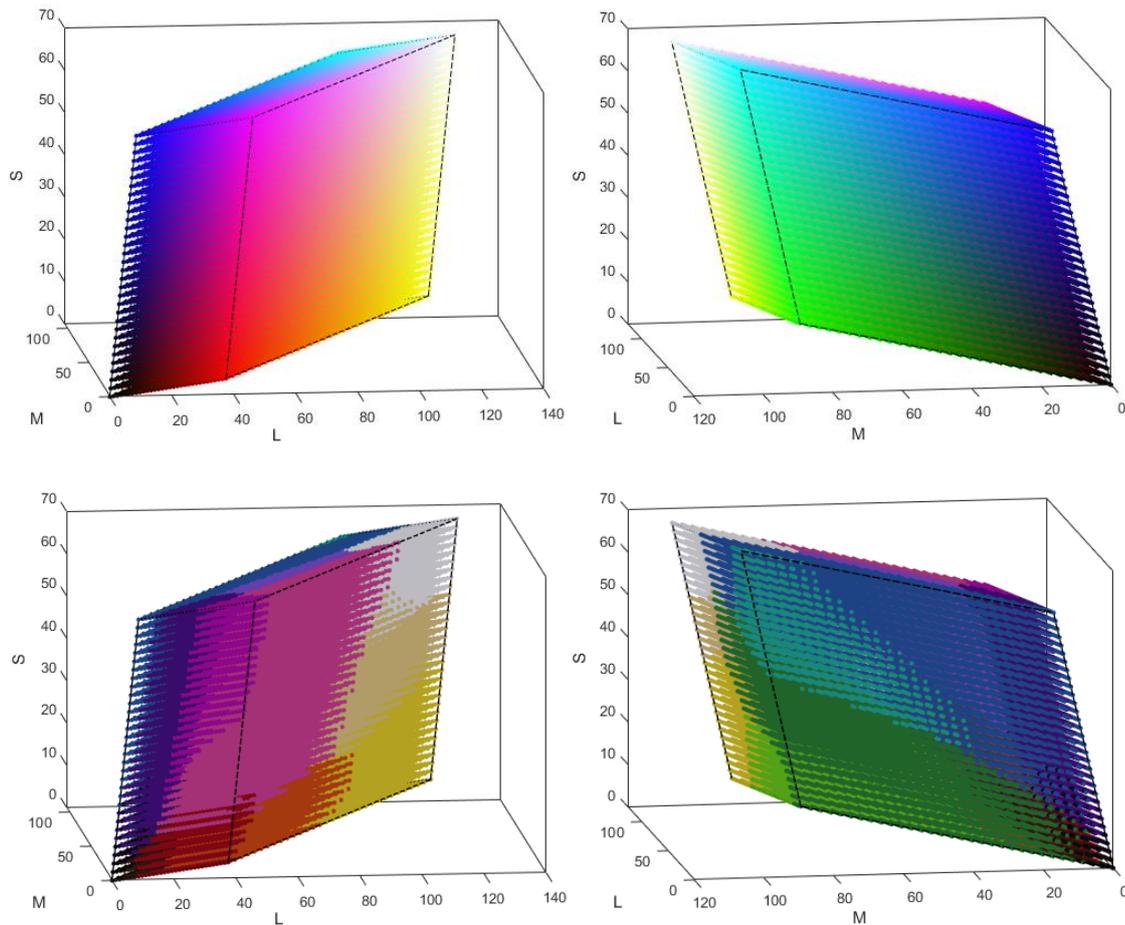
Figure 6.18 Warm (top-left) and cool (top-right) side-views of uniform RGB grid of 32768 points in LMS of the CRT gamut. Segregated warm (bottom-left) and cool (mottom-right) side-views of RGB grid of 32768 points by MAP into n=20 dominant names. Coordinates of the centroids of the dominant colour names were used to colour each category.

The regions with the highest purity of L, M and S cone excitations were assigned to *red*, *green* and *blue* colour names respectively. *Yellow* was assigned to regions with contributions mainly from L and M cones. S cone contribution to *red* was minimal. *White* requires contribution from all three cones and *black* from none. *Turquoise* covers the region that requires strong contribution from M and S cones, while *magenta* was localised to areas with substantial contribution of L and S cones. None of the chromatic corners of the LMS space were the most typical samples of the six colour names.

For more clarity, we visualise the categories on a 2D plane. Similarly to 5.10. , we created a synthetic image by taking a cross section of a conic representation of the HSL colour solid where the additive and secondary primaries (*red, green, blue* and *yellow, cyan, magenta*) are arranged around the outside edge of the solid at maximum Saturation of 1 and Lightness of 0.5. The HSL coordinates were then converted to LMS cone excitation

space via RGB of the CRT monitor and the cone fundamentals of Sharpe & Stockman (2000). The MAP colour naming model was trained again by dominant names identified with >=50% of consensus and segmented the synthetic image in 16 (*black, dark brown, dark green* and *aubergine* were not assigned to any pixels) lexical colour categories. Coordinates of the centroids of the dominant colour names were used to colour each category. We show the results of segmenting the synthetic image in HSL (Figure 6.19) and in the cone chromaticity space of MacLeod and Boynton (1979) in Figure 6.20.
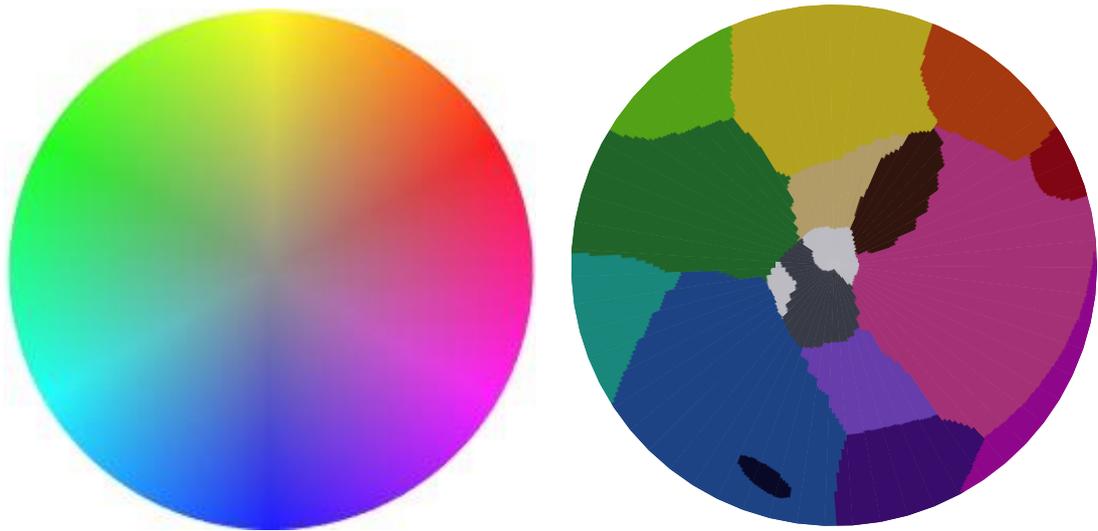
Figure 6.19 Synthetic cone image in HSL (left) segregated by MAP trained by dominant names with >=50% of consensus (right). Coordinates of the centroids of the dominant colour names were used to colour each category.

The largest number of pixels of the synthetic image (Figure 6.19) were assigned to *pink*, followed by *blue* and *green*. *Dark blue*, *red* and *magenta* ware the smallest categories. Qualitative features include a near circular achromatic region, with *brown*, *cream*, *lilac* and *dark blue* nested in the interior of the space.
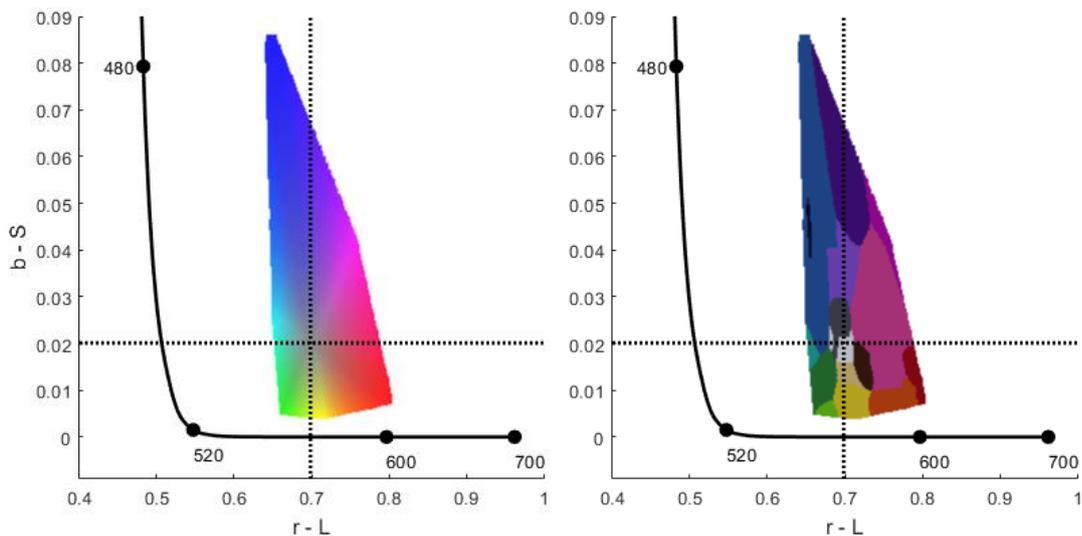
Figure 6.20 Synthetic image in MacLeod and Boynton diagram (left) segregated by MAP trained by dominant names with >=50% of consensus (right). Dotted horizontal and vertical lines correspond to the adapting background chromaticities.

In the MB diagram the achromatic colours *white* and *grey* are the centre of the space surrounded by chromatic colours. The synthetic image did not cover *black* regions. Along the horizontal axis of L/(L+M) with S cone values smaller than the horizontal axis originating from the chromaticities of the background, we found a greenness to redness dimension with low to high values respectively. *Lime green, yellow, cream, brown* and *orange* are localised between *green* and *red*. As the S cone contribution increases above the horizontal axis of the background *green* turns to *turquoise* and then to *blue* with a small *dark blue* region at low L/(L+M) values. As L/(L+M) increases colours change to *lilac*, *purple* and then to *pink* and *magenta* at the very high L/(L+M) values. Some categories share straight boundaries while for others the boundaries form a curvature.

The chromaticity diagram of MB distorts significantly the colour space because the S dimension is arbitrary set by definition. For example, the size of the purple category in the diagram appears disproportionally larger than the size of green category in HSL space. This gets clearer when we plot the results in the Derrington, Krauskopf and Lennie (DKL; 1984) cone contrast space in Figure 6.21.
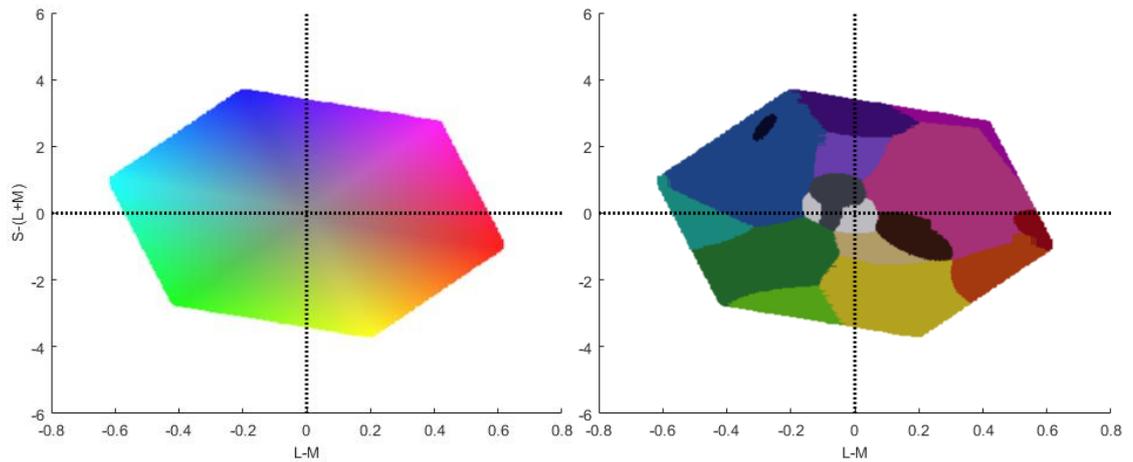
Figure 6.21 Synthetic image in DKL (left) segregated by MAP trained by dominant names with >=50% of consensus (right). Dotted horizontal and vertical lines correspond to the adapting background chromaticities. Axis are scaled by the pooled contrast to produce unit response for each mechanism.

Despite the unevenness of lightness levels in the synthetic image, in DKL the proportions of the categories are better represented than in the MB diagram. *Yellow* and *purple* cross the vertical- and *blue* and *pink* cross the horizontal- axes originating from the background chromaticity. *Green* is confined at the negative values of S-(L+M) and L-M while *red* is nearly confined at the positive L-M and the negative S-(L+M) values.

6.11. Summary and Discussion

A criticism that often arise is whether the uncontrolled colour reproduction and viewing conditions of online colour naming experiments meet the requirements for rigorous colour research. Laboratory-based experiments are also often judged as not being able to generalise their results to real world monitor settings. In this chapter, we assessed the precision of our uncalibrated colour naming experiment conducted over the Internet against a calibrated experiment – using the same sample set and background – performed in a laboratory environment, and the ability of the later to estimate colour naming functions in real-life monitor settings. We found a better correspondence between the loci of the BCTs in our online and offline experiments ($>5\Delta E_{00}$) than between previous lab-based studies ($<7\Delta E_{00}$; Boynton & Olson, 1987; Sturger & Whitfield, 1995); while we showed larger variations between the BCT centroids for individual observers ($<7\Delta E_{00}$). Overall, these findings suggest that online and offline colour naming experiments produce consistent results and support the validity of both methods in estimating colour naming functions in laboratory and real-world monitor settings.

Considering that inter-experimental differences were found to be smaller than the intra-experimental differences among individuals, it has been conjectured that languages gravitate to an optimal set of categories and return to them despite departures from the norm by individual speakers (Bimler, 2005; Griffin, 2006; Regier et al., 2007). In Chapter 3, we reported a maximum mean $\Delta E_{00} = 5$ between the BCT centroids of British English and Thai speakers. In this chapter, we confirm the findings of earlier studies (Berlin & Kay, 1969/1991; Webster & Kay, 2007) in that intra-language differences among individuals are larger than inter-language differences. It can be now convincingly argued that averaging a small number of colour naming responses from a large number of participants in a crowdsourcing experiment offers a better agreement between the underlying categories responsible for colour naming ($\Delta E_{00} = 1$ between British and American English BCTs) compared with averaging a large number of responses from a small number of different individuals in controlled viewing conditions ($\Delta E_{00} = 8;$ Shapiro, Carl & Varian, 1998; Surowiecki, 2005; Yi, Steyvers, Lee & Dry, 2012).

In regard to the range of possible cone responses as an explanation for the mechanisms that pressure colour categories to optimality, we examined a widely cited account of this type suggesting that colour categories are determined by optimising the division of an irregular perceptual colour space to maximise similarity within a category and minimise similarity across categories (Jameson & D'Andrade, 1997; Regier et al., 2007). In our assessment, as to whether perceptual structure can explain the consistent location of the BCT centroids, we used a *k-means* algorithm operating on CIELAB distances to construct a set of imaginary colour naming systems (Zaslavsky et al., 2018). A comparison of observed BCT centroids against a set of BCTs from the hypothetical colour naming systems ($\Delta E_{ab} = 15$) showed that perceptual structure embedded in the stimuli set alone cannot explain the agreement ($\Delta E_{ab} = 7$) between our online and offline experiments.

A second aim of this chapter was to map for the first-time unconstrained colour names in the physiologically-based cone excitation space adopted recently by the CIE (Stockman & Sharpe, 2000; CIE 170-1:2006; CIE 170-2: 2015). This new cone excitation space includes a better representation of the spectral sensitivities of the long-, middle- and short- wavelenth cones (L, M, and S) than the widely used CIE XYZ 1931 colour matching functions – especially at short wavelengths – but its relationship with higher-order cognitive processes is uncertain. Here, we contribute an optimised transformation matrix that allows future researchers to transform the Munsell Renotation Dataset from the CIE XYZ 1931 under illuminant C to the new CIE XYZ 2015 colour matching functions under the more natural and widely used daylight illuminant D65. Furthermore, we

contribute to the research community a calibrated colour naming dataset for 600 simulated Munsell samples consisting in total of 7,400 unconstrained naming responses from 10 English observers. Our dataset will be be useful for testing hue uniformity for the next generation of colour appearance models that will be based on the new CIE XYZ 2015 colour matching functions.

In the naming responses, 50% of the data included single terms, 42% two-word and 8% three-word descriptions. The eleven basic colour terms (Berlin & Kay, 1969/1991) occurred in 35% of the responses. This is comparable with the number of words found in the multilingual datasets of our online colour naming experiment described in Chapter 3. We argue that constraining the responses in colour naming experiments (Berlin & Kay, 1969/1991; Boynton & Olson, 1987; Olkkonen et al., 2009; Kay et al., 2010) produces a rather over-simplified picture of the complexity found in natural colour lexicons by overlooking in some cases more than 65% of other possible responses that can be very useful in basic and applied research (Zeki, 1983; Mylonas & MacDonald, 2012).

The eleven BCTs were used frequently by our observers but *lilac*, a non-basic term, was used more frequently than the basic *grey*, *yellow* and *white*. The latter was the least frequent BCT, followed closely by *turquoise* and *magenta*. Considering the response time of the vocal responses, the rank of the colour names with the shortest latencies included both basic and non-basic colour names. *Black*, *white* and *pink* were the fastest named categories but the non-basic terms, *olive* and *navy* were found in the 4th and 7th position respectively. The slowest to name basic term was *grey*, ranked in the 19th position. In terms of consistency between responses of observers that repeated the experiment twice, the eleven BCTs plus *lilac* were found at the top 12 positions. We also identified the centroids and focal colours of dominant lexical colour categories. Only *black*, *white* and *orange* were used with 100% consensus. In the consensus level of >=75%, we found all eleven BCTs plus *cream*. In the lowest level of consensus (>=50%), we found eight additional names: *turquoise, lilac, magenta, aubergine, lime green,* and the *dark* modifiers of *blue, green and brown*. These findings contradict earlier reports of constrained colour naming studies (Boynton & Olson, 1987; Uchikawa and Boynton 1987) that report BCTs are used more quickly, more consistently and with greater consensus than any other colour name. Instead our results support the view that basicness is a gradual rather than a discrete characteristic of lexical colour categories, and the eleven BCTs do not constitute the upper limit of basic categories (Moss et al., 1990; Mylonas & MacDonald, 2016; Paramei et al., 2018; Witzel, 2018 for a review). This is not to say that the eleven BCTs are not important in the English colour lexicon – they ranked in the top positions in most measures – but that *cream, lilac* and *turquoise* also

constitute strong BCT candidates in British English (Sturges & Whitfield, 1995; Mylonas & MacDonald, 2016). The strong candidacy of *cream* is also supported by our examination for a possible 12th basic colour term in Chapter 4, where the addition of *cream* to the 11 BCTs produced a perfect coherence score for the Basic class.

In regard to the perceptual attributes of dominant colour names, overall focal colours were more saturated than centroids. This confirms the importance of saturation in the selection of the best examples of chromatic categories (Berlin & Kay, 1969/1991; Rosch Heider, 1972; Regier et al., 2005; Olkkonen et al., 2010; Lindsey et al., 2015). The vertical axis of S/(L+M) in MB diagram originating from the background chromaticity coincides with the centroids of purple and yellow at high and low values respectively. The horizontal axis of L/(L+M) is positioned between red and pink at high values, and between blue and green at low values. These results confirm not only the large discrepancies between colour discrimination mechanisms and the axes of colour appearance mechanisms (Abramov & Gordon, 1994; Webster et al., 2000; Valberg, 2001; Wuerger et al., 2005), but also the discrepancies between the sensitivity of these second stage mechanisms to hue differences and the boundaries of colour categories (Malcoc et al., 2005, Witzel & Gegenfurtner, 2013, 2018; Shepard et al., 2017; Witzel, 2018).

Considering the relationship between colour names and unique hues, the foci of the landmark colour names usually associated with unique hues, *red* and *green* (Boynton & Olson, 1987) were not colinear with *white* but *yellow* was colinear through *white* with *blue*. *Red* was nearly colinear with *turquoise* and *green* with *magenta* through *white*. The foci of *cyan* could also align with *red* but *turquoise* was offered more frequently and with higher consensus to describe this region. Except for *blue*, there was a good correspondence of the landmark colour names with the location of unique hue settings in CIELUV space (Xiao et al., 2011). These results support the strong relationship between the four landmark colour names and the associated colour-opponent mechanisms of the hypothetical third stage of colour appearance. For the relative larger hue differences between the location of the blue unique hue and the foci-centroid pair of the *blue* colour term, we argue that the collinearity between *blue*, *yellow* and *white* in our colour naming data as well as their more vertical alignment in MB diagram corresponds better to variation in the signal of the short-wave sensitive cones than the larger modulation of the ratio of the long- and middle- wave signals of the more oblique line of the unique blue and yellow settings (Mollon, 2006). Failures of colinearity of red and green unique hues settings suggest either a single non-linear mechanism or multiple unipolar mechanisms; while the curved blue and yellow unique hue vectors suggest

some additivity failure (Stockman & Brainard, 2010). Our findings suggest that considering three bipolar or six unipolar chromatic mechanisms would produce a system with complementary relationships between the opponent pairs (Helmholtz, 1852; Pridmore, 2008; Shepard et al., 2017). On the whole, the good correspondence with a large unique hue dataset further supports the usefulness of our data to evaluate colour appearance models.

Considering the overlapness of colour categories in cone excitation space, *purple* and *green* were the most connected categories, followed by *grey*, *brown* and *blue*. The large number of connections of *purple* and *green* can be explained by their larger extent in colour space. *Lime green, yellow, turquoise* and *magenta* exhibited the smallest number of links with other names as they were located closer to spectrum locus than other colour terms. Basic colour terms tend to share a small number of common samples between them, with the strongest link found between *red* and *pink*. *Cream* was found to be a crucial node linked with *yellow*, *orange*, *white*, *pink* and *green*; and our findings support the suggestion of a missing BCT in this area (Boynton & Olson, 1987; Sturges & Whitfield, 1995). Strong links were also found between turquoise and blue and between lilac, pink and purple implying that there is a considerable overlap between the suggested additional BCTs turquoise and lilac and their neighbours (Mylonas & MacDonald, 2016). As expected from Hering's opponent theory, *green* was not linked to *red*, and *blue* was not linked to yellow. However, *red* was not connected to *yellow*, and *orange* serves as an important node between them. *Yellow* and *orange* were not connected to the achromatic categories. Contrary to earlier results (Boynton & Olson, 1987) we report a weak link between *red* and *blue*. Other unpaired BCTs were *green-purple, blue-orange* and *brown-blue*. In the practice of many artists, *green* complements *purple*, and *blue* complements *orange* while *brown* can be seen as a muted darker orange (Goethe, 1840; Gage, 1993).

For the generalisation of our observations from the 600 stimuli used in the experiment to the entire 3D gamut of the CRT monitor in cone excitation space, we employed a Maximum a Posteriori estimator (Mylonas et al., 2010). The regions with the highest purity of L, M and S cone excitations were assigned to *red, green* and *blue* colour names respectively. These findings support traditional accounts (Maxwell, 1872) for the labelling of cone excitations while it contradicts the results of recent studies that assigned the highest values of L excitation to *orange* (Cao et al., 2005) or to *yellow* (Pridmore, 2011,

2013). The combined signals from L and M cones at maximum intensity were assigned to *yellow*, from L and S cones to *magenta* and from M and S cones to *turquoise*.

The visualisation of MAP's classification on a 2D plane of MB's chromaticity diagram confirms our previous conclusions that the vertical axis of S/(L+M) originating from the background chromaticity is crossing *purple* and *yellow* while the horizontal axis of L/(L+M) crosses *turquoise* and *pink*. At very high values the L/(L+M) coincides with the boundary between *pink* and *red*. In DKL colour space the positive pole of the S-(L+M) dimension coincides with the centroid of *purple* and it's negative with an area close to the boundaries of *yellow***. The negative pole of the L-M dimension coincides with *turquoise* and the positive with *pink*. The L-M axis coincides also with the boundary between *blue* and *green* and the boundary between *pink* and *red* at low values of the negative pole and high values of the positive pole respectively. It is clear that any interpretation that colour categories are constrained within the cardinal axes of colour discrimination mechanisms should be treated with caution, especially when the hue sampling is too coarse and the saturation of the samples varies (Skelton et al., 2017; Witzel & Franklin, 2014; Witzel, 2018). It is unclear why the proposed inherent, hard-wired, categories are constrained within the axes of colour discrimination in infants (Skelton et al., 2017) but in adults *purple* and *yellow* categories cross over the axes (Malcoc et al., 2005, Witzel & Gegenfurtner, 2013, 2018; Witzel, 2018) while unique hues are overall stable across the life span (Schefrin & Werner, 1990; Wuerger, 2013). In addition, the size and shape and of lexical colour categories varied in the three colour spaces (LMS, MB and DKL) and the question of which is the most appropriate space to measure these features of colour names remain open.

Supplementary qualitative features include a near circular achromatic region, with *brown*, *cream, lilac* and *dark blue* nested in the interior of the cone chromaticity diagram. *Yellow*, *orange, red, magenta, purple, turquoise* and *lime green* cover mainly the very saturated areas while *blue, green* and *pink* extend from the achromatic to the most chromatic regions. Some categories share straight boundaries (*blue-purple, pink lilac*), while for others (*red-pink, orange-yellow, brown*) the boundaries form a curvature. The existence of borders with curvature raise questions about whether this is the true form of categories in colour space, or an artefact of the models and training data (Moroney, 2008; Cao et al., 2015).

Chapter 7

The indispensability of basic colour terms

Unconstrained colour naming experiments are able to capture a great deal of the large colour lexicons found in many languages of the world. Yet, establishing which colour names are shared and well comprehended by most speakers in each language has proved to be a non-trivial task that requires multiple criteria and combinations of associated measures. In this chapter, we employ information theoretic analysis in the context of language games to define a simple metric that identifies basic colour terms from unconstrained colour naming data in different languages.

7.1.   The identification of Basic Colour Terms

The number of colour names in languages is often large, as that makes colour communication easier and improves the accuracy of colour naming (Lantz & Stefflre, 1964). Yet, only a small number of colour names are shared and comprehended well by most speakers in each language (Brown & Lenneberg, 1954).

In a seminal study, Berlin and Kay (1969/1991) proposed a total universal inventory of eleven basic colour categories, the Basic Colours Terms (BCTs), corresponding to the English *black, white, red, yellow, green, blue, brown, orange, purple, pink* and *grey*. Berlin and Kay did not regard all basic terms as equivalent; the first six were described as 'primary basic' and the remaining five terms as 'derived' or 'secondary basic'. Berlin & Kay's criteria for the identification of BCTs was based on multiple factors (e.g. single word terms that are not the name of an object, see section 2.3. for all eight criteria) judged by experts as not being equally applicable across languages (Crawford, 1982; Saunders & van Brakel, 1997; Levinson, 2000; Biggam, 2012). Others have segregated basic colour categories on more rigorous behavioural criteria such as frequency, response time and consistency, but not without applying language-specific criteria such as restricting the responses to single word terms (Boynton & Olson, 1987; Sturges & Whitfield, 1995; Corbet & Davies 1997; Lindsey & Brown, 2014, Mylonas & MacDonald, 2016). These criteria have been applied to determine a limited number of basic colour terms within different languages across the world, but current consensus suggest that basicness is not a simple dichotomous but more a continuous gradual characteristic of

lexical colour categories, and the quest for a more cross-culturally legitimate approach remains unsettled (Witzel, 2018; for a review).

The question underlying this chapter is what is meant by 'basicness' in information technology-enabled communication systems with an aim to establish a language-independent scale of basicness from unconstrained colour naming responses in different languages. In the context of colour naming, basic colour terms refer to the degree that linguistic signifiers are shared and comprehended by most speakers in each language to communicate their categories of colour sensations. This depends broadly on the responses of the human visual system (Berlin & Kay, 1969/1991; Kay & McDaniel, 1978; Griffin, 2001; Regier et al., 2007), the referents in the environment (Mollon 1982; Webster & Mollon, 1997; Yendrikhovskij, 2001) and the communication needs of the social group (Brown & Lenneberg, 1954; Lucy & Shweder, 1979; Davidoff et al., 1999; Levinson, 2000; Gibson et al., 2017).

Along these lines, in Chapter 3 we presented families of lexical, behavioural, and geometric features of unconstrained colour names in different languages, but none was sufficient to demarcate alone BCTs from non-basics. In Chapter 4, a robust classifier required training by more than one family of these features to produce perfect coherence between members of the Basic class while the coherence for members of the Primary class was much lower for all available features. This multiplicity of measures is subject to high risk in being applied differently by different researchers for demarcating BCTs in different languages. At the same time, consistent with recent studies (Lindsey et al., 2015; Regier et al., 2015, Gibson et al., 2017), our findings imply that informativeness provide a better framework to advance our understanding in colour naming than the hypothetical primary colours of the opponent theory (Hering, 1878/1964). Here, we employ information theoretic analysis in the context of language games to propose a simple metric to determine the degree of basicness of unconstrained colour names in different languages.

7.2. Methods

Recently colour names have been studied by measuring their communication effectiveness using information theoretic analysis in the context of language games (Lindsey et al., 2015, Regier et al., 2015; Gibson et al., 2017). For example, in Figure 7.1 'Alice' is presented with a colour (step 1), which she names (step 2). 'Bob' then attempts to guess the colour from the name (step 3). How many tries will Bob on average take to guess the colour?
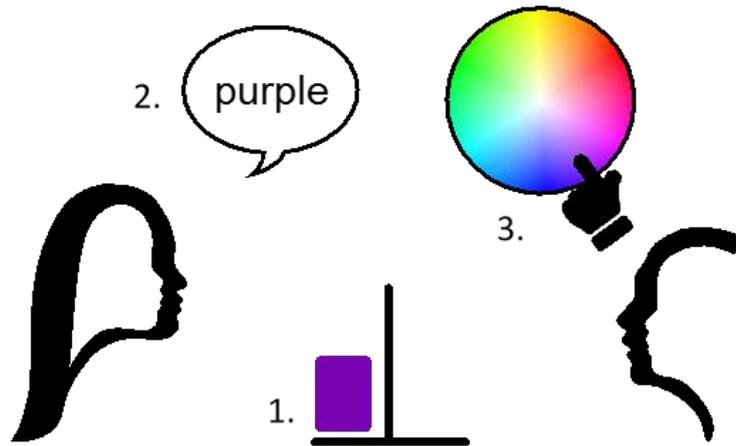
Figure 7.1 Communication game for colour chips. 'Alice' is presented with a colour (step 1), which she names (step 2). 'Bob' then attempts to guess the colour from the name (step 3).

Assuming Alice names colours like the population average, and Bob guesses optimally, then performance is computable from colour naming data as:

$$surprisal(c) := -\sum_{n} P(n|c) \log_2 P(c|n) \qquad (7.1)$$

where *P(n|c)* is the conditional probability that name *n* will be chosen for a colour *c*, and *P(c|n)* that colour *c* was the cause of a naming *n*. In Figure 7.2, we show the surprisal of *c* = 600 colour samples named in the online colour naming experiment using *n* = 478 distinct colour names in British English offered by at least two observers.
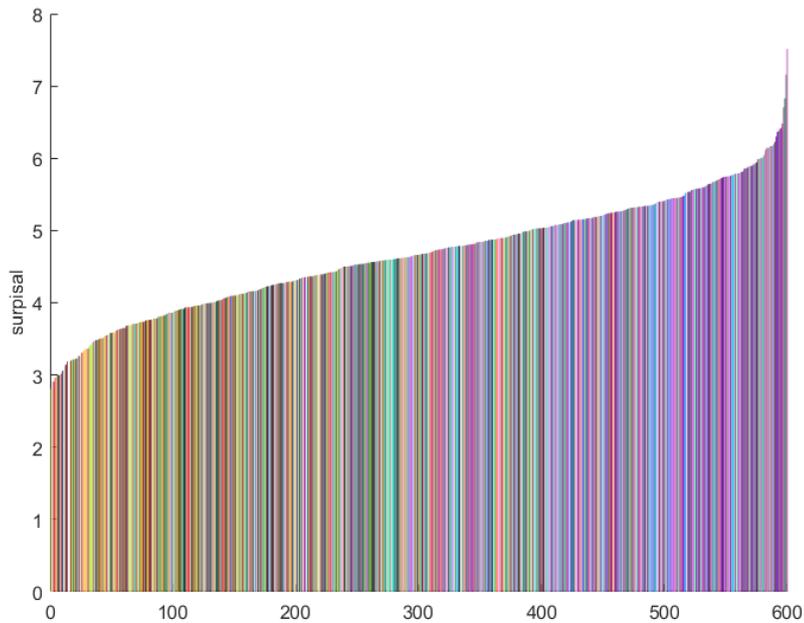
Figure 7.2 Surprisal of colour samples named in the online colour naming experiment by British English speakers.

Overall, surprisal tends to be higher for cooler than warmer colours (Gibson et al., 2017) reflecting in our view the larger perceptual extents of categories in the cool region of colour space (see for example the extent of BCTs mapped on Munsell array by Berlin & Kay in Figure 2.10). Our claim is further supported by the strong Pearson positive correlation between surprisal and the extent of the category across samples, *r = 0.56, p < 0.001* compared to the weak negative Pearson correlation *r = -0.14, p < 0.001* between surprisal and ambiguity in naming, measured by the entropy of colour naming distributions over samples (Figure 7.3).
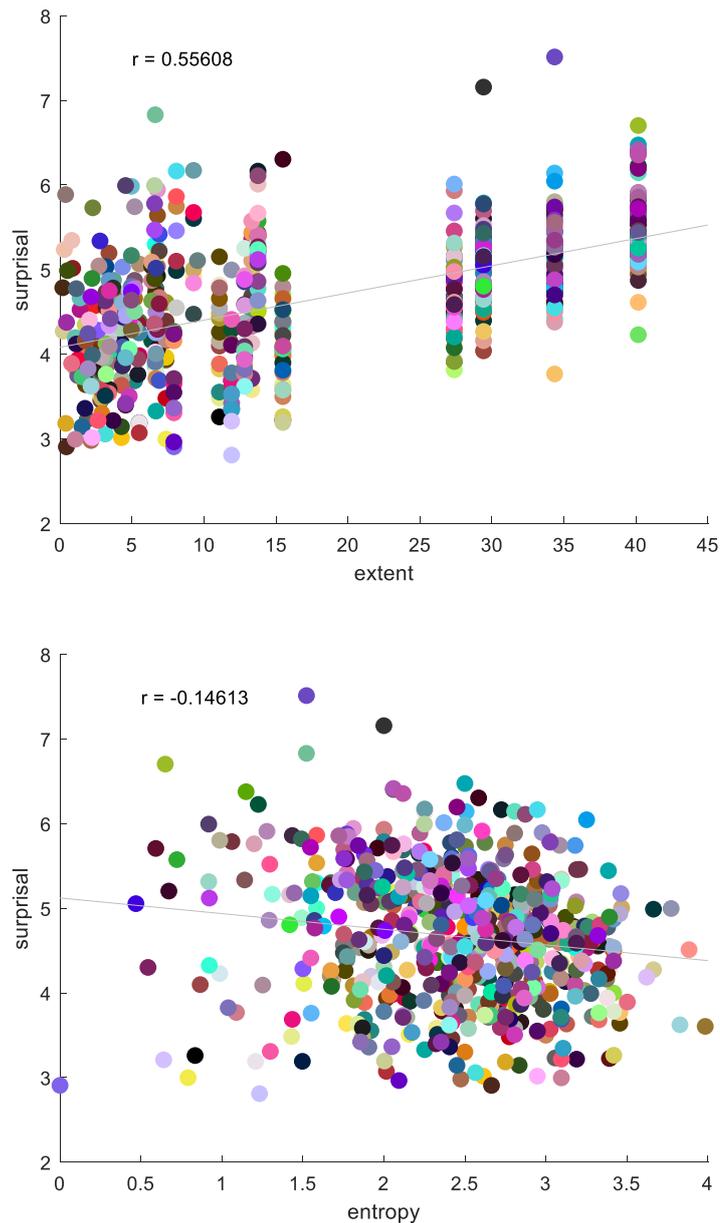
Figure 7.3 Correlation between extent and surprisal (top); entropy and surprisal (bottom)
of colour naming distributions.

Having determined the surprisal for each chip, we were interested in defining a measure of basicness for lexical colour categories. For inspiration, consider Figure 7.4 which shows the conditional probability of *yellow* – a basic term – and *coral* – a non-basic term – over chips in the online colour naming experiment. First, we observe that *yellow* is more frequent overall than *coral*. Also, for a number of samples, *yellow* was the most likely colour name whilst for *coral* there are always a substantial fraction of other colour names that have been used to describe them. However, neither overall frequency nor consensus is sufficient to fully separate BCTs from non-basic colour names (see Figure 7.5). In terms of frequency measured here as the sum of P(n|c) for each category, *lilac*

and *turquoise* are more frequent than *orange*, *yellow*, *grey*, *red*, *black* and *white* (Lindsey & Brown, 2014; Mylonas & MacDonald, 2016). While, in terms of consensus measured as the peak response rate of P(n|c) for each category, *lilac* was again found earlier than *white* and *green*; *lime green* and *royal* blue are equal to *green*. Therefore, an alternative measure is needed to capture the importance of BCTs.
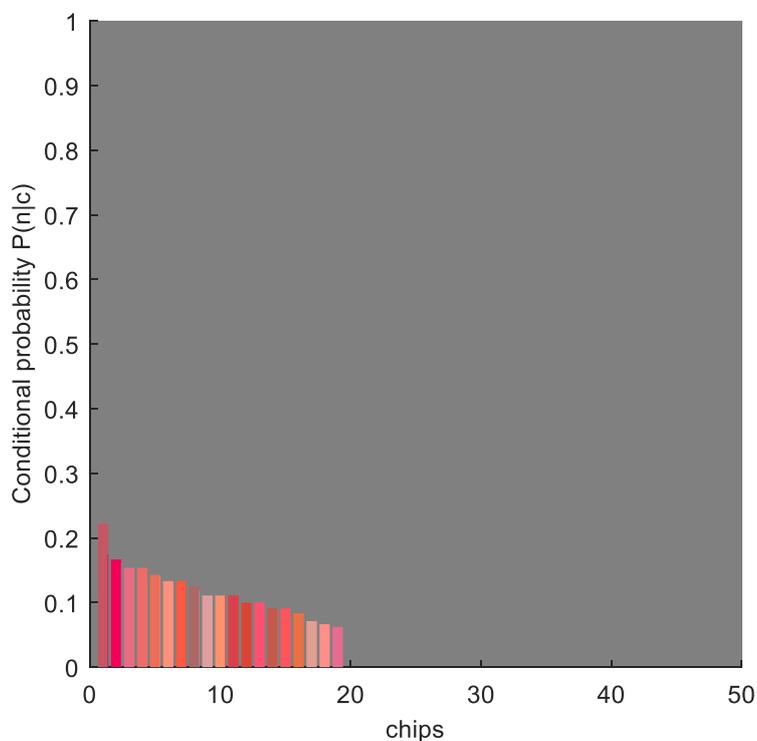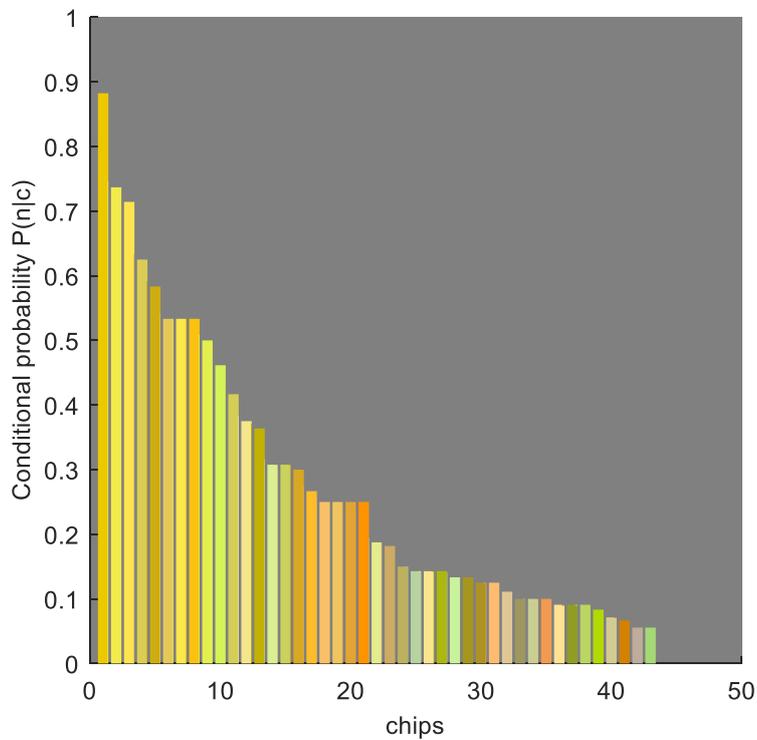
Figure 7.4 Conditional probability of names n: yellow (top) and coral (bottom), given colours c, over chips.
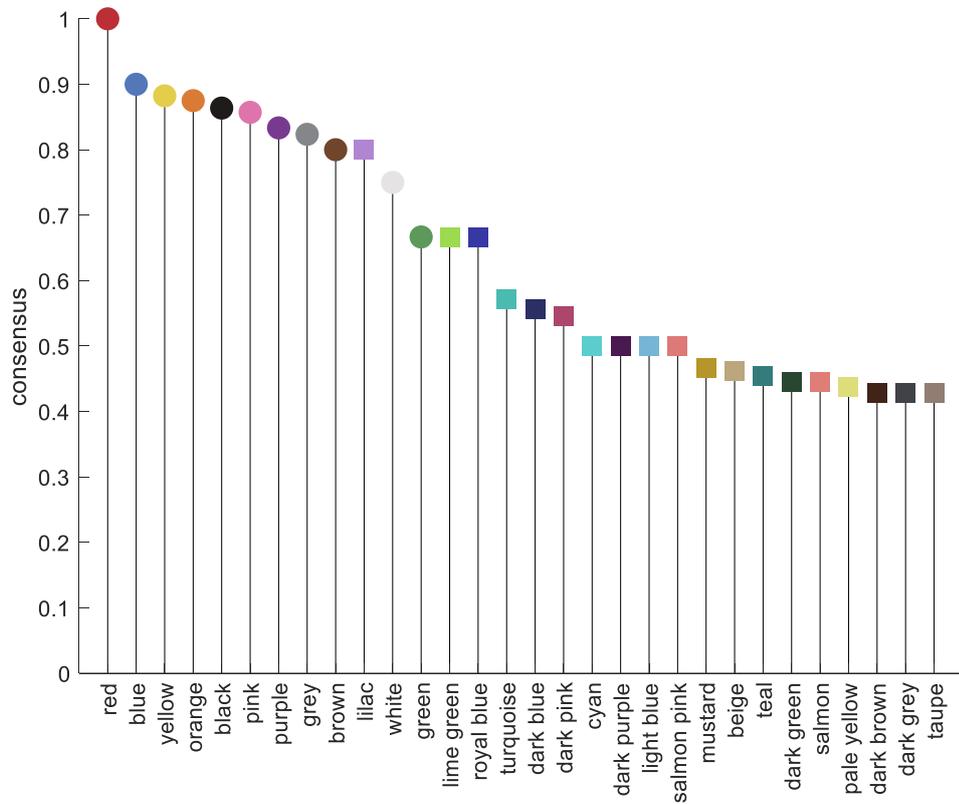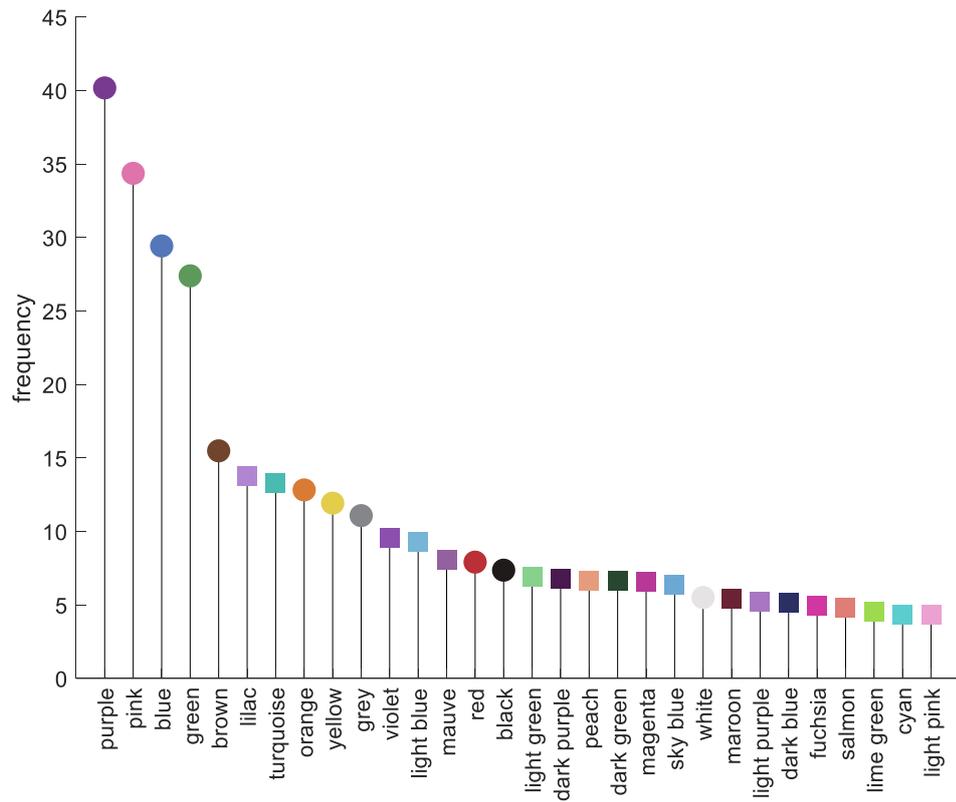
Figure 7.5 Frequency (top) and consensus (bottom) of colour names. BCTs are drawn with disks and non-basic names with squares.

Here, we unite frequency and consensus with a novel information theoretic measure, which we call *dispensability*, for colour categories that we hypothesised would predict basicness. Our *dispensability* measure is an analogue of the surprisal measure (Figure 7.6) but instead of measuring the importance of colour chips (Gibson et al., 2017), it is concerned with measuring the importance of colour names. Alice is given a colour name (step 1) and points at a colour (step 2) which could give rise to the name. Bob then attempts to guess the name (step 3) from the indicated colour. How many tries will Bob on average take to guess the name?
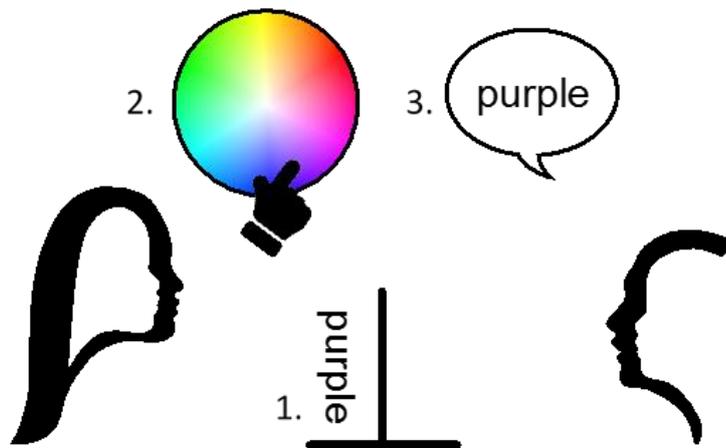
Figure 7.6 Communication game for colour categories. Alice is given a colour name (step 1) and points at a colour (step 2) which could give rise to the name. Bob then attempts to guess the name (step 3) from the indicated colour.

If Alice chooses colours for names according to colour naming data, and Bob is optimal in his guessing, then performance can be computed as

$$dispensability(n) := -\sum_{c} P(c|n) \log_2 P(n|c) \qquad (7.2)$$

Dispensability will be low for a name, if a fraction of the chips that can be so named, are named by that term. Dispensability takes into account for each colour name, the colours for which all other names are rarely used. In other words, dispensability determines basicness by identifying which are the superordinate colour names that cannot be replaced with any other name – hence the title we have given to our measure.

We computed dispensability using all available colour naming datasets from the unconstrained online colour naming experiment in British English, American English,

Russian, Greek, Turkish and Thai presented in detail in Chapter 3 and in Appendix A. We have also considered the English dataset obtained in the offline colour naming experiment presented in Chapter 6.

For the c=600 colour samples in the online colour naming experiment, 500 British English observers offered 7,405 responses using n=478 distinct colour names shared by at least two observers. In American English, we obtained 8,948 responses from 500 observers using n=483 distinct colour names. In Greek, 500 participants gave 5,870 responses, in which n=314 colour names were distinct. From 500 Russian observers, we obtained n=342 distinct colour names from 7,802 responses. In Turkish, we obtained n=285 distinct colour names from 4,727 responses of 309 observers. In Thai, 255 observers gave 3,516 responses using n=284 distinct colour names. Finally, in the offline experiment, 10 English speakers offered 4,684 responses using n=246 unique colour names. In all language datasets, only colour names that were offered by two or more observers were considered.

7.3. Results

In Figure 7.7, we show the 30 most indispensable colour names ordered from low to high dispensability in all datasets. We chose 30 because non-expert observers are able to identify 30 colour names in their native language without training (Derefeldt & Swartling, 1995).

Dispensability varies with colour name and remarkably, for all three datasets in English, all BCTs (Berlin & Kay, 1969/1991) had lower dispensability scores than all non-BCTs. For example, in British English the score for *yellow* was 1.39, for the American English was 1.46 and for the laboratory-based English dataset was 1.28; while for *mustard*, the scores were 2.32, 3.10 and 2.11 respectively. The range of dispensability for basic terms was 1.39-1.90 for the British English speakers, 1.14-1.89 for the American English and 0.81-1.40 for English speakers of the laboratory-based experiment. For non-basic terms, the ranges were 2.04-4.49, 2.01-4.49 and 1.75-3.25 for British, American and laboratory English speakers respectively. In British English, the 11 BCTs were followed closely by *turquoise* and *lilac*; but there was a considerable jump in dispensability value to the following non-basic term, *beige*. In American English, the last BCT *green* was followed firmly by *peach, salmon*, and *maroon* while the separation was bigger with the 15th, *dark green*. In the English dataset from the laboratory-based experiment, there was a clear separation between the 11 BCTs and the non-basics *lime green, beige* and *dark green* in terms of dispensability scores.

In terms of dispensability, the two proposed basic terms that describe the unitary English *blue* category in Greek (*ble* in 7th and *galazio* in 11th), Russian (*sinij* in 5th and *goluboj* in 8th) and Thai (*fa* in 3rd and *namngen* in 10th) were ranked in the top 11 positions but in Turkish the first *blue* term (*mavi*) was ranked in the 6th and the second *blue* term (*lacivert*) was ranked in the 13th position. In Greek, we identified an additional indispensable colour term for the *olive* category (*ladi*) that was ranked in the 12th position earlier than the last basic term *grey* (*gri*) that was ranked in the 13th position. *Olive green* was also ranked in the 17th position in the dispensability rank of the American English and in 20th in the English of the offline experiment.

English - Offline
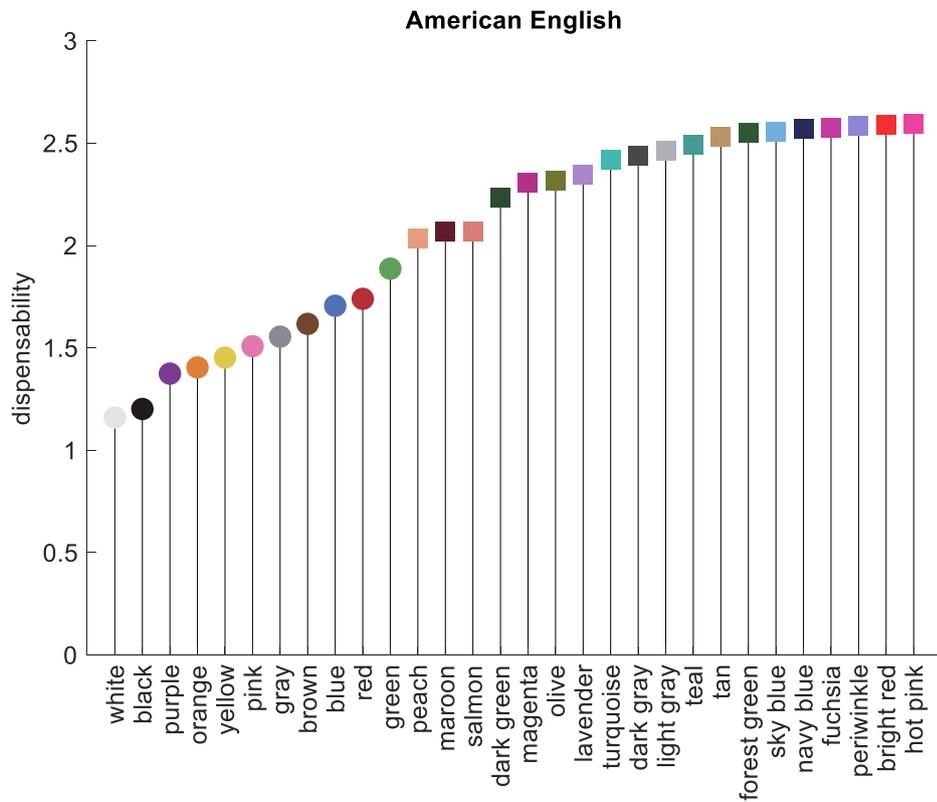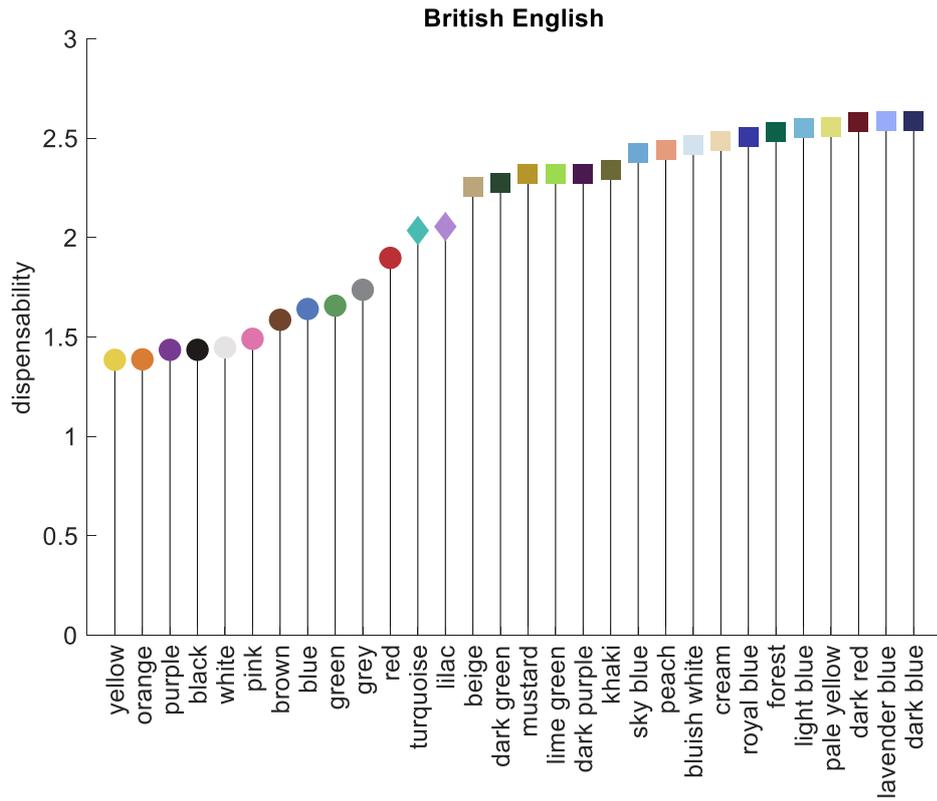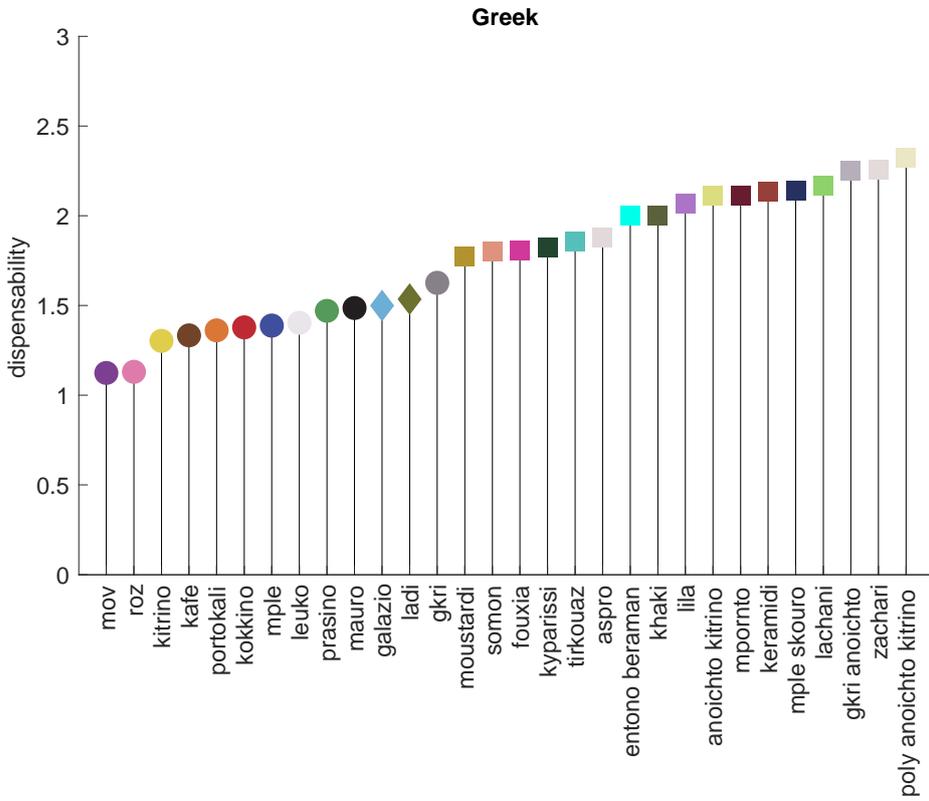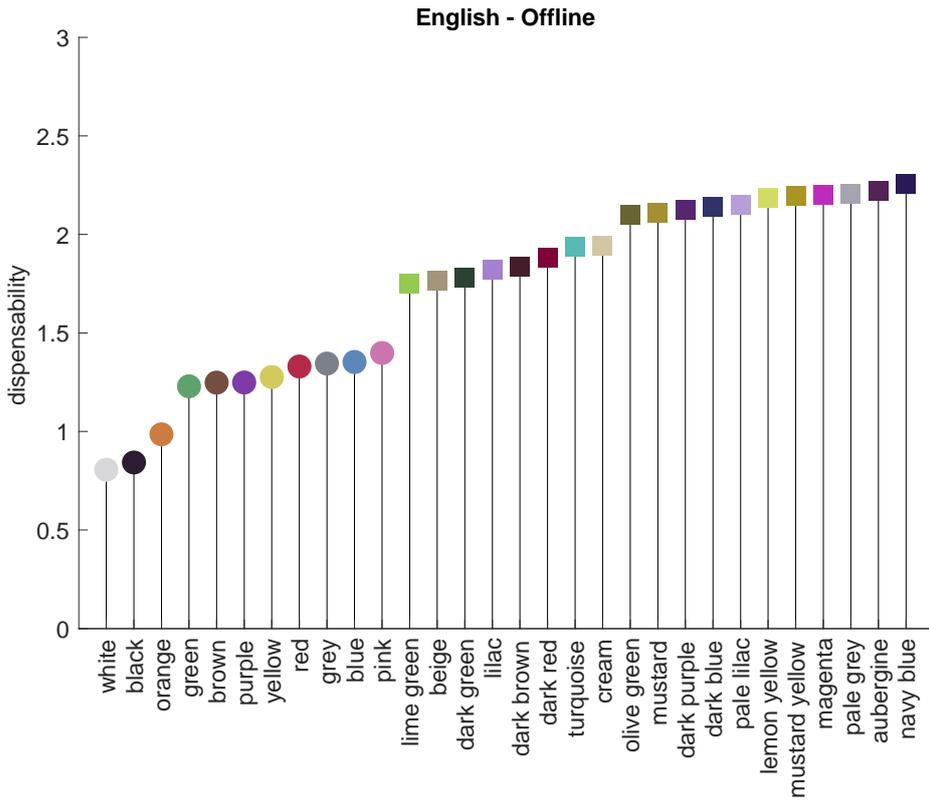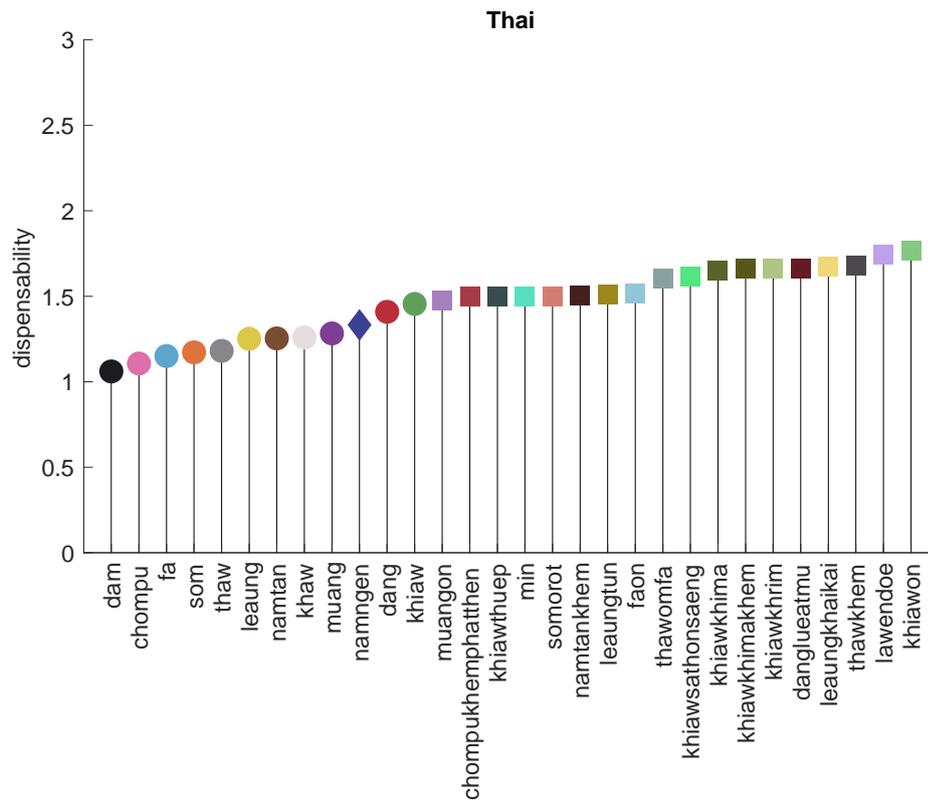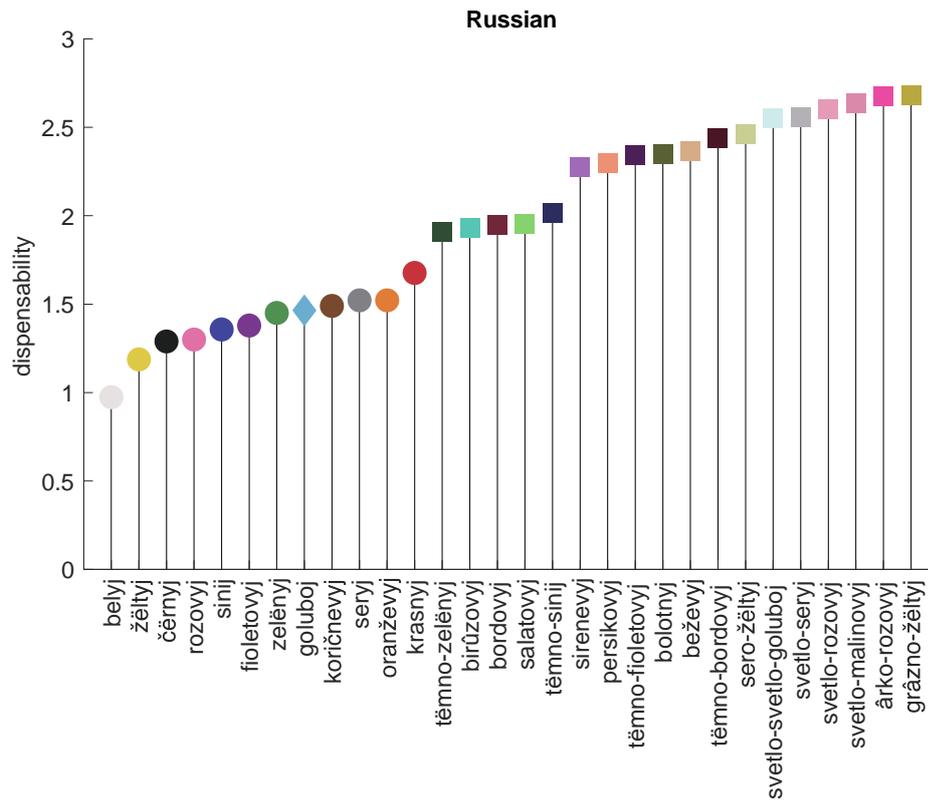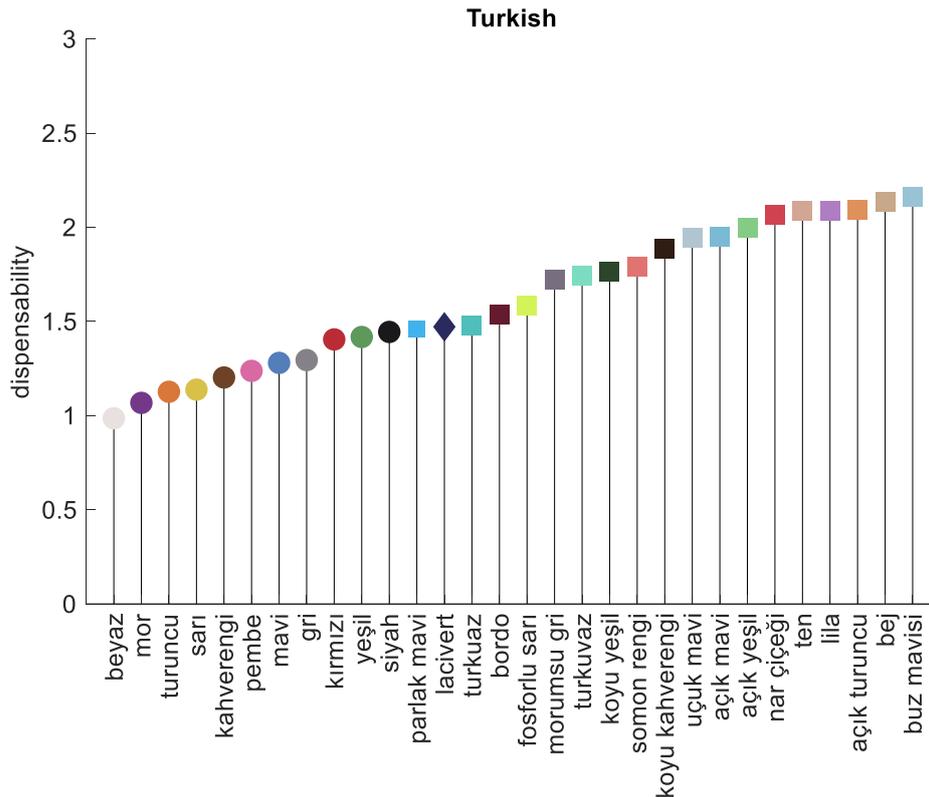
Greek

Russian

Thai

Figure 7.7 Dispensability of colour names within British English, American English, English-Offline experiment, Greek, Russian, Thai and Turkish (from top to bottom). The 11 Basic Colour Terms of Berlin & Kay (1969/1991) are drawn with disks, proposed additional basic terms with diamonds and non-basic colour names with squares.

To explore whether there is quantintative evidence for the upper limit to the number of BCTs within each language in Table 7.1 we express the dispensability 'gap' between last BCT and first non-BCT as a fraction of the standard deviation of dispensability scores for the 11 BCTs of Berlin and Kay (1969/1991), and of the proposed BCTs (see 2.4. The largest fraction was found in the laboratory-data for the 11 BCTs while for the proposed BCTs the largest fraction was found for the Russian data. The lowest fraction was observed in the American English, Thai and Turkish dispensability scales. On the whole, the low number of statistic scores imply dispensability is a gradual scale.

Table 7.1 Dispensability fraction of the step between last BCT and first non-BCT and variance of BCTs for the 11 BCTs of Berlin & Kay and proposed BCTs (Engchuan, 2003; Ozgen & Davies, 1998; Androulaki et al., 2006; Mylonas & MacDonald, 2016; Paramei et al., 2018).

	11 BCTs	**Proposed BCTs**
British English	0.35	0.84
American English	0.11	NA
English - Offline	1.86	0.02
Greek	1.00	1.01
Russian	1.22	1.27
Thai	0.18	0.19
Turkish	0.09	0.06

Overall, our dispensability analysis shows that basicness is not a simple dichotomous but a continuous gradual characteristic of lexical colour categories. The order of the most indispensable colour names varies in each language. Colour names with the low dispensability scores neither resemble the evolutionary hierarchical order (Berlin & Kay, 1969/1991) nor give a special status to the primary colour categories. Languages with two basic blue terms are not an exception in this regard, but rather form a majority in the colour naming datasets of this thesis.

7.4. Discussion

The identification of colour names that are shared and well comprehended among speakers in each language has proved to be a non-trivial task that requires multiple criteria and combinations of associated measures. These criteria have been strongly criticised as not being equally applicable across languages (Saunders & van Brakel, 1997; Levinson, 2000; Biggam, 2012) while their multiplicity is vulnerable to high risk of being applied differently by different researchers for demarcating BCTs in different languages.

In this chapter, we argue that basic colour categories are indispensable and propose a simple information theoretic measure of basicness that combines the frequency and consensus of colour names between speakers in each language. Dispensability determines basicness by identifying which are the superordinate colour names that cannot be replaced with any other name – hence the title we have given to our measure.

To test the performance of our information theoretic measure, we considered unconstrained colour names in British and American English, Greek, Russian, Thai and Turkish from an online experiment, along with a calibrated English dataset from a laboratory-based experiment. Dispensability varied with category and produced a graded scale of basicness. Most remarkably, for all three datasets in English the 11 BCTs (Berlin & Kay, 1969/1991) had lower dispensability scores than all non-BCTs while also being able to capture the indispensability of the proposed second blue basic term in Greek, Russian, Thai and to a lesser degree in Turkish. Dispensability does not identify the BCTs because they are commonly used (e.g. dark green is more frequent in our datasets than white). Rather it works because for each there are colours for which all other names are rarely used. These findings suggest that information theory is well-suited to provide a simple language-independent measure to determine the degree of basicness of unconstrained colour names within different languages, and can reveal the emergence of unknown basic categories that may not meet the conceptual criteria of Berlin & Kay (1969/1991).

In our assessment of dispensability for British English; the 11 BCTs were followed closely by *turquoise* and *lilac*, but there was a considerable jump in dispensability score to the following non-basic term *beige*. The candidacy of *lilac* and *turquoise* as BCTs was also supported in our earlier study (Mylonas & MacDonald, 2016), where we analysed colour naming responses from the online colour naming experiment and computed the mean of the ranks for each colour term across six different measures (frequency, consensus, response time, consistency, volume and inter-experimental agreement) to obtain an gradual index of basicness. Both terms appear to reduce the uncertainty of colour naming from using only the 11 BCTs, as *lilac* partitions the large colour category of *purple* into light and dark segments; while *turquoise* stabilises the large boundary area between *green* and *blue*. *Beige* appears in the area between *white, yellow, pink* and *orange* which attracts a large number of different names, such as *cream, peach, tan* and *salmon*, and has been claimed as a region with a missing BCT (*peach* in Boynton & Olson, 1987; *cream* in Sturges & Whitfield, 1995). This is also supported by the high consensus of *cream* obtained in our laboratory-based experiment of Chapter 6 and the examination for a possible 12th basic colour term in Chapter 4, where the addition of *cream* to the 11 BCTs produced a perfect coherence score for the Basic class.

Considering the results in American English, the last BCT *green* was followed firmly by *peach, salmon* and *maroon* while the separation was bigger with the 15th colour, *dark green*. Lindsey & Brown (2014) applied Zipf's law (1935) to colour naming frequencies and reported a steep decrease in frequency of colour terms beyond the 15th term in

American English colour lexicon. They also found that *peach, teal, lavender* and *maroon* were named with high consensus across observers. The low dispensability of *peach* and *salmon* provides additional evidence for a missing BCT in this hard-to-name region. The high indispensability of *maroon* was not found in the responses of British English speakers who more often offered the dark modifier of *red* to describe this region. We note the close similarity of *teal* to *turquoise*, and also *lavender* to *lilac* which were ranked in the 23rd, 19th, 18th and 41st position by our dispensability measure. On the whole, our dispensability measure ranked in the top 11 positions the same BCTs for British and Americans English speakers. However, the highly indispensable colour names closely following the BCTs were different between the two groups. This suggests that the 11 BCTs in English are more universally shared between English speakers than non-basic terms. This is also evident from the remarkably small mean colour difference between the BCTs in American and British English (ΔE_{00}=1.49) reported in Chapter 3. The universality between English speakers of the 11 BCTs is also supported by the analysis of the international English dataset of the lab-based experiment. Here, we found a clear separation between the 11 BCTs and the non-basics *lime green, beige, dark green, lilac, dark brown, dark red, turquoise* and *cream*; with a second jump in terms of dispensability score at the 20th *olive green*.

In regard to the dispensability index of the two proposed basic terms that describe the unitary English *blue* category in Greek (*ble/blue* in 7th and *galazio/sky blue* in 11th), Russian (*sinij/blue* in 5th and *goluboj/sky bue* in 8th) and Thai (*fa/sky blue* in 3rd and *namngen/blue* in 10th); these terms were ranked in the top 11 positions, but in Turkish the first *blue* term (*mavi/blue*) was ranked in the 6th and the second *blue* term (*lacivert/navy blue*) was ranked in the 13th position. These results are generally consistent with previous reports on the 12 BCTs in Greek, Russian and Thai as well as on the uncertainty about the basicness of the second *blue* term in Turkish (Androulaki et al., 2006; Morgan & Corbett, 1989; Prasithrathsint, 1988; Ozgen & Davies, 1998).

In Modern Greek, Androulaki and her colleagues (2006) reported twelve BCTs including two *blues* (*ble* and *galazio*). The differentiation of the *blue* English category into two basic *blue* categories in Greek was also supported by the semantic shifts of category prototypes with different levels of Greek-English bilingualism (Athanasopoulos, 2009). The two basic blue categories in Russian (*sinij* and *goluboj)* were reported by Berlin and Kay (1969/1991). This exemption to their universal inventory of the eleven BCTs and the possibility of an additional evolutionary stage in the development of colour lexicons (Stage VIII) triggered a large number of studies that confirmed the second basic *blue* in Russian (Morgan & Corbett, 1989; Paramei, 2005; Paramei et al. 2018). In Thai, Berlin

& Kay (1969/1991) reported only 10 BCTs and a single *sky blue* basic term (*fa*) but recent studies identified twelve BCTs including *grey* (*thaw*) and two *blues* (*fa* and *namngen*) (Prasithrathsint, 1988; Engchuan, 2003). Thai is the only language in our investigation where the *sky blue* term (*fa*) has lower dispensability than the *blue* term (*namngen*). The status of the proposed *navy blue* category as a BCT (*lacivert)* is less certain in Turkish. Ozgen and Davies (1998) reported that *navy blue* (*lacivert)* term was offered frequently and with high consensus but the term violated the non-inclusion criterion of Berlin & Kay (1969/1991) with regard to the main *blue* category (*mavi*). A consequent study (Ekici et al., 2006) supported its basicness by reporting similar response time for this *navy blue* term and the other 11 BCTs. Contrary, a more recent study reported a low consensus for the *navy blue* term (Rätsep, 2011) but in the analysis of the Turkish data of this study, we found that *lacivert* was the fourth term with the highest consensus score (Ulusoy et al., 2017).

In Chapter 3, we showed that in Russian, Greek and Thai the loci of their two basic *blues* correspond well and their centroids deviated from the centroid of English *blue*. This was not the case in Turkish. The range of the basic *blue* term (*mavi*) in Turkish overlapped with the *blue* term in British English while the *navy blue* term (*laçikvert*) appear to differ from the *blue* term (*mavi*) mainly in the lightness dimension with no obvious differences in the hue dimension. Except for *blue*, the comparison between British English and American English, Greek, Thai, and Turkish for the loci of the other 10 BCTs showed a very good correspondence with a mean ΔE_{00} of 1.83, 2.13, 2.39 and 2.23 respectively. An equivalent second basic *blue* term has been reported recently also in Italian (Paggetti, Menegaz, & Paramei, 2015). Taking into account the results of this chapter in conjunction with the evidence described above we confirm the postulation of the existence of an additional evolutionary Stage VIII in the development of colour lexicons for Greek, Russian and Thai, but for Turkish further investigation is needed.

In Greek, we identified an additional indispensable colour term for the *olive* category (*ladi*) that was ranked after the second basic *blue* term (*galazio*) in the 12th position but earlier than the last basic term *grey* (*gri*). Olive trees and their by-products play an important role in the diet and culture of classical and modern Greeks (Boardman 1976; Trichopoulou & Lagiou, 1997). *Olive green* was also ranked in the 17th position in the dispensability rank in the American English and in 20th in the English of the offline experiment; while in Chapter 4, the examination of *olive* as 12th member of the Basic class in British English produced the second highest coherence score after *cream*. Contrary to English, in our Greek dataset, *olive* was used as a modifier of green in less than 0.5% of its total occurrences while frequently attracting both light and dark modifiers.

These results are consistent with previous reports on the basic colour terms in Modern Greek (Androulaki et al., 2006) where *olive* (*ladi*) and *sky blue* (*galazio*) where also found with higher basic status than *grey*. In case that these findings will be further confirmed by consequent studies the Greeks may be postulated having 13 BCTs.

With regard to the order of dispensability ranks across languages, we found that the order of the most indispensable colour names varied in each language. White and black were often ranked in the top positions as it would be expected by the evolutionary sequence hypothesis (Berlin & Kay, 1969/1991) but this was also true for *yellow, purple, orange, brown* and *pink*. *Red* was never found earlier than the 6th position and *green* occupied more often the last positions of the BCTs. *Yellow* was found earlier than *red* in all test languages. Clearly, our dispensability analysis cannot be used to deduce a fixed universal sequence across languages (Berlin & Kay, 1969/1991, Loreto et al., 2012). This is not to say that the evolutionary sequence hypothesis is valid or not since our data can only provide a snapshot in time of these fully developed colour languages and the importance of colour categories in different languages may not necessarily follow their order of emergence (Lindsey & Brown, 2006). Nevertheless, we found no evidence for the priority of primary basics over secondary basics in our dispensability scales. These results reinforce earlier conclusions that primary basics do not play a fundamental role in colour naming systems (Zeki, 1980; Boynton & Olson, 1987; Malkoc et al., 2005; Bosten & Boehm, 2014).

Instead, our results support growing evidence that communication efficiency provides a better framework to understand colour naming than opponent theory (Jameson & D'Andrade, 1997; Lindsey et al., 2015; Regier et al., 2015; Abbot et al., 2016; Gibson et al., 2017). This hypothesis, following in part the categorisation principles of Rosch (1978), proposes that colour naming systems are based on optimising the balance between simplicity and informativeness. On one hand, a simple colour naming scheme with a small number of names would be easier to use. On the other hand, an informative scheme with a large number of names would maximise the precision of colour names in colour communication. Our dispensability measure provides the average amount of information conveyed by each colour name in bits and based on this scale one could in principle optimise the granularity of colour naming schemes in different languages described in Chapter 5.

Chapter 8

Discussion

The work presented in this thesis constitutes part of ongoing interdisciplinary research in which we explore the intriguing activity of communicating a large number of discriminable colours using a smaller set of colour names, with an ultimate aim to facilitate colour communication within different languages.

8.1. Online and offline colour naming experiments

We started our investigation in Chapter 3 by asking thousands of people over the Internet to name freely a small number of colours in American and British English, Greek, Russian, Thai and Turkish. We extended previous cross-cultural studies which used only the most saturated colour samples on the surface of the Munsell system (Berlin & Kay, 1969/1991; Kay et al., 2010) by sampling also the interior of the colour solid. In addition, we departed from usual methods which would use a small number of observers and/or the use of only a restricted set of single-word names (Berlin & Kay, 1969/1991; Boynton & Olson, 1987; Sturges & Whitfield, 1995; Benavente et al., 2006; Lindsey & Brown, 2014; Parraga & Akbarinia, 2016). Instead, thousands of volunteers from linguistically and demographically diverse populations named freely a large number of colours online (Moroney, 2003; Mylonas & MacDonald, 2010; Munroe, 2010). We argued that participating in an online experiment in your own familiar environment, with your own equipment, and without the physical attendance of the examiner would give more ecological validity to the underlying categories responsible for colour naming.

Despite the uncalibrated colour reproduction of the online experimental methodology and the linguistic diversity of the observers; except for *blue*, we found a good correspondence between the location of their proposed BCTs (<5 ΔE_{00}). Consistent with earlier studies (Berlin & Kay, 1969/1991; Boynton & Olson, 1987; Uchikawa & Boynton, 1987; Sturges & Whitfield, 1995; Regier et al. 2005; Lindsey & Brown, 2009; Paramei et al., 2018), these results demonstrate that different languages tend to categorise colours into BCTs similarly but with some important differences. The mean colour difference between British and American English (ΔE_{00} = 1.5) was about 3 times smaller than the mean colour differences between British English and Greek (ΔE_{00} = 4.3), Russian (ΔE_{00} = 4.8), Thai (ΔE_{00} = 5.0) and Turkish (ΔE_{00} = 4.3) indicating that speakers of similar linguistic

groups at a population level agree more on the location of BCTs than speakers of different languages (Davidoff, 2015). From our earlier assessment of the colour reproduction error in uncalibrated monitors against sRGB reported in Chapter 2 (ΔE_{00} = 8.0 for desktops and ΔE_{00} = 6.0 for mobiles), we were expecting moderately larger differences and we speculate that the similar stimuli and background configuration in the interface of the experiment led to some kind of perceptual stabilisation of colours (Webster, 2011; Foster, 2011; Zeki et al., 2019).

To respond to the criticism whether the uncalibrated colour reproduction and viewing conditions of online colour naming experiments meet the requirements for rigorous colour research, as well as to criticism of generalizing from laboratory-based experiments estimates of colour naming functions in real-world monitor settings, we also collected a large number of unconstrained colour naming responses from a small number of participants using a calibrated CRT monitor in laboratory settings. The agreement for the location of the basic terms between our web- and laboratory- based experiments (ΔE_{00} = 4.6) was satisfactory and superior of the agreement (ΔE_{00} = 7.4) between previous laboratory-based studies (Boynton & Olson, 1987; Sturges & Whitfield, 1995). The larger differences between earlier studies can be explained by the usage of different colour order systems to map the colour naming distributions. A comparison between the loci of BCTs between individual observers of our offline experiment showed that inter-experimental and inter-language differences in the online experiment were smaller than the intra-experimental differences (ΔE_{00} = 7.9) among individuals. These findings are in agreement with the reports of previous studies that intra-language differences among individuals are larger than inter-language differences (Berlin & Kay, 1969/1991; Webster & Kay, 2007). The intra-language measure was not available for the online datasets because each individual named only 0.3% of the 600 total colour samples of the experiment. These results suggest that online and offline colour naming experiments produce fairly consistent results and support the validity of both methods in estimating colour naming functions in calibrated and uncalibrated monitor settings.

Considering the question of which experimental methodology (online or offline) produces the best agreement for the underlying categories responsible for colour naming, we found that averaging a small number of colour naming responses from a large number of participants in an uncalibrated crowdsourcing experiment produced a better agreement (ΔE_{00} = 1.5, colour difference between British and American English BCTs) than averaging a large number of responses from a small number of different individual in a calibrated laboratory-based experiment (ΔE_{00} = 7.9, colour difference between BCTs of individuals) (Shapiro et al., 1998; Surowiecki, 2005; Webster & Kay, 2007; Yi et al.,

2012). Nevertheless, in the calibrated dataset from the laboratory-based experiment, we mapped for the first-time unconstrained colour naming distributions in the physiologically-based cone excitation space recently adopted by the CIE that allowed us to examine with more confidence the relationships between colour naming and colour vision mechanisms (Stockman & Sharpe, 2000; CIE 170-1:2006; CIE 170-2: 2015).

8.2. Colour naming and 1st stage of colour vision mechanisms

It has been hypothesised that languages gravitate to an optimal set of categories and return to them despite departures from the norm by individual speakers driven by cultural, biological or behavioural mechanisms (Bimler, 2005; Regier et al., 2007; Griffin, 2006). Considering the range of possible cone responses as an explanation for the mechanisms that pressure colour categories to optimality, in Chapter 6 we examined a widely cited account of this type suggesting that colour categories are determined by optimising the division of an irregular perceptual colour space to maximise similarity within a category and minimise similarity across categories (Jameson & D'Andrade, 1997; Regier et al., 2007). In our assessment, as to whether perceptual structure can explain the consistent location of the BCT centroids between our online and offline experiments, we used a k-means algorithm to construct a set of imaginary colour naming systems based on Euclidean distances between the 600 approximately uniformly distributed simulated Munsell samples in CIELAB (Zaslavsky et al., 2018). A comparison of observed BCTs loci of the lab-based experiment against optimal sets of BCTs from the hypothetical colour naming systems (ΔE_{ab} = 14.61) showed that perceptual structure embedded in the stimuli set alone cannot explain the agreement (ΔE_{ab} = 7.30) with the online data. These results are consistent with previous reports on model-based colour categories where consensus between agents was achieved through the communication process (Steels & Belpaeme, 2005). Perceptual structure could be a prominent account if the focus of the research would be artificially constrained on the six primary basics mapped only on the uneven, in terms of saturation, surface of the Munsell system, but it cannot account for the consistency of all BCTs when mapped on an approximately uniformly distributed grid of samples. This highlights the necessity of sampling the surface as well as the interior of the colour solid in colour naming studies (Buchsbaum & Bloch, 2002; Paramei, 2005; Mylonas & MacDonald, 2010; Ocelák, 2014; Witzel, 2018).

In regard to labelling long-, medium- and short- wavelength cone excitation mechanisms for facilitating their communication to wider audiences; a probabilistic colour naming model trained by participant responses assigned the regions with the highest purity of L, M and S cone excitations in the gamut of a CRT monitor to the traditional *red, green* and

blue colour terms respectively, but these were not the most typical samples of the three colour categories (Maxwell, 1872; Mylonas et al., 2010). In addition, the region with the highest contributions from L and M cones was assigned to *yellow*. *Turquoise* was assigned to the region with the maximum contributions from M and S cones and *magenta* was localised to areas with the strongest contributions from L and S cones. We note that single cones are colour blind and applying these labels in different settings – e.g. wavelength peak cone sensitivity – can be misleading (Rushton, 1972; Stockman & Brainard, 2010).

An examination in Chapter 4, for whether the three cone excitation mechanisms leave their traces in the properties of their associated colour names, compared to other colours, showed that *red, green* and *blue* do not form a coherent class. This supports the ongoing argument against the special role of first stage mechanisms in the formation of colour categories (Wuerger et al., 2005; Witzel & Gegenfurtner, 2013).

8.3. Colour naming and 2nd stage of colour vision mechanisms

Consistent with earlier studies was also the discrepancy between the bipolar cardinal directions of the second stage mechanisms with the landmark colour names usually associated with the unique hues in DKL colour space (Abramov & Gordon, 1994; Valberg, 2001; Wuerger, Atkinson & Cropper, 2005). The negative side of the L-M axis coincided with the boundaries between *blue* and *green*, and the positive with the boundary between *red* and *pink*. The negative S-(L+M) axis crossed *yellow*, and the positive crossed over the centroid of *purple*. These were also evident in the presentation of the results in the MacLeod & Boynton (1979) cone chromaticity diagram. These results confirm not only the large discrepancies between colour discrimination mechanisms and the axes of colour appearance mechanisms (Abramov & Gordon, 1994; Webster et al., 2000; Valberg, 2001; Wuerger et al., 2005), but also the discrepancies between the sensitivity of the second stage mechanisms to hue differences and the boundaries of colour categories (Malcoc et al., 2005, Witzel & Gegenfurtner, 2013, 2018; Shepard et al., 2017; Witzel, 2018).

Considering whether the second stage mechanisms play a fundamental role in colour naming and leave measurable traces to their associated colour names compared to others, in Chapter 4 we assessed the coherence of the proposed labels for six (Shepard et al., 2017) and five (Skelton et al., 2017) inherent mechanisms. We found that both subsets of colour names do not form a coherent class, and support earlier conclusions that subcortical and higher-order second stage mechanisms are not the basis of colour

naming systems (Abramov & Gordon, 1994; Webster et al., 2000; Valberg, 2001; Wuerger et al., 2005; Malcoc et al., 2005, Witzel & Gegenfurtner, 2013, 2018; Witzel, 2018). However, the examination of multiple higher-order colour discrimination mechanisms in addition to, or beyond the four cardinal dimensions showed an improved coherence between their associated colour names and it is an important step towards the right direction in understanding the relationship between perceptual and linguistic aspects of colour (Hansen & Gegenfurtner, 2006; Eskew, 2009; Shepard et al., 2017).

These findings highlight also the discrepancy between infant and adult colour categories in English (Skelton et al., 2017; Witzel, 2018). It is unclear why the proposed inherent, hard-wired, categories are constrained by the axes of colour discrimination in infants, but here the *purple* and *yellow* categories of adults, cross over the axes (see also Malcoc et al., 2005, Witzel & Gegenfurtner, 2013, 2018; Witzel, 2018). Unique hues have been found overall stable across the life span (Schefrin & Werner, 1990; Wuerger, 2013).

8.4. Colour naming and 3rd stage of colour vision mechanisms

In our assessment of the relationship between colour naming and the opponent mechanisms of colour appearance (Hurvich & Jameson, 1957), except for *blue*, we found a good coincidence between the location of colour terms *red*, *green* and *yellow* and the corresponding unique hue settings (Xiao et al., 2011; Kuehni, 2005). The centroids of *red* and *green* were not colinear with *white* in the cone chromaticity diagram, but *yellow* was colinear through *white* with *blue,* confirming the consistent failure of linearirty for the opponent pair of unique red and green mechanisms (Chichilnisky & Wandell, 1999; Wuerger et al., 2005). *Red* was nearly colinear to *turquoise*, and *green* to *magenta*. The location of *cyan* would also align with *red* but *turquoise* was used by our observers more frequently and with higher consensus to describe this region.

An explanation for the misalignment of *red* and *green* and the alignment of *blue* and *yellow* in chromaticity space is that their location coincides to sensations that have a simpler relation to their underlying reflectance than other colours (Philipona & O'Regan, 2006). Failures of colinearity between *red-green* or *blue-yellow* suggest either single non-linear mechanisms or multiple unipolar mechanisms of cone absorption combinations (Chichilnisky & Wandell, 1999; Stockman & Brainard, 2010; Shepard et al., 2017). For the relative larger hue differences between the location of the blue unique hue and the foci-centroid pair of the *blue* colour term, the collinearity between *blue*, *yellow* and *white* in our colour naming data as well as their more vertical alignment in the cone chromaticity diagram corresponds better to variation in the signal of the short-wave sensitive cones

than the larger modulation of the ratio of the long- and middle- wave signals of the more oblique line of the unique blue and yellow settings (Mollon, 2006).

The location of the unique blue (Xiao et al., 2011) corresponds better with the second *sky blue* basic terms in Greek (*galazio*), Russian (*guloboj*) and Thai (*fa*) reported in Chapter 3, but the existence of two basic *blues* in these languages that split the unitary blue category in English (see also Newton, 1730) challenges physiological as well as ecological explanations for the basis of the blue mechanism (Kuehni, 2005; Regier et al., 2007; Philipona & O'Regan, 2006; Mollon, 2006). Furthermore, the colour differences between the primary basic terms across languages were considerable larger ($ΔE_{00}$ = 6.25) than differences between secondary basic terms ($ΔE_{00}$ = 3.31) and there is no evidence that these so-called landmark colours are acting as such (Boynton & Olson, 1987; Steels & Belpaeme, 2005).

Similarly, our examination in Chapter 4 on the coherence of classes of colours showed that the six members of the primary class (*white, black, red, green, yellow* and *blue*) produced a similar coherence score with an equally sized class of secondary basics colours (*brown, orange, purple, pink* and *grey* plus one of Hering's primaries). These findings are consistent with recent studies (Malkoc, et al., 2005; Bosten & Boehm, 2014) that found no differences between unique-hue judgments of non-primary (i.e., *teal, orange, purple* and *lime*) and primary hues (i.e. *red, yellow, green and blue*). Hue cancelation procedures may be used to derive colour opponent mechansisms as charted quantitatively by Hurvich and Jameson (1957) but say little about which colour categories play a fundamental role in the development of colour naming systems. The relationship between peak wavelength sensitivity of cells and psychophysical colour opponency is also too loose to account for colour appearance in terms of the physiological one (De Valois et al., 1966, Derington et al., 1984; Valberg, 2001; for a review). Our findings indicate that primaries are not a completely haphazard class but are not more coherent than classes of secondary colours; consistent with the results of earlier studies in adults (Boynton & Olson, 1987), infants (Franklin, et al., 2008) and monkeys (Zeki, 1980).

Collectively, this thesis provide evidence against Hering's primaries playing a fundamental role in the development of colour categories and challenge explanations based on this claim (Berlin & Kay, 1969/1991; Kay & MacDaniel, 1978; Kuehni, 2005; Philipona & O'Regan, 2006; Regier et al., 2007). In agreement with an earlier study (Bosten & Boehm, 2014), we argue that unique hue settings should be accepted for what they are; examples in lexical colour categories that have the least trace of the anchor colour names. Accordingly, we claim that unconstrained colour naming procedures

provide a better estimation for the location of the opponent mechanisms than constrained hue cancelation tasks and suggest that considering three bipolar (*red-turquoise, green-magenta and blue-yellow*) or six unipolar chromatic mechanisms (*red, turquoise, green, magenta, blue* and *yellow*) that produce complementary relationships between the opponent pairs would better account for colour appearance phenomena than the primaries of the opponent theory (Chevreul, 1839; Helmholtz, 1852; Hurvich & Jameson, 1957; Pridmore, 2008).

8.5.   Colour naming and colour constancy

Colours named in both online- and offline- experiments were perceived relative to their spatial context, and the overall good correspondence between BCTs across languages and in experiments conducted in variable viewing conditions could be explained by their similar stimulus and background configuration. Light emitted from each test stimulus was compared with the light emitted from the surrounding neutral grey background and the visual system of the observers either adapted to the background chromaticity or a ratio between the two was taken to assign a constant colour name to each test sample. The net result of these operations, which cannot be separated given the experimental set up of this thesis, is that colour perception becomes largely independent of the uncontrolled viewing conditions and the uncalibrated monitors of the online experiment, thus leading to a perceptual stabilisation of colours. This stabilisation of colour can be due to short-range, long-time-course adaptation, in the receptors or due to nearly instant, long-range, ratio-taking operations in the cortex or, most likely due to a combination of both processes (von Kries, 1905; Land, 1974; Zeki, 1980; Brill & West, 1986; Fairchild & Lennie 1992; Fairchild & Reniff 1995; Rinner & Gegenfurtner, 2000, 2002). Understanding what retinal information is being used by the visual system to stabilise colours in the cortex and then to connect it with the first and second mechanistic sites of adaptation is of critical importance but only partially developed research (Uchikawa et al., 1989; Zaidi, 1998; Golz & MacLeod, 2002; Stockman & Brainard, 2010).

It is difficult to generalise findings based on spatially uniform backgrounds used in the experiments of this thesis to account for natural scenes that have complex spatial structure (Stockman & Brainard, 2010; Conway, Eskew, Martin, Stockman, 2018). To investigate colour categorization in complex scenes, we asked in a parallel study whether subjects of different linguistic and ethnic backgrounds categorize different colours when light reflected from patches of different colour reflect the same wavelength-energy composition in a Land Colour Mondrian display (Figure 1 in Zeki et al., 2019). For each test colour sample, we adjusted the amount of long, middle and short-wave light reflected

from each patch to a constant ratio of 60% long-, 20% middle- and 20% short- wave light. Participants were asked to match the colour of the eight test patches in the Mondrian display with one of 44 colour chips from the Munsell Book of Color (Figure 2 in Zeki et al., 2010). Our experimental set up replicated Land's (1974) experiment but we were interested in the colour category instead of hue, to which matches were made. Similar to the results presented in Chapter 7 of this thesis, the variability in the responses was lower for *reddish* than *bluish* hues reflecting the smaller perceptual extents of categories in the warm region rather than in the cool region of colour space (Figure 3 in Zeki et al., 2019; Berlin & Kay, 1969/1991; Mylonas & MacDonald, 2016; Gibson et al., 2017). In terms of colour categories, except for *blue*, we found very little variability across subjects and colours, consistent with the results presented here in Chapter 3 and with the strong correlation between naming consistency across illuminants and across observers reported by Olkkonen et al., (2009). Collectively, these results suggest a strong link between categorical colour constancy and consistent colour communication.

To examine whether chromatic adaptation or spatial ratio-taking operations can explain colour constancy in a second parallel study, we asked participants to name and match the after-image colour of central patches in Land Colour Mondrian displays that reflected the same wavelength-energy composition but appeared as different colours (Figure 1 in Zeki et al., 2017). For example, would a surface that reflects more long-wave light but appears *green* in a complex Mondrian scene would result in a *green* as chromatic adaptation model would predict or a *red* after-image as a spatial computational model would predict? Our results showed that the colour of the after-images belonged to the family of colours complementary to the perceived colour of the viewed patches (Figures 3 and 4 in Zeki et al., 2017). Therefore, the colour of the after-image – like the colour itself – depended on the ratio of light of different wavebands reflected from each test patch and their surrounds. In addition, the uniformity of the colour naming responses for the after-images followed again the relative size of these terms in our colour naming experiment presented in Chapter 3. Overall, the findings of both studies using complex spatial scenes demonstrated the close linkage between long-range spatial ratio-taking operations in the cortex and colour categorisation (Zeki et al., 2017, 2019).

These cortical ratio-taking operations have been described earlier as colour constancy (Land, 1974; Zeki, 1980; Brill & West, 1986; Foster, 2011); but describing the end product as *constant colour category* is preferable because what does not change as a result is in fact the colour category, not the hue – the latter changes when surfaces are viewed in different viewing conditions (Zeki et al., 2017, 2019). In a proposed cognitive architecture (MacDonald & Mylonas, 2016), these incoming constant colour categories are generated

in V4 and compared to stored categorical object colours in long-term memory by a parallel processing network for the identification of the colour. For a verbal response, the cognitive match recalls the underlying colour name for articulation. In accordance with Hunt (2004), "the basis of judgement is usually a comparison between the colour perceptions aroused by the reproduction, and a mental recollection of the colour perceptions previously experienced when looking at similar objects". In other words, we argue that while the process of colour categorisation is an inherent mechanism, the stored lexical colour categories are acquired through learning. Our claim is further supported by the overall good correspondence of colour terms loci across languages presented in Chapter 3 and 5, while the number of colour names in each language varied. In agreement with recent studies, our results support a reconciliation between the opposite views that called into question whether colour categories are formed under the influence of perceptual mechanisms, or whether language influences the structure of colour categories (Regier & Kay, 2009; Kemp & Regier, 2012; Gibson et al., 2017). Contrary to recent suggestions (Regier & Kay, 2009; Witzel, 2018), however, we believe that this debate is not-resolved yet and can be still fruitful in advancing our understanding on inherited and acquired colour mechanisms (Zeki, 2009; Zeki et al., 2019).

8.6. Computational colour naming models

In our evaluation of several supervised nonparametric colour naming models using cross-validation, a Rotated Split Trees (RST) approach performed best. RST chooses each attribute and split at random (Geurts et al., 2006) while the random rotation of its decision space infuse diversity within the constructed forest and improves its accuracy at determining colour categories in a three-dimensional space (Blaser & Fryzlewicz, 2016; Andrews, Jaccard, Rogers & Griffin, 2017). When we trained the RST model with colour naming responses in different languages the output of the model showed universal patterns, but these were not without language-specific differences. For example, the predicted categories were consistent with earlier reports on the existence of a second blue basic term in Greek, Russian and Thai while this category was absent when the model was trained by the British, American and International English datasets (Androulaki et al., 2006; Athanasopoulos, 2009; Corbett & Morgan, 1988; Moss, 1988; Morgan & Corbett, 1989; Paramei, 2005; Paramei et al. 2017; Prasithrathsint, 1988; Engchuan, 2003; Mylonas & MacDonald, 2016). Furthermore, except in Thai, the model predicted a well-formed, non-basic, turquoise category in all other colour naming datasets in agreement with our earlier study on the importance of this term in colour naming (Mylonas & MacDonald, 2016).

The evaluation of RST in several colour spaces showed that the model performed best in the approximately uniform colour space of CIELUV while its performance deteriorated in the non-uniform CIE XYZ and RGB spaces. This is in agreement with a recent study that suggested CIELUV as the best space for performing colour clustering algorithms (Douven, 2017). It also implies that our colour naming model performed best when connected to the output of second stage mechanisms, rather than the output of first stage mechanisms. In this thesis, we considered only the classification of single colours viewed against a uniform grey background. For the automatic assignment of colours to names in complex images, an earlier study showed improvements when their colour naming model was connected with the output of a Retinex type of algorithm that models long-range spatial operations (Benavente, 2006); while more recent approaches fuse colour naming descriptors with higher-level shape information and colour attention representations (Khan et al., 2012; Weijer et al., 2013).

Our comparison of the performance of RST on predicting the distribution of the 11 BCTs on the surface of the Munsell array against existing computational colour naming models (Lammen, 1994; MacLaury; 1992; Benavente & Vanrell, 2004; Seaborn, 2005; Benavente et al, 2008; Weijer et al., 2007; Mylonas et al., 2010; Parrage & Akbarinia, 2016) showed that RST achieved state-of-the-art performance for the psychophysical results of Sturges & Whitfield (1995); while identifying 16 categories in total (11 BCTs + *turquoise, teal, lilac, mauve* and *maroon*). It is important to note that excepting our earlier MAP model (Mylonas et al., 2010), all other models constrained their predictions only to the 11 BCTs. The strong assumption that BCTs can name all colours is not supported by empirical findings (Boynton & Olson, 1987; Sturges & Whitfield, 1995; Mylonas & MacDonald, 2016). However, the predictions of RST were equally good with these models and in many cases better without this constrain. We argue that the use of the surface colours of the Munsell array and the focus on the distribution of the 11 BCTs is of limited value and out-of-date for comparing the performance of modern computational colour naming models. People use in their native language 30-50 colour names without training to describe both saturated colours and also paler colours in the interior of the colour space (Chapanis, 1965; Derefeldt & Swartling, 1995). We suggest instead the use of a synthetic image described in Chapter 5 and 6 to present the performance of computational models trained by unconstrained colour naming responses (see also Griffin & Mylonas, 2019).

8.7. Information theory and Basic Colour Terms

Existing methods for the identification of BCTs within different languages require multiple criteria and combinations of associated measures (Berlin & Kay, 1969/1991; Boynton & Olson, 1985; Corbet & Davies 1997; Lindsey & Brown, 2014; Mylonas & MacDonald, 2016). These criteria have been strongly criticised as not being equally applicable across languages, while their multiplicity is vulnerable to high risk of being applied differently by different researchers (Saunders & van Brakel, 1997; Levinson, 2000; Biggam, 2012; Witzel, 2018). In Chapter 3, we confirmed earlier studies that no single lexical, behavioural, or geometric feature available from online colour naming data was sufficient to identify BCTs from non-basic colour terms. This was also the case for the calibrated data from our laboratory-based experiment described in Chapter 6. Furthermore, in Chapter 4, the well-established Random Forest classifier (Breiman, 2001) required training by more than one family of these features to produce perfect coherence between members of the Basic class (Berlin & Kay, 1969/1991). Therefore, the quest for a simple and a cross-culturally legitimate approach for demarcating BCTs from non-basics was unsettled (Witzel, 2018).

In our view, basicness refers to the degree that linguistic signifiers are shared and comprehended by most speakers in each language to communicate their categories of colour sensations. Accordingly, in Chapter 7, we claimed that basic colour categories are indispensable and proposed a simple information theoretic measure of basicness that we call dispensability. The evaluation of our measure using unconstrained colour names in British and American English, Greek, Russian, Thai and Turkish from an online experiment along with a calibrated English dataset from a laboratory-based experiment showed that dispensability varied with category and produced a graded scale of basicness. Critically, for all three datasets in English the 11 BCTs (Berlin & Kay, 1969/1991) had lower dispensability scores than all non-BCTs, while the metric was also able to capture the indispensability of the proposed second *blue* basic term in Greek, Russian, Thai as well as the uncertainty about the basicness of the second *blue* term in Turkish (Androulaki et al., 2006; Paramei et al. 2017; Prasithrathsint, 1988; Ozgen & Davies, 1998). In Chapter 3, we showed that the loci of the two basic *blue* terms in Greek, Russian and Thai corresponded well, and their centroids deviated from the centroid of English *blue*, but this was not the case in Turkish. These findings support the postulation of the existence of an additional evolutionary Stage VIII in the development of colour lexicons (Berlin & Kay, 1969/1991) for Greek, Russian and Thai but further research is needed for the second *blue* term Turkish.

In British English, the 11 BCTs were followed closely by *turquoise* and *lilac* but there was a considerable jump in dispensability score to the following non-basic term, *beige*. This is in agreement with our recent study where we found that these terms are strong candidates to be included in the English basic inventory based on an aggregated scale of six measures (Mylonas & MacDonald, 2016). Both terms appear to reduce the uncertainty of colour naming from using only the 11 BCTs as *lilac* partitions the large colour category of *purple* in light and dark segments while *turquoise* stabilises the large boundary area between *green* and *blue*. In American English, the last BCT *green* was followed firmly by *peach*, *salmon*, and *maroon* while the separation was bigger with the 15th, *dark green*. Lindsey & Brown (2014) applied Zipf's law (1935) to colour naming frequencies and reported a steep decrease in frequency of colour terms beyond the 15th term in American English colour lexicon. They also found that *peach*, *teal*, *lavender* and *maroon* were named with high consensus across observers. In the international English dataset of the laboratory-based experiment, we found a steep step between the 11 BCTs and the non-basics *lime green, beige, dark green, lilac, dark brown, dark red, turquoise* and *cream*. These findings show that while the 11 BCTs are universally shared between English speakers, the following indispensable colour names may vary.

Our dispensability measure does not identify the BCTs because they are the maximally spaced in colour space, nor because they are commonly used (Regier et al., 2007; Lindsey & Brown, 2014); but instead combines the frequency and consensus of colour names between speakers in each language in an informative way to identify which are the superordinate colour names that cannot be replaced with any other name – hence the title we have given to our measure.

Considering the hierarchical order of the BCTs at each evolutionary stage (Berlin &Kay, 1969/1991), the ranks of the most indispensable colour names varied in each language. *White* and *black* were often ranked in the top positions but this was also true for *purple*, *orange*, *brown* and *pink*. *Red* was never found earlier than the 6th position and *green* more often occupied the last positions of the BCTs. The low indispensability score of *orange, pink* and *brown* cannot be predicted by the unevenness in terms of saturation of the colour space. Consistent with our earlier findings previously discussed, we found no evidence for the priority of primary basics over secondary basics in terms of dispensability. Instead, our results support growing evidence that communication efficiency provides a better framework to understand colour naming than opponent theory (Jameson & D' Andrade, 1997; Lindsey et al., 2015; Regier et al., 2015; Abbot et al., 2016; Gibson et al., 2017). In closing, our findings suggest that information theory is

well-suited to provide a simple language-independent measure to determine the degree of basicness of unconstrained colour names within different languages.

Chapter 9

Conclusions

The aim of this thesis is to advance the field of colour communication within different languages. First, we reviewed previous work on colour vision and its relationship to colour naming. We also surveyed a range of colour spaces and earlier colour naming experiments, models and information theoretic approaches on which our work is based on. Second, we showed that large colour lexicons in different languages can be crowdsourced through an online colour naming experiment. Third, we employed classification theory to access the coherence of achromatic, primary and basic classes of colours based on their linguistic, behavioural and geometric features. Fourth, we evaluated a range of computational models trained by speakers of different languages of the colour naming experiment to automate the assignment of colour names across the full three-dimensional gamut. Fifth, we compared the findings between online- and laboratory- based colour naming experiments and mapped colour names in a physiologically-based cone excitation space. Finally, instead of multiple conceptual criteria and combination of variable measures, we proposed a novel information theoretic measure – called dispensability – to identify basic colour terms from unconstrained colour naming data across languages. The main findings of this thesis are given below:

- Online experimental methodologies offer considerable advantages over traditional approaches to obtain unconstrained colour naming responses in different languages from thousands of observers. The location of colour categories corresponds well between web- and laboratory- based experiments and our findings support the validity of both methods in estimating colour naming functions in controlled and real-world monitor settings.

- The application of machine learning methods to discover criteria in linguistic, behavioural and geometric features of colour names and to distinguish classes of colours, showed that achromatic and basic colours are coherent classes but not primaries. These results reinforce the ongoing argument against the special role of primaries in the formation of colour categories.

- An evaluation of several computational colour naming models and colour spaces using cross-validation showed that a Rotated Split Trees (RST) approach applied

in CIELUV produced the best performance. A comparison against earlier methods showed that our approach achieves state-of-the-art performance while it identifies five additional categories on the surface of the Munsell system. Our tools and data in operation were able to demonstrate the automation of the colour naming task across the full three-dimensional colour gamut in different languages.

- A laboratory-based colour naming experiment using a calibrated CRT monitor allowed us to accurately map for the first-time unconstrained colour names in the physiologically-based cone excitation space adopted recently by the CIE. In cone chromaticity diagram, *yellow* was colinear through *white* with *blue*, *red* was complementary to *turquoise*, and *green* to *magenta*. The regions with the highest purity of L, M and S cone excitations were assigned to *red*, *green* and *blue* colour names respectively while he combined signals from L and M cones at maximum intensity were assigned to *yellow*, from L and S cones to *magenta* and from M and S cones to *turquoise*.

- The dispensability measure produced a graded scale of basicness where basic colour categories within different languages and experimental methods had lower scores than non-basics. These findings suggest that information theory is pertinent to provide a simple language-independent measure to determine the degree of basicness of unconstrained colour names within different languages and reveal the presence of additional basic categories. Our results confirm the existence of two basic *blue* terms in Greek, Russian and Thai.

9.1. Future work

There are many ways to further improve colour communication within different languages. Above all, in the collection and analysis of unconstrained naming responses in additional languages for colours on the surface but also in the interior of the colour space. The recently redesigned interface of the online colour naming experiment has been already translated in 13 languages and gathered more than 100,000 responses in the first year of operation. Also, the design of a simple offline interface for collecting vocal colour naming responses using calibrated monitors, offers the opportunity to obtain responses in fieldwork and or in controlled laboratory conditions. Cross-cultural comparisons could shed more light on the relationship between perception and language.

The work described in the present thesis on the mechanistic linkage between how people perceive colours and how they communicate their perceptions through language created more questions than answers. What is certain is that mapping linguistic to perceptual aspects of colour will allow us to augment colour communication within and across different languages and between humans and machines. The implication of increasing the capability of humans to articulate and comprehend colour percepts with technological aids can enhance the understanding of our and other people's perceptions and bring us together to make the world a more colourful place.

References

Abbott, J. T., Griffiths, T. L., & Regier, T. (2016). Focal colors across languages are representative members of color categories. *Proceedings of the National Academy of Sciences*, *113*(40), 11178–11183. https://doi.org/10.1073/pnas.1513298113

Abramov, I., & Gordon, J. (1994). Color Appearance: on Seeing Red-or Yellow, or Green, or Blue. *Annual Review of Psychology*, *45*(1), 451–485. https://doi.org/10.1146/annurev.ps.45.020194.002315

Andrews, J. T. A., Jaccard, N., Rogers, T. W., & Griffin, L. D. (2017). Representation-learning for anomaly detection in complex x-ray cargo imagery. *Anomaly Detection and Imaging with X-Rays (ADIX) II*, *10187*, 101870E. https://doi.org/10.1117/12.2261101

Androulaki, A, Gômez-Pestañ, N., Mitsakis, C., Jover, J., Coventry, K., Davies, I. (2006). Basic colour terms in Modern Greek: Twelve terms including two blues. *Journal of Greek Linguistics*, *7*, 3–47.

Aristotle. (350 B.C.E.). *On Sense and the Sensible*. (J. I. Beare, Trans.). Adelaide: The University of Adelaide Library.

Athanasopoulos, P. (2009). Cognitive representation of colour in bilinguals: The case of Greek blues. *Bilingualism: Language and Cognition*, *12*(01), 83–95. https://doi.org/10.1017/S136672890800388X

Barbur, J. L., Harlow, J., & Plant, G. T. (1994). Insights into the different exploits of colour in the visual cortex. In *Proceedings of the Royal Society London. B.* (Vol. 258, pp. 327–334).

Basser, P. J., & Pierpaoli, C. (2011). Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *Journal of Magnetic Resonance*, *213*(2), 560–570. https://doi.org/10.1016/j.jmr.2011.09.022

Belpaeme, T., & Bleys, J. (2005). Explaining Universal Color Categories Through a Constrained Acquisition Process. *Adaptive Behavior*, *13*(4), 293–310. https://doi.org/10.1177/105971230501300404

Benavente, R., Vanrell, M., & Baldrich, R. (2006). A data set for fuzzy colour naming. *Color Research & Application*, *31*(1), 48–56. https://doi.org/10.1002/col.20172

Benavente, R., Vanrell, M., & Baldrich, R. (2008). Parametric fuzzy sets for automatic color naming. *Journal of the Optical Society of America A*, *25*(10), 2582–2593. https://doi.org/10.1364/JOSAA.25.002582

Berk, T., Brownston, L., & Kaufman, A. (1982). A New Color-Naming System for Graphics Languages. *Computer Graphics and Applications, IEEE*, *2*(3), 37–44.

Berns, R. S. (2000). *Billmeyer and Saltzman's Principles of Color Technology, 3rd Edition* (3rd ed.). Wiley-Interscience.

Berlin, B., & Kay, P. (1969/1991). *Basic color terms : their universality and evolution*. Stanford, Calif.: Center for the Study of Language and Information.

Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society*, *35*, 99–109.

Bichard, J.-A., & Gheerawo, R. (2011). The ethnography in design. In A. J. Clarke (Ed.), *Design Anthropology: Object Culture in the 21st Century* (pp. 45–55). https://doi.org/10.1007/978-3-7091-0234-3_4

Biggam, C. (2012). *The Semantics of Colour: A Historical Approach*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139051491

Bimler, D. (2005). Are color categories innate or internalized? Hypotheses and implications. *Journal of Cognition and Culture*, *5*(3), 265–292. https://doi.org/10.1163/156853705774648518

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. Beijing ; Cambridge Mass.: O'Reilly Media.

Bird, C. M., Berens, S. C., Horner, A. J., & Franklin, A. (2014). Categorical encoding of color in the brain. *Proceedings of the National Academy of Sciences*, 201315275. https://doi.org/10.1073/pnas.1315275111

Blaser, R., & Fryzlewicz, P. (2016). Random Rotation Ensembles. *Journal of Machine Learning Research*, *17*(4), 1–26.

Boardman, J. (1976). The olive in the Mediterranean: its culture and use. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*. https://doi.org/10.1098/rstb.1976.0080

Boring, E. G. (1949). I. - A color solid in four dimensions. *L'année Psychologique*, *50*(1), 293–304. https://doi.org/10.3406/psy.1949.8454

Bosten, J. M., & Boehm, A. E. (2014). Empirical evidence for unique hues? *Journal of the Optical Society of America A*, *31*(4), A385. https://doi.org/10.1364/JOSAA.31.00A385

Boynton, R. M., & Olson, C. X. (1987). Locating basic colors in the OSA space. *Color Research & Application*, *12*(2), 94–105. https://doi.org/10.1002/col.5080120209

Brakel, J. V. (1993). The Plasticity of Categories: The Case of Colour. *British Journal for the Philosophy of Science*, *44*(1), 103–135.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and Regression Trees*. Taylor & Francis.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Brill, M. H., & West, G. (1986). Chromatic adaptation and color constancy: A possible dichotomy. *Color Research & Application*, *11*(3), 196–204. https://doi.org/10.1002/col.5080110306

Brouwer, G. J., & Heeger, D. J. (2013). Categorical Clustering of the Neural Representation of Color. *The Journal of Neuroscience*, *33*(39), 15454–15465. https://doi.org/10.1523/JNEUROSCI.2472-13.2013

Brown, R. W., & Lenneberg, E. H. (1954). A study in language and cognition. *The Journal of Abnormal and Social Psychology*, *49*(3), 454–462. https://doi.org/10.1037/h0057814

Buchsbaum, G., & Bloch, O. (2002). Color categories revealed by non-negative matrix factorization of Munsell color spectra. *Vision Research*, *42*(5), 559–563. https://doi.org/10.1016/S0042-6989(01)00303-0

Cao, D., Pokorny, J., & Smith, V. C. (2005). Associating color appearance with the cone chromaticity space. *Vision Research*, *45*(15), 1929–1934. https://doi.org/10.1016/j.visres.2005.01.033

Chapanis, A. (1965). Color names for color space. *American Scientist*, *53*, 327–346.

Cheng, Z., Li, X., & Loy, C. (2016). Pedestrian Color Naming via Convolutional Neural Network. In S.-H. Lai, V. Lepetit, K. Nishino, & Y. Sato (Eds.), *Computer Vision – ACCV 2016*. Taipei, Taiwan: Springer.

Chevreul, M. E. (1839). *De la loi du contraste simultané des couleurs*. Paris: Pitois-Levraux.

Chichilnisky, E. J., & Wandell, B. A. (1999). Trichromatic opponent color classification. *Vision Research*, *39*(20), 3444–3458.

Chuang, J., Hanrahan, P., & Stone, M. (2008). A Probabilistic Model of the Categorical Association Between Colors (pp. 6–11). In *16th Color Imaging Conference*, Portland, Oregon: IS&T.

CIE-TC 1-36. (2006). *CIE 170-1:2006 Fundamental Chromaticity Diagram With Physiological Axes* (No. ISBN: 978 3 901906 46 6).

CIE-TC 1-36. (2015). *CIE 170-2:2015 Fundamental Chromaticity Diagram with Physiological Axes – Part 2: Spectral Luminous Efficiency Functions and Chromaticity Diagrams* (No. ISBN: 978-3-902842-05-3).

Conway, B. R., & Tsao, D. Y. (2006). Color Architecture in Alert Macaque Cortex Revealed by fMRI. *Cerebral Cortex*, *16*(11), 1604–1613. https://doi.org/10.1093/cercor/bhj099

Conway, B. R., Eskew, R. T., Martin, P. R., & Stockman, A. (2018). A tour of contemporary color vision research. *Vision Research*, *151*, 2–6. https://doi.org/10.1016/j.visres.2018.06.009

Corbett, G. G., & Davies, I. R. L. (1997). Establishing Basic Colour Terms: Measures and Techniques. In C. L. Hardin & L. Maffi (Eds.), *Color Categories in Thought and Language* (pp. 197–223). Cambridge: Cambridge University Press.

Crawford, T. D. (1982). Defining "Basic Color Term." *Anthropological Linguistics*, *24*(3), 338–343.

Cutler, A., & Zhao, G. (2001). PERT - perfect random tree ensembles. *Computing Science and Statistics*, 497.

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random Forests for Classification in Ecology. *Ecology*, *88*(11), 2783–2792. https://doi.org/10.1890/07-0539.1

Damasio, A. R., McKee, J., & Damasio, H. (1979). Determinants of performance in color anomia. *Brain and Language*, *7*(1), 74–85. https://doi.org/10.1016/0093-934X(79)90007-5

D'Andrade, R. G., & Romney, A. K. (2003). A quantitative model for transforming reflectance spectra into the Munsell color space using cone sensitivity functions and opponent process weights. *Proceedings of the National Academy of Sciences*, *100*(10), 6281–6286. https://doi.org/10.1073/pnas.1031827100

Davidoff, J. (1996). Lewandowsky's Case of Object-Colour Agnosia. In *Classic Cases in Neuropsychology* (pp. 145–158). Psychology Press.

Davidoff, J., Davies, I., & Roberson, D. (1999). Colour categories in a stone-age tribe. *Nature*, *398*(6724), 203–204. https://doi.org/10.1038/18335

Davidoff, J. (2015). Color categorization across cultures. In A. J. Elliott, M. D. Fairchild, & A. Franklin (Eds.), *Handbook of Color Psychology* (pp. 259–278). Cambridge: Cambridge University Press.

Davies, I. L., & Corbett, G. G. (1995). A practical field method for identifying probable basic colour terms. *Languages of the World*, *9*(1), 25–36.

Davies, I. R. L., & Corbett, G. G. (1994). A Statistical Approach to Determining Basic Color Terms: An Account of Xhosa. *Journal of Linguistic Anthropology*, *4*(2), 175–193. https://doi.org/10.1525/jlin.1994.4.2.175

De Valois, R. L., Abramov, I., & Jacobs, G. H. (1966). Analysis of Response Patterns of LGN Cells. *Journal of the Optical Society of America*, *56*(7), 966–977. https://doi.org/10.1364/JOSA.56.000966

De Valois, R. L., & De Valois, K. K. (1993). A multi-stage color model. *Vision Research*, *33*(8), 1053–1065. https://doi.org/16/0042-6989(93)90240-W

Dedrick, D. (1998). *Naming the rainbow : colour language, colour science, and culture*. Dordrecht; Boston: Kluwer.

Derefeldt, G., & Swartling, T. (1995). Colour concept retrieval by free colour naming. Identification of up to 30 colours without training. *Displays*, *16*(2), 69–77. https://doi.org/10.1016/0141-9382(95)91176-3

Derefeldt, G., Swartling, T., Berggrund, U., & Bodrogi, P. (2004). Cognitive color. *Color Research & Application*, *29*(1), 7–19. https://doi.org/10.1002/col.10209

Derhak, M. W., & Berns, R. S. (2012). Analysis and Correction of the Joensuu Munsell Glossy Spectral Database. *Color and Imaging Conference*, *2012*(1), 191–194.

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

Dimmick, F. L. (1925). Discussion: The Series of Blacks, Grays, and Whites. *Psychological Review*, *32*(4), 334–336.

Douven, I. (2017). Clustering colors. *Cognitive Systems Research*, *45*, 70–81. https://doi.org/10.1016/j.cogsys.2017.05.004

Ekici- Şahin, E., Yener, C., & Camgöz, N. (2006). Colour naming. *Optics & Laser Technology*, *38*(4–6), 466–485. https://doi.org/10.1016/j.optlastec.2005.06.007

Engchuan, S. (2003). *Color terms and the concept of color of the Thais in Sukhotai period and at the present time*. In *AIC 2003 Midterm Meeting*, Bangkok, Thailand.

Eskew Jr., R. T. (2009). Higher order color mechanisms: A critical review. *Vision Research*, *49*(22), 2686–2704. https://doi.org/10.1016/j.visres.2009.07.005

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*(5), 429–448. https://doi.org/10.1017/S0140525X0999094X

Fairchild, M. D., & Lennie, P. (1992). Chromatic adaptation to natural and incandescent illuminants. *Vision Research*, *32*(11), 2077–2085. https://doi.org/10.1016/0042-6989(92)90069-U

Fairchild, M. D., & Reniff, L. (1995). Time course of chromatic adaptation for color-appearance judgments. *Journal of the Optical Society of America A*, *12*(5), 824–833. https://doi.org/10.1364/JOSAA.12.000824

Fairchild, M. D. (2005). Color Appearance Models (2nd ed.). Wiley.

Fletcher, R. (1978). Recent experiences with the City spot test. *Modern Problems in Ophthalmology*, *19*, 142–143.

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

Franklin, A., Pitchford, N., Hart, L., Davies, I. R. L., Clausse, S., & Jennings, S. (2008). Salience of primary and secondary colours in infancy. *British Journal of Developmental Psychology*, *26*(4), 471–483. https://doi.org/10.1348/026151007X256672

Gärdenfors, P. (2004). *Conceptual Spaces: The Geometry of Thought*. Cambridge, Mass.: MIT Press.

Gage, J. (1993). *Color and culture : practice and meaning from antiquity to abstraction*. Boston: Little, Brown and Company.

Gegenfurtner, K. R., & Kiper, D. C. (1992). Contrast detection in luminance and chromatic noise. *Journal of the Optical Society of America. A, Optics and Image Science, 9*(11), 1880–1888.

Gegenfurtner, K. R., Kiper, D. C., & Fenstemaker, S. B. (1996). Processing of color, form, and motion in macaque area V2. *Visual Neuroscience*, *13*(01), 161–172. https://doi.org/10.1017/S0952523800007203

Gegenfurtner, K. R., & Kiper, D. C. (2003). Color vision. *Annual Review of Neuroscience*, *26*(1), 181–206. https://doi.org/10.1146/annurev.neuro.26.041002.131116

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, *63*(1), 3–42. https://doi.org/10.1007/s10994-006-6226-1

Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., Gibson, M., Piantadosi, S. T., Conway, B. R. (2017). Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences*, 201619666. https://doi.org/10.1073/pnas.1619666114

Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random Forests for land cover classification. *Pattern Recognition Letters*, *27*(4), 294–300. https://doi.org/10.1016/j.patrec.2005.08.011

Goethe, J. W. von. (1840). *Goethe's Theory of Colours* (C. L. Eastlake, Trans.). J. Murray.

Golz, J., & MacLeod, D. I. A. (2002). Influence of scene statistics on colour constancy. *Nature, 415*(6872), 637. https://doi.org/10.1038/415637a

Golz, J., & MacLeod, D. I. A. (2003). Colorimetry for CRT displays. *JOSA A*, *20*(5), 769–781. https://doi.org/10.1364/JOSAA.20.000769

Grassmann, H. (1853). Zur Theorie der Farbenmischung. *Annalen Der Physik*, *165*(5), 69–84. https://doi.org/10.1002/andp.18531650505

Green, P. (2002). Colorimetry and colour difference. In P. Green & L. MacDonald (Eds.), *Colour Engineering: Achieving Device Independent Colour* (pp. 49–77). UK: Wiley.

Griber Y., Paramei G.V., Mylonas, D. (2017). Gender Differences in Russian Colour Naming. In *13th AIC Congress*, South Korea.

Griber, Y. A., Mylonas, D., & Paramei, G. V. (2018). Objects as culture-specific referents of color terms in Russian. *Color Research & Application*, *43*(6), 958–975. https://doi.org/10.1002/col.22280

Griffin, L. D. (2001). Similarity of psychological and physical colour space shown by symmetry analysis. *Color Research & Application*, *26*(2), 151–157. https://doi.org/10.1002/1520-6378

Griffin, L. D. (2006). Optimality of the basic colour categories for classification. *Journal of The Royal Society Interface*, *3*(6), 71–85. https://doi.org/10.1098/rsif.2005.0076

Griffin L.D., Mylonas, D. (2016). Why so much talk about "red"? In *AVA Christmas Meeting 2016*. Queen Mary University of London, London, UK.

Griffin, L. D., & Mylonas, D. (2019). Categorical colour geometry. *PLOS ONE, 14*(5), e0216296. https://doi.org/10.1371/journal.pone.0216296

Guest, S., & Laar, D. V. (2000). The structure of colour naming space. *Vision Research*, *40*(7), 723–734. https://doi.org/10.1016/S0042-6989(99)00221-7

Hansen, T., & Gegenfurtner, K. R. (2006). Higher level chromatic mechanisms for image segmentation. *Journal of Vision*, *6*(3), 5–5. https://doi.org/10.1167/6.3.5

Hardin, C. L. (2005). Explaining Basic Color Categories. *Cross-Cultural Research*, *39*(1), 72–87. https://doi.org/10.1177/1069397104267891

Hardin, C. L., & Maffi, L. (1997). Introduction. In C. L. Hardin & L. Maffi (Eds.), *Color Categories in Thought and Language*. Cambridge University Press.

Harnad, S. (1987). Psychophysical and cognitive aspects of categorical perception: A critical overview. In S. Harnad (Ed.), *Categorical Perception: The Groundwork of Cognition* (pp. 1–52). Cambridge University Press.

Hays, D. G., Margolis, E., Naroll, R., & Perkins, D. R. (1972). Color Term Salience. *American Anthropologist*, *74*(5), 1107–1121. https://doi.org/10.1525/aa.1972.74.5.02a00050

Heer, J., & Stone, M. (2012). Color naming models for color selection, image editing and palette design. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems* (pp. 1007–1016). New York, NY, USA: ACM. https://doi.org/10.1145/2208516.2208547

Helmholtz, H. von. (1911). *Handbuch der physiologischen Optik* (3rd ed., Vol. 2). Hamburg: Voss.

Hering, E. (1878/1964). *Outlines of a theory of the light sense*. (L. M. Hurvich & D. Jameson, Trans.). Cambridge, Mass.: Harvard University Press.

Householder, A. S. (1958). Unitary Triangularization of a Nonsymmetric Matrix. *Journal Association Computing Machinery (ACM)*, *5*(4), 339–342. https://doi.org/10.1145/320941.320947

Hubel, D. (1988). Color Vision. In *Eye, Brain, and Vision.* New York: W. H. Freeman & Company.

Hunt, R. W. G. (2004). *The Reproduction of Colour* (6th ed.). Wiley, Chichester, UK, pp. 32–33.

Hurvich, L. M., & Jameson, D. (1957). An opponent-process theory of color vision. *Psychological Review, 64, Part 1*(6), 384–404.

IEC. (1999). *Multimedia systems and equipment - Colour measurement and management - Part 2-1: Colour management - Default RGB colour space - sRGB".* IEC 61966-2-1.

Jameson, D. (1983). Some Misunderstandings about Color Perception, Color Mixture and Color Measurement. *Leonardo, 16*(1), 41–42.

Jameson, K. (2010). Where in the World Color Survey is the Support for Color Categorization Based on the Hering Primaries. In J. D. Cohen & M. Matthen (Eds.), *Color Ontology and Color Science* (pp. 179–202). MIT Press.

Jameson, K., & D'Andrade, R. (1997). It's not really red, green, yellow, blue: an inquiry into perceptual color space. In C. L. Hardin & L. Maffi (Eds.), *Color Categories in Thought and Language* (pp. 197–223). Cambridge: Cambridge University Press.

Judd, D. B. (1949). Response Functions for Types of Vision According to the Müller Theory, Research Paper RP1946. *Journal of Research of the National Bureau of Standards, 42*, 357–371.

Kaiser, P. K., & Boynton, R. M. (1996). *Human color vision.* Washington, DC: Optical Society of America

Kang, H. R. (2006). *Computational Color Technology.* Bellingham, Wash: SPIE Press.

Katemake, P., Mylonas, D. MacDonald, L., Prasithrathsint, A. (2015). Comparison Among Three Methods for Thai Colour Naming. In *AIC 2015 Midterm Meeting.* Tokyo, Japan

Kay, P., Berlin, B., Maffi, L., Merrifield, W., & Cook, R. (2010). *The World Color Survey.* Berkley, CA: CSLI Publications.

Kay, P., & McDaniel, C. K. (1978). The Linguistic Significance of the Meanings of Basic Color Terms. *Language, 54*(3), 610–646. https://doi.org/10.2307/412789

Kay, P., & Regier, T. (2003). Resolving the question of color naming universals. *Proceedings of the National Academy of Sciences, 100*(15), 9085–9089. https://doi.org/10.1073/pnas.1532837100

Kelly, K. L., & Judd, D. B. (1955). *The ISCC-NBS Method of Designating Colors and a Dictionary of Color Names.* U.S. Department of Commerce, National Bureau of Standards.

Kerttula, S. (2007). Relative basicness of colour terms. In R. E. MacLaury, G. V. Paramei, & D. Dedrick (Eds.), *Anthropology of Color: Interdisciplinary Multilevel Modeling* (p. 151). John Benjamins Publishing.

Khan, R., van de Weijer, J., Shahbaz Khan, F., Muselet, D., Ducottet, C., & Barat, C. (2013). Discriminative Color Descriptors. *2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2866–2873. https://doi.org/10.1109/CVPR.2013.369

Kiper, D. C., Fenstemaker, S. B., & Gegenfurtner, K. R. (1997). Chromatic properties of neurons in macaque area V2. *Visual Neuroscience*, *14*(6), 1061–1072.

Komatsu, H., Ideura, Y., Kaji, S., & Yamane, S. (1992). Color selectivity of neurons in the inferior temporal cortex of the awake macaque monkey. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *12*(2), 408–424.

Kuehni, R. G. (2003). *Color Space and Its Divisions: Color Order from Antiquity to the Present* (1 edition). Hoboken, N.J: Wiley-Interscience.

Kuehni, R. G. (2005). Focal Color Variability and Unique Hue Stimulus Variability. *Journal of Cognition and Culture*, *5*(3), 409–426. https://doi.org/10.1163/156853705774648554

Kraft, J. M., & Brainard, D. H. (1999). Mechanisms of color constancy under nearly natural viewing. *Proceedings of the National Academy of Sciences*, *96*(1), 307–312. https://doi.org/10.1073/pnas.96.1.307

Krauskopf, J., Williams, D. R., & Heeley, D. W. (1982). Cardinal directions of color space. *Vision Research*, *22*(9), 1123–1131.

Krauskopf, J., & Gegenfurtner, K. (1992). Color discrimination and adaptation. *Vision Research*, *32*(11), 2165–2175. https://doi.org/10.1016/0042-6989(92)90077-V

Kuhn, H. W. (1955). The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, *2*(1–2), 83–97. https://doi.org/10.1002/nav.3800020109

Lammens, J. (1994). *A computational model of color perception and color naming* (PHD). (State University of New York at Buffalo.

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

Land, E. H. (1974). The retinex theory of color vision. *Proceedings of the Royal Institution of Great Britain*, *47*, 23–58.

Lennie, P., Krauskopf, J., & Sclar, G. (1990). Chromatic mechanisms in striate cortex of macaque. *The Journal of Neuroscience*, *10*(2), 649–669.

Lafer-Sousa, R., & Conway, B. R. (2013). Parallel, multi-stage processing of colors, faces and shapes in macaque inferior temporal cortex. *Nature Neuroscience*, *16*(12), 1870–1878. https://doi.org/10.1038/nn.3555

Lantz, D., & Stefflre, V. (1964). Language and cognition revisited. *The Journal of Abnormal and Social Psychology*, *69*(5), 472–481. https://doi.org/10.1037/h0043769

Levinson, S. C. (2000). Yélî Dnye and the Theory of Basic Color Terms. *Journal of Linguistic Anthropology*, *10*(1), 3–55. https://doi.org/10.1525/jlin.2000.10.1.3

Li, Y.-Y., & Lee, H.-C. (2014). DLAB: a class of daylight-based uniform color space. *JOSA A*, *31*(8), 1876–1885. https://doi.org/10.1364/JOSAA.31.001876

Li, C., Li, Z., Wang, Z., Xu, Y., Luo, M. R., Cui, G., Melgosa, M., Brill, M., Pointer, M. (2017). Comprehensive color solutions: CAM16, CAT16, and CAM16-UCS. *Color Research & Application*, *42*(6). https://doi.org/10.1002/col.22131

Lin, H., Luo, M. R., MacDonald, L. W., & Tarrant, A. W. S. (2001a). A cross-cultural colour-naming study. Part I: Using an unconstrained method. *Color Research & Application*, *26*(1), 40–60.

Lin, H., Luo, M. R., MacDonald, L. W., & Tarrant, A. W. S. (2001b). A cross-cultural colour-naming study: Part II: Using a constrained method. *Color Research & Application*, *26*(3), 193–208. https://doi.org/10.1002/col.1017

Lin, Y., & Jeon, Y. (2006). Random Forests and Adaptive Nearest Neighbors. *Journal of the American Statistical Association*, *101*(474), 578–590. https://doi.org/10.1198/016214505000001230

Lindner, A., Li, B. Z., Bonnier, N., & Süsstrunk, S. (2012). A Large-Scale Multi-Lingual Color Thesaurus. *Color and Imaging Conference*, *2012*(1), 30–35.

Lindsey, D. T., & Brown, A. M. (2006). Universality of color names. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(44), 16608–16613. https://doi.org/10.1073/pnas.0607708103

Lindsey, D. T., & Brown, A. M. (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. https://doi.org/10.1073/pnas.0910981106

Lindsey, D. T., & Brown, A. M. (2014). The color lexicon of American English. *Journal of Vision*, *14*(2), 17. https://doi.org/10.1167/14.2.17

Lindsey, D. T., Brown, A. M., Brainard, D. H., & Apicella, C. L. (2015). Hunter-Gatherer Color Naming Provides New Insight into the Evolution of Color Terms. *Current Biology*, *25*(18), 2441–2446. https://doi.org/10.1016/j.cub.2015.08.006

Livingstone, M., & Hubel, D. (1988). Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science*, *240*(4853), 740–749. https://doi.org/10.1126/science.3283936

Lloyd-Jones, T. J., & Nakabayashi, K. (2014). Long-term repetition priming and semantic interference in a lexical-semantic matching task: tapping the links between object

names and colors. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00644

Loreto, V., Mukherjee, A., & Tria, F. (2012). On the origin of the hierarchy of color names. *Proceedings of the National Academy of Sciences*, *109*(18), 6819–6824. https://doi.org/10.1073/pnas.1113347109

Lucy, J. A., & Shweder, R. A. (1979). Whorf and His Critics: Linguistic and Nonlinguistic Influences on Color Memory. *American Anthropologist*, *81*(3), 581–615. https://doi.org/10.1525/aa.1979.81.3.02a00040

Luo, M. R., Cui, G., & Li, C. (2006). Uniform colour spaces based on CIECAM02 colour appearance model. *Color Research & Application*, *31*(4), 320–330. https://doi.org/10.1002/col.20227

Lyons, J. (1995). Colour in Language. In T. Lamb & J. Bourriau (Eds.), *Colour: Art & Science.* Cambridge University Press.

MacAdam, D. L. (1942). Visual Sensitivities to Color Differences in Daylight. *Journal of the Optical Society of America*, *32*(5), 247–273. https://doi.org/10.1364/JOSA.32.000247

MacDonald, L., Mylonas, D. (2016). Colour Naming: Linking Vision and Speech. In *3rd International Conference 'Colour, Culture, Science'*. Kraków, Poland.

MacLaury, R. E. (1992). From brightness to hue: An explanatory model of color-category evolution. *Current Anthropology*, *33*(2), 137–186. https://doi.org/10.1086/204049

MacLeod, D. I. A., & Boynton, R. M. (1979). Chromaticity diagram showing cone excitation by stimuli of equal luminance. *Journal of the Optical Society of America*, *69*(8), 1183–1186. https://doi.org/10.1364/JOSA.69.001183

Maier, M., & Abdel Rahman, R. (2018). Native Language Promotes Access to Visual Consciousness. *Psychological Science*, *29*(11), 1757–1772. https://doi.org/10.1177/0956797618782181

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

Martin, A., Haxby, J. V., Lalonde, F. M., Wiggs, C. L., & Ungerleider, L. G. (1995). Discrete Cortical Regions Associated with Knowledge of Color and Knowledge of Action. *Science*, *270*(5233), 102–105. https://doi.org/10.1126/science.270.5233.102

Maxwell, J. C. (1872). On colour vision. In *Royal Institution Proceedings VI. Scientific Papers* (Vol. 1, pp. 260–271).

McKeefry, D. J., & Zeki, S. (1997). The position and topography of the human colour centre as revealed by functional magnetic resonance imaging. *Brain*, *120*(12), 2229–2242. https://doi.org/10.1093/brain/120.12.2229

Menegaz, G., Le Troter, A., Sequeira, J., & Boi, J. M. (2007). A Discrete Model for Color Naming. *EURASIP Journal in Advances in Signal Processing*, *2007*(1), 113–113. https://doi.org/10.1155/2007/29125

Miceli, G., Fouch, E., Capasso, R., Shelton, J. R., Tomaiuolo, F., & Caramazza, A. (2001). The dissociation of color from form and function knowledge. *Nature Neuroscience*, *4*(6), 662–667. https://doi.org/10.1038/88497

Mollon, J. D., & Jordan, G. (1997). On the Nature of Unique Hues. In *John Dalton's Colour Vision Legacy* (pp. 381–392). London: Taylor & Francis.

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

Mollon, J. D. (2009). A neural basis for unique hues? *Current Biology*, *19*(11), R441–R442. https://doi.org/10.1016/j.cub.2009.05.008

Mojsilovic, A. (2005). A computational model for color naming and describing color composition of images. *Image Processing, IEEE Transactions on*, *14*(5), 690–699.

Morgan, G., & Corbett, G. (1989). Russian colour term salience. *Russian Linguistics*, *13*(2), 125–141. https://doi.org/10.1007/BF02551669

Moroney, N. (2003). Unconstrained web-based color naming experiment. *Color Imaging: Processing, Hardcopy and Applications VIII (Proceedings of SPIE/IS&T)*, *Volume 5008*, 36–46.

Moroney, N., Obrador, P., & Beretta, G. (2008). Lexical Image Processing. *Sixteenth Color Imaging Conference*, 268–273.

Moss, A., Davies, I., Corbett, G., & Laws, G. (1990) Mapping Russian basic colour terms using behavioural measures. *Lingua*, *82*, 313–332.

Moss, V. (1998). Russian blues and purples: a tentative hypothesis. *Quinquereme*, *11*, 164–177.

Motomura, H. (1997). Categorical Color Mapping for Gamut Mapping. In *Fifth Color Imaging Conference: Color Science, Systems and Applications* (Vol. 5, pp. 50–55). Arizona: IS&T/SID.

Munroe, R. (2010). Color survey results. Retrieved from https://blog.xkcd.com/2010/05/03/color-survey-results/

Munsell, A. H. (1905). *A Color Notation* (1st ed.). Geo. H. Ellis Co. Boston, USA.

Mylonas, D., & MacDonald, L. (2010). Online Colour Naming Experiment Using Munsell Samples. In *5th European Conference on Colour in Graphics, Imaging, and Vision and 12th International Symposium on Multispectral Colour Science*. Joensuu, Finland: IS&T.

Mylonas, D., MacDonald, L., & Wuerger, S. (2010). Towards an Online Colour Naming Model. In *18th Color and Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*. San Antonio, Texas: IS&T.

Mylonas, D., Karatzas, D., & Wuerger, S. (2010). *Error Assessment of Colour Reproduction in Mobile Display Devices* (No. TR 17). Liverpool, UK: University of Liverpool & TruColour Ltd.

Mylonas, D., & MacDonald, L. (2012). Colour naming for colour communication. In J. Best (Ed.), *Colour Design: Theories and Applications* (pp. 254–270). Cambridge, UK: Woodhead Publishing.

Mylonas, D., Purver, M., Sadrzadeh, M., MacDonald, L., & Griffin, L. D. (2015). The Use of English Colour Terms in Big Data. *Midterm AIC Conference*, Tokyo, Japan.

Mylonas, D., & MacDonald, L. (2016). Augmenting basic colour terms in English. *Color Research & Application*, *41*(1), 32–42. https://doi.org/10.1002/col.21944

Mylonas, D., Andrews, J., Griffin L.D. (2016). Variability in free colour naming across the full colour gamut. In *AVA (Applied Vision Association) Christmas Meeting 2016*. Queen Mary University of London, London, UK.

Mylonas, D., MacDonald, L. (2017). *Colour Naming for Colour Design*. In Colour Design, ed. Best, J. Elsevier

Mylonas, D., MacDonald L., Griffin, L.D. (2017). Differences in Color Naming between British and American English Speakers. In *13th AIC Congress*, South Korea.

Mylonas D., Stockman A., Griffin L.D. (2018) Basic Colour Terms are Indispensable. In *AVA Christmas Meeting 2018*, Birkbeck University of London, London, UK.

Mylonas D., Griffin L.D., (2019-submitted) Coherence of Achromatic, Primary and Basic Classes of Colour.

Naor-Raz, G., Tarr, M. J., & Kersten, D. (2003). Is Color an Intrinsic Property of Object Representation? *Perception*, *32*(6), 667–680. https://doi.org/10.1068/p5050

Newhall, S. M., Nickerson, D., & Judd, D. B. (1943). Final Report of the O.S.A. Subcommittee on the Spacing of the Munsell Colors. *Journal of the Optical Society of America*, *33*(7), 385–411. https://doi.org/10.1364/JOSA.33.000385

Newton, I. (1730). *Opticksor, a Treatise of the Reflections, Refractions, Inflections,and Colours of Light* (Fourth Edition). London, UK: William Innys. (pg. 186).

Nickerson, D. (1981). OSA uniform color scale samples: A unique set. *Color Research & Application*, *6*(1), 7–33. https://doi.org/10.1002/col.5080060105

Olkkonen, M., Hansen, T., & Gegenfurtner, K. R. (2009). Categorical color constancy for simulated surfaces. *Journal of Vision*, *9*(12), 6–6. https://doi.org/10.1167/9.12.6

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

Ocelák, R. (2014). The Myth of Unique Hues. *Topoi*, 1–10. https://doi.org/10.1007/s11245-014-9249-4

Özgen, E., & Davies, I. R. L. (1998). Turkish color terms: Tests of Berlin and Kay's theory of color universals and linguistic relativity. *Linguistics*, *36*(5), 919–956. https://doi.org/10.1515/ling.1998.36.5.919

Paggetti, G., Menegaz, G., & Paramei, G. V. (2015). Color naming in Italian language. *Color Research & Application*, *41*(4). https://doi.org/10.1002/col.21953

Panorgias, A., Kulikowski, J. J., Parry, N. R. A., McKeefry, D. J., & Murray, I. J. (2010). Naming versus matching and the stability of unique hues. *Ophthalmic and Physiological Optics*, *30*(5), 553–559. https://doi.org/10.1111/j.1475-1313.2010.00736.x

Paramei, G. V. (2005). Singing the Russian Blues: An Argument for Culturally Basic Color Terms. *Cross-Cultural Research*, *39*(1), 10–38. https://doi.org/10.1177/1069397104267888

Paramei, G. V., Griber, Y. A., & Mylonas, D. (2018). An online color naming experiment in Russian using Munsell color samples. *Color Research & Application*, *43*(3), 358–374. https://doi.org/10.1002/col.22190

Parraga, C. A., & Akbarinia, A. (2016). NICE: A Computational Solution to Close the Gap from Colour Perception to Colour Categorization. *PLoS ONE*, *11*(3). https://doi.org/10.1371/journal.pone.0149538

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

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, *108*(9), 3526–3529. https://doi.org/10.1073/pnas.1012551108

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Pointer, M. R., & Attridge, G. G. (1998). The number of discernible colours. *Color Research & Application*, *23*(1), 52–54.

Prasithrathsint, A. (1988). *A comparative study of Zhaung and Thai kinship systems*. Presented at the 4th International Conference on Thai Studies, Yunnan, China: Kunming.

Pridmore, R. W. (2008). Chromatic induction: Opponent color or complementary color process? *Color Research & Application*, *33*(1), 77–81. https://doi.org/10.1002/col.20363

Pridmore, R. W. (2011). Complementary colors theory of color vision: Physiology, color mixture, color constancy and color perception. *Color Research & Application*, *36*(6), 394–412. https://doi.org/10.1002/col.20611

Pridmore, R. W. (2013). Cone Photoreceptor Sensitivities and Unique Hue Chromatic Responses: Correlation and Causation Imply the Physiological Basis of Unique Hues. *PLOS ONE, 8*(10), e77134. https://doi.org/10.1371/journal.pone.0077134

Probst, P., & Boulesteix, A.-L. (2017). To tune or not to tune the number of trees in random forest? *ArXiv:1705.05654 [Cs, Stat]*. Retrieved from http://arxiv.org/abs/1705.05654

Quinn, P. C., Wooten, B. R., & Ludman, E. J. (1985). Achromatic color categories. *Perception & Psychophysics, 37*(3), 198–204. https://doi.org/10.3758/BF03207564

Rafegas, I., & Vanrell, M. (2018). Color encoding in biologically-inspired convolutional neural networks. *Vision Research, 151*, 7–17. https://doi.org/10.1016/j.visres.2018.03.010

Rätsep, K. (2011). Preliminary research on Turkish basic colour terms with an emphasis on blue. In C. P. Biggam, C. A. Hough, C. J. Kay, & D. R. Simmons (Eds.), *New Directions in Colour Studies* (pp. 133–145). John Benjamins Publishing.

Regier, T., Kay, P., & Cook, R. S. (2005). Focal colors are universal after all. *Proceedings of the National Academy of Sciences of the United States of America, 102*(23), 8386–8391. https://doi.org/10.1073/pnas.0503281102

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

Regier, T., Kemp, C., & Kay, P. (2015). Word Meanings across Languages Support Efficient Communication. In BriancWhinney & W. O'Grady (Eds.), *The Handbook of Language Emergence* (pp. 237–263).

Reips, U.-D. (2000). The Web Experiment Method: Advantages, Disadvantages, and Solutions. In *Psychological Experiments on the Internet*. Academic Press.

Rinner, O., & Gegenfurtner, K. R. (2000). Time course of chromatic adaptation for color appearance and discrimination. *Vision Research, 40*(14), 1813–1826.

Rinner, O., & Gegenfurtner, K. R. (2002). Cone contributions to colour constancy. *Perception, 31*(6), 733–746.

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

Rosch Heider, E. (1972). Probabilities, Sampling, and Ethnographic Method: The Case of Dani Colour Names. *Man, 7*(3), 448. https://doi.org/10.2307/2800917

Rosch, E., & Mervis, C. B. (1975). Family Resemblances: Studies in the Internal Structure of Categories. *Cognitive Psychology, 7*(4), 573–605.

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

Saunders, B. A. C., & van Brakel, J. (1997). Are There Nontrivial Constraints on Colour Categorization? *Behavioral and Brain Sciences*, *20*(02), 167–179.

Schefrin, B. E., & Werner, J. S. (1990). Loci of spectral unique hues throughout the life span. *JOSA A*, *7*(2), 305–311. https://doi.org/10.1364/JOSAA.7.000305

Schein, S. J., Marrocco, R. T., & de Monasterio, F. M. (1982). Is there a high concentration of color-selective cells in area V4 of monkey visual cortex? *Journal of Neurophysiology*, *47*(2), 193–213.

Schein, S., & Desimone, R. (1990). Spectral properties of V4 neurons in the macaque. *The Journal of Neuroscience*, *10*(10), 3369–3389.

Seaborn, M., Hepplewhite, L., & Stonham, J. (2005). Fuzzy colour category map for the measurement of colour similarity and dissimilarity. *Pattern Recognition*, *38*(2), 165–177. https://doi.org/10.1016/j.patcog.2004.05.001

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shapiro, C., Carl, S., & Varian, H. R. (1998). *Information Rules: A Strategic Guide to the Network Economy*. Harvard Business Press.

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

Shepard, R. N. (1992). The perceptual organization of colors: An adaptation to regularities of the terrestrial world? In *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 495–532). New York, NY, US: Oxford University Press.

Shepard, T. G., Swanson, E. A., McCarthy, C. L., & Eskew, R. T. (2016). A model of selective masking in chromatic detection. *Journal of Vision*, *16*(9), 3. https://doi.org/10.1167/16.9.3

Shepard, T. G., Lahlaf, S. I., & Eskew, R. T. (2017). Labeling the lines: A test of a six-mechanism model of chromatic detection. *Journal of Vision*, *17*(13), 9–9. https://doi.org/10.1167/17.13.9

Skelton, A. E., Catchpole, G., Abbott, J. T., Bosten, J. M., & Franklin, A. (2017). Biological origins of color categorization. *Proceedings of the National Academy of Sciences*, *114*(21), 5545–5550. https://doi.org/10.1073/pnas.1612881114

Smith, V. C., & Pokorny, J. (1975). Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm. *Vision Research*, *15*(2), 161–171. https://doi.org/10.1016/0042-6989(75)90203-5

Sorabji, R. (1972). Aristotle, Mathematics, and Colour. *The Classical Quarterly*, *22*(2), 293–308.

Steels, L., & Belpaeme, T. (2005). Coordinating perceptually grounded categories through language: a case study for colour. *The Behavioral and Brain Sciences*, *28*(4), 469-489-529. https://doi.org/10.1017/S0140525X05000087

Stockman, A., & Brainard, D. H. (2010). Color vision mechanisms. In *OSA Handbook of Optics* (3rd edition, M. Bass, ed, p. 11.1-11.104). New York: McGraw-Hill.

Stockman, A., & Sharpe, L. T. (2000). The spectral sensitivities of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype. *Vision Research*, *40*(13), 1711–1737. https://doi.org/16/S0042-6989(00)00021-3

Stockman, A., & Sharpe, L. T. (2006). CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams. *ISS/CIE Expert Symposium "75 Years of the CIE Standard Colorimetric Observer*, 1–10. Ontario, Canada.

Stoughton, C. M., & Conway, B. R. (2008). Neural basis for unique hues. *Current Biology*, *18*(16), R698–R699. https://doi.org/16/j.cub.2008.06.018

Sturges, J., & Whitfield, A. (1995). Locating basic colours in the Munsell Space. *Color Research and Application*, *Volume 20*(Number 6), 364–376.

Surowiecki, J. (2005). *The Wisdom of Crowds: Why the Many Are Smarter Than the Few* (New Ed edition). London: Abacus.

Tanigawa, H., Lu, H. D., & Roe, A. W. (2010). Functional organization for color and orientation in macaque V4. *Nature Neuroscience*, *13*(12), 1542–1548. https://doi.org/10.1038/nn.2676

Tominaga, S. (1985). A color-naming method for computer color vision. In *Proceedings of IEEE International Conference on Cybernetics and Society* (pp. 573–577).

Topper, D. (1990). Newton on the number of colours in the spectrum. *Studies in History and Philosophy of Science Part A*, *21*(2), 269–279. https://doi.org/10.1016/0039-3681(90)90026-5

Trichopoulou, A., & Lagiou, P. (1997). Healthy Traditional Mediterranean Diet: an Expression of Culture, History, and Lifestyle. *Nutrition Reviews*, *55*(11), 383–389. https://doi.org/10.1111/j.1753-4887.1997.tb01578.x

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

Uchikawa, K., & Boynton, R. M. (1987). Categorical color perception of Japanese observers: comparison with that of Americans. *Vision Research*, *27*(10), 1825–1833.

Uchikawa, K., Uchikawa, H., & Boynton, R. M. (1989). Partial color constancy of isolated surface colors examined by a color-naming method. *Perception*, *18*(1), 83–91. https://doi.org/10.1068/p180083

Ulusoy B., Griffin L.D., Mylonas, D. (2017). Turkish Colour Naming on the Net. In *13th AIC Congress*, South Korea.

Valberg, A. (2001). Unique hues: an old problem for a new generation. *Vision Research*, *41*(13), 1645–1657. https://doi.org/10.1016/S0042-6989(01)00041-4

Von Kries, J. (1905). Die Gesichtsempfindungen. In W. Nagel (Ed.), *Handbuch der Physiolgie des Menschen* (pp. 205–212). Braunschweig: Vieweg und Sohn.

Varnhagen, C. K., Gushta, M., Daniels, J., Peters, T. C., Parmar, N., Law, D., Hirsch, R., Takach, B.S. & Johnson, T. (2005). How informed is online informed consent? *Ethics & Behavior*, *15*(1), 37–48. https://doi.org/10.1207/s15327019eb1501_3

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

Wade, A., Augath, M., Logothetis, N., & Wandell, B. (2008). fMRI measurements of color in macaque and human. *Journal of Vision*, *8*(10). https://doi.org/10.1167/8.10.6

Wachtler, T., Sejnowski, T. J., & Albright, T. D. (2003). Representation of Color Stimuli in Awake Macaque Primary Visual Cortex. *Neuron*, *37*(4), 681–691. https://doi.org/16/S0896-6273(03)00035-7

Webster, M., & Kay, P. (2007). Individual and population differences in focal colors. In R. E. MacLaury, G. V. Paramei, & D. Dedrick (Eds.), *Anthropology of Color: Interdisciplinary multilevel modeling* (pp. 29–53). Amsterdam, Netherlands: John Benjamins Publishing.

Webster, M. A. (2011). Adaptation and visual coding. *Journal of Vision*, *11*(5), 3–3. https://doi.org/10.1167/11.5.3

Weijer, J.V., Schmid, C., & Verbeek, J. (2007). Learning Color Names from Real-World Images. In *IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR '07* (pp. 1–8). https://doi.org/10.1109/CVPR.2007.383218

Weijer, J. van de, & Khan, F. S. (2013). Fusing Color and Shape for Bag-of-Words Based Object Recognition. In S. Tominaga, R. Schettini, & A. Trémeau (Eds.), *Computational Color Imaging* (pp. 25–34). Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-36700-7_3

Whelan, R. (2008). Effective Analysis of Reaction Time Data. *Psychological Record*, *58*(3), 475–482.

Wittgenstein, L. (2009/1953). *Philosophical Investigations*. (P. M. S. Hacker & J. Schulte, Eds.) (4th Revised ed. edition). Chichester, West Sussex, U.K. ; Malden, MA: Wiley-Blackwell.

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

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

Witzel, C. (2018). The role of saturation in colour naming and colour appearance. In W. L. MacDonald, C. P. Biggam, & G. Paramei (Eds.), *Progress in Colour Studies: Cognition, language and beyond* (pp. 41–58). John Benjamins.

Witzel, C. (2018). Misconceptions About Colour Categories. *Review of Philosophy and Psychology*, 1–42. https://doi.org/10.1007/s13164-018-0404-5

Witzel, C., & Gegenfurtner, K. R. (2018). Are red, yellow, green, and blue perceptual categories? *Vision Research.* https://doi.org/10.1016/j.visres.2018.04.002

Wool, L. E., Komban, S. J., Kremkow, J., Jansen, M., Li, X., Alonso, J.-M., & Zaidi, Q. (2015). Salience of unique hues and implications for color theory. *Journal of Vision, 15*(2), 10–10. https://doi.org/10.1167/15.2.10

Wuerger, S. M., Atkinson, P., & Cropper, S. (2005). The cone inputs to the unique-hue mechanisms. *Vision Research, 45*(25–26), 3210–3223. https://doi.org/10.1016/j.visres.2005.06.016

Wuerger, S. (2013). Colour Constancy Across the Life Span: Evidence for Compensatory Mechanisms. *PLOS ONE, 8*(5), e63921. https://doi.org/10.1371/journal.pone.0063921

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

Xiao, K., & Mylonas, D. (2010). *Assessment of Colour Reproduction Error in Uncorrected Displays* (No. TR 10). Liverpool, UK: University of Liverpool & TruColour Ltd.

Xiao, K., Wuerger, S., Fu, C., & Karatzas, D. (2011). Unique hue data for colour appearance models. Part I: Loci of unique hues and hue uniformity. *Color Research & Application, 36*(5), 316–323. https://doi.org/10.1002/col.20637

Xiao, K., Fu, C., Mylonas, D., Karatzas, D., & Wuerger, S. (2011). Unique hue data for colour appearance models. Part II: Chromatic adaptation transform. *Color Research & Application, 38*(1). https://doi.org/10.1002/col.20725

Xiao, K., Mylonas, D., Fu, C., Karatzas, D., & Wuerger, S. (2011). Locating unique hues under mixed illumination conditions in CIECAM02. *19th Color Imaging Conference*, 94–97. San Jose, California, USA: IS&T.

Yang, Y., Yang, J., Yan, J., Liao, S., Yi, D., & Li, S. Z. (2014). Salient Color Names for Person Re-identification. *ECCV.* https://doi.org/10.1007/978-3-319-10590-1_35

Yendrikhovskij, S. (2001). A computational model of colour categorization. *Color Research & Application, 26*(S1), S235–S238.

Yi, S. K. M., Steyvers, M., Lee, M. D., & Dry, M. J. (2012). The Wisdom of the Crowd in Combinatorial Problems. *Cognitive Science*, *36*(3), 452–470. https://doi.org/10.1111/j.1551-6709.2011.01223.x

Young, T. (1802). The Bakerian Lecture: On the Theory of Light and Colours. *Philosophical Transactions of the Royal Society of London*, *92*, 12–48. https://doi.org/10.1098/rstl.1802.0004

Yu, L., Zhang, L., van de Weijer, J., Khan, F. S., Cheng, Y., & Parraga, C. A. (2018). Beyond Eleven Color Names for Image Understanding. *Machine Vision and Applications*, *29*(2), 361–373. https://doi.org/10.1007/s00138-017-0902-y

Zaidi, Q. (1998). Identification of illuminant and object colors: heuristic-based algorithms. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, *15*(7), 1767–1776.

Zaslavsky, N., Kemp, C., Regier, T., & Tishby, N. (2018). Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(31), 7937–7942. https://doi.org/10.1073/pnas.1800521115

Zeki, S. M. (1973). Colour coding in rhesus monkey prestriate cortex. *Brain Research*, *53*(2), 422–427. https://doi.org/10.1016/0006-8993(73)90227-8

Zeki, S. (1980). The representation of colours in the cerebral cortex. *Nature*, *284*(5755), 412–418. https://doi.org/10.1038/284412a0

Zeki, S. (1983). Colour coding in the cerebral cortex: The responses of wavelength-selective and colour-coded cells in monkey visual cortex to changes in wavelength composition. *Neuroscience*, *9*(4), 767–781. https://doi.org/10.1016/0306-4522(83)90266-X

Zeki, S., & Marini, L. (1998). Three cortical stages of colour processing in the human brain. *Brain*, *121*(9), 1669–1685. https://doi.org/10.1093/brain/121.9.1669

Zeki, S., Cheadle, S., Pepper, J., & Mylonas, D. (2017). The Constancy of Colored After-Images. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00229

Zeki, S., Javier, A., & Mylonas, D. (2019). The Biological Basis of the Experience and Categorization of Colour. *European Journal of Neuroscience (EV).*

Zipf, G. (1935). *The Psychobiology of Language: An Introduction to Dynamic Philology*. Cambridge, Mass.: M.I.T. Press.
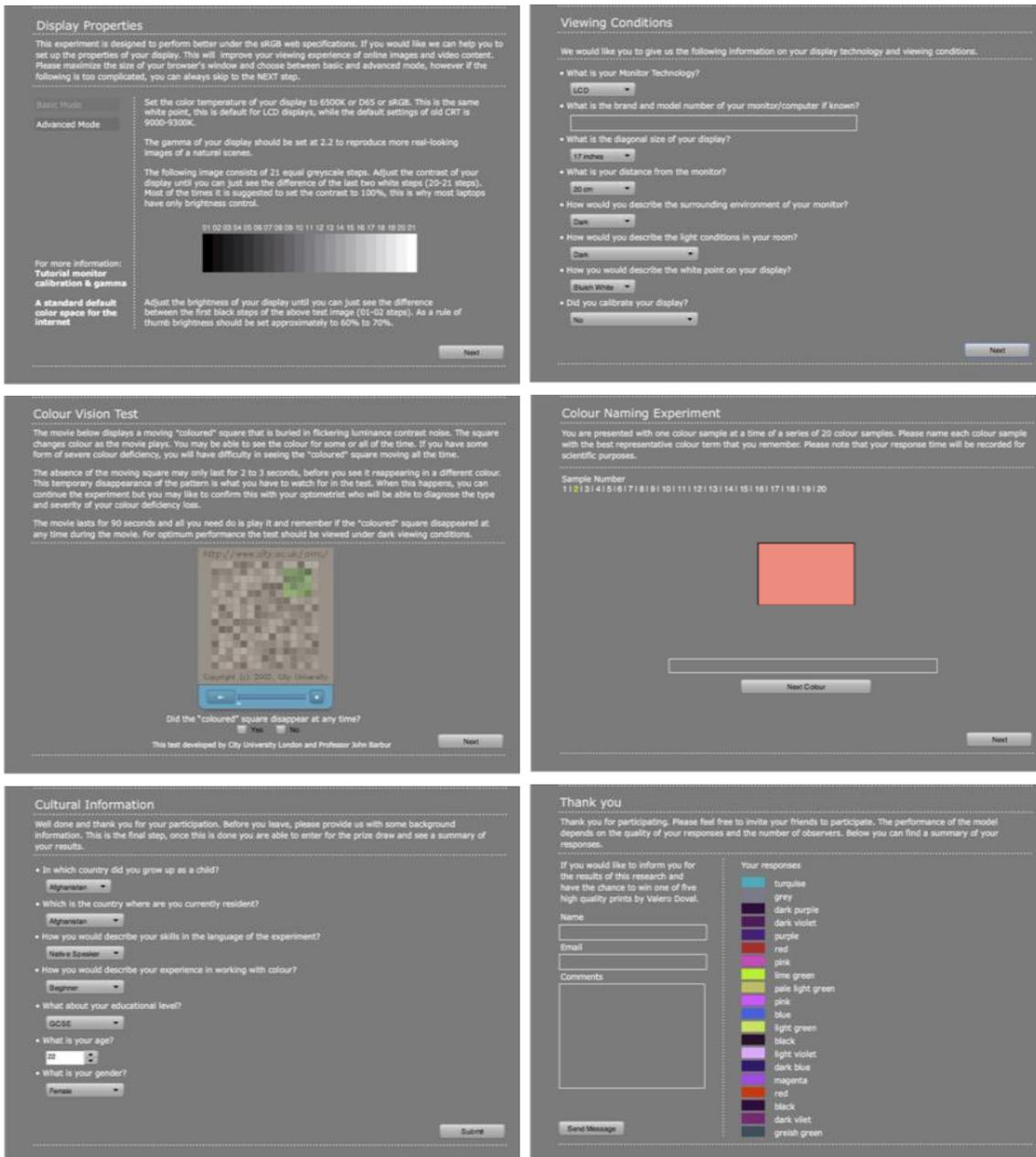
Appendix A Supplementary Material for Chapter 3

Figure A.1 Interface of the online colour naming experiment in the period between 2009 and 2018 consisting of six steps. Snapsots of the website can be found at: https://web.archive.org/web//colornaming.net*

Appendix A.1 Colour naming datasets

The following description of the dataset follow an alphabetical order of the languages in question.

Apendix A.1.1 American English

For the American English dataset, we retrieved 10,000 raw responses from 600 observers. Excluding disruptive observers (0.5%) and observers with possible colour deficiency (10.4%) resulted in a dataset for 448 American English speakers. Their mean age was 33 years old (SD=14 years). Females offered 58% and males 42% of the responses. Excluding unique responses from single observers resulted in 7,546 responses with 436 distinct colour descriptors. The occurrence of colour descriptors with varying word number was: monolexemic BCT 30%; monolexemic non-BCT 31%; colour terms with one modifier 36% and colour descriptors containing ≥ 3 words 4% (Figure 4.3). Due to rounding, percentages may not add up to 100% throughout this section.
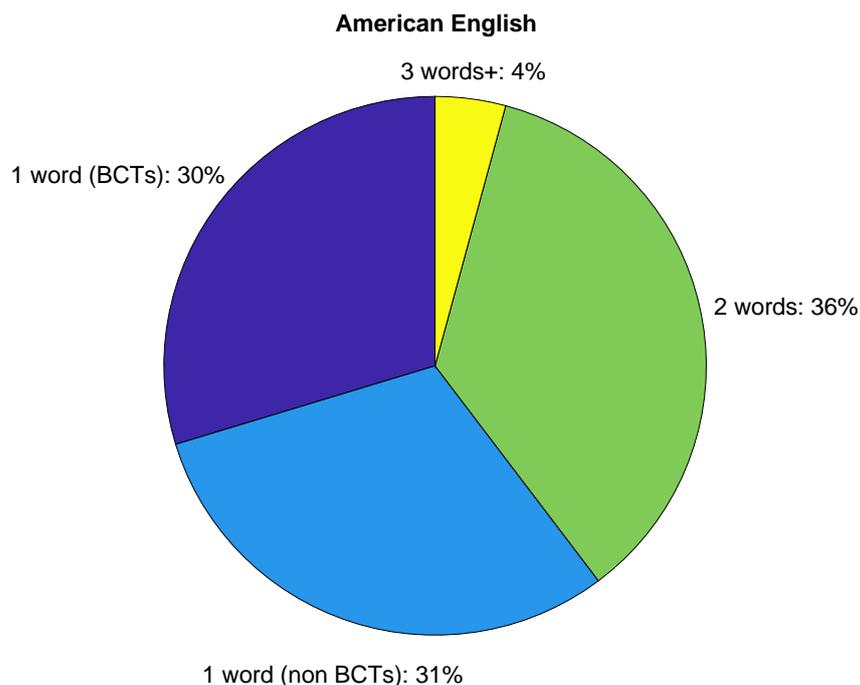
American English

Figure A.2 Number of words in colour descriptors for American English speakers.

Apendix A.1.2 Greek Dataset

For the Greek dataset we considered 10,000 raw responses from 600 observers. Excluding disruptive observers (1%) and observers with possible colour deficiency (18.8%) left 324 observers. Their mean age was 32 years old (SD=9 years). Females offered 64% and males 36% of the responses. Excluding unique responses from single observers resulted in 5,871 responses with 313 distinct colour descriptors. Of these, 37%

involved BCTs (n=12); 34% monolexemic non-BCTs; 26% two-word responses and 2% three or more words (Figure 4.5).

Greek

3 words+: 2%

2 words: 26%

1 word (BCTs): 37%

1 word (non BCTs): 34%
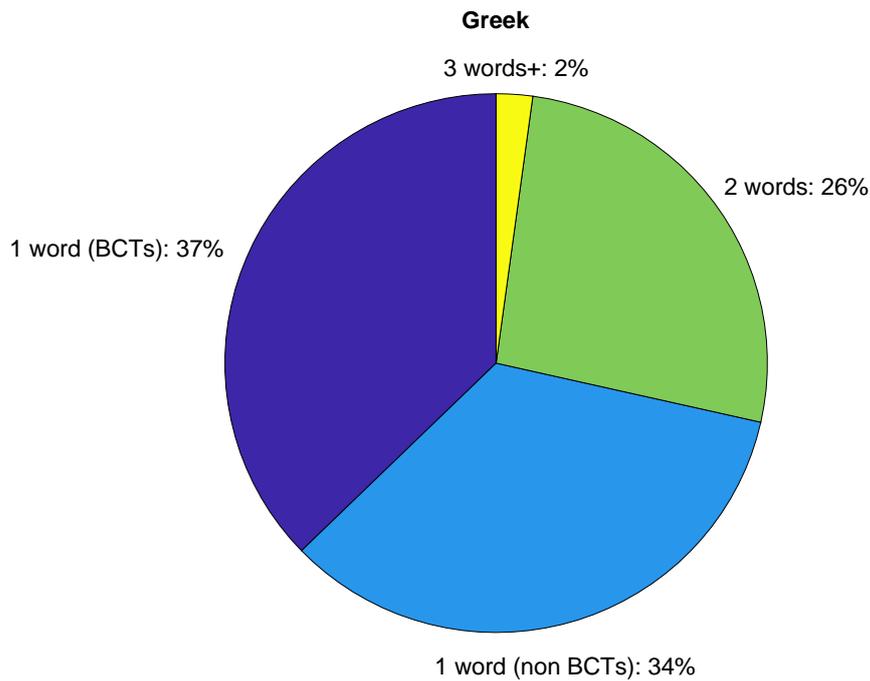
Figure A.3 Number of words in colour descriptors for Greek speakers.

Apendix A.1.3 Russian

The Russian dataset consisted of 10,000 raw responses from 600 Russian speakers. Excluding observers with possible colour deficiency (10.2%) left 449 observers. Their mean age was 24 years old (SD=9 years). Females offered 62% and males 38% of the responses. Excluding unique responses from single observers resulted in 7,802 responses with 342 distinct colour descriptors. The occurrence of colour descriptors with varying word number was: monolexemic BCT (n=12) 38%; monolexemic non-BCT 26%; colour terms with one modifier 34% and colour descriptors containing ≥ 3 words 2% (Figure 4.6).
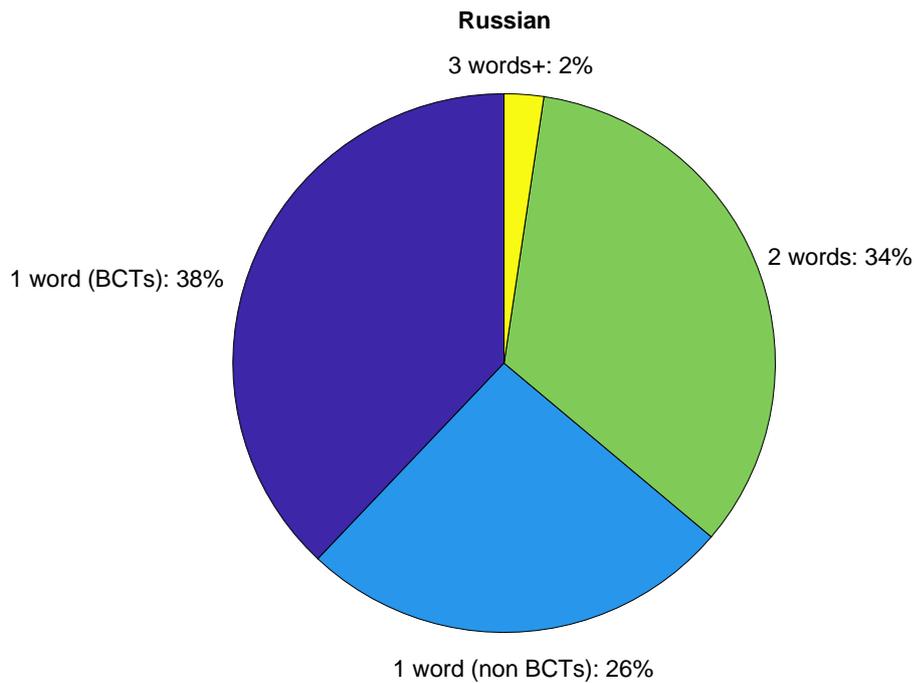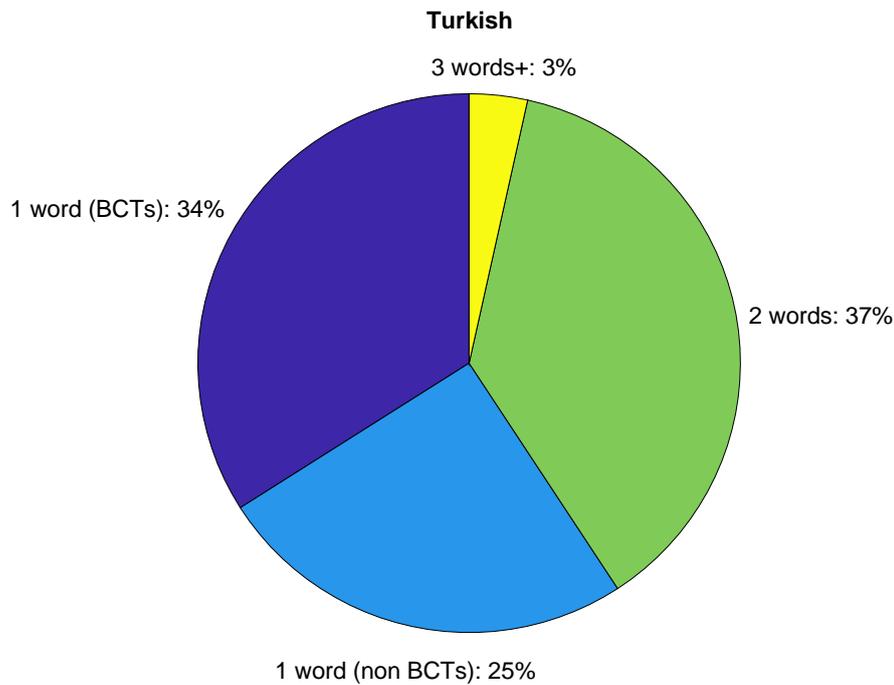
Figure A.4 Number of words in colour descriptors for Russian speakers.

Apendix A.1.4 Thai

The Thai dataset was smaller and consisted of 5,100 raw responses from 255 observers. Excluding observers with possible colour deficiency (16.1%) left 202 observers. Their mean age was 29 years old (SD=10 years). Females offered 73% and males 27% of the responses. Excluding unique responses from single observers resulted in 3,516 responses with 287 distinct colour descriptors. The task of counting the number of words in Thai responses automatically is challenging due to the fact that there are no spaces between words. Hence, we report here only the percentage of monolexemic BCTs (33%) and all the other responses combined (67%). In Thai, when people name colours, often put the word 'สี' in front of its name that means colour; this was excluded from the analysis.

Apendix A.1.5 Turkish

The Turkish dataset consisted of 6,180 raw responses from 309 observers. Excluding observers with possible colour deficiency (6.5%) left 273 observers. Their mean age was 28 years old (SD=8 years). Females offered 68% and males 32% of the responses. Excluding unique responses from single observers resulted in 4,727 responses with 285 distinct colour descriptors. The occurrence of colour descriptors with varying word number was: monolexemic BCT (n=12) 34%; monolexemic non-BCT 25%; colour terms with one modifier 37% and colour descriptors containing ≥ 3 words 3% (Figure 4.7).

Figure A.5 Number of words in colour descriptors for Turkish speakers.

Table A.1 Euclidean distances ΔE$_{ab}$ between centroids of BCTs in British English (Br) and American English (Am), Greek (Gr), Russian (Ru), Thai (Th) and Turkish (Tu). The last column shows the mean distance across languages per term.

	Br vs Am	Br vs Gr	Br vs Ru	Br vs Th	Br vs Tu	mean per term
white	1.49	3.64	1.55	2.55	2.57	2.61
black	1.35	1.89	3.48	3.88	3.18	3.18
red	2.36	3.31	3.90	1.16	2.09	3.16
yellow	1.97	0.84	1.69	2.47	5.98	2.68
green	3.98	2.76	3.77	4.33	0.86	3.73
blue	3.73	19.26	25.43	26.14	4.27	22.12
brown	1.05	1.24	2.95	3.35	1.26	2.58
purple	0.97	2.57	0.89	1.93	3.25	2.35
pink	2.35	4.13	2.96	3.37	4.33	4.52
orange	1.75	3.86	2.57	8.13	5.01	4.43
grey	2.52	4.04	3.06	2.03	3.17	2.66
blue (2)	NA	27.83	29.45	26.97	33.04	20.75
mean	2.14	6.28	6.81	7.19	5.75	6.23

Appendix B Supplementary Material for Chapter 4

In this section we present three sets of behavioural, geometric and linguistic features computed for each common colour name given more than 20 times in the British English dataset. These sets of features were used in the assessment of the coherence between members of classes of colours presented in Chapter 4 and supplemented here.

Appendix B.1 Behavioural Features

The behavioural features include frequency of occurrence, consensus and response time.
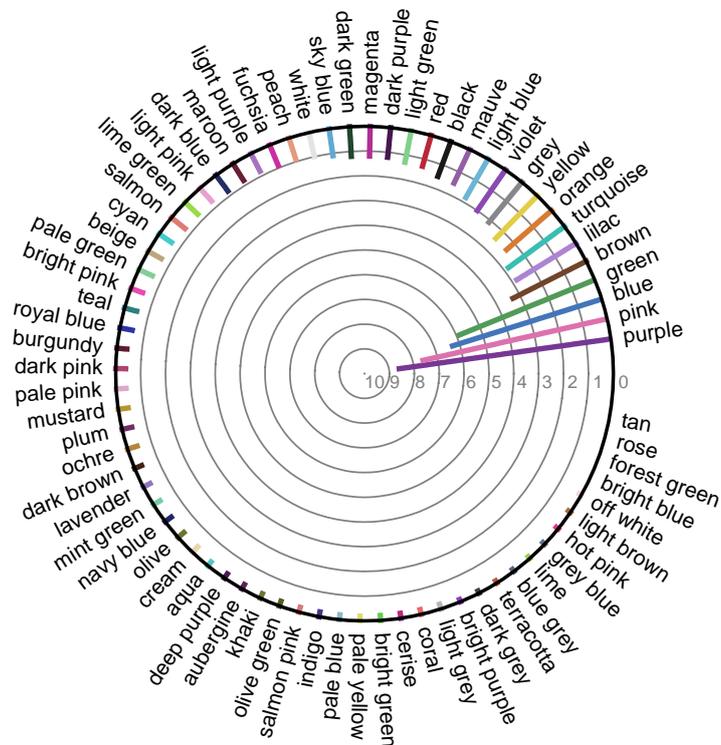
Figure A.6 Frequency of colour names in online experiment. Radial length indicates percentage of all responses that were exactly the indicated term.
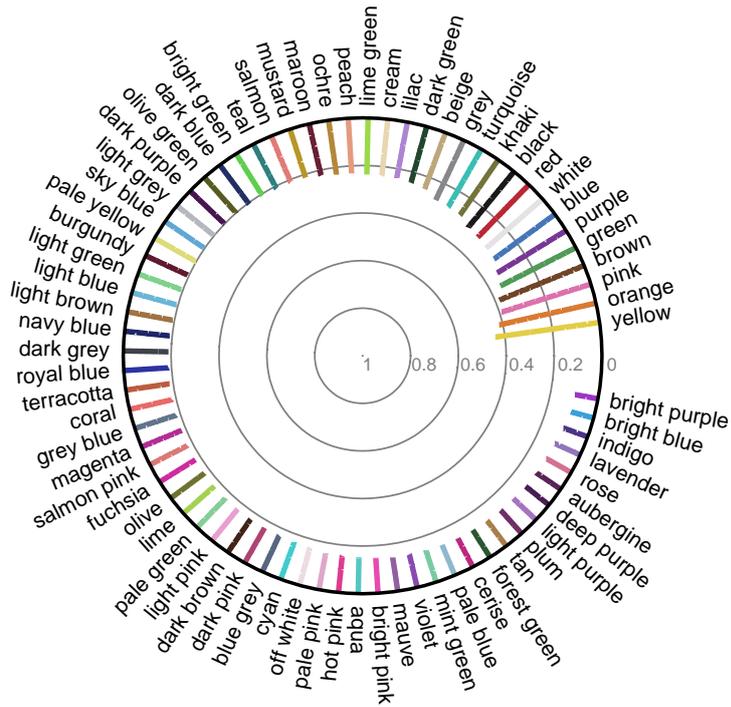
Figure A.7 Colour names with consensus (radial scale) across samples and observers.
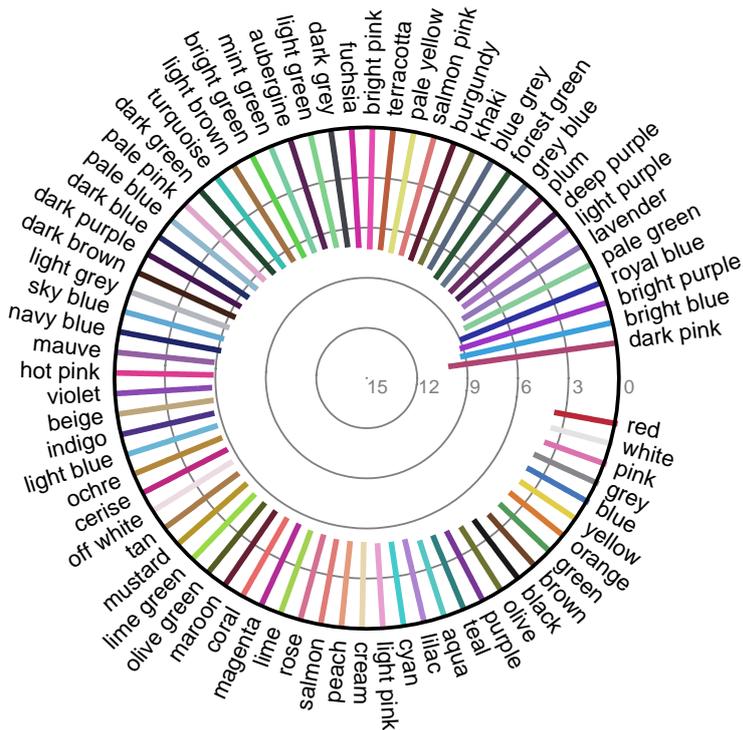
Figure A.8 Median response time (radial scale, secs) of responding colour names.

Appendix B.2 Geometric features

The geometric features include the size (volume) and shape (anisotropy) in colour space. For the location see of centroids of colour categories in colour space see Figure 0.1.
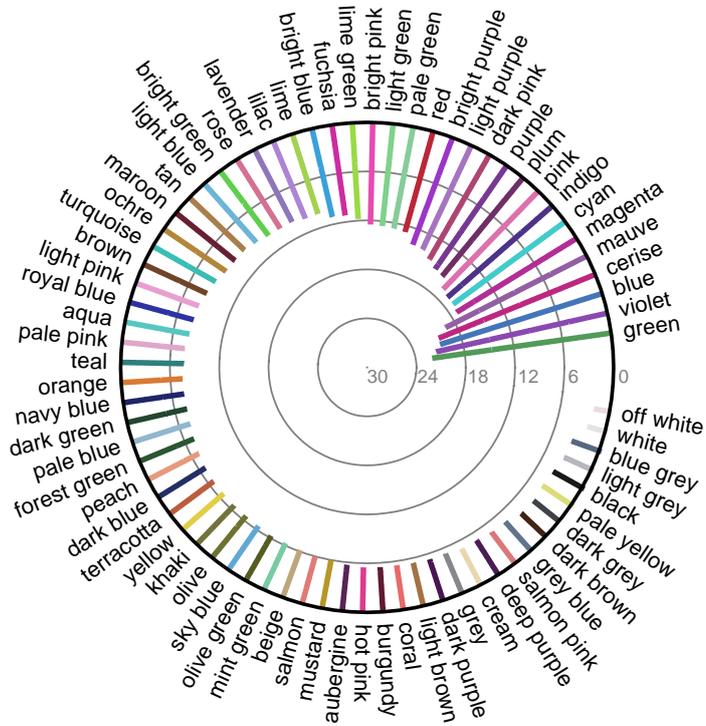
Figure A.9 Volume (radial scale, units are cubic (ΔE$_{ab}$) of lexical colour categories in colour space.
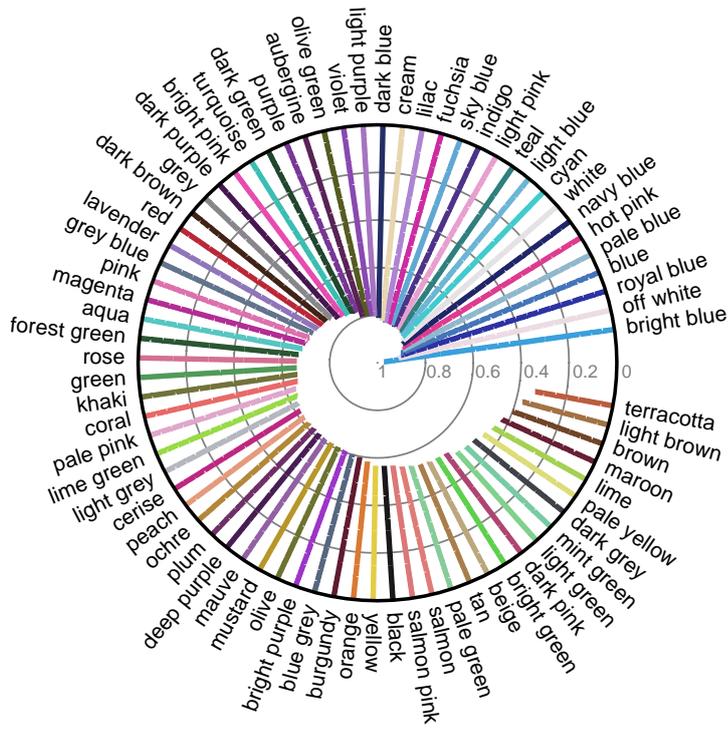
Figure A.10 Fractional anisotropy (radial scale, high values indicate non-spherical shape) of colour categories in colour space.

Appendix B.3 Linguistic features

The linguistic features include the frequency in ordinary communication, the length of the words and the number of derivative forms.
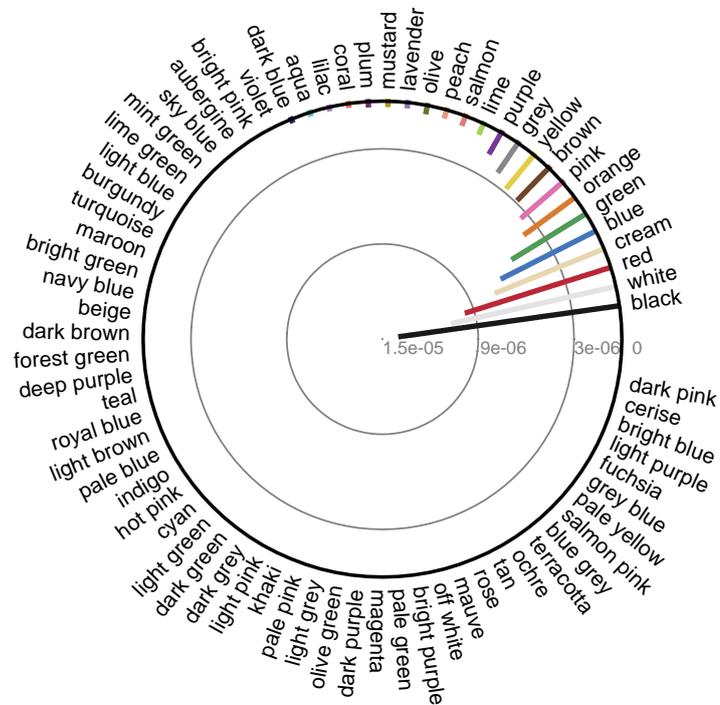
Figure A.11 Frequency of colour names in Twitter messages. Note that the 13 last colour names were not found in the Twitter dataset.

Figure A.12 Name length (radial scale is number of letters) of colours.

Figure A.13 Number of derivative forms (radial scale) of colours.

Appendix B.4 Coherence of alternative primary classes

The assessment of coherence of the seven pure or regular colour categories as described by Aristotle (350 B.C.E.) produced a MAP score of 0.57, higher than that for Hering primaries but still distinctly less than the ideal of 1. As with the Hering primaries the leading problem terms were low in-class confidence for *yellow*, and high in-class confidence for out-of-class *pink* (Figure A.14).
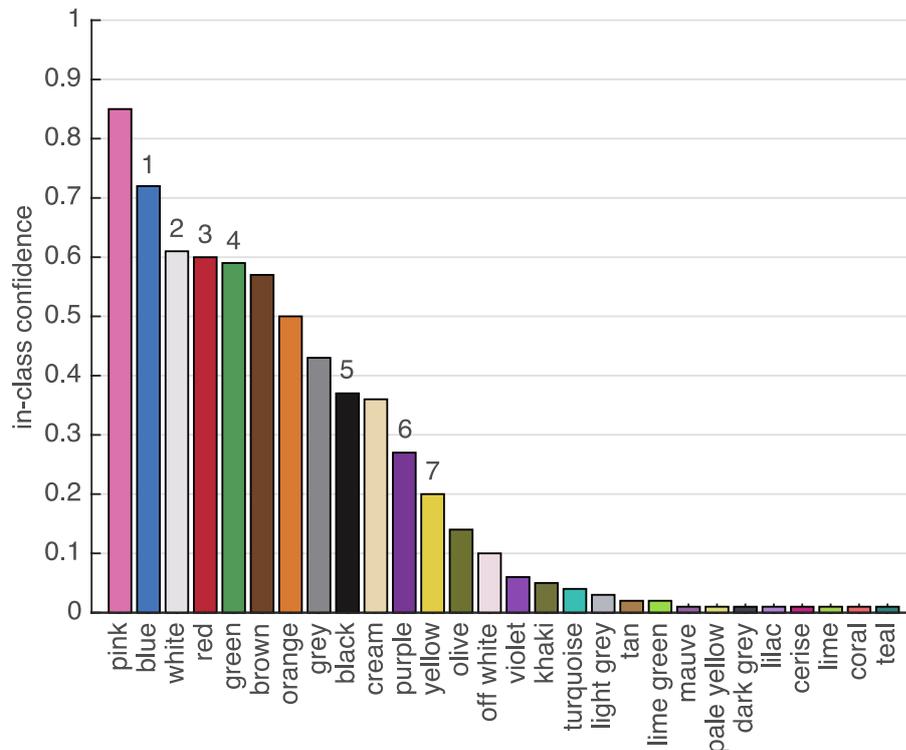
Figure A.14 Coherence of colour names for being members of primary class (n=7).

We also considered the 7 colour categories of the spectrum named by Newton (1730). The MAP score for these members of the class was 0.57. *Pink* and *brown* were found at the top of the confidences ranking. *Indigo* was given a class confidence of 0 and it is not shown in the figure (Figure A.15). The outlier status of *indigo*, amongst Newton's spectral colours, is consistent with explanations for its inclusion (McLaren, 1985).
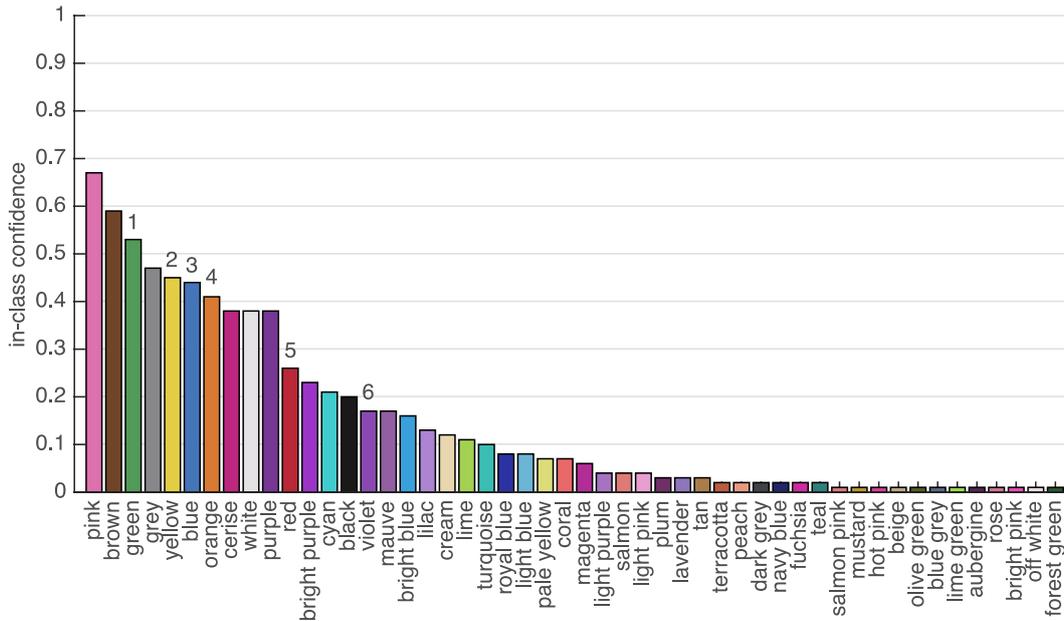
Figure A.15 Coherence of colour names for being members of primary class (n=7b).

The third primary class considered was the three primaries of Maxwell (1872): *red*, *green* and *blue*. Coherence was quantified by MAP score as 0.33. *Pink* was a false positive found at the top of the confidences ranking, with *green* in the 2nd place and *blue* in the 4th. *Red* was found in the 17th position (Figure A.16).
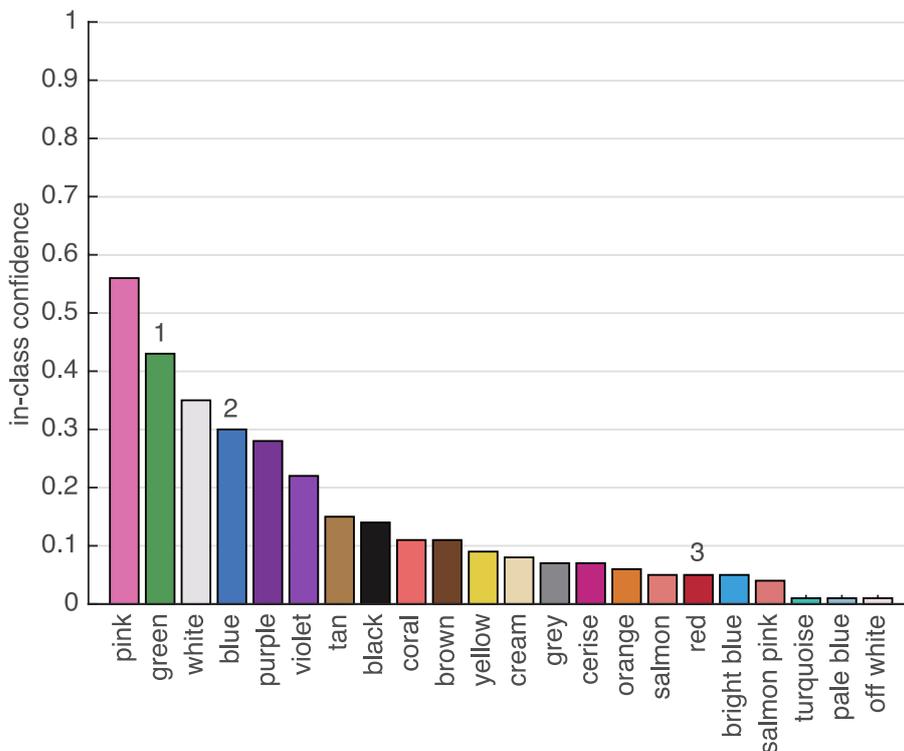
Figure A.16 Coherence of colour names for being members of Maxwellian primary class (n=3).

The fourth primary class considered was the more recent suggestion by Eskew (2009) for six unipolar labelled mechanisms: *red, orange, yellow green, blue* and *purple*. The classifier produced a MAP score of 0.62. *Pink* and *brown* were found again at the top of the coherence rank and *grey* in the 6th position (Figure A.17).
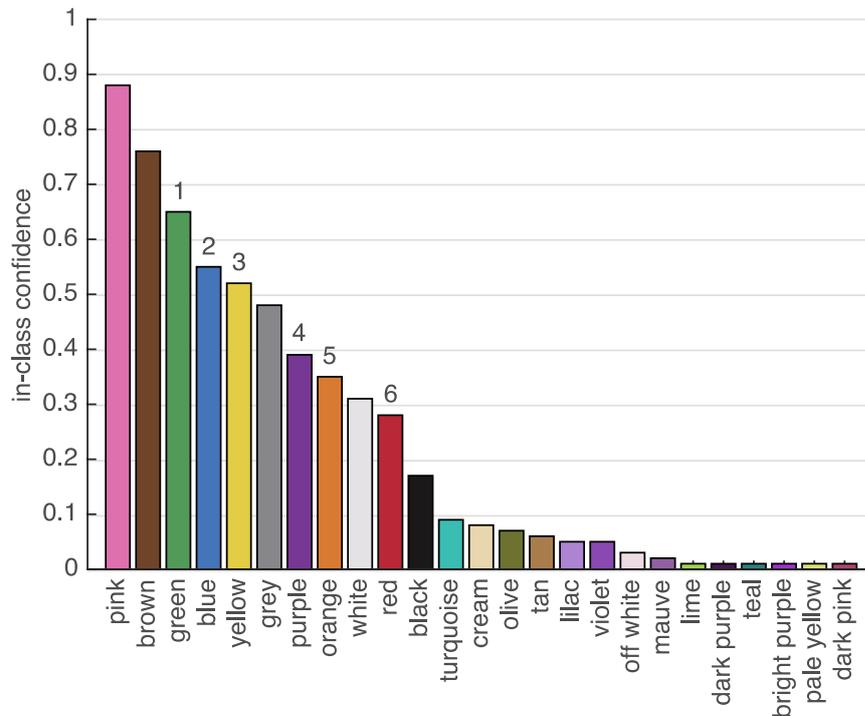
Figure A.17 Coherence of Eskew's primary class (n=6b).

The final primary class considered was the recent suggestion (Skelton et al., 2017) for five biological mechanisms – *red, yellow, green, blue* and *purple* - reported in infants' categorization of colour. The coherence assessment of this class produced a MAP score of 0.40 with *green* and *blue* found in the 3rd and 4th position, *purple* in the 8th, *yellow* in the 9th and *red* in the 11th position. *Pink* and *brown* were found at the top of the rank (Figure A.18).
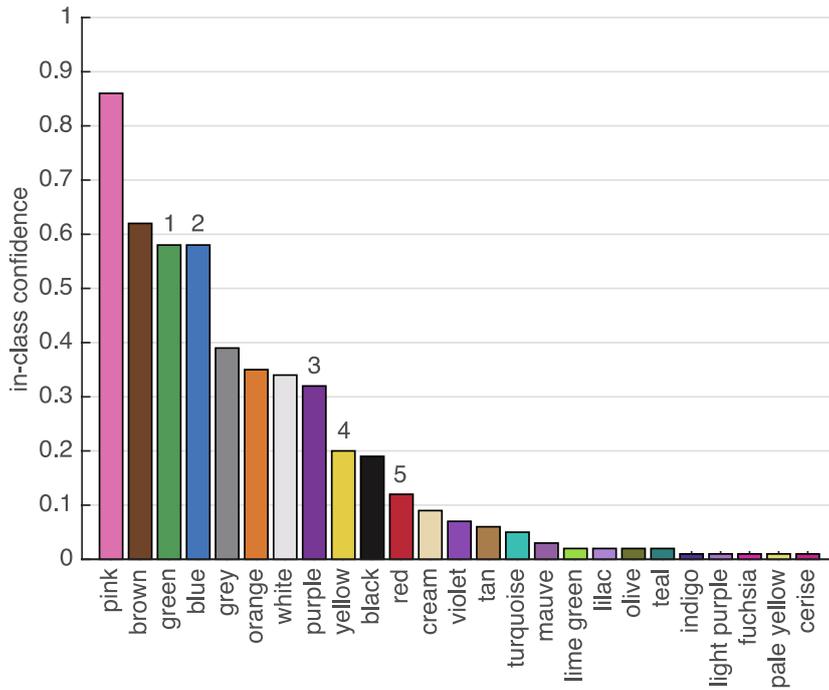
Figure A.18 Coherence of Skelton's primary class (n=5).

To determine the typical MAP scores for classes consisting of six member colours, we selected random sets of six colours from the 73 most common and measured the coherence of these random classes. The mean MAP score for five random classes was 0.13. In Figure A.19, we show for example the random class with the highest MAP score (0.33) consisting of *dark brown, bright blue, sky blue, bright purple, terracotta* and *indigo*. *Bright blue* and *bright purple* were found in the two top positions whilst *terracotta* was given a class confidence of 0 and it is not shown in the figure.
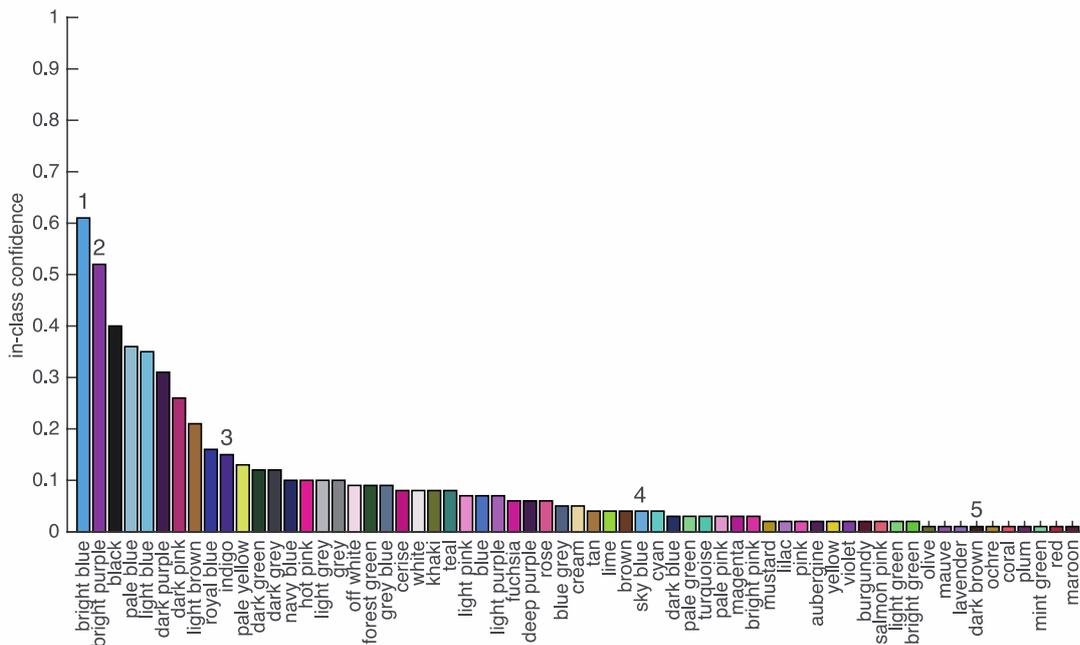
Figure A.19 Coherence of a random class (n=6).

We also accessed the coherence of equally sized secondary classes consisting of five secondary basic terms (*brown, purple, pink, orange* and *grey*) plus one of Hering's primaries. The mean MAP score of these classes was 0.53 similar, if not higher, to the MAP score of primaries. For example, in Figure A.20 we show the coherence for a secondary basics plus *green* class that produced a MAP score of 0.67. *Yellow* and *blue* were given the highest confidence followed by *pink, brown* and *green. Grey* was the in-class member with the lowest confidence.
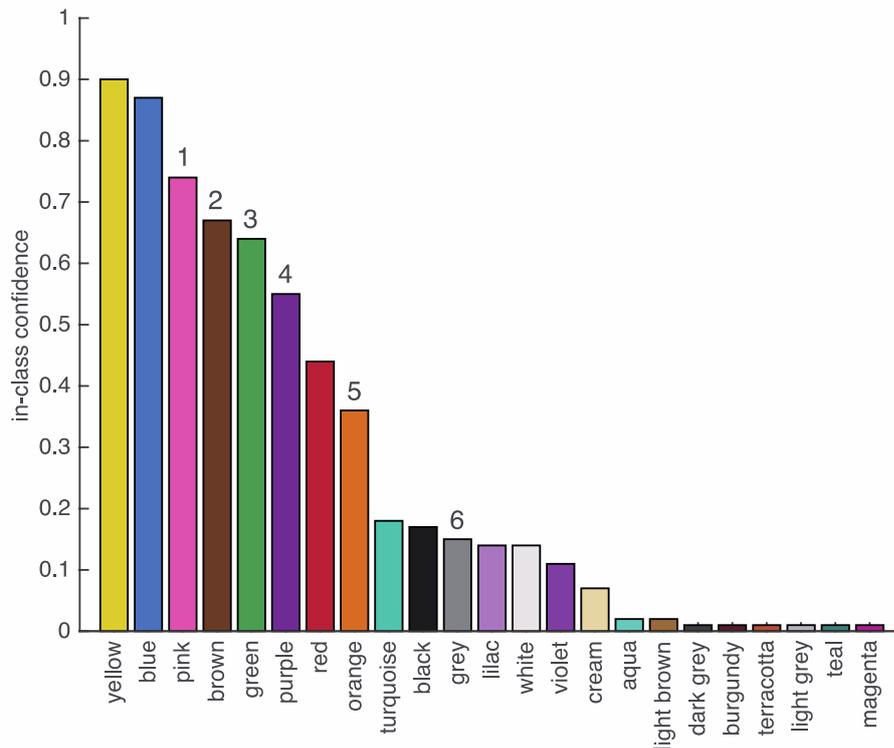
Figure A.20 Coherence of secondary basic terms class plus Green (n=6).

Appendix B.5 Coherence of basic classes

To examine the coherence of the basic class if we would just add the 12th term with the highest in-class confidence we assessed the coherence of the basics plus *olive* (see Figure 14). The coherence deteriorated from a MAP score of 1.00 to a MAP score of 0.917 because *cream* is assessed as more confidently in class than *olive* (Figure A.21). We also examined all 62 non-basics as an additional 12th basic colour. *Cream* was the only term that produced a MAP score of 1.00 but with a much lower confidence value than the other 11 basic terms (Figure A.22). All other colour names produced a MAP score of 0.917 or less by predicting all the 11 basic terms correctly except their self.
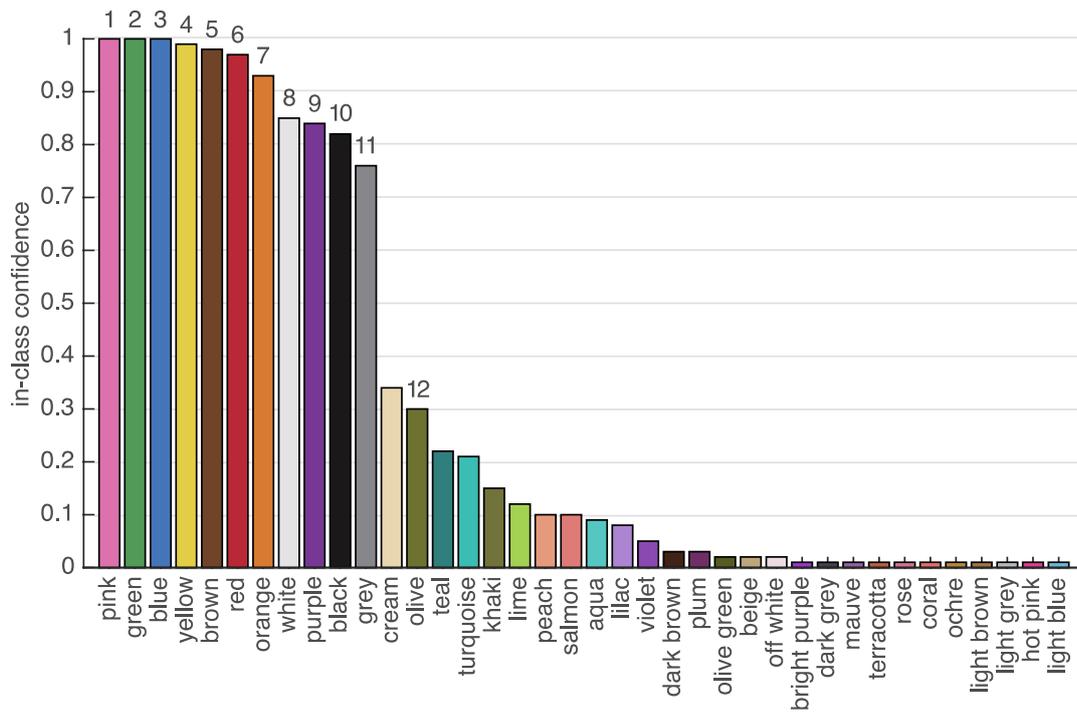
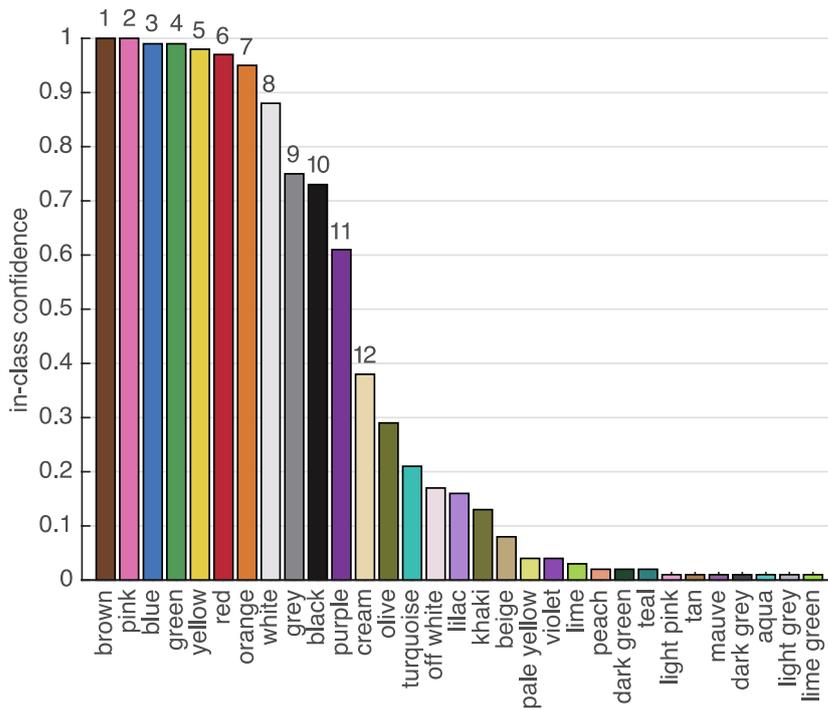*Figure A.21 Coherence of the basics plus olive (n=12).*

Figure A.22 Coherence of the basics plus cream (n=12).

Appendix C Supplementary Material for Chapter 5

In this section, we present the performance of the MAP model (Mylonas, MacDonald & Wuerger, 2010).
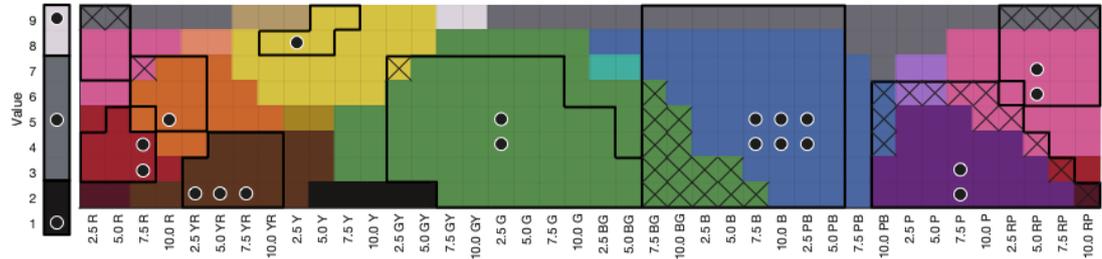

Figure A.23 Segmentation of simulated Munsell array into monolexemic colour terms by MAP model (Mylonas et al., 2010). Berlin and Kay's foci of BCTs in American English are drawn with dots and their distribution with black boxes.
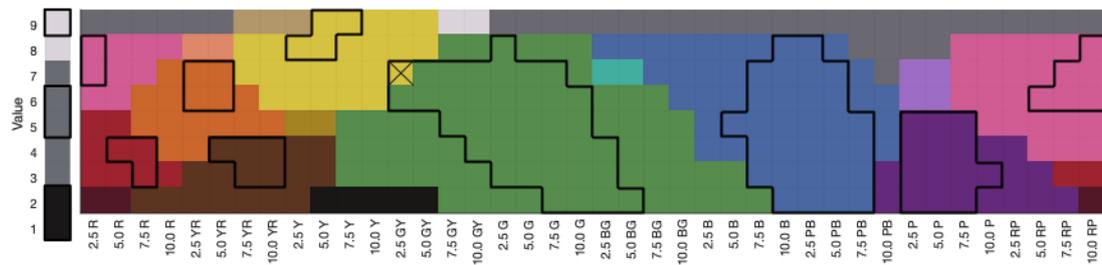

Figure A.24 Segmentation of simulated Munsell array into monolexemic colour terms by MAP model (Mylonas et al., 2010). Sturges & Whitfield's mapping of BCTs in British English are drawn with black boxes.

Appendix D Supplementary Material for Chapter 6

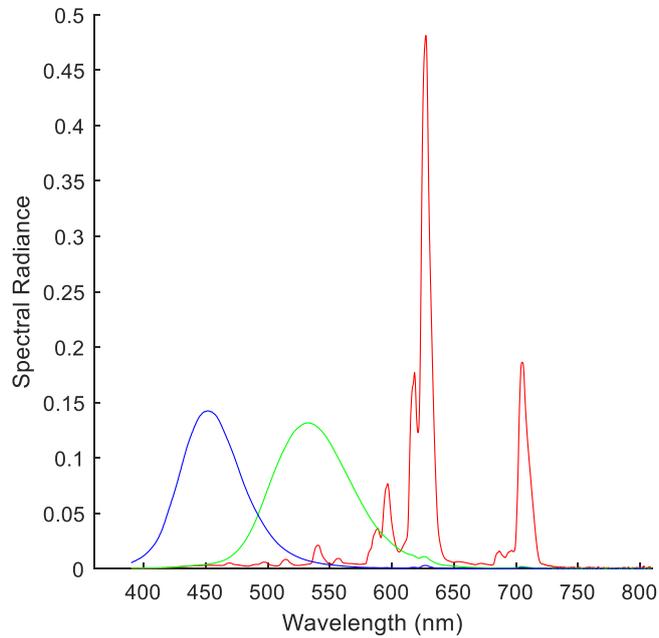In this section, we present the colourimetry and linearity test of the CRT display monitor.

Figure A.25 Spectra power distribution of red, green and blue phosphors of CRT monitor
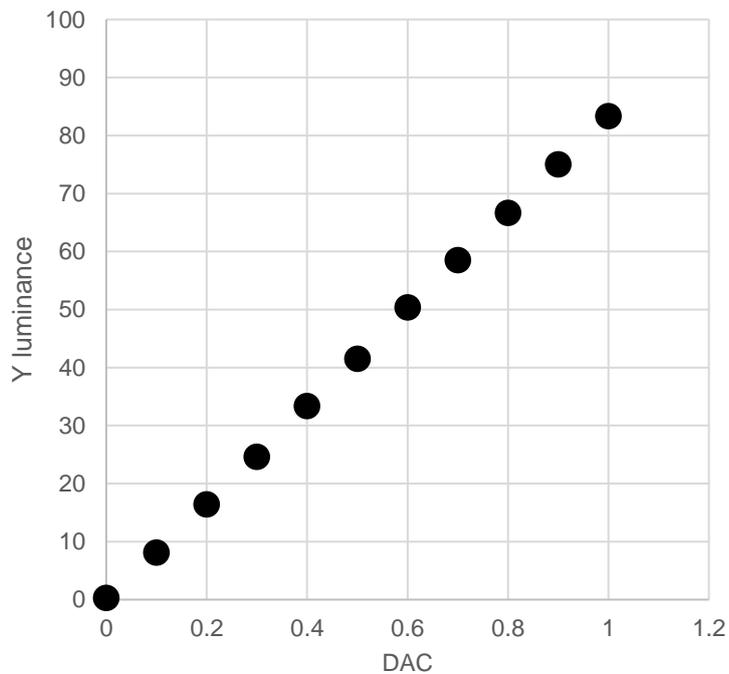
using a Radoma spectro-radiometer.

Figure A.26 Linearity test of calibrated CRT monitor measured by a ColorCal CRS

colorimeter produced a linear fit of R^2= 0.9999.