The Origin and Nature of Perceptual Colour Categories

Oliver Wright

Thesis submitted for the award of Doctor of Philosophy
August 2006

Department of Psychology
University of Surrey
Abstract

This thesis investigates the relationship between colour perception and colour language. Three questions are addressed. The first relates to the influence of colour category membership on colour perception, the second to potential influences of colour language on colour perception. The third question asks whether non-linguistic differences in colour perception may influence differences in colour language.

Experimental studies provide evidence for categorical colour perception. However, unlike previous studies, the experiments reported here found no evidence that linguistic colour categories underpin, or influence, colour perception. The reasons for this difference are discussed. The findings support the view that linguistic colour categories reflect non-linguistic perceptual colour categories. Finally evidence is presented suggesting that physiological differences in colour perception may influence the colour categories encoded by languages.
Acknowledgements

I would primarily like to acknowledge the generous support, help, advice and encouragement received from the people who have supervised this thesis. Ian Davies was my principal supervisor for all three years. Anna Franklin helped supervise the last two years and Emre Özgen the first year. The form and content of this thesis owe much to their contributions.

I am also grateful to many other people who have contributed, in greater or lesser degree and in varying ways, to the work presented here. Space precludes naming more than a few who have provided assistance of a direct and tangible kind. The following spring to mind: Christine Daoutis, Kemuu Jackarama, Humberto Moreira, Nigel Woodger, Terry Thacker, Kazuvire Veil, Verity Hobbs, Ben Dean, Natalie Rundell and Chris O'Neill.

The work was supported by various grants and bursaries. Surrey University Psychology Department provided a bursary in the first year. Subsequently the work was funded by the Economic and Social Research Council. An additional grant from the British Psychological Society helped fund some of the experiments. I am grateful to all of these institutions.

Finally I would like to thank my family, especially my parents, for their patience with me over all these years.
## Chapter 1

**General Introduction: Categories in Colour Language and Perception**

1.1. Categorical Perception

1.1.1. Theories of CP

1.2. Colour Language and Categorisation

1.2.1. Historical Perspectives

1.2.2. Recent Developments

1.2.2.1. Cross-Cultural Evidence Challenging the Universalistic View

1.2.2.2. Further Cross-Cultural Evidence for Universalism

1.2.2.3. Evidence from Infants and Young Children

1.3. Colour Categories and Colour CP

1.4. Other Potential Cross-Cultural Influences on Colour Perception

1.5 Summary

1.6. Overview of the Physiological Basis of Colour Vision

1.7. Colour Spaces

1.8. Outline of Experimental Chapters

## Chapter 2

**Colour Categorical Perception**

2.1. Introduction

2.1.1. Evidence for Colour Categorical Perception from Tasks with an Inter Stimulus Interval
2.1.1.1. Verbal Labelling 21
2.1.1.2. Category Adjustment Model 22
2.1.2. Evidence for Colour Categorical Perception from Other Tasks 26
2.1.3. Summary and Overview of Experiments 1 and 2 29

2.2. Experiment 1: No-Delay X-AB Task Measuring Hue Just Noticeable Differences 32
2.2.1. Introduction 32
2.2.2. Method 33

Participants
Apparatus, Stimuli and Design
Procedure
2.2.3. Results 35
2.2.4. Discussion 37

2.3. Experiment 2: Temporal 2-AFC Detection Task Measuring Hue Just Noticeable Differences 39
2.3.1. Introduction 39
2.3.2 Method 39

Participants
Apparatus, Stimuli and Design
Procedure
2.3.3. Results 41
2.3.4. Discussion 44

2.4. General Discussion of Experiments 1 and 2 45

Chapter 3
Colour Language and Colour Categorical Perception 48
3.1. Introduction 48
3.1.1. Perspectives on the Debate 49
3.1.2. Evidence for Cross-Linguistic Differences in Colour Cognition and Perception 51
3.1.3. Summary and Overview of Experiments 3-7 54
3.2. Experiment 3: X-AB task: English and Himba Two-to-Four Year Olds 56
3.2.1. Introduction 56
3.2.2. Method 57
Participants
Stimuli, Apparatus and Design
Procedure
3.2.3. Results 59
3.2.4. Discussion 61

3.3. Experiment 4: Target Detection Task: English and Himba
Two-to-Four Year Olds 61
3.3.1. Introduction 61
3.3.2. Method 62
Participants
Apparatus and Experimental Set Up
Stimuli and Design
Procedure
3.3.3. Results 63
3.3.4. Discussion 65

3.4. Experiment 5: Target Detection Task: English and Himba
Adults 65
3.4.1. Introduction 65
3.4.2. Method 67
Participants
Apparatus and Experimental Set Up
Stimuli and Design
Procedure
3.4.3. Results 67
3.4.4. Discussion 69

3.5. Experiment 6: Just Noticeable Difference Task: English and Himba Two-to-Four Year Olds 70
3.5.1. Introduction 70
3.5.2. Method 71
Participants
Apparatus
Chapter 3

Stimuli and Design
Procedure
3.5.3. Results 75
3.5.4. Discussion 77

3.6. Experiment 7: Just Noticeable Difference Task: English and Himba Adults 77

3.6.1. Introduction 77
3.6.2. Method 78

Participants

Apparatus

Stimuli and Design
Procedure
3.6.3. Results 80
3.6.4. Discussion 82

3.7. General Discussion of Experiments 3 to 7 82

Chapter 4

Lens Brunescence and Colour Perception 86

4.1. Introduction 86

4.1.1. Changes in Colour Perception and Changes in Ocular Media Associated with Aging and Exposure to UV-B light 86
4.1.2. Overview of Experiments 91

4.2. Experiments 8a-e 93

Participants

Apparatus

4.2.1. Experiment 8a: Temporal Achromatic Measurement

Luminance Adjustment Task (AMLA) 93
4.2.1.1. Introduction 93
4.2.1.2. Method 94

Stimuli and Design
Procedure
4.2.1.3. Results 98
4.2.1.4. Discussion 99
Chapter 5

General Discussion

5.1. Summary of Research Questions and Findings

5.1.1. The influence of colour category membership on colour perception

5.1.2. Potential influences of colour language on colour perception

5.1.3. Do differences in colour perception influence differences in colour language?

5.2. Implications for Differing Theoretical Perspectives

5.2.1. Categorical perception and perceptual learning

4.2.2. Colour categories and universalist-relativist debate

5.3. Suggestions for Future Research

5.4. Conclusions

References

Appendices

Appendix 1
Appendix 2
Appendix 3
Appendix 4
Appendix 5
## List of Tables

**Table 2.1** Mean Munsell hue JNDs from Experiment 1.  
**Table 2.2** Mean Munsell hue JNDs from Experiment 2.  
**Table 3.1** Mean errors for across and within-category trials in Experiment 3.  
**Table 3.2** Mean response latencies in milliseconds for measurement locations used in Experiment 4.  
**Table 3.3** Mean response latencies in milliseconds for measurement locations used in Experiment 5.  
**Table 3.4** Mean JNDs in L*a*b* units for measurement locations used in Experiment 7.  
**Table 4.1** CIE (1976) u* v* values of stimuli used in Experiment 8a.  
**Table 4.2** u* v* coordinates of measurement locations in Experiment 8b.  
**Table 4.3** Mean ΔE JNDs for over 50 and under 30 groups in Experiment 8b.  
**Table 4.4** ΔE distances (L*u*v*) between adjacent stimuli pairs used in Experiment 8c.  
**Table 4.5** Mean proportional use of term blue to name stimuli in Experiment 8d.
List of Figures

**Figure 1.1** Berlin and Kay's (1969) proposed sequence of basic colour terms acquisition in languages.  
7

**Figure 2.1** The effect of shift towards prototype.  
23

**Figure 2.2** The effect of shift towards prototype on across and within-category trials in an X-AB task.  
23

**Figure 2.3** An example showing how shift towards prototype predicts order effects for within-category trials on a same-different task.  
25

**Figure 2.4** JNDs at different wavelengths.  
30

**Figure 2.5** Graphic representation of a $4^{th}$ order polynomial (quartic) polynomial function.  
32

**Figure 2.6** Bi- and uni-directional JND measurements.  
34

**Figure 2.7** Experiment 1: JNDs in Munsell hue units for X-AB task.  
36

**Figure 2.8** Mean JNDs from Experiment 1 with $4^{th}$ order polynomial function fitted.  
37

**Figure 2.9** Stimuli display in temporal 2-AFC detection experiment.  
40

**Figure 2.10** Schematic representation of trial in temporal 2-AFC detection experiment.  
42

**Figure 2.11** Experiment 2: JNDs in Munsell hue units for temporal 2-AFC detection task.  
43

**Figure 2.12** Mean JNDs from Experiment 2 with $4^{th}$ order polynomial function fitted.  
44

**Figure 3.1** Within-category and across-category stimuli pairs used in Experiment 3.  
58

**Figure 3.2** English and Himba two-to-four year olds. Mean errors for within and across-category pairs, in Experiment 3.  
60

**Figure 3.3** Mean target detection task response latencies in milliseconds (±1 SE) for Himba and English two-to-four year olds in Experiment 4.  
64

**Figure 3.4** Mean target detection task response latencies in milliseconds (±1 SE) for Himba and English adults in Experiment 5.  
68

**Figure 3.5** Monochromatic example of an item used in Experiment 6.  
72
Figure 3.6 Conceptual diagram illustrating the four stimulus sets used in Experiment 6.

Figure 3.7 Example of up-down procedure used to measure JNDs in Experiment 6.

Figure 3.8 Mean JNDs for English and Himba two-to-four year olds for within-category and category boundary sets in Experiment 6.

Figure 3.9 Example of an item from Experiment 7.

Figure 3.10 JND measurement sets used in Experiment 7.

Figure 3.11 Mean JNDs for Himba and English adults in Experiment 7.

Figure 4.1 Stimuli used Experiment 8a plotted on CIE (1976) u* v* diagram.

Figure 4.2 Trial design in Experiment 8a.

Figure 4.3 Results of temporal AMLA task used in Experiment 8a.

Figure 4.4 JND measurement locations used in Experiment 8b.

Figure 4.5 Mean JNDs for over 50 and under 30 groups in Experiment 8b.

Figure 4.6. Hypothetical example of subject weight plot for two-dimensional INDSCAL solution.

Figure 4.7. Location of stimuli used in Experiment 8c in CIE (1976) u*v* space at three different L* levels.

Figure 4.8. Example of display used to obtain similarity judgement data in Experiment 8c.

Figure 4.9. Stimulus space derived from INDSCAL analysis of participant similarity matrices in Experiment 8c.

Figure 4.10. Plot of participant weights from INDSCAL analysis of similarity matrices: Experiment 8c.

Figure 4.11. Location of stimuli used in Experiment 8d in CIE (1976) u*v* space at three different L* levels.

Figure 4.12. Estimated optical density for a twenty-seven year old in simulated lens brunesence colour naming experiment (8e) with and without filter.

Figure 4.13. u* v* values of stimuli named in Experiment 8e as viewed through yellow filter.

Figure 4.14. Proportional use of 'blue' term to name stimuli in Experiment 8e.
Chapter 1
General Introduction: Categories in Colour Language and Perception

Colour cognition and perception is of great interest to psychologists because it offers an opportunity to address a number of issues fundamental to human cognition. Amongst these are issues relating to the nature of categories — how is it that we are able to segment the stream of sensory input into distinct and meaningful groups? Related to this is the relationship between thought and language — do languages influence mental processes to the extent that different languages lead to different mental processes? Studying colour perception and cognition also embraces the issue of the flexibility of neural processes — does learning to chunk stimuli into categories, for example by learning to name a particular range of colours with the same word, alter low level perceptual processes?

The reason that colour offers such potentially valuable insights stems from the way in which the continuously varying spectrum of electromagnetic radiation comes to be perceived, not as a continuum, but as a series of distinct groups. English speakers name these groups using terms like blue, red, green, purple and so on. Other languages divide colour space in different ways, using different numbers of colour terms. The most appropriate interpretation of this fact, and its connection to the categories we actually see when looking at, say, a rainbow lies at the heart of the work presented here. It is an oversimplification, but one that serves useful purposes, to suggest those who have investigated the issue have taken one of two opposed perspectives. On the one hand relativists suggest cross-cultural variations in colour language not only reflect but cause real differences in colour perception. On the other, universalists suppose that colour vision is the same across languages and cultures.

Bearing this in mind, this thesis may be seen as addressing three main questions. The first relates to the influence of colour category membership on colour perception. The phrase colour category membership used here includes not only the idea of linguistic colour categories, but also universal, ‘hardwired’ colour categories. The second question relates to the potential influence of colour language on colour perception. Both of these questions connect with a further issue — that of categorical perception,
of which more will be said shortly. The third question to be addressed relates to the effects of certain physiological processes on colour perception and the possible links between them and cross-cultural differences in colour perception and colour language.

The first chapter provides a more detailed outline of the issues summarised above, along with an overview of relevant research and competing theories. Chapter 1 also gives a brief outline of the physiological basis of colour vision and the colour order systems, or colour spaces, used in studies of colour perception. Chapter 2 addresses the first of the three questions stated above: do colour categories influence colour perception? In Chapter 3 the second question, relating the possible influence of colour language on colour perception, is investigated in cross-cultural studies that compare speakers of languages with different colour language structures. In Chapter 4 speakers of different age groups are compared to assess the last of the questions posed above.

1.1. Categorical Perception

Categorical perception (hereafter CP) refers to a qualitative difference in how things appear that depends on category membership. In experimental terms CP typically manifests itself as faster and/or more accurate discrimination between stimulus pairs drawn from different categories than between pairs belonging to the same category (Harnad, 1987).

\[ A_1 \leftrightarrow A_2 \leftrightarrow B \]

Here stimuli A1 and A2 belong to the same category, while stimulus B belongs to a different category. Stimulus A2 is equally separated from both A1 and B, and CP would be demonstrated if more accurate/faster discrimination occurs for pair A1 – A2 than A2 – B. Note that CP as defined here implies neither 0% within-category discriminability nor 100% identification accuracy for stimuli from different categories; merely that performance for stimulus pairs straddling the boundary will be superior to that of within-category stimulus pairs.

CP has been demonstrated in a wide range of domains, for example in perception of phonemes (e.g., Liberman, Harris, Hoffman & Griffith, 1957; Burnham, Earnshaw &
Clark, 1991), of non-speech sounds (e.g., Cutting & Rosner, 1974; Pastore, Li & Layer, 1990), line length (Tajfel & Wilkes, 1963), familiar objects (Newell & Bülthoff, 2002) and in connection to various aspects of perception relating to faces, (e.g., Etcoff & Magee, 1992; Campbell, Woll, Benson & Wallace, 1999; Levin & Angelone, 2002; Rossion, Shiltz, Lawrence, Pirenne & Grommelinck, 2001, Campanella, Chrysochoos & Bruyer, 2002).

CP has also been demonstrated in colour perception using a variety of different tasks such as X-AB (e.g., Roberson & Davidoff 2000; Roberson, Davies and Davidoff, 2000; Pilling, Wiggett, Özgen & Davies, 2003) and same-different (e.g., Bornstein and Korda, 1984; Boynton, Fargo, Olson & Smallman, 1989) tasks. On same-different tasks, a target stimulus is shown for a limited time, removed and after an interval, a test stimulus shown. The task is to decide if test and target stimuli are same or different. X-AB tasks are similar, except the test stimulus is always identical to the target stimulus and is accompanied by a foil which is different. The task here is to identify the test stimulus. Results of ‘odd-one-out’ triad tasks have also been taken to provide evidence of colour CP (e.g., Kay & Kempton, 1984, Laws, Davies and Andrews, 1998; Davies and Corbett, 1997). Here the task is to choose, from a set of three stimuli, the one that is the most different. Further evidence of colour CP comes from various target detection tasks. Colour visual search tasks involve searching among objects for a specific object, for example, a blue target object in a field of green and purple distractor objects (Yokoi & Uchikawa, 2005; Daoutis, Franklin, Riddett, Clifford & Davies, 2006a; Daoutis, Pilling & Davies, 2006b). More simply, a target object of one colour may be embedded against a uniform background of differing colour (e.g., Franklin, Pilling and Davies, 2005a).

1.1.1. Theories of CP

Although a broad stream of evidence indicates that CP is a genuine phenomenon, there is no clear consensus as to its origin and nature. However, several accounts seeking to explain CP have been put forward.

CP may be an innate phenomenon. Work with young infants has shown they can perceive colour in a categorical fashion (Bornstein, Kessen & Weiskopf, 1976,
Franklin & Davies, 2004) and discriminate between coloured stimuli better when they belong to different categories (Franklin et al. 2005a). Macaque monkey also seem to categorise the spectrum in a similar way to humans (Sandell, Gross & Bornstein, 1979). Similarly, chinchillas appear to show CP for human speech phonemes (Kuhl & Miller, 1978). Findings such as these imply natural discontinuities exist in colour space and audition, which serve as convenient locations for category boundaries.

A second view is that learning may ‘warp’ the representation of perceptual space, perhaps by stretching of perceptual space around category boundaries, compression of perceptual space at the centre of categories, or both (Harnad, 1987). This would produce CP. ‘Warping’ of perceptual space would aid categorisation by making objects in different categories more distinct from one another. Indeed, an AI simulation of category learning found that warping of similarity space was necessary for a neural network to learn to categorise (Harnad, Hanson & Lubin, 1991). Perceptual learning experiments have also provided evidence supporting the possibility that perceptual space may be ‘warped’ (e.g., Goldstone, 1994; Livingstone, Andrews & Harnad, 1998, Guenther, Husain & Shinn-Cunningham, 1999, Özgen & Davies, 2002). In adult primates perceptual learning experiments result in observable neural change in the visual cortex (Recanzone, Schreiner & Merzenich, 1993).

A third view is that CP is caused, at least in part, by labelling of stimuli (e.g., Kay and Kempton, 1984; Roberson & Davidoff, 2000; Rosen & Howell, 1987). Discriminations between stimuli are, by this account, based on comparison either of labels, or of a label, to a perceptual representation. Discrimination between perceptually different stimuli belonging to the same category is less likely to succeed than when stimuli belong to different categories because in the former instance the labels are more likely to match.

None of the three accounts outlined above can satisfactory account for all the data relating to CP. The emergence of CP through category learning is difficult to explain if CP is entirely innate. For example, Özgen and Davies (2002) taught participants to split existing colour categories into two new categories and subsequently found CP effects across the newly learned boundary. Theories that implicate learning of
categories as the source of CP struggle to explain findings such as colour CP in infants and CP for phonemes in chinchillas. Labelling theories have similar difficulties, particularly if the labels are assumed to be verbal. Labelling theories of CP also differ from the other two theories in that they do not directly implicate perceptual processes in CP. However, unless indicated otherwise, use of the term CP in subsequent sections may be taken to indicate behavioural differences, specifically better performance on across than within-category trials, and will not be used to imply theoretical differences.

1.2. Colour Language and Categorisation

As stated in the general introduction, different languages divide colour space in different ways using different numbers of colour terms. In this section differing universalist and relativist interpretations will be examined in more detail and a link will be made between issues of colour language/categorisation and CP.

1.2.1. Historical Perspectives

Variation in colour language was brought to prominence in the 19th century by several authors, amongst them Geiger (1871) who noted that the Rig-Veda, a collection of Sanskrit hymns composed between 1500-1300 BC, contained extremely few words for colour. A similar paucity of colour language was apparent in other ancient texts, a pattern which led Geiger to suppose that only a vague sense of colour existed in the earliest literary times. It was subsequently discovered that many current societies used a colour vocabulary considerably more restricted than those of European languages. Magnus (1877, 1880, cited in Woodworth, 1910) took the important step of making initial comparisons between such groups and Europeans, finding that despite the differences in colour vocabulary, the limits of colour vision did not appear to show corresponding differences. Somewhat later, Rivers (1901a, 1901b), recorded a range of different 'stages' in a languages encoding of colour terms, an idea later to be extended by Berlin and Kay (1969). Rivers also claimed that sensitivity to certain colours, in particular to blue, did differ in the societies he tested.
Both of Rivers’ claims appear to have been largely forgotten for a number of years. Woodworth (1910) concluded that variations in colour language were explicable in terms other than variations in colour perception, whilst by the 1960s there was a general consensus that languages partitioned colour space on an arbitrary basis. This view stemmed naturally from interpreting earlier observations of colour language differences in the light of the linguistic relativity hypothesis (hereafter LRH) most strongly associated with Whorf (1956). Though influential, the LRH was never clearly defined in Whorf’s work but what he suggests is that different languages, rather than serving to passively map ideas in fact play an active part in shaping them. Different languages lead to different conceptual structures and thus an idea expressed in one language may be impossible to translate meaningfully and accurately into another language. In other words, there is no one-to-one correspondence between concepts in different languages. In essence then our reality is shaped by the language we speak. Expressing the idea in its strongest terms, Whorf writes, *This fact [that we cannot talk except by subscribing to the organisation and classification of data which the agreement (i.e. linguistic structure) decrees] is very significant for modern science, for it means that no individual is free to describe nature with absolute impartiality.¹* In this interpretation, the LRH suggests that language determines thought. Under such circumstances, it is impossible to imagine how comparisons between languages can be made and thus difficult to see how the theory could be tested. A weaker version of the hypothesis is that the language we speak may influence our perception. This leads to the prediction that variations in colour language ought to lead to variations in colour perception. This echoes Rivers earlier claims though, crucially, linguistic differences cause here perceptual differences rather than passively mapping them. A weaker claim still is that linguistic relativity only influences higher cognitive functions. The two weaker versions of the LRH can and have been tested and it is easy to see why colour perception, which transforms the continuous spectrum of electromagnetic radiation into a discrete perceptual categories has been used as a vehicle to do so.

Earlier studies using colour stimuli appeared to confirm weaker versions of the LRH. Brown and Lenneberg (1954) found a correlation between the codability (a measure

¹Science and Linguistics (In Language, Thought and Reality: selected writing of B. L. Whorf, pp135-137)
of inter-participant naming agreement among other things) of coloured stimuli and recognition accuracy. Lantz and Steffle (1964), rather than measuring codability using intra-participant agreement, defined it in terms of communication accuracy and found it too was a good predictor of colour memory. Again, this was taken as evidence in support of the LRH.

In 1969 though Berlin and Kay's (1969) theory led to a fundamental shift in which relativist accounts of colour language were eclipsed by universalist accounts. Of great importance to the theory was a division of colour terms into those that were considered basic (BCTs), and those that were not. Basic colour terms, by Berlin and Kay's definition, exhibited four distinguishing characteristics; monolexemity, signification, application and psychological salience. Armed with this definition eleven basic colour terms were identified, corresponding to English terms; black, white, grey, blue, red, green, yellow, purple, orange, pink and brown. Examination of colour terms used in a range of other languages, containing between two and eleven BCTs suggested that the order in which languages acquired BCTs was constrained, proceeding though a number of distinct stages until the full complement of eleven BCTs was incorporated. In addition, best examples of BCTs, called focal colours, were also found to be similar across languages.

Figure 1.1 Berlin and Kay's (1969) proposed sequence of basic colour terms acquisition in languages.
The proposed order in which Berlin and Kay proposed languages acquire BCTs is shown in Figure 1.1. Languages at the first of the proposed evolutionary stages would always have terms corresponding closely to a lightness division, with a term encompassing red the next to appear and so on. As can be seen, even in the original model, there is limited scope for variation: yellow may precede green and vice-versa. Subsequently this sequence has been revised to allow for increased variation in the order of acquisition (see for example, Kay and McDaniel, 1978; Kay and Maffi, 1999), but the central idea – that colour term acquisition is constrained – remains. This suggests that linguistic segmentation of colour space is far from arbitrary, thereby constituting a significant challenge to the relativist position.

Subsequently, a series of studies by Rosch built on Berlin & Kay’s findings (Rosch Heider, 1972; Rosch & Oliver, 1972; Rosch, 1973). Rosch studied Dani speakers, a group whose language seemed to contain only two colour terms. In one experiment (Rosch Heider, 1972) showed that Dani speakers’ memory for the focal colours identified by Berlin and Kay was superior to that of other colours. This implied these colours were salient even if not marked in a participant’s language and therefore could be considered universal foci. In another experiment (Rosch, 1973) Dani speakers learned to name colour categories not marked in their own language. In some sets the focal colour was central, in others at the periphery of the to be learned categories. Dani speakers learned to name sets where the focal colour was central more easily than when it was located at the periphery. This led Rosch to claim that focal colours made learning new categories easier: further evidence for universalism in colour categorisation.

Both Berlin and Kay’s and Rosch’s work has been criticised. Berlin and Kay’s original data were all gathered in the USA and many languages were represented by only a single, bi-lingual speaker. Furthermore, seventeen of the 20 languages originally studied were written languages from industrialised societies, which detracted from the cross-cultural validity of the study. The definition of BCTs is also problematic, as it removes most of the world’s colour vocabulary (see e.g., Saunders and van Brakel (1997); Levinson (2001) for detailed critiques of BCTs). Rosch’s work has also been criticised, in particular with regard to experimental design (Lucy 1997, Ratner, 1989, Saunders and van Brakel (1997) and the interpretation of the data.
Roberson et al. (2000) for example argue Rosch & Oliver’s (1972) findings may equally well be interpreted to provide support to the relativist viewpoint. In spite of these criticisms, the net effect of Berlin and Kay’s and of Rosch’s work was that universalist accounts of colour cognition became widely accepted.

1.2.2. Recent Developments
Recently, there have been a number of challenges to the universalist account of colour categories and also responses to those challenges. Here, three separate strands of evidence are examined in turn: first there is a body cross-cultural work that challenges the universalist view. The second strand consists of a contrasting set of cross-cultural work that bolsters the universalist view. Finally work with infants and young children has also tended (though see Roberson, Davidoff, Davies and Shapiro (2004) for an exception) to support universalist claims.

1.2.2.1. Cross-Cultural Evidence Challenging the Universalist View
Davies and Corbett (1997) found that Setswana speakers, whose language has no separate blue and green colour terms, were more likely to group these colours together on a colour grouping task than English speakers. Similar differences between the same language groups were found using a triadic judgement task (Davies, Sowden, Jerrett, Jerrett & Corbett, 1998), providing evidence for a weak form of the LRH.

Roberson et al. (2000) compared performance of Berinmo, whose language has five basic colour terms, and English speakers, on a number of tasks similar to those used by Rosch (1972). Berinmo speaking participants showed no advantage in recognition memory for focal colours, nor did use of focal colours aid paired associate learning. CP effects were found, but only across boundaries marked in the participant’s language. English speakers showed CP effects across an English colour category boundary, but not across a category boundary marked only in Berinmo. The reverse was true for Berinmo speakers. On the basis of this Roberson et al. (2000) suggested that linguistic colour categories influenced perceptual judgements. A subsequent study (Roberson, Davidoff, Davies, & Shapiro, 2005) tested another group speaking a language containing only five BCTs, the Himba. Unlike Berinmo speakers who live in rain-forest, Himba people live in arid savannah. Himba speakers’ performance was unlike that of both English and Berinmo speakers, lending further weight to the
conclusions drawn by Roberson et al's (2000) earlier study. What is suggested by these studies is a strong form of the LRH, which to Davidoff (2001), "...indicates that perceptual categories are derived from the words in the speaker's language."  

1.2.2.2. Further Cross-Cultural Evidence for Universalism
In contrast to the above is the work of Kay and colleagues (Kay and Regier, 2003; Kay, 2005; Regier, Kay and Cook, 2005), largely based on data collected by the World Color Survey\(^3\) from 110 unwritten languages in non-industrialised societies. The aim of the World Color Survey was to address shortcomings in Berlin and Kay's original work. Analysis of the location of colour category centroids (the location calculated to represent the centre of a given colour category) in these languages suggests they share colour category centroids with each other, and with centroids for terms from industrialised languages far more than would be expected to occur by chance (Kay and Regier, 2003). Focal colours representing the best examples of named categories show a similar tendency to cluster around particular points in colour space (Regier, Kay and Cook, 2005). For example, Turton (1980) found that the Mursi people from Ethiopia whose colour terms are based on cattle colour naming tended to identify focal colours as best examples of their own colour terms despite, the author claims, probably never having seen such artificial stimuli before. Should languages arbitrarily divide colour space, it is difficult to explain these apparent similarities.

1.2.2.3. Evidence from Infants and Young Children
Children too young to speak lack knowledge of linguistic colour categories and investigations of colour perception and categorisation in young children are therefore relevant. Bornstein, Kessen and Weiskopf (1976), and Franklin and Davies (2004) used dishabituation and novelty preference techniques to show that infants at four-months old respond categorically to colour. Franklin Pilling and Davies (2005b) have shown infants fixate more rapidly on a green target presented against a blue background than against a background of a different green. In another experiment with toddlers, using an X-AB task, Franklin et al (2005a) found that toddlers speaking a language lacking separate blue and green colour terms made fewer errors when

---

\(^2\) Davidoff, 2001, pp 382
\(^3\) Data at http://www.icsi.berkeley.edu/wcs/data.html
stimulus pairs straddled the blue-green category boundary than when the stimuli in a pair belonged to the same (blue) category. These findings imply that even infants are able to perceive colours categorically. This suggests the perception of different colour categories cannot wholly be due to learning colour language. On the basis of a cross-cultural study examining colour cognition in children from two language groups with different colour language structures Roberson et al. (2004) suggested something different, namely that, "...an initially perceptually driven color continuum is progressively organized into category sets appropriate to each culture and language." 

Further evidence supporting the view that colour language may alter colour perception comes from Daoutis et al. (2006a) who compared performance on visual search tasks of children from cultural groups differing in their colour language. Participants performed searches in which targets belonged to a different category to distractors faster than searches where targets and distractors belonged to the same category, but the categorical advantage was only present if the boundary was marked in their language.

1.3. Colour Categories and Colour CP

Theories seeking to account for linguistic colour categories and theories of colour CP are closely intertwined. If learning colour language can alter perception, then this would support theories which suppose that learning is responsible for CP (e.g., Harnad, 1987), at least with respect to colour perception. Conversely evidence that named colour categories are constrained dovetails with theories of CP stressing natural discontinuities in perceptual space. Although the ideas presented in the two preceding sentences appear mutually exclusive, this is not necessarily the case. It is possible that an innate form of colour CP exists, as studies with infants would suggest, and that this may then be modified, to a greater or lesser extent, by later learning, as category training studies imply, resulting in 'perceptual reorganisation'. As with CP theories that stress the role of learning, theories of perceptual reorganisation

*Roberson et al., 2004, pp 554*
presuppose some plasticity in colour perception. Where they differ is in allowing the possibility of innate forms of CP.

A third theory of CP which implies that labelling plays a role is compatible with a weak form of the LRH in which language can influence higher cognitive functions. By this third account, cross-cultural differences in performance of tasks using coloured stimuli are real but do not reflect differences in low-level colour perception caused by different colour language. Differences corresponding to colour language differences would however be expected on tasks that do not measure perception directly, for example tasks that either require or are facilitated by use of memory and/or language. Performance of tasks that measure perception more directly should, if the above is assumed to be true, and if in addition there are no other cross-culturally varying influences on colour perception, be similar in spite of differences in colour language.

1.4. Other Potential Influences on Colour Perception

Assessment of the possible role of colour language on colour perception is complicated by the possible involvement of other factors. For example, although the basic structure of colour vision is believed to be common to all individuals, variations in spectral sensitivities of the three cone types (Jameson, Highnote & Wasserman, 2001) and the ratios of L to M cones (Wesner, Pokorny, Shevell & Smith, 1991) are known to occur although there is no evidence that these variations covary with culture. It has also been suggested different chromatic environments might lead to different chromatic processing (Macleod, 2003, Yendrikhovskij, 2001), though again there is no direct evidence to support this assertion.

A more likely cause of differences in colour vision and one that may covary between cultures lies in possible differences in sensitivity to short-wavelength light. Rivers (1901b) speculated that a mechanism similar to that responsible for melanin production in skin might cause increased pigmentation of the macula lutea, a retinal pigment. This would give it a yellowish appearance causing absorption of short-wavelength light. This explanation appears unlikely for the following reasons. First,
no evidence exists that *macula lutea* density increases closer to the equator and second, *macula lutea* density is thought to be controlled primarily by diet (Nussbaum, Pruett, & Delori, 1981).

Alternatively, decreased sensitivity to short-wavelength light may be due to lens brunescence — a natural accumulation of yellow pigment in the lens as it ages. Support for the lens brunescence hypothesis comes from two studies. Lindsey and Brown (2002) conducted a colour naming study in which lens brunescence was simulated and found that young participants’ ability to distinguish between green and blue stimuli decreased as the level of simulated brunescence increased. Bornstein (1973) showed that the occurrence of languages not encoding separate blue and green colour terms increases at tropical latitudes. This is significant because the process of lens brunescence may be speeded if the eye is exposed to increasing levels of UV-B light (Javitt & Taylor, 1994) and sunlight in equatorial regions is relatively higher in UV-B than that at other latitudes. Consequently it is reasonable to suggest, as Lindsey and Brown (2002) have done that decreased sensitivity to short-wavelength light caused by lens brunescence might account for the lack of separate green and blue colour terms in many languages. What counts against the lens brunescence hypothesis is the lack of separate blue and green colour terms in some languages spoken at northerly latitudes, such as Welsh (Lazar-Meyn, 2004) and Hardy, Frederick, Kay and Werner’s (2004) study indicating that both old and young observers name blue and green stimuli using virtually the same colour terms.

1.5. Summary

It would appear that there are three different ways to account for the relationship between colour language and colour perception.

First, some evidence suggests that colour perception varies between cultures and that these variations are linked to differences in colour language. Here, colour CP is either caused or moderated by colour language and differences in colour perception between different language groups ought to be apparent.
Second, other evidence implies that colour perception may be universal. The apparent diversity in colour naming across languages is misleading; in fact languages acquire colour terms in a constrained fashion consistent with a universal account of colour perception, as universal underlying colour categories are successively given names. In addition, the locations of particularly salient areas of colour space show consistency across languages. Colour CP is driven by the universal colour categories that drive the acquisition of colour names in languages.

Third, colour perception differences between different groups do exist, but these differences are caused by physiological rather than linguistic factors. Colour naming may reflect, rather than cause, these differences in colour perception.

1.6. Overview of the Physiological Basis of Colour Vision

Estimates as to the number of colours a normal human can distinguish vary between 1 million (Kaiser and Boynton, 1996) and 7 million (Zeki, 1993). This process begins when light enters the eye and passes in turn through the cornea, aqueous humour, lens and vitreous humour before reaching the retina where it is absorbed by three classes of cone photoreceptors differing in their spectral sensitivities. These three cone types form the basis of trichromatic colour vision, their existence first being proposed by Young (1801/1948), later by Helmholtz (1909/1962). Subsequent experiments appear to confirm that cones contain one of three different kinds of photopigment (Boynton, 1979). All three are sensitive to broad range of wavelengths; the total range covered being 380 - 750 nm. One of the photopigments is maximally sensitive to light of around 430nm (S-cones), another 530nm (M-cones) and the final one to light at 560nm (L-cones).

The next stage of colour vision is thought to involve the signals from the three kinds of cones being combined in various ways for processing by opponent mechanisms like those proposed by Hering in the late 19th century. Hering suggested there were four indivisible chromatic sensations; blue, red, green and yellow. Green and red could be imagined as opposite ends of one antagonistic mechanism, blue and yellow at opposite ends of another. Since they occupied opposite ends a single mechanism,
blue and yellow could not be signalled at the same time, the same for green and red. This explanation neatly accounts for why it does not seem possible to imagine either a yellowish-blue or a greenish-red and also why afterimages are of opponent colours. Thus if a green colour is stared at for long enough and the eyes are then diverted to a white background, a red afterimage will be perceived. Hue cancellation experiments have provided psychophysical evidence for Hering’s conjectures (Hurvich and Jameson, 1957). In such experiments the amount of light of one hue needed to cancel out perception of light of another hue is measured. The fact that pairs of lights corresponding to Hering’s primaries cancelled each other is strong evidence for colour opponency. In addition, evidence of opponent processing cells in ganglion cells and in cells in the lateral geniculate nucleus has also been found (De Valois, Abramov & Jacobs, 1966; De Valois & De Valois, 1975).

Although the evidence for trichromacy and some kind of opponent process organisation in the visual system is generally accepted, it is unlikely that there is a direct correspondence between opponent cells in the LGN and perception. If there were, the opponent end points of the cells identified should correspond to the unique hues, for example a red that appears to contain no trace of either yellow or blue. Instead, peak activity in these cells corresponds to colours described as, ‘cherry-red’, ‘chartreuse’, ‘teal’ and ‘violet’ (Jameson & D’Andrede, 1997). Consequently, more complex models of colour processing have been developed most of which suggest that colour processing takes place at different localities. De Valois & De Valois (1993), for example, revised their earlier opponent process theory to incorporate opponent processes in cells in the cortex. Zeki, (1980, 1983) proposed that areas V4 and V4a of the prestripate cortex were responsible for assimilating colour information from different localities to provide the ultimate experience of colour.

Despite much progress, understanding of the processes and physiological mechanisms surrounding colour vision is still uncertain. This uncertainty makes accounts seeking to tie linguistic accounts of colour categories to physiology unsure. Kay and MacDaniel (1978) for example proposed that regularities in early visual perception, particularly the wavelength selective opponent cells in the LGN may represent the basis of colour categories. But as is outlined above, peak activity of LGN cells does not match the perception of primary hues. More generally as Abramov (1997) points
out, physiology alone cannot tell us how visual information gives rise to the categorically defined percept of colour we experience.

1.7. Colour Spaces

Colour spaces perform two major functions. The first function is to allow colour appearance to be specified and therefore reproduced accurately. The second function is to allow quantification of the magnitude of perceptual differences between colours. A number of such systems exist, all of which have particular strengths and weaknesses. Examples are the Commission Internationale de l’Eclairage (CIE) colour spaces, the Munsell system, The Optical Society of America Colour Space (OCS) and the Natural Colour System (NCS). Stimuli used in the experiments presented in later chapters of this thesis were specified in either Munsell or CIE colour spaces. Appendix 1 gives more detail about both systems and also contains information about light sources and colour temperature. A few general comments about the two systems are given below.

Stimulus specifications in Munsell are described in terms of three attributes: hue, chroma and value (lightness). Hue is represented as an angular co-ordinate, value is a linear scale and chroma is a radial scale. In CIE colour spaces stimuli are defined using three Cartesian co-ordinates. One axis defines lightness, the other two define chromaticity. Euclidian distances between pairs of stimulus in two of the CIE colour spaces (L*a*b* and L*u*v*) represent their perceptual discriminability. With stimuli defined in Munsell space, the situation is more complicated. Differences along any one scale are perceptually uniform. Thus the perceptual difference between two stimuli of value 4 and value 5 is the same as between two stimuli of value 5 and value 6, but only provided the other pair of attributes, in this case hue and chroma, remain constant. An important consequence of this is that the perceptual uniformity of Munsell breaks down when changes are made to more than one of the attributes, which is not the case for CIE spaces.

In spite of this, Munsell colour space has an important advantage over CIE colour space deriving from the renotation of the Munsell system carried out by Newhall,
Nickerson and Judd (1943). Scaling experiments involving over forty observers and some three million colour judgements were carried out, which make Munsell, within the limits described above, highly uniform perceptually. CIE spaces lack such rigorous empirical grounding. Because of this, the experiments presented in this thesis use stimuli defined in Munsell colour space wherever possible.

It is also possible to define colour stimuli using wavelength based systems, which equate differences in colour with differing wavelengths of light. Such stimuli are not used here for several reasons. They do not contain broadband light and represent the most saturated instances of a colour and thus do not represent the majority of naturally occurring colour stimuli. Neither is it possible to create some colours, such as purples or browns with such stimuli. Finally, the human eye is not uniformly sensitive to fixed intervals in wavelength (Wright & Pitt, 1934).

1.8. Outline of Experimental Chapters

Chapter 2 provides a more detailed description experimental evidence for colour CP. It is suggested that because some of the tasks used to measure CP do not measure perception directly, the CP effects observed do not necessarily reflect a perceptual component in CP. Two experiments are then presented which provide evidence for colour CP using tasks that rely on low-level perceptual processes.

If colour language influences colour perception, then differences in colour perception between the linguistic groups ought to correspond to differences in colour language. This idea is tested in Chapter 3. In the introduction to the chapter, previous evidence suggesting that such differences exist is assessed. Particular attention is paid to the nature of the tasks used by previous studies and doubts are raised as to whether these tasks can actually be claimed to measure perceptual processes. Performance of two groups was compared, Himba speakers from Namibia and English speakers from the UK. The tasks used were similar to those used in Chapter 2 in that they measured low-level perceptual processes. There were however some differences between the tasks used in Chapters 2 and 3. In part these differences were unavoidable. It was not possible for example to use such complex data collection methods during fieldwork in
Namibia as were used to collect data in laboratory conditions in the UK. Using a range of different tasks also had advantages. Foremost amongst these was in demonstrating the robustness of experimental findings to provide converging evidence that there is a perceptual component to CP.

In Chapter 4 an alternative explanation seeking to account for differences in colour language is explored – that is the idea that differences in colour language may be influenced by differences in colour perception. In particular the idea that increased relative absorption of short-wavelength light by the eye may affect perception of green and blue colours is investigated. The experiments reported compared over 50 year old and under 30 year old samples of English speakers living in the UK. It is believed that the processes causing decreases in short-wavelength light reaching the retina are associated both with normal aging and with exposure to high levels of UV-B light, associated with equatorial regions. Consequently, differences in the performance of the groups (of UK residents) are cautiously referred to theories which link colour language differences to differences in colour perception on which language has no influence.
Chapter 2
Colour Categorical Perception

2.1. Introduction

Chapter 1 briefly outlined three competing theories of CP; those that suggest it is an innate phenomenon, those that stress the role learning plays and those that imply labelling plays a role in CP. Studies demonstrating that the colour category to which stimuli belong influences task performance were also discussed. These tasks provide evidence of colour CP and, because different tasks involve different demands, they have been used to evaluate different theories of CP. In the section below experimental evidence for colour CP is examined in more detail. The key question addressed is whether the results of these tasks are best explained by perceptual differences between stimuli, or whether they may be influenced by other factors. It is argued that because of the demands of many of the tasks, participants may employ strategies that do not involve making direct perceptual comparisons between stimuli. The term ‘strategy’ is used here neutrally with respect to intent. The strategies proposed may be automatic and used unconsciously. Because it is possible that strategies not involving direct perceptual comparisons are used, claims that colour CP is perceptual in nature may require qualification.

In the remainder of the chapter, two experiments are presented. The aim was to investigate whether CP is found on tasks that measure low-level perceptual processes. One was an X-AB task with no delay between test and target phases. The second was a temporal 2-AFC detection task. Evidence for CP from both tasks would provide converging evidence that CP can be found on tasks that tap perception more directly than many of the tasks previously used.

2.1.1. Evidence for Colour Categorical Perception from Tasks with an Inter-Stimulus-Interval

In one of a series of same-different experiments carried out by Bornstein and Korda (1984), participants were shown pairs of Munsell colours drawn from the blue and green regions of colour space. An inter-stimulus-interval (ISI) of 3000 ms was used.
Participants were required to respond ‘same’ if the two stimuli appeared to be physically the same and ‘different’ if they appeared physically different. ‘Different’ responses were made more rapidly when stimulus pairs belonged to different colour categories (a blue and green), than to the same colour category (two blues or two greens), even when perceptual spacing between the stimuli in each pair was equivalent. Bomstein and Korda suggested their results supported a model in which processing of stimulus information proceeds in two codes; physical and categorical. A similar model to account for performance in phoneme matching was earlier proposed by Pisoni and Tash (1974). However, Bomstein and Korda’s (1984) model differs in proposing that physical and categorical processing is carried out in parallel, rather than in series. Judgement in across-category trials is faster than in within-category trials, by this account, because the simultaneously available categorical information is useful in making the required discrimination. In within-category trials, this information is not useful and judgement must proceed on the basis of physical information alone.

Roberson and Davidoff (2000) used an X-AB task with an ISI of 300, 5000 or 10000 ms between target and test phases. Equally spaced pairs of Munsell colours were again used, but rather than measuring response latency, response accuracy was recorded. Length of delay did not affect accuracy. Regardless of delay, accuracy was higher when stimulus pairs straddled the blue-green category boundary (across-category trials) than when they belonged to the same category (within-category trials).

Both of the above experiments provide evidence for colour CP. But because these experiments involved ISIs of varying duration, it is not possible to rule out the possibility that performance reflects effects of memory and/or language during the ISI rather than perception. Posner and Keele (1967) found that on a task involving matching latter identities, same case letter pairs (e.g. A-A) were matched faster than different case letter pairs (a-A) with ISIs of less than 1500 ms. When ISIs were increased beyond this duration, the advantage disappeared. The code representing a stimulus’ physical properties thus seems to decay relatively rapidly, at least in category matching tasks. In tasks like that used by Roberson and Davidoff (2000) which involve physical rather than categorical judgements, there is more incentive to preserve the physical code. This may explain why length of delay did not influence
performance. But it still seems possible that the physical code decays faster than the
categorical code. After 5000-10000 ms ISIs, categorical codes may play a role in
performance, even of tasks that require physical, rather than categorical judgements.
It is thus problematic to claim that the categorical effects found in these experiments
arise directly from perceptual processes.

2.1.1.1. Verbal Labelling
In another experiment, Roberson and Davidoff (2000), filled ISIs of 5000 and 10000
ms with either verbal or visual interference. In visual interference conditions results
consistent with CP were again found. In verbal interference conditions however, CP
was abolished. This led the authors to claim that in some circumstances verbal rather
than visual coding is required to produce colour CP. Verbal interference presumably
disrupts memory for a verbal label, so if CP is caused by memory for verbal label,
verbal interference during an ISI may be presumed to interfere with CP. Thus when
shown a stimulus, participants retain a verbal label during the ISI, which is then
compared to the test and foil stimuli. If test and foil stimuli (on an X-AB task) belong
to different categories, the stored label should aid task performance. But if test and
foil belong to the same category, the label is less likely to allow a correct choice
because it is more likely to be shared by both test and foil stimuli. This would result in
better performance of across than within-category trials. The same explanation – that
participants remember verbal labels during the ISI – can also be used to account for
Bomstein and Korda’s (1984) results. This account is similar in many respects to the
parallel processing model put forward by Bornstein and Korda’s (1984) though places
more emphasises on a linguistic nature for the categorical code. Several authors have
reached similar conclusions to those drawn by Roberson and Davidoff (2000).
Munnich and Landau (2003) also suggest that CP may result from verbal labelling of
stimuli, whilst Boynton et al. (1989) suggest colour categorisation occurs when
colours must be remembered.

Cross-cultural experiments have shown performance appears to be influenced by
colour language. Roberson et al., (2000) and Roberson et al., (2005) has shown that
on X-AB tasks CP is only found when stimuli straddle a boundary marked in the
participant’s language. Such findings are consistent with the idea that labelling plays a
role in CP. However, they can equally be accounted for if it is assumed that
differences in perception exist between the groups tested. Pilling et al., (2003) also examined the idea that verbal labelling causes CP by replicating Roberson and Davidoff's (2000), experiment. In Pilling et al.'s (2003) study, the type of interference (verbal or visual) used on each trial was not predictable. The effect of this manipulation was that CP survived verbal interference. However the same researchers also showed that participants were able to retain target colour names across verbal interference filled ISIs. They concluded, like Roberson and Davidoff (2000), that colour CP shown on tasks involving an ISI may involve labelling.

2.1.1.2. Category Adjustment Model

Huttenlocher and colleagues, (Huttenlocher & Hedges, 1992; Huttenlocher, Hedges and Vevea, 2000) suggest a category adjustment (or shift towards prototype) model in which memory for a target stimulus is determined by a combination of the stimulus and the best example of the category to which stimuli belong. By this model, during an ISI representation of a stimulus in memory shifts towards the prototypical example of the category to which the stimulus belongs, as shown in Figure 2.1. Prototypical colours lie away from category boundaries (Berlin and Kay, 1969).

In X-AB tasks, performance on across-category trials would be improved because the dissimilarity between the target and foil would be exaggerated. Discrimination on within-category trials would depend both on the order of presentation and the location of the stimulus relative to the prototype. Representation of a target lying closer to the prototype than the foil, but still between prototype and boundary would be shifted in memory away from the foil towards the prototype. As above, dissimilarity between target and foil would be exaggerated, making discriminations easier, as shown in Figure 2.2.
Direction of shift in stimulus representation during ISI

Figure 2.1. The effect of shift towards prototype. The experimental stimulus is B1. During an ISI, representation of the stimulus in memory, shown as 'B1*' is shifted towards the category prototype.

Location of category prototype

Location of category boundary

Figure 2.2. The effect of shift towards prototype on across and within-category trials in an X-AB task. On a within-category trial (a.) the memory representation (B1*) of the target stimulus B1 has shifted towards foil (B2). Correct performance of the task is more difficult than in across-category trials, (b.) where the difference between B1* and the foil (C) is always relatively greater than the difference between B1* and B1.
However when the target lies closer to the category boundary than the foil, but still on the same side of the prototype, representations would be shifted in memory towards the foil. This would make discrimination relatively more difficult. If target and foil are in the same colour category, but on opposite sides of the prototype, the effect would make stimuli more difficult to discriminate than on equivalently spaced across-category trials. On average, the effect would result in better performance on across-category than on within-category trials, the hallmark of CP. On same-different tasks too the overall effect would be similar. The shift of stimulus representation in memory would make test and target stimuli appear more dissimilar in across-category than in within-category trials.

Although the category adjustment model appears capable of explaining results of same-different and X-AB tasks showing colour CP as an effect of memory, there is some doubt as to whether such a model is adequate. The effect of the category adjustment model depends on the order of stimulus presentation. An example of this effect is shown in Figure 2.3.

Pilling et al. (2003) tested the category adjustment model with green and blue stimuli using X-AB and same-different tasks. No evidence of the predicted order effects on within-category trials was found. On the same-different task moreover, order effects were found to be the opposite of those predicted by the category adjustment model. On trials in which target and test stimuli were not physically the same, but were within-category, performance was better when the stimulus closer to the boundary was the target stimulus than when it was the test stimulus. Such a finding can be explained if the prototypes towards which stimuli shift are not green and blue colour category prototypes but a single temporary prototype representing the average of the stimuli presented during the experiment.
Figure 2.3. How shift towards prototype may predict order effects for within-category trials on a same-different task. Figure 2.3a: target stimulus is B1, the test stimulus is B2. Representation of the target stimulus shifts in memory during the ISI to B1*. An Incorrect ‘same’ response is more likely than in Figure 2.3b, where order of presentation is reversed and representation of target stimulus is relatively dissimilar to test stimulus.

Sailor and Antoine (2005) show that judgements of the size of objects from two categories with different mean object size (blue or red squares) are biased toward the mean of the combined distribution of the two categories rather than to the mean of either category. In the case of Pilling et al’s (2003) experiments, and also those of Roberson and Davidoff (2000) and Bornstein and Korda (1984), the temporary prototype formed in this way would be centred on the category boundary. The direction of the shift in stimulus representation in memory would be towards the category boundary. This would have the effect of making across-category trials more difficult than within-category trials. Therefore, if stimulus representation in memory shifts towards a temporary category composed of the average of experimental stimuli, the category adjustment model suggested by Huttenlocher and Hedges (1992),
Huttenlocher et al. (2000) cannot account for the colour CP found in X-AB and same different experiments.

2.1.2. Evidence for Colour Categorical Perception from Other Tasks

Although it appears there is little evidence to support the category adjustment model, it seems verbal labelling of stimuli could play a role in colour CP. However it is unlikely that colour CP can be entirely explained in this way. A variety of studies suggest colour CP may persist even on tasks in which the use of verbal labels seems unlikely. As is discussed in Chapter 1, several studies demonstrate that infants and toddlers show effects consistent with colour CP (Bornstein, Kessen and Weiskopf, 1976; Franklin and Davies, 2004; Franklin et al., 2005b). Toddlers lacking colour term knowledge show CP across the purple-blue boundary on an X-AB task (Franklin et al., 2005a). Bornstein, Kessen and Weiskopf’s (1976) tested infants using a habituation technique and monochromatic light. After habituation to a stimulus, infants looked for longer at a second stimulus if it came from a different colour category to the first stimulus than if it came from the same category. Franklin and Davies (2004) showed a similar pattern of results using reflective Munsell stimuli and a novelty preference technique. The categorical patterns of performance found in these experiments cannot be explained by use of verbal labels. Thus it appears even young infants may possess perceptual colour categories similar to those of adults.

However, the habituation and novelty preference techniques used to assess colour CP in infants in the two studies described above (Bornstein et al., 1976; Franklin & Davies, 2004) are rather different from those used to assess CP in adults. For example, no significant within-category dishabituation or novelty preference effects were found. Such findings imply nothing about discriminability. Snyder, Blank, Cheek, Kuefner and Marsolek (2004) found that a group of infants showing no novelty preference were nevertheless, based on ERP patterns, capable of discriminating a novel stimulus. Equally, it may be that infants failed to show significant within-category effects because they are unable to discriminate between stimuli belonging to the same category. Both studies therefore only demonstrate that infants prefer to look at colours from a category different to that previously displayed. This implies merely that colour categories are salient to infants. What the two studies (Bornstein et al., 1976; Franklin & Davies, 2004) do have in common with many
demonstrations of colour CP in adults is that task performance depends partly on being able to remember the original stimuli. As a result, it is not possible to claim the effect is solely perceptual. Similarly, although Franklin et al. (2005a) do demonstrate categorical performance on an X-AB task that cannot be accounted for by use of verbal labels, the use of an ISI means that the involvement of memory processes in performance cannot be ruled out. Again it is not possible to claim the task measures perception directly. Although strategies involving use of verbal labels may be discounted, other effects of memory cannot be.

Nevertheless, the target detection task used by Franklin et al. (2005b) provides evidence suggesting both that infants show CP in the absence of labels and that the effect is likely to be based in low-level perception. In this experiment, which measured target detection times using eye-tracking equipment, participants were shown a green target against either a background of another shade of green, or against a blue background. Although within and across-category chromatic differences were equated in CIE (L* u* v*, 1976) colour space both groups detected the (green) target more quickly when it belonged to a different colour category to the (green or blue) background. Since it is presented simultaneously with the background, the target does not have to be memorised before being searched for. Consequently, it is unlikely the task involves a memory component resembling that found in same-different and X-AB tasks, or some other kinds of visual search tasks in which the target is known prior to search (see below). Thus the category effect found by Franklin et al. (2005b) may reasonably be assumed to be perceptual in nature.

Other experiments using colour stimuli in visual search also show effects of category (Daoutis, Franklin, Riddett, Clifford, & Davies, 2006a; Daoutis, Pilling & Davies, 2006b; Yokoi and Uchikawa, 2005). These search experiments differ from Franklin et al. (2005b) search experiments in that a target or targets are to be found amongst a number of distractor stimuli. In general, search time in such experiments depends on the similarity between target and distractors and the number of distractors. As similarity between target and distractors decreases, so do search times. Above a certain level of dissimilarity, ‘pop-out’ occurs. ‘Pop-out’ search is characterised by rapid search times, with no increase in search times as the number of distractors increases (Wolfe et al., 1992). Daoutis et al. (2006b) used searches with distractors of
two kinds and non-linearly separable targets – that is where target stimuli lie along a straight line drawn in CIE colour space between the two distractor types. Previous research has shown that searches of this kind are considerably more difficult than searches where target and distractors are linearly separable – that is where target and distractors can be separated by a single line drawn in stimulus space (D’Zmura, 1991; Bauer, Jollicoeur & Cowan, 1996). However in the study carried out by Daoutis et al. (2006b), non-linearly separable searches were made easier when target and distractor stimuli belonged to three different colour categories than when all belonged to the same colour category. This suggests categorical differences between target and distractors can reduce search time in non-linearly separable searches. Furthermore, because search times were fast, the role of verbal labelling in performance of the task is probably, the authors suggest, indirect.

Even if search performance is not influenced by labelling, the results of Daoutis et al.s. (2006b) search experiments cannot be taken as evidence that perceptual space around category boundaries is ‘warped’ (see Harnad, 1987). It may be that the demanding nature of non-linearly separable searches prompts the use of top-down strategies. Thus categorical information might serve to ‘damp down’ the otherwise large effects of non-linear separability on search time. Since participants had knowledge of the target prior to each trial, it is plausible this knowledge may be used to assist search (Duncan & Humphreys, 1989). If such knowledge takes the form of a categorical code, it could be used to assist search in trials where the target belongs to a different category to distractors. Categorical information would not however be useful in trials where target and distractors belong to the same category.

In searches where prior knowledge of target is not available, use of prior information concerning the target is obviously not available either. Yokoi and Uchikawa (2005) have conducted such experiments, though using a different visual search paradigm to that used by Daoutis et al. (2006b). In Yokoi and Uchikawa’s (2005) experiments, the number of distractors sharing the colour category of the target was varied (though stimulus separations were kept equivalent). Increasing the proportion of distractors that shared the target’s colour category resulted in more difficult searches, a broadly similar finding to that of Daoutis et al. (2006b). An additional manipulation involved an uncued search condition. In this condition, performance was similar to that of cued
conditions. Searches were quicker in trials where the target shared a colour category with fewer distractors. Since no foreknowledge of target was possible in uncued trials but category effects persisted, it is difficult to see how strategies involving linguistic labelling or any other kind of memory process could account for performance. Yokoi and Uchikawa’s (2005) study appears therefore to provide evidence that CP effects may be driven by perceptual, as opposed to top-down, processing mechanisms, at least in some circumstances.

2.1.3. Summary and Overview of Experiments 1 and 2
Evidence of colour CP comes from a range of tasks and the different demands involved in these tasks make it difficult to establish the source of CP. Demonstrations of CP using X-AB and same-different tasks appear vulnerable to effects of memory and strategies involving the use of verbal labels. The weakening or loss of the physical code representing a stimulus may also encourage responding based on categorical information representing the stimulus. CP may therefore arise from the use of categorical information stored in memory as Borstein and Korda (1984) suggest. An alternative, but similar, suggestion is that categorical information in the form of verbal labels produces CP (Roberson & Davidoff, 2000). On the other hand, tasks like visual search are believed to be less vulnerable to such strategies and effects. Because they place less demand on memory, performance is more likely to reflect low-level perception. Evidence for CP found on these tasks (e.g., Daoutis et al., 2006b, Yokoi & Uchikawa, 2005) suggests that CP may also stem from low-level perceptual mechanisms. So whilst the results of different kinds of tasks may be interpreted as providing evidence for CP, it is not established whether the CP found in, say, a same-different task is equivalent to that found in a visual search task. In short, CP may arise from several sources. The nature of the tasks used to test for CP though makes it establish which of the potential sources of CP contribute to the CP found.

The two experiments presented in the remainder of this chapter test theories of CP using two slightly different tasks. Performance of both tasks though reflects the operation of low-level perceptual mechanisms. Experiment 1 used an X-AB task with no-delay between target and test phases. Experiment 2 was a temporal 2-AFC detection task. Both tasks inevitably involved a memory component. In Experiment 1, participants were required to store a representation of a stimulus. But since there was
no ISI performance was expected to closely reflect perceptual processes. The task used in Experiment 2 required direct perceptual comparisons and thus judgement did not require memory of a stimulus.

Both tasks were computerised and rather than measuring accuracy or reaction times, were configured to measure just noticeable differences (JNDs). If perceptual space is ‘warped’, either by expansion at category boundaries or within-category compression, measuring JNDs assesses this effect more directly than measurement of response times or accuracy.

JNDs measured using single wavelength stimuli (Wright & Pitt, 1934) provide partial evidence that JNDs vary according to colour category. Figure 2.4. summarises the pattern of JNDs found by Wright and Pitt. It shows apparent JND minima at some category boundaries, for example, the purple-blue (around 440 nm), blue-green, indicated by dashed line (around 490 nm) and red-yellow (around 590 nm) boundaries. Regions of poor discrimination also correspond to hue category centres (Bornstein, 1979; Graham, 1965). However, equal variations in wavelength at different parts of the visible spectrum do not correspond to equal perceptual separations. Stimuli used in Experiments 1 and 2 were therefore specified in the Munsell colour metric, into which considerable efforts have been made to achieve perceptual uniformity (Newall, 1939; Newall et al., 1943).

![Figure 2.4. JNDs at different wavelengths (adapted from Wright & Pitt, 1934). Dashed line indicates location of blue-green boundary.](image-url)
Experiment 1 asked the question: given the demands of the task, are JNDs smaller measured at the boundary between green and blue categories than at within-category green and blue locations? Experiment 2 asked a similar question though using a temporal 2-AFC detection task. The participants tested were native English speaking adults. JNDs in both experiments were measured at seven locations in the blue-green region of colour space. One of the measurement locations was at the boundary between green and blue colour categories, three were within the green named region of colour space and three within the region of colour space named blue. Findings constituting evidence for CP in these experiments would be provided if the JNDs were found to be smaller when measured around category boundary locations in colour space than when measured at within-category locations, particularly those within-category locations adjacent to the category boundary.

If CP has a perceptual component and this is caused by ‘warping’ of colour space the JND differences between the seven measurement locations should conform to a particular pattern. JNDs at the blue-green category boundary should be lower than at the centre of those categories. JNDs should also be lower at the extreme blue and green locations assessed since these locations are away from category centres towards the purple-blue and green-yellow boundaries. Expressed graphically this would involve a line with three turning points, as shown in Figure 2.5. This pattern would also correspond to a 4th order polynomial trend. Fourth order (quartic) polynomial functions can also expressed as graphs with three turning points. Evidence for CP, in adults would therefore be provided if the 4th order polynomial trend shows significant correspondence with the patterns of JNDs observed. Thus of the three turning points found in a 4th order polynomial, one should reflect the blue-green category boundary and the other two the centres of blue and green colour categories.

Findings consistent with CP would extend the conclusions suggested by search experiments, (e.g., Yokoi & Uchikawa's, 2005; Daoutis et al., 2006b) to other tasks. Because of the demands of the tasks used, this would suggest that CP has a perceptual component.

It would be possible to have a 4th order polynomial function that fitted the data, but did not provide support for colour CP. The graphic representation of such a function would be like an upside-down version of Figure 2.1, with a local maxima at the boundary location and local minima around central category measurement locations.
Failure to find CP would imply, first, that the CP found in search tasks may be due to specific task demands and that findings from search tasks cannot be generalized, even to other tasks that are thought to place little emphasis on language and/or memory. In addition failure to find CP would imply that factors such as memory and verbal labelling may be responsible for CP.

2.2. Experiment 1: No-Delay X-AB Task Measuring Hue Just Noticeable Differences

2.2.1. Introduction
Experiment 1 used a no-delay X-AB task to measure JNDs in the blue-green region of colour space. In each trial a target stimulus was displayed briefly, then removed and immediately replaced by test and foil. Removing the delay ought to allow more direct assessment of perceptual processes since there is less scope for the stimulus representation to be affected by memory effects such as shift to prototype. The physical representation of the stimulus is also unlikely to decay significantly. Thus there is less incentive for category information, either in the form of explicitly verbal labels, or a non-verbal categorical code to be used. Evidence of smaller JNDs at the category boundary would suggest that perceptual space is ‘warped’ in this region. This would imply that such ‘warping’ may contribute to CP found on other tasks.
2.2.2. Method

Participants
Fifteen women and three men, a total of seventeen participants (mean age = age 26 years, SD = 9 years) took part in the experiment. All were either students or staff from the University of Surrey psychology department and were native English speakers. Some received course credits for taking part in the study. All had normal colour vision as assessed by the City University Colour Vision Test (Fletcher, 1980).

Apparatus, Stimuli and Design
The experiment was run using a personal computer and a calibrated AOC sixteen-inch monitor. The monitor was calibrated to reproduce colours accurately using the phosphor values of red, green and blue guns and three separate gamma functions to calculate luminance values for each of the guns. Colour readings were made using a Cambridge research systems colourCAL colorimeter and cross-checked with a Minolta CS-100 colorimeter. The display was controlled using an intel 82845G graphics controller. Stimuli were Munsell colours of fixed value (6) and chroma (7.9), varying only in hue, displayed as 50 mm squares against a neutral grey background of the same Munsell value. These subtended to a visual angle of approximately 5.5° x 5.5° from a viewing distance of 500 mm. Colour conversions were made using ‘Colour Science Library 2.0’ software. The task was X-AB task with no-delay between target and test phases, configured to calculate the minimum hue separation between target and foil necessary for accurate performance.

Procedure
JNDs were calculated separately at seven locations in Munsell colour space, (7.5GY, 4.5G, 1.25BG, 7.5BG, 2.5B, 7.5B and 5PB). JND measurements were made bidirectionally around each of the seven locations, as shown in Figure 2.6.

---

6 CIE (x, y, Y) luminance value of the white point was, Y = 62.5
7 Chromaticity values in CIE 1931 chromaticity coordinates were: x = 0.310, y = 0.316
Stimulus dimension (e.g. hue)

a. Bi-directional JND measurement

b. Uni-directional JND measurement

Figure 2.6. Bi- and uni-directional JND measurements. Figure 2.6a, shows a bi-directional measurement procedure. The stimuli used in each trial straddle the measurement location. Easier trials entail the two stimuli being further from the measurement location. On more difficult trials the two stimuli are moved closer to the measurement location. A stimulus corresponding to the measurement location is never shown. In uni-directional measurements (Figure 2.6b), one of the stimuli in every trial has the value of the measurement location, the other varies in location, but is always the same side of the measurement location.

The first three measurement locations were examples of English green, whilst the latter three locations were blue (Berlin & Kay, 1969; Roberson et al., 2000). The central location, 7.5BG is the boundary between blue and green colour categories (e.g., Bornstein & Korda, 1984; Bornstein & Monroe, 1980; Roberson, Davidoff & Braisby, 1999). At each of the seven locations, three JND estimates were calculated using three randomly interleaving staircases of thirty-two trials each. On each trial, the colourimetric properties of the stimuli to be used were calculated using a ZEST algorithm (King-Smith, Grigsby, Vingrys, Benes & Supowit, 1994) converging on an estimate of threshold 75% correct. The first trial in each staircase displayed stimuli
separated by four Munsell units. Eight units was specified as the maximum separation range because at some of the measurement locations stimuli at larger separations were out of the monitor gamut and so could not be produced. Event timing was controlled using a high-resolution timer, DLL, produced by Exactics.

On each trial, a target stimulus was presented for 1000 ms at the centre of the display. Immediately following this test and foil stimuli appeared in flanking locations on each side of the target, remaining in location until a response was made. There was an unfilled blank interval of 1000 ms between trials.

Participants sat at approximately 500 mm and at right-angles to the display in a dark room. The experimenter explained the task and practice trials were run using monochromatic stimuli. The task was to identify which of the two stimuli appearing in the test phase of each trial was identical to the target. Responses were made using a mouse pad. A left key press was used to indicate the stimulus on the left was the test stimulus, a right key press the one on the right. After each response test and foil stimuli were removed. The seven blocks of ninety-six trials were taken in a random order. Following each experimental block a message appeared informing how many blocks remained. A key press was required to initiate each block.

2.2.3. Results
The median of the three JNDs estimates calculated at each location for each participant was used in the analysis. For three participants, JNDs at one or more of the locations tested were at ceiling (eight hue units). This may indicate elevated thresholds. Because none of the participants showed colour vision abnormalities as assessed by the colour vision test used, it is more likely these results indicate difficulty understanding the task or else inattentiveness. Consequently, these data were eliminated, leaving data from fourteen participants to be included in the analysis.

Table 2.1 gives mean JNDs across participants for each of the measurement locations calculated in this way. The same information is displayed graphically in Figure 2.7. There appears to be a local JND minima at the blue-green boundary. JNDs at the two extremities of the region tested also appear to be smaller than JNDs measured at other locations.
Table 2.1. Mean Munsell hue JNDs from Experiment 1 (X-AB). Figures in brackets give standard errors. Bold type shows blue-green boundary location.

<table>
<thead>
<tr>
<th>Measurement location</th>
<th>7.5GY</th>
<th>5G</th>
<th>1.25BG</th>
<th>7.5BG</th>
<th>2.5B</th>
<th>7.5B</th>
<th>5PB</th>
</tr>
</thead>
<tbody>
<tr>
<td>JND (Munsell hue)</td>
<td>1.54</td>
<td>2.48</td>
<td>2.59</td>
<td>1.88</td>
<td>2.16</td>
<td>2.32</td>
<td>1.16</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.30)</td>
<td>(0.21)</td>
<td>(0.17)</td>
<td>(0.26)</td>
<td>(0.19)</td>
<td>(0.17)</td>
</tr>
</tbody>
</table>

Figure 2.7. Experiment 1: JNDs in Munsell hue units for X-AB task (± 1 within-subjects 95% confidence interval (Loftus & Masson, 1994, 2003)). Dashed line indicates blue-green category boundary.

Results were analysed using an analysis of variance (ANOVA) examining the effect of measurement location as a repeated measure. The sphericity assumption was not met so the Huynh-Feldt correction was applied. There was a main effect of measurement location, $F(3.41, 44.31) = 9.92$, $MSE = 0.682$, $p < 0.001$, indicating
significant variation in JNDs across the seven measurement locations. The 4\textsuperscript{th} order polynomial trend was also significant, $F(1, 13) = 20.33$, $MSE = 0.404$, $p < 0.005$.

Figure 2.8. shows the data fitted with this function, with one of the three turning points at a local minima around the blue-green boundary and the other two turning points towards the centre of the green and blue categories.

![Figure 2.8](image)

**Figure 2.8.** Mean JNDs from Experiment 1 with 4\textsuperscript{th} order polynomial function fitted (± 1 within-subjects 95\% confidence interval (Loftus & Masson, 1994, 2003)).

Planned comparisons showed that JNDs at the blue-green boundary (7.5BG) were significantly smaller than at the adjacent within-category green location (1.25BG), $F(1, 13) = 17.15$, $p < 0.005$, but not significantly smaller that the adjacent within-category blue location (2.5B), $F(1, 13) = 1.93$, $p = 0.188$.

2.2.4. Discussion

The results of Experiment 1 indicate that Munsell hue JNDs show considerable variation in the blue-green region of colour space, as is clearly shown in Figure 2.7. The pattern of these variations suggests that effects consistent with CP are present on the task. There was a significant 4\textsuperscript{th} order polynomial trend and JNDs measured at the
category boundary were found to be significantly lower than at the neighbouring within-category green location. JNDs at the within-category blue location adjacent to the boundary (1.25B) were also larger than those at the boundary, but this difference was not significant. Overall, the results support the idea that there is a perceptual component to CP, and to the idea that, on some tasks, 'warping' of perceptual space may play a role in CP.

Despite there being no delay between target and test phases of the task, the task still involved a memory component. To complete the task participants must have compared a perceptual representation of test and foil to a representation in memory of the target, rather than the direct perceptual representation of that stimulus. Removing the delay phase may have minimised the difference between the representation in memory of the stimulus and the direct perceptual representation of the stimulus. Representation of stimuli in a physical code is also likely to have been available during the test phase because decay of the physical code takes at least 1500 ms (Posner & Keele, 1967). Consequently, the influence of categorical code information or verbal labelling in completing the task is likely to have been minimal. It is also unlikely that shifts in stimulus representation in memory towards either category prototypes (Huttenlocher & Hedges, 1992; Huttenlocher, Hedges & Vevea, 2000) or the mean of the combined distribution of the stimulus (Sailor and Antoine, 2005) significantly influenced performance. It therefore seems likely that the task reflects the operation of low-level perceptual processes more directly than tasks involving a delay. But it is still not possible to claim that the task used is immune to effects of language and/or memory. To measure perceptual processes more directly a task that requires direct comparison between stimuli is required. A task fitting this criterion more closely was used in Experiment 2. The generalizability of findings indicating CP can also be examined by manipulating the separation between stimuli. Danilova and Mollon (2006) found that chromatic discrimination thresholds varied depending on the spatial separation of stimuli, being optimal when there was a small separation between the edges of stimuli. In Experiment 2 there was no separation between stimuli, in contrast to the separation between stimuli displayed in the test phase of Experiment 1.
2.3. Experiment 2: Temporal 2-AFC Detection Task Measuring Hue Just Noticeable Differences

2.3.1. Introduction
Experiment 1 used a no-delay X-AB task to measure JNDs in the blue-green region of colour space. Evidence that CP effects persist when low-level perceptual processes are measured was found. But as was pointed out in section 2.2.4, doubts may be raised as to how directly the task used measures perceptual processes. In Experiment 2 JNDs were measured using a temporal 2-AFC detection task. This task requires direct perceptual comparisons between stimuli. This makes it very unlikely that labelling strategies, categorical code information or shift to prototype influences performance. A temporal 2-AFC detection task was used because technical difficulties made use of a simultaneous X-AB task problematic. In addition, the tasks used in Experiments 1 and 2 may have differed in their demands. Evidence of CP from both tasks would therefore provide evidence that any perceptual component to CP is a robust phenomenon.

Each trial involved successive rapid presentation of two rectangles, one composed of uniform hue, the other divided into regions of different hue, the task being to identify which of the two rectangles was composed of two hues. Although such a task may involve a memory component, this memory component is quite different to that involved in Experiment 1 or in the same-different (Bornstein & Korda, 1984), and X-AB (Roberson & Davidoff, 2000) tasks used to provide evidence for colour CP described earlier. This is because a difference in hue between two patches involves a direct perceptual comparison, not comparison of a stimulus with a stimulus represented in memory. Consequently the temporal 2-AFC detection task used in Experiment 2 may be considered to measure perceptual processes more directly than the task used in Experiment 1.

2.3.2 Method
Participants
Twelve women and four men (mean age = 24 years 3 months, SD = 8 years) took part in the experiment. Most were psychology undergraduates and were given course
credits for participation. All were native English speakers and had normal colour vision as assessed by the City University Colour Vision Test (Fletcher 1980).

**Apparatus, Stimuli and Design**

The experiment was carried out using the same equipment used in Experiment 1. Stimuli colorimetric properties were also the same as those used in Experiment 1. Stimuli were displayed either singly as rectangular (130 x 60 mm) patches or in abutting pairs (65 x 60 mm), as shown in Figure 2.9. Patches containing a single stimulus subtended a visual angle of approximately 14.0° x 6.5° viewed from a distance of 500 mm.

![Figure 2.9. Stimuli display in temporal 2-AFC detection experiment: either alone (a.) or paired with another stimulus (b.) Colour of stimulus in a. is same as one of two stimuli in b.](image)

**Procedure**

The task used measured Munsell hue JNDS at the same measurement locations and using the same methods employed in Experiment 1. The form of an individual trial is shown in Figure 2.10. Trials were preceded by a 250 ms fixation cross then an unfilled blank interval of 250 ms before the display of target 1 for 120 ms. This was followed by a 250 ms mask, consisting of a monochromatic texture pattern, then another blank interval (250 ms), then target 2 (120 ms), another mask and finally a third blank interval which remained until a response was made. Masks were used because pilot studies without a mask obtained very low JNDS. Stimulus production at such small separations was not reliable with the equipment used. A mask was therefore inserted to elevate JND.
In any trial one of the two targets comprised of a rectangle comprising a single hue, the other was composed of two hues. The task on each trial was to identify which of the targets contained two hues. More difficult trials entailed reducing separation between the abutting pair of stimuli. In easier trials, this separation was increased.

Participants sat at approximately 500 mm and at right-angles to the display in a dark room. The task was explained using a demonstration version of the program and practised with monochromatic stimuli. Responses were made using a mouse. A left key press indicated that the first of the targets contained two stimuli, a right press that the second target contained two stimuli. Responses also triggered a 500ms delay, after which the next trial began. During practice, feedback was given automatically. Practice terminated after five consecutive correct trials. Experimental measurements were then made in seven blocks of trials.

2.3.3. Results
The median of the three JNDs estimates calculated at each location for each participant was used. None of the values indicated performance at ceiling. Table 2.2. gives mean JNDs across participants for each of the measurement locations calculated in this way. The same information is displayed graphically in Figure 2.11. As can be seen, there appears to be a minimum at 7.5BG, whilst JNDs measured at the two extremes also appear to be smaller than those of adjacent locations.

Visual angles for each stimulus in abutting pairs were approximately 7°.
Figure 2.10. Schematic representation of trial in temporal 2-AFC detection experiment.
Table 2.2. Mean Munsell hue JNDs from Experiment 2 (Temporal 2-AFC detection). Figures in brackets give standard errors. Bold type shows blue-green boundary location.

<table>
<thead>
<tr>
<th>Measurement location</th>
<th>7.5GY</th>
<th>5G</th>
<th>1.25BG</th>
<th>7.5BG</th>
<th>2.5B</th>
<th>7.5B</th>
<th>5PB</th>
</tr>
</thead>
<tbody>
<tr>
<td>JND (Munsell hue)</td>
<td>1.69</td>
<td>2.23</td>
<td>2.38</td>
<td>2.02</td>
<td>2.76</td>
<td>3.19</td>
<td>1.79</td>
</tr>
<tr>
<td></td>
<td>(0.21)</td>
<td>(0.25)</td>
<td>(0.22)</td>
<td>(0.18)</td>
<td>(0.31)</td>
<td>(0.28)</td>
<td>(0.23)</td>
</tr>
</tbody>
</table>

Figure 2.11. Experiment 2: JNDs in Munsell hue units for temporal 2-AFC detection task (± 1 within-subjects 95% confidence interval (Loftus & Masson, 1994, 2003)). Dashed line indicates green-blue category boundary.
Results were analysed as in Experiment 1. There was a main effect of measurement location, $F(6, 90) = 12.97, \text{MSE} = 0.356, p < 0.001$, indicating that JNDs differed depending on the measurement location. The 4th order polynomial contrast was also significant, $F(1, 15) = 31.97, \text{MSE} = 0.341, p < 0.001$. Figure 2.12 shows the data fitted with this function.

![Figure 2.12](image)

**Figure 2.12.** Mean JNDs from Experiment 2 with 4th order polynomial function fitted $\pm 1$ within-subjects 95% confidence interval (Loftus & Masson, 1994, 2003).

Planned comparisons showed JNDs at the blue-green boundary (7.5BG) were significantly lower than at either the adjacent within-category green location (1.25 BG), $F(1, 15) = 8.18, p < 0.025$ or the adjacent within-category blue location (2.5 B), $F(1, 15) = 8.94, p < 0.01$.

2.3.4. Discussion

The evidence for colour CP presented in Experiment 2 provides further evidence that CP is found on tasks that measure perception directly. The pattern of variation in JNDs shows the predicted minima round the blue-green boundary and maxima towards the centre of blue and green colour categories. Similarly, JNDs at the
boundary location were significantly lower than at either of the two neighbouring within-category locations.

The result also appear to corroborate Yokoi and Uchikawa’s (2005) conclusion that under some circumstances colour category effects cannot be due to colour memory and are mediated by perceptual processes. In the case of Yokoi and Uchikawa’s (2005) study this conclusion was reached on the basis of a category effect persisting in visual searches in which target information was not available prior to the start of trials. Experiment 2 would also appear to support this conclusion because the judgement required in each trial required a direct perceptual comparison.

In addition since the task used in Experiment 2 assessed colour CP by measuring JNDs, rather than measures of speed or accuracy, the results seem to lend more direct support to theories which suggest CP is caused by ‘warping’ of perceptual space (Harnad, 1987), than previous studies. The pattern of JNDs found in Experiment 2 is consistent either with within-category compression of perceptual space, across-category stretching or both.

2.4. General Discussion of Experiments 1 and 2

The results of the two experiments presented here suggest there is a perceptual component to CP. Munsell colour space is perceptually uniform, and thus no variations in JNDs should be present. However, in both experiments, JNDs were found to vary, with minima present at the category boundary and maxima in category-centre regions. This variation closely mirrored that predicted to occur if perceptual space was altered such that discriminations around category boundaries were easier than equivalently spaced within-category discriminations. The tasks used assessed perceptual processes more directly than many previous tasks. Because of this, it seems unlikely that verbal labelling, shift to prototype, use of a categorical processing code or any other effect associated with retention of a stimulus in memory can account for the CP effects found. If these effects do reflect the operation of low level perception, as seems likely, they provide evidence that colour CP may be caused by ‘warping’ of perceptual space.
However, it cannot be claimed that the colour CP found in other experiments is produced by the same mechanism as that which appears to underlie performance in Experiments 1 and 2. Filling the ISI in an X-AB task with verbal interference has been found to block CP if the interference kind is predictable (Roberson & Davidoff, 2000). This strongly suggests that under some circumstances verbal labelling does play a role in producing CP. Filling an ISI with visual interference though does not abolish CP. Visual interference might be expected to degrade a perceptual representation of a stimulus stored in memory. That CP survives such interference further implies the use of either verbal labelling or categorical information, even on a task that measures discrimination as opposed to categorical identity.

Consequently, it appears that CP may be caused by a number of factors. First, perception may be altered such that stimuli around category boundaries appear relatively more different than stimuli at category centres (Hammad, 1987). Second, verbal labelling of stimuli may occur, again producing better performance on across rather than within-category trials (Roberson & Davidoff, 2000). Finally, use of categorical, though not necessarily verbal information would also produce CP (Bornstein & Korda, 1984). These three factors could all contribute to colour CP. On an X-AB task with a delay for example, CP could be caused either by ‘warping’ of perceptual space, verbal labelling, use of a non-verbal categorical code or a combination of all three. Because English language colour categories coincide with the proposed universal set of colour categories (Berlin and Kay, 1969) these three factors are all aligned. This makes it difficult to assess them independently. Whilst the two experiments presented here suggest CP can be a perceptual in nature, they do not make clear the origin of this effect. Other research suggests that labelling and/or categorical information can also produce colour CP (Roberson & Davidoff, 2000; Pilling et al., 2003).

According to a universalist interpretation, the CP found in Experiments 1 and 2 reflects the influence of the proposed universal perceptual colour categories. From a relativist perspective, learning to divide colour space into green and blue named regions can be argued to cause the CP effects found, even at a perceptual level. Indeed, if ‘warping’ of perceptual space occurs, then learning categories could likely
drive that process, as perceptual learning studies imply (e.g., Goldstone, 1994; Livingstone, Andrews & Harnad, 1998, Guenther, Husain & Shinn-Cunningham, 1999, Özgen & Davies, 2002). One means of resolving this issue, examined in the next chapter, is through comparisons of speakers whose colour language differs.

It is interesting to note that the pattern of performance in the two experiments differed. In the X-AB task, JNDs at within-category blue measurement locations, 2.5B and 7.5B and 5PB, appear to be lower than in the temporal 2-AFC detection task. It is not clear why this should be the case though the differences may reflect the different demands the two tasks place on participants. Another feature of the pattern of JNDs is the presence of minima not only at the blue-green boundary, but also at the two extreme measurement locations, 7.5GY and 5PB. Both of these locations are within respectively green and blue named categories. However, 7.5GY lies between the focal point of the green category and the yellow-green boundary, whilst 5PB is between the focal blue and the blue-purple boundary (see Rosch Heider and Oliver, 1972). In consequence, it seems likely these minima represent decreasing JNDs towards the yellow-green and blue-purple boundaries, though to test this would require further experiments measuring JNDs around the region of these two boundaries.
Chapter 3  
Colour Language and Colour Categorical Perception

3.1. Introduction

In Chapter 2 two experiments were presented which imply that there may be a component to CP that is purely perceptual. What is not established in Chapter 2 is the origin of this perceptual effect. In Chapter 3 therefore, claims that colour CP is influenced by colour language are examined. Five cross-cultural experiments comparing performance of Himba and English speakers, whose languages differ in their segmentation of colour space, are presented.

One of the experiments was an X-AB task involving an ISI. Such tasks may be vulnerable to the influence of the various non-perceptual mechanisms, described earlier, which may be responsible for producing CP. Participants in this experiment though were young children from two language groups. Because of the language they spoke, children from one of these groups could not possess separate verbal labels to name the stimuli used. Consequently, for this group, strategies involving labelling of stimuli were not readily available on the task. The remaining four experiments assessed CP using tasks which, like those used in Chapter 2, attempt to measure low level-perceptual processes. This is considered important, because although Experiments 1 and 2 suggest a perceptual component to CP, they do not provide information relating to the origins of colour CP. The distinction of immediate interest in this chapter is whether colour CP is a universal phenomenon. If CP arises from learning processes, learning linguistic categories is thought to play an important role in that learning. Because the participants used in Experiments 1 and 2 were all native English speakers, and because the proposed universal colour categories correspond closely with those of the English language\(^9\) (Berlin and Kay, 1969), it was not possible to establish whether the origin of CP effects found are linguistic. By comparing speakers of different languages, in particular those whose colour language

\(^9\) Some research suggests that the region of colour space named blue in English may be divided into two basic categories (Davies & Corbett, 1994; Özgen & Davies, 1998); the same for red.
does not categorise the proposed universal categories using separate terms, the experiments presented in this chapter should help to resolve this issue.

The following sections examine whether cross-cultural differences in performance of tasks involving coloured stimuli that have been shown to exist by previous research provide strong evidence for differences in colour perception originating in colour language differences. Subsequent sections detail the five experiments carried out.

3.1.1. Perspectives on the Debate
Differences in the way languages segment up colour space, noted originally in the 19th century, have come to be explained in several different ways and from a variety of perspectives. From an extreme relativist stance, this variation is essentially arbitrary and points towards a world carved up by language (Whorf, 1956). By this argument, the continuum of electromagnetic radiation is only resolved into distinctive colour categories by language. Others would suggest the variation in colour language is ordered: languages acquire colour terms in a methodical fashion suggestive of underlying, universal colour categories (Berlin & Kay, 1969). However, in spite of much work, particularly involving the collection of colour naming data from different languages, the disagreement over basic issues relevant to the debate appears intractable (see Saunders and Van Brakel, 1997). It does not seem likely that further collections and analysis of naming data will resolve the issue.

It may therefore be useful to take a different approach and pose a slightly different question: do differences in colour language lead to measurable differences in colour perception? If they do, and they are large scale differences, this would suggest not only that the division of colour space by languages is arbitrary, but that this division is more than a matter of words. It would suggest, in fact, that perception is strongly influenced by language. On the other hand, a failure to find large scale differences in perception that correspond to linguistic colour categories would imply the linguistic division of colour space does not influence colour perception. A third possibility exists – that differences in perception caused by differences in colour language do exist, but these are relatively small in scope. For example, it is possible that colour categories are universal, based on one or a combination of say, asymmetries in the behaviour of surfaces with respect to light (Philopona & O'Regan, in press), the
properties of ambient light sources (Shepard, 1992), colour appearance (Kay & Maffi, 1999), the topography of colour space (Jameson & D’Andrade) or because they represent some optimum combination of economy and discriminability (Roberson et al., 2004). In any of these cases, it is still possible to imagine differences in linguistic coding of colour categories resulting in perceptual differences. Goldstone (1998) suggests, ‘Category learning systematically distorts perception of category members by shifting their perceived dimension values away from members of opposing categories.’ This is similar to Harnad’s (1987) description of CP as resulting from a ‘warping’ of perceptual space. Özgen & Davies’ (2002) study provides experimental support for this possibility. Participants were taught to split existing colour categories into two new categories and subsequently CP effects were found across the newly learned boundary. This suggests that learning to divide an existing category into two separate categories may cause CP. However, this interpretation requires qualification. The task used to assess CP was a same-different task with an ISI. Consequently it is not possible to be certain that the CP found involved a change in perception. It may merely have involved the learning of new verbal labels for stimuli. Second, training was accomplished using massed practice trials spaced over a period of days, a situation that does not appear to mirror the learning of colour language in childhood.

If learning linguistic colour categories can result in changes to low-level perceptual function, differences in colour perception should exist between speakers of languages with differing colour category structures. The extent of these proposed differences would then depend on whether perceptual colour categories are caused by learning colour language. If it is supposed that language divides the continuum of visible light into categories, colour CP ought only to be found in conjunction with a linguistic category boundary. But linguistic differences in segmentation of colour space might only denote the successive linguistic encoding of a set of universal categories. In this eventuality colour CP (as evidenced by tasks that measure perceptual processes) might occur regardless of colour language. Critically though, CP should be enhanced if the boundary is linguistically marked and/or decreased if the boundary is not marked linguistically. Besides the evidence from perceptual learning studies which appears to show such changes are possible, (e.g., Goldstone, 1994; Livingstone,

Andrews & Harnad, 1998, Guenther, Husain & Shinn-Cunningham, 1999, Özgen & Davies, 2002) there is some evidence in other domains that supports this idea. For example, infants' ability to discriminate some speech contrasts and conceptual distinctions may be lost if the contrasts are not encoded linguistically (Werker & Lalonde, 1988; Hespos & Spelke, 2004). However, direct evidence that colour perception may be altered in such a way is lacking.

One way to address the question of whether differences in colour language lead to measurable differences in colour perception is through cross-cultural investigations. Such studies are relevant to the two issues outlined above. The first issue relates to the influence of category learning on perception, the second to the more fundamental problem of whether perceptual categories arise from language learning. A number of previous cross-cultural studies are relevant to these topics. These are discussed below.

3.1.2. Evidence for Cross-Linguistic Differences in Colour Cognition and Perception
Although Rivers (1901b) speculated that differences in colour perception exist between members of different cultural groups, he supposed any such differences originated in physiological differences between different groups of people. It is therefore important to distinguish between various kinds of theories concerned with cross-cultural differences in colour perception. Some make claims similar to those of Rivers or claim that colour vision is affected by other kinds of environmental influence, such as the effects of ultra-violet light (e.g., Bornstein, 1973). Other theories are more concerned with differences in colour cognition and perception caused by differences in colour language. The discussion below confines itself to cross-cultural research that has investigated the possible role language plays in influencing colour cognition and perception.

Discussion of Berlin and Kay's (1969) theory and subsequent studies by Rosch-Heider, is found in the Introduction. The conclusions of this work -- that the perceptual-cognitive division of colour space was universal -- were widely accepted until a number of new studies examining cross-cultural variations in colour cognition reopened the debate.
The focus here will be on claims that differences in colour perception exist and these are caused by differences in colour language. Criticisms of these claims will centre, as in Chapter 2, on the nature of the tasks used and whether they can be considered to accurately assess basic perceptual processes. More specifically, it will be argued that although recent cross-cultural studies have demonstrated differences in performance of different language groups, it may be a mistake to attribute these differences to differences in perception rather than to differences in language and/or memory processes.

Studies of colour grouping (e.g., Davies & Corbett, 1997; Davies, 1998) in which participants are required to sort coloured chips into piles, have found small but reliable differences between groups that correspond to differences in colour language. Whilst differences in grouping may be caused by differences in perception, they may equally be influenced by the colour names given to stimuli. Thus the tendency of Setswana speakers, whose language lacks separate blue and green colour terms (Davies, MacDermid, Corbett, McGurk, Jerrett, Jerrett, & Sowden, 1992) to place blue and green stimuli in the same group more often than English speakers may merely reflect the influence of verbal category labels on the sorting task. Such findings do provide evidence in support of a weak form of the LRH, one where only higher cognitive functions are influenced, but support for stronger versions of the LRH is limited. It is also salient that, as the authors of the above studies stress, the differences found between groups, though consistent, are small. If perceptual colour categories are created by language learning and language divides colour space in an arbitrary manner, it is hard to explain this pattern of similarity.

Performance of triadic judgements also indicates reliable differences in performance that covary with colour language. In these tasks, participants are shown three coloured chips and required to choose an ‘odd one out.’ Davies et al., (1998) again found reliable differences between Setswana and English speakers. Kay and Kempton (1984) also found differences between English speakers and speakers of Tarahumara, a language spoken in Mexico, corresponding to differences in colour language.

Roberson et al. (2000, 2005) provide further triad data, comparing performance of English, Berinmo and Himba speakers. Again stimuli were judged more similar when
they came from the same linguistic category. Furthermore, if all three chips belonged
to the same named category, the pair of chips furthest from the boundary was
consistently chosen as most similar. The authors argue that the triad task used taps
perceptual judgements. This claim is based on the finding that, in trials where all three
chips belonged to the same named category, the pair of chips furthest from the
boundary was consistently chosen as most similar. Because the triad task has no ISI, it
is immune to the effects of memory which make claims that X-AB and same-different
tasks reflect low-level perceptual processes difficult to substantiate. But on the other
hand, Kay and Kempton (1984) argue that triadic tasks are susceptible to a labelling
strategy and "...it operates at a level that is not only out of consciousness but which
cannot easily be brought under conscious control." Furthermore, in an attempt to
prevent the use of labelling strategies, the same authors manipulated a triadic
judgement task such that participants were required to make the same judgement, but
could only view two of the three chips at a time. With these changes, differences
between performance of English and Tarahumara speakers were abolished. This
indicates that although triadic judgement tasks do tap perceptual judgements they are
also influenced by stimulus labelling. Consequently the claim that performance
differences found between speakers of different languages reflect differences in
perception cannot be supported by triad tasks – the differences may just as likely
reflect an influence of labelling strategies, not necessarily to the exclusion of
perceptual strategies, on the judgements made.

Evidence from X-AB tasks has also been interpreted as supporting the idea that colour
perception is affected by colour language. Roberson et al., (2000, 2005) have shown
that Himba, English and Berinmo speakers only show CP across colour category
boundaries marked in their language. If colour language causes changes in colour
perception such results are to be expected. But such results can equally be interpreted
as supporting the argument made by Roberson & Davidoff (2000), discussed in the
previous chapter, that verbal labels play a role in performance of X-AB tasks.

Visual search experiments carried out by Daoutis et al (2006a) have also shown
differences in responding between different language groups. In these experiments,

---

English children found targets more rapidly when they belonged to a different colour category to distractors, than when targets and distractors came from the same colour category. However Kwanyama and Himba speaking children of comparable age, whose languages lack separate terms for the stimuli used in the ‘across-category’ searches, failed to show a similar category effect. Such findings appear to provide stronger evidence of language dependent differences in colour perception than the colour grouping, triadic judgement and X-AB tasks described above. But as the authors acknowledge, participants could have performed the task by labelling search stimuli. In these circumstances English speakers, with separate labels for target and distractor stimuli, might be expected to show a category effect absent in Himba and Kwanyama. Thus the differences in performance need not reflect differences in perception.

Other studies using cross-cultural children have reached firmer conclusions than Daoutis et al., (2006a). Roberson et al., (2004) carried out a longitudinal study investigating children’s colour term acquisition in two languages. Results of colour naming, colour comprehension and colour memory tasks led the authors to conclude, ‘...an initial perceptually driven continuum progressively organised into category sets appropriate to each culture and language.” On the other hand, Franklin et al., (2005a) found that groups of Himba speaking toddlers, who knew no colour terms, and English toddlers both showed CP across the blue-purple boundary. The task used was an X-AB task, with an ISI, so it is possible to explain the CP shown by English groups in terms of verbal labelling strategies. An explanation involving linguistic labelling would not appear viable for the Himba group. If it is assumed performance is driven by universal colour categories, then the findings are easier to explain. Himba toddlers may have shown a categorical effect either because of perceptual differences across a boundary between universal colour categories, or else through use of a categorical, though non-verbal, code reflecting the proposed universal categories.

3.1.3. Summary and Overview of Experiments 3-7
Cross-cultural studies have demonstrated differences in performance corresponding to differences in colour language on a range of tasks involving coloured stimuli.

---

12 Roberson et al. (2004), pp 554.
However these differences do not, it is argued here, provide convincing evidence that differences in low-level colour perception corresponding to differences in colour language exist. Although this is possible, it is also possible to explain the differences in terms of differences in higher cognitive functions, particularly those in which linguistic labels are likely to play a role. Indeed, if labelling plays a role in producing CP directly (as opposed to indirectly by altering perception) the differences observed amongst different groups are to be expected.

The aim of the experiments presented in this chapter is to compare performance of Himba and English speaking participants using tasks that measure perceptual processes more directly than those used previously.

Himba people are semi-nomadic pastoralists living in a remote semi-arid region of Africa around the Namibian-Angolan border. The Himba language is a dialect of Herero. Himba people have retained a traditional lifestyle and have a strong cultural identity. These factors contribute to Himba people having little contact with other cultures (see Crandell (2000) for a more detailed account of the Himba people).

All the experiments compared performance in blue-green region of colour space, a region named by Himba language using a single colour term (Roberson et al., 2004, Androulaki, 2003). Experiment 3 compared performance of two-to-four year olds on an X-AB task, to replicate and extend the finding of Franklin et al. (2005a). Although the use of X-AB tasks was criticised earlier, one is used here to test whether performance of both groups is similar. If colour CP is a product of colour language learning, then CP should not be found in the Himba two-to-four year olds tested.

Experiments 4 and 5 used a target detection task similar to that used by Franklin et al. (2005a). Two-to-four year old and adults groups were tested. Experiments 6 and 7 measured JNDs in the two age and language groups. The tasks used in these experiments (4-7) do not involve an ISI and therefore the role of memory effects is minimised. In addition, they are unlikely to be influenced by verbal labelling strategies as it is suggested the triadic judgement tasks mentioned above are. This is because the tasks used here involved detection of a stimulus, involving different
demands to those required to make a judgement relating to the relative appearance of three stimuli viewed simultaneously.

Since the stimuli used in Experiments 3 – 7 cross a linguistic colour category boundary marked in the English, but not the Himba language, the relative performance of the two groups should help in assessing the role of language in colour CP. If differences in performance of the tasks correspond with differences in colour language, this would suggest that learning colour terms can influence lower level perception. The larger the degree of difference, the larger would be the involvement of language. For instance, evidence of CP across the blue-green boundary for English but not for Himba speakers (who name the blue-green region a single colour term), would suggest that learning colour terms is the cause of colour CP. Should both language and age groups show comparable levels of colour CP, this would imply that colour perception is not affected by language. In this scenario Himba speakers would show CP across a boundary not marked in their language equivalent to that shown by English speakers across a linguistically signified boundary. Such a pattern of performance is not compatible with a hypothesis that suggests colour language learning can influence colour perception. If CP is present in both language groups, but is greater in English speakers this would suggest learning colour language influences perception, though more as a process of ‘fine-tuning’ than of basic organisation.

3.2. Experiment 3: X-AB task: English and Himba Two-to-Four Year Olds

3.2.1. Introduction
Franklin et al. (2005a) used an X-AB task to show that Himba speaking toddlers who knew no colour terms responded categorically across the purple-blue boundary. Participants were shown a test stimulus and after a five-second delay during which the test stimulus was removed were shown the same stimulus and a foil which could belong either to the same or to a different colour category. Fewer errors were made using a stimulus pair drawn from different colour categories (blue 2 and purple) than using a stimulus pair drawn from the same colour category (blue 1 and blue 2), even though separations between the stimuli in each pair were equated. This finding was interpreted as supporting a universalist model of colour perception. If categorical
responding is caused by learning colour terms, findings like this are difficult to explain, because although a Himba category boundary (see Roberson et al., 2005, Figure 1.) lies close to the region tested, the Himba toddlers tested knew no colour terms. If the universalist interpretation is true, Himba toddlers should also show CP across other boundaries not marked in their language. The blue-green boundary is not marked in the Himba language, the blue-green region being named with a single colour term, [mburu] (Roberson et al., 2005; Androulaki, 2003). Experiment 3 therefore examined whether Himba two-to-four year olds, whose knowledge of colour terms is limited, and English two-to-four year olds would show categorical patterns of response across the blue-green boundary, using a procedure identical to that used in the original study by Franklin et al. (2005a).

Categorical patterns of responding across the blue-green boundary and no differences between the two language groups would add support to the idea that even young children respond categorically to colour and that these responses may be ‘hardwired’ rather than linguistic in origin.

3.2.2. Method

Participants

Twenty-six English (mean age = 38 months, S.D. = 3 months) and twenty-five Himba two-to-four year olds took part in the study. The English group were recruited from and tested in local nurseries. Himba two-to-four year olds were tested in villages in the Kunene region of north-western Namibia with the aid of an experienced translator.

Stimuli, Apparatus and Design

Experimental stimuli were Munsell glossy papers varying only in hue (value = 5, chroma = 10). The following hues: were used 7.5G, 5BG and 2.5B. Stimuli were used to make two equally spaced pairs, one ‘within-category green’ (7.5G and 5BG),

13 Pilot work suggested that the majority of Himba two-to four year olds had, in comparison to English two-to four year olds, less knowledge of the colour terms in their language. See Roberson et al. (2004) for more detailed description of Himba and English toddlers colour term knowledge.

14 Himba speakers keep no record of their ages. Ages were estimated from discussions with participant’s immediate family members. Those in the youngest Himba group were probably all under 5 years of age as none could touch their ear with their hand passing across the top of their head. Children under 5 years do not pass this test due to the size of the head relative to the length of their arms (Gabriel, 2001).
the other 'across-category' (5BG, 2.5B), as shown in Figure 3.1. Training stimuli were two identical sets of eleven Color aid papers similar in appearance to the eleven colours identified by Berlin and Kay (1969) as representing the best examples of each of the basic colour categories.\(^{15}\) Stimuli were cut into the shape of sweaters that could be placed on one of two identical outlines of a child (Child A and Child B) made from white card with outline and facial features drawn on in black ink. The design was an X-AB task adapted for use with young children, identical to that developed by Franklin et al. (2005).

![Munsell Hue](image)

**Figure 3.1.** Within-category and across-category stimulus pairs used in Experiment 3.

**Procedure**

The task was carried out on a table at which participants were seated. Himba two-to-four year olds completed the task outside, avoiding deep shade and direct sunlight (Colour temperature range measured with a Gossen 3F colormaster: 5500 - 8000°K). English speakers were tested indoors using a D65 illuminant (6500°K) lamp. Illumination ranged from 1000 – 3000 lux.

\(^{15}\) Color Aid papers used were: BLACK, WHITE, RO Hue, G Hue, Y Hue, B Hue, O S3, R T4, V Hue, YO Hue and Gray 4. Two sweaters of each colour were used.
Prior to the main experimental trials, training was given. The two outline figures of children were laid side by side on the table on a large sheet of white paper. Participants were shown the two child outlines and told they were brothers who always wore the same sweater. One set of Color aid sweaters was placed face up around Child A. A sweater from the other set was then placed on Child B. The participant was asked to make Child A wear the same sweater as Child B by choosing one of the eleven sweaters and placing it on Child A. Correct matches were praised. After incorrect matches the instructions were repeated and another choice encouraged. The process continued until reliable matches were made. Practice trials involving a delay were then attempted. The set of eleven Color aid sweaters was covered. A sweater was placed on Child B. After it had been inspected, Child B was covered. Following a five second delay, the other set of eleven Color aid sweaters around Child A was uncovered. The participant was asked to choose a sweater the same as the one Child B was wearing and place it on Child A. Child B was then uncovered and the participant was encouraged to evaluate their response. After three successive correct matches had been achieved using the delayed choice, the experimental trials were attempted.

The structure of experimental trials was identical to that of training trials which used a five second delay, except that participants were given a choice of only two sweaters, a correct choice (target) and an incorrect choice (foil). Each stimulus pair was used on four trials, with the stimuli in each pair appearing as target and foil on two trials each, making a total of eight trials. The five second delay between test and target phases of each trial was timed by the experimenter. The number of errors made in trials involving each pair was recorded. During experimental trials, feedback was given. The number of incorrect matches was recorded.

3.2.3. Results
Error scores (maximum possible = 4) made by each participant for within and across-category pairs were calculated. Mean error scores across participants are shown in Figure 3.2. and Table 3.1. Although the English group made fewer errors, the pattern of performance appears similar in both groups, with fewer errors made in across than within-category trials.
Data were analyzed using an analysis of variance examining the effects of category (Within and Across) and language (Himba and English). Category was a repeated measure.

<table>
<thead>
<tr>
<th>Stimulus set</th>
<th>Within-category</th>
<th>Across-category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Himba</td>
<td>2.04 (0.93)</td>
<td>1.48 (1.08)</td>
</tr>
<tr>
<td>English</td>
<td>1.62 (0.85)</td>
<td>1.15 (1.00)</td>
</tr>
</tbody>
</table>

Table 3.1. Mean errors (SD) for across and within-category trials (maximum score = 4) in Experiment 3.

Figure 3.2. English and Himba two-to-four year olds. Mean errors (±1 SE) for within and across-category pairs, in Experiment 3. Bold dashed line indicates level of performance at chance. Bold bar in centre gives ±1 within-subjects 95% confidence interval (Loftus & Masson, 1994, 2003).
There was a significant effect of category, (mean, (S.D.): Within = 1.83 (0.91), Across = 1.32 (1.04)), $F (1, 49) = 7.91, MSE = 0.925, p < 0.025$, indicating significantly fewer errors were made in trials using across-category than within-category pairs. Differences in overall accuracy between the two language groups approached, but did not reach significance, $F = 3.711, p = 0.060$. The language by pair interaction was not significant, $F = 0.067, p = 0.797$.

3.2.4. Discussion

The results indicate a significant category effect across the blue-green boundary, with more errors made on within-category than on across-category trials. Although the overall difference in performance between the two groups approached significance (this may reflect the unfamiliarity of the Himba sample with formalized memory tasks) the difference in the pattern of performance between the two groups was not significant. This replicates and extends the findings of Franklin et al. (2005a) and suggests that the categorical responding found in two-to-four year olds here is not caused by learning linguistic colour terms. The result however supports the view that colour CP may reflect the influence of universal colour categories.

3.3. Experiment 4: Target Detection Task: English and Himba Two-to-Four Year Olds

3.3.1. Introduction

Although it is hard to account for the results of Experiment 3 in terms of linguistic labelling strategies, it is not clear whether the pattern of categorical responding found in toddlers across the blue-green boundary reflects perceptual processes, or else derives from memory processes during the delay. For instance, representation of stimuli in memory may shift to a categorical, though non-linguistic, code, possibly reflecting universal colour category prototypes. Experiment 4 examined whether the categorical effect persisted on a task that measures perceptual processes more directly. A target detection task was used in which a target of one hue was shown against a uniform background of differing hue. In such a task, the target does not have to be memorised, merely detected against the background. Performance therefore depends on perceptual processes to a greater degree than on X-AB tasks. An additional
difference is that the target detection task measured response latency rather than accuracy.

Such tasks have already been used to show that when target and background are on different sides of the blue-green boundary infants fixate the target more rapidly than a green target shown against a green background (Franklin et al., 2005b). In Experiment 4 stimulus pairs from both sides of the blue-green boundary were used, allowing comparison of response latency of searches at the blue-green boundary region with those at within-category blue and green regions of colour space.

3.3.2. Method

Participants
Thirty-nine Himba and fourteen English (mean age = 46 months, S.D. = 3 months) two-to-four year olds took part in the study. Himba speakers were tested in their home villages. English speakers were tested in nursery schools in southern England.

Apparatus and Experimental Set Up
Data were collected using a laptop computer with a touch screen LCD display* viewed at 90° from a distance of 300 – 400 mm. In Namibia the experiment took place in a large canvas tent which cut out almost all external light. Data from the English sample were collected in a darkened room. Participants sat on an adjustable chair within comfortable reach of the display.

Stimuli and Design
The task used was a target detection task recording response latencies. Stimuli were three pairs of Munsell colours of constant value (7) and chroma (7). Each pair was separated by four hue units and straddled one of three measurement locations. Two of the locations 10B and 10G were chosen to lie equidistant (2.5 hue units) from the focal hues of blue (2.5PB) and green (7.5G) identified by Berlin and Kay (1969). Stimulus pairs around these two locations were within-category green and within-category blue. The third location, 7.5BG is the boundary between blue and green colour categories identified in numerous studies (e.g., Bornstein & Korda, 1984; Screen white point CIE x, y, Y values were: x = 0.31, y = 0.33, Y = 41.0, approximating CIE standard illuminant D65. Screen size was 270mm x 200mm.
Roberson, Davidoff, & Braisby, 1999) and stimuli around this location were category boundary stimuli. Displays consisted of a circular target (diameter 32mm) superimposed at a randomly selected location in a uniform coloured background (270mm x 200mm). The colour of target and background consisted of the colours of a given stimulus pair.

Procedure
In experimental trials, stimulus pairs from the three measurement locations (within-green, within-blue and boundary) were displayed eight times each, with each stimulus appearing as target and background equally often, making twenty-four trials in total. The task was to identify target location. Responses were made by touching the target location using a stylus. Correct location of target resulted in the cartoon image of a sheep appearing at the target location, accompanied by a tone. After each trial the screen turned a neutral shade of grey, remaining so for a minimum of 1000 ms until the following trial was initiated. Trials were initiated manually by the experimenter using a mouse pad attached to a long lead.

The task was explained as a game with the aim of finding a sheep hiding in the display. A demonstration was given and practice trials attempted using stimulus pairs selected from other regions of colour space. Participants were encouraged to make responses as rapidly and accurately as possible. Once it was clear that the task was understood and responses were being made rapidly, the experimental trials were taken.

3.3.3. Results
For each participant, median response latencies for trials involving each of the three stimulus pairs were calculated. Mean response latencies (in milliseconds) across participants are shown in Figure 3.3. and Table 3.2. Performance of both language groups appears similar, with quicker response latencies for trials involving the blue-green boundary stimulus pair than for either of the within-category stimulus pairs.
Table 3.2. Mean response latencies in milliseconds (SD) for measurement locations used in Experiment 4.

<table>
<thead>
<tr>
<th>Measurement location</th>
<th>Within-boundary</th>
<th>Within-green</th>
<th>Within-blue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Himba</td>
<td>4411 (2675)</td>
<td>2673 (1883)</td>
<td>5398 (4199)</td>
</tr>
<tr>
<td>English</td>
<td>4636 (2983)</td>
<td>2705 (1104)</td>
<td>5784 (4200)</td>
</tr>
</tbody>
</table>

Figure 3.3. Mean target detection task response latencies in milliseconds (±1 SE) for Himba and English two-to-four year olds in Experiment 4. Bold bar on right of figure gives ±1 within-subjects 95% confidence interval (Loftus and Masson, 1994, 2003).

Data were evaluated using an analyses of variance examining effects of measurement location, (Within-category green, Within-category blue and Category boundary) and language group (English or Himba). To test the main hypothesis, comparisons of performance at the boundary with performance at within-category locations and of
performance at the two within-category locations were made using planned comparisons.

There was a significant main effect of measurement location, (Means (S.D.): Within-green = 4470 (2731), Boundary = 2682 (1701), Within-blue = 5500 (4612)), $F(2, 102) = 10.86, MSE = 8176877, p < 0.001$, indicating that response latencies varied depending on measurement location. There was no overall difference in response latencies between the two groups, $F(1, 51) = 0.12, p = 0.73$, nor was the interaction between location and language group significant, $F = 0.039, p = 0.961$. Planned comparisons indicated response latencies at the boundary to be significantly lower than response latencies at either of the within-category locations, minimum $F(1, 51) = 16.93, maximum p < 0.001$. Response latencies for the trials at the two within-category pairs were not significantly different, $F(1, 51) = 2.10, p = 0.153$.

3.3.4. Discussion
The performance of both groups of two-to-four year olds was very similar, with response latencies significantly faster for searches at the blue-green boundary location than at either of the within-category locations. Franklin et al. (2005b) found a similar effect with infants and English speaking adults, using a similar task, although they only tested on one side of the blue-green boundary. The results of Experiment 4 provide additional evidence that categorical responding observed in two-to-four year olds is a perceptual effect that occurs even in a group of young children whose language does not encode colour terms marking the boundary across which the effect is assessed. This in turn implies that learning colour terms cannot be the cause of the CP effects found in this experiment.

3.4. Experiment 5: Target Detection Task: English and Himba Adults

3.4.1. Introduction
Together the results of Experiments 3 and 4 suggest that learning linguistic colour terms is unlikely to be the sole cause of colour CP found on perceptual tasks. If colour CP were caused by learning colour terms, it is difficult to account for the results of these two experiments. It is also difficult to explain earlier studies demonstrating
categorical responding in infants (Bornstein et al., 1976, Franklin et al., 2005a) and toddlers (Franklin et al., 2005b) in terms of language. Second, as Experiment 4 demonstrates, two-to-four year olds show CP effects on a perceptual task, implying such effects are unlikely to result from just memory processes. The results are more consistent with the idea that colour perception is governed by universal colour categories (Kay & Regier, 2003), or by universal foci around which colour categories are organized (Regier, Kay & Cook, 2005).

If this were the case, it is still possible that learning colour terms might influence colour perception at later stages in development. As is discussed earlier, cross-cultural studies have shown differential task performance in line with this possibility (Roberson et al., 2000; Roberson et al., 2004; Roberson et al., 2005), whilst category training studies (Özgen & Davies, 2002) suggest that colour perception may be malleable. What is lacking though is confirmation that the effects observed in both cross-cultural and category training studies reflect perceptual processes.

Experiment 5 therefore compared performance of Himba and English adults on a task with high perceptual demand. The task used was the same task used in Experiment 4. However, because chromatic thresholds are known to improve until young adulthood is reached (Knoblauch, Vital-Durand & Barbour, 2001) separations between stimulus pairs were reduced from four to two Munsell hue units.

Should learning linguistic colour terms alter perception then performance of Himba and English adults ought to differ across the blue-green region tested. Specifically, Himba adults showing relatively less CP across the blue-green boundary than English adults, or no evidence of CP across that boundary, would provide evidence both that learning colour terms alters perception and of the strength of any effect. On the other hand, similar patterns of performance by the two language groups would suggest that colour language learning has little or no effect on colour perception. Roberson et al., (2005) have previously found that Himba speakers do not show CP across the blue-green boundary. But the task used was an X-AB task with a delay. It is not clear whether the lack of CP in Himba adults on an X-AB task reflects perceptual processes or the influence of verbal labelling on task performance.
3.4.2. Method

Participants

Twenty-nine English speaking adults (mean age = 27 years 10 months, S.D. = 9 years 2 months) and seventy-two Himba speaking adults (estimated age range 16-50 years) took part in the study. Himba speakers do not keep an exact record of birthdates, so ages were estimated from physical appearance or from conversation with participants. Himba speakers were tested in villages in Namibia, English speakers at the University of Surrey in England. All participants had normal colour vision as assessed using the City Colour Vision Test (Fletcher, 1980).

Apparatus and Experimental Set Up

The same equipment as in Experiment 4 was used.

Stimuli and Design

Experimental design was as in Experiment 4, although stimulus pairs were separated by two rather than four Munsell hue units. The image of a sheep and the accompanying tone were removed. Viewing distances were increased to 500 mm.

Procedure

The task was explained and practice trials were presented. Participants were asked to respond as rapidly as possible. Following the practice trials, experimental trials were presented.

3.4.3. Results

For each participant, median response latencies (in milliseconds) for each of the three measurement locations (Within-category green, Within-category blue and Category boundary) were calculated. Figure 3.4. and Table 3.3. give mean response latencies across participants. Comparison of Figure 3.4. with Table 3.3. shows the pattern of performance across conditions for adults appears similar to that of two-to-four year olds.
Table 3.3. Mean response latencies in milliseconds (SD) for measurement locations used in Experiment 5.

<table>
<thead>
<tr>
<th>Measurement location</th>
<th>Within-</th>
<th>Boundary</th>
<th>Within-</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>green</td>
<td></td>
<td>blue</td>
</tr>
<tr>
<td>Himba</td>
<td>6421 (4117)</td>
<td>4096 (3291)</td>
<td>8457 (5087)</td>
</tr>
<tr>
<td>English</td>
<td>4512 (3566)</td>
<td>2881 (1674)</td>
<td>6439 (3946)</td>
</tr>
</tbody>
</table>

Figure 3.4. Mean target detection task response latencies in milliseconds (±1 SE) for Himba and English adults in Experiment 5. Bold bar on right of figure gives ±1 within-subjects 95% confidence interval (Loftus and Masson, 1994, 2003).

Although Himba adults' response latencies are larger overall than those of English adults, the pattern of responding appears similar in both groups. Response latencies appear lower for trials involving the boundary stimulus pair, than for trials using either of the within-category stimulus pairs.
An analysis identical to that used in Experiment 4 was carried out. This confirmed there was a significant main effect of measurement location (Means (S.D.): Within-green = 5873 (4043), Boundary = 3747 (2963), Within-blue = 7877 (4855)), F(2, 198) = 19.63, MSE = 16512799, p < 0.001. Overall response latencies were significantly quicker in the English than in the Himba group (Means (S.D.): Himba = 6324 (2330), English = 4611 (1740)), F(1, 99) = 12.78, MSE = 14251704, p < 0.005. Crucially, there was no interaction between measurement location and language group, F = 0.237, p = 0.789. Planned comparisons showed response latencies at the boundary to be significantly lower than response latencies at either of the within-category locations, minimum F(1, 99) = 18.41, maximum p < 0.001. Response latencies for trials involving the within-category blue stimulus pair were significantly larger than for trials involving the within-category green pair F(1, 99) = 7.77, p < 0.01.

3.4.4. Discussion
Adults in both language groups showed a similar pattern of performance, although Himba adults' response latencies were overall higher than those of English adults. This may reflect differences between the two groups unrelated to colour perception since there is no evidence to suggest colour perception in Himba speakers is uniformly poorer than amongst English speakers. Response latencies were overall significantly quicker at the blue-green boundary than at either of the within-category locations. This suggests that colour CP is at least in part perceptual. However because both groups show lower response latencies across the boundary, it is difficult to see how the underlying cause can be colour language. Similarly, the effect does not appear to be any larger in the English than in the Himba group, which suggests that if colour language does affect colour perception, the effect is too small to be detailed here.

There was also a significant difference between response latencies at the two within-category locations. Searches using the within-category green stimulus pair were quicker than for the within-category blue stimulus pair. It is not clear why this should be the case. Since both within-category stimulus pairs were positioned the same distance from the respective hue values of the universal green and blue category foci, identified by Berlin and Kay (1969), it appears unlikely that the difference can be
accounted for in terms of the influence of category prototypes. The difference may
however reflect the relatively greater distance of the within-category blue stimulus
pair from the blue-green category boundary than of the within-category green
stimulus pair. If proximity to the category boundary has a greater influence on
response latencies than proximity to the category foci, then trials involving within-
category green stimuli should be faster than trials involving within-category blue
stimuli, as was found in this experiment.

3.5. Experiment 6: Just Noticeable Difference Task: English and Himba
Two-to-Four Year Olds

3.5.1. Introduction
Experiments 4 and 5 (and to a lesser extent Experiment 3) suggest that there is a
perceptual component to CP, but find no evidence that language can affect colour
perception. To assess the robustness of these findings, two further experiments were
carried out comparing Himba and English participants from the two age groups.
Whereas the target detection task measured response latencies for supra-threshold
stimuli, Experiment 6 used a different task, measurement of JNDs, thereby providing
an additional means of assessing CP in the blue-green region of colour space.
Measurement of JNDs also allows more direct assessment of theories of CP which
suggest perceptual space may become ‘warped’ (Hamad, 1987).

The tasks used in the Experiment 6 estimated JNDs using a paper based task and a
procedure simpler than that used in Experiments 1 and 2. Paper stimuli were used and
a simplified task was devised to take account of the limited attention span of two-to-
four year olds.

Patterns of performance on the JND task mirroring those on the target detection tasks
would provide additional evidence that colour CP is at least in part a perceptual
phenomenon. A different pattern of performance would imply the results the target
detection task should not be generalized and that supra-threshold performance
patterns may differ from those at threshold.
3.5.2. Method

Participants
Forty-two Himba and thirty-two English (mean age = 39 months, S.D. = 7 months) two-to-four year olds took part in the study. Himba speakers were tested in their home villages, English speakers in nursery schools in southern England.

Apparatus
A solar powered light box, similar to that described in Roberson et al. (2000) was used to provide illumination. This provided approximate illuminant C (natural shaded daylight). Stimuli were created using a Hewlett-Packard design jet 10PS colour proofer using a six-ink, 2,400-dpi system. Colorimetric measurements were made using an Avantes Spectrocam.

Stimuli and Design
Experimental items consisted of coloured rectangular backgrounds (250mm x 170mm) printed on gloss finish A4 photo paper. Margins (28mm top, 14mm bottom, 19mm right, 28mm left) were left blank. On each item, a target was embedded close to one of the four corners. The target was the outline of a dog (maximum length 66mm, maximum height 57mm) differing from the background in colour. To minimize the visibility of the border between the target and background, a five-pixel wide Gaussian blur was inserted on the border between target and background. Items were covered with a sheet of transparency to protect them from damage. Figure 3.5. shows an example of an item used in the experiment.

Colorimetric properties were specified in CIE (1976) L*a*b* colour space. Four sets of experimental items were produced. In each the separation between background and target was varied systematically to produce a range of separations between 2.5 and 40 L*a*b* units (ΔE). In each set co-ordinates of the target remained constant. Differing separations were achieved by varying the co-ordinates of the background. Items at 11 different ΔE separations from the background (approximately 2.5, 5, 7.5, 10, 12.5, 15, 20, 25, 30, 35 and 40 ΔE units) were produced in each set. Two items were produced for each separation level, the difference between them being that the target location differed. Separations between backgrounds and target were defined primarily by a*b* chromaticity differences. Luminance was kept constant, though was subject to small
variations brought about by constraints in the equipment used to create the items. Each of the four sets spanned a different region of colour space.

Two of the sets measured within-category JNDs, in green and blue regions respectively (within-category sets). The other two measured JNDs around the boundary between green and blue colour categories (category boundary sets). Figure 3.6. is a conceptual diagram showing how, in each set, background colour varied. L*a*b* co-ordinates of item backgrounds and targets are given in Appendix 2. A number of practice items were also produced. In these, targets identical to those in each of the sets were printed against backgrounds of differing primary colours.

Figure 3.5. Monochromatic example of an item used in Experiment 6. Location of the target (the outline of a dog) can be in any of the corner locations.

At separations of 10 ΔE units or smaller, average contribution of luminance to overall ΔE separations was less than 10% of total separation distance. Luminance differences were approximately equal for within-category and category boundary sets. (For separations less than 10 ΔE units, average luminance contribution in within-category sets was 9.75%, in category boundary sets, 6.75%.)
Figure 3.6. Conceptual diagram illustrating the four stimulus sets used in Experiment 6. Two sets measured within-category JNDS, the other two sets measured JNDS at the blue-green category boundary. Colorimetric properties of target were the same in each set. Target background separation was manipulated by changing colorimetric properties of the background.

Procedure

Items were displayed in the light box. Himba speakers completed the experiment outside, in shaded conditions. English speakers were tested indoors with the light box again used to provide illumination. For each participant, JNDS were measured twice, once using a within-category set and once with the category boundary set. These were taken in a counterbalanced order.

The experiment began with a number of trials using practice items. Participants were told a dog was 'hiding' in the coloured background and asked to indicate its location.

---

\(^{18}\) The category boundary set taken was always the one in which target colour belonged to the same category as the target colour in the within-category set taken. Thus if the within-category blue set was used to assess within category JNDS, the across category set used was the one where target colour lay to the blue side of the blue-green boundary.
by pointing. After correct responses, praise was given and a new item presented: following rare incorrect responses the child was reminded of the task and instructed to look again. If still unable to identify target location the experimenter indicated it by pointing. Practice continued until a minimum of three successive correct responses were recorded. The first of the test sets was taken next. Instructions were as before, though it was added that every item contained a target (dog) which might be difficult to see. The item was displayed until a response was made. If the participant suggested there was either no target present, or that they could not see it, they were encouraged to continue looking. If they still made no pointing response, an incorrect response was recorded. Correct responses were recorded when the appropriate corner of the display was indicated.

A simple adaptive procedure, essentially a 1-up, 1-down staircase converging on an estimate of threshold 50% correct, was used to determine the item shown in each trial. The procedure was based on the method described by Wetherill & Levitt (1965) and evaluated by Wetherill, Chen & Vasudeva (1966). The first item shown in each set was that with maximum separation between target and background (40 ΔE units). With each subsequent correct response, ΔE separation was halved until the first incorrect response was recorded. ΔE separation was subsequently increased for incorrect responses or decreased for correct responses by the smallest step available. The task continued until six reversals of direction was achieved, resulting in successive 'runs' of correct and incorrect responses, as shown in Figure 3.7. Three separate estimations of JND were made, each being based on the mid-point of every run of incorrect responses. To prevent participants from learning a pattern of responding the two items available in each set at each separation size were used interchangeably as required.
3.5.3. Results

Ten Himba two-to-four year olds either failed to complete the experimental task or else failed to identify target location in trials with separation of 20 ΔE units or more. This could indicate either colour vision abnormalities or inattentiveness to the task. Their data were excluded from the analysis, as was that of five English participants for similar reasons. The task produced two JNDs for each participant: one for a within-category set, the other from a category boundary set. A pronounced floor effect was encountered in category boundary sets, with 43.3% of participants correctly identifying target location in these sets. For within-category sets, the equivalent figure was 3.3%. The floor effect was strongest amongst English participants with 74% performing at floor in category boundary sets. Where targets were correctly identified on trials involving the smallest available ΔE separation, a JND value of 2.5 ΔE units was assigned. Figure 3.8. gives JND values calculated using this procedure. It appears that JNDs in both groups are smaller measured at category boundary than at within-
category locations. The category effect appears considerably larger for the English than the Himba group.

![Bar Graph](image)

**Figure 3.8.** Mean JNDs (±1 SE) for English and Himba two-to-four year olds for within-category and category boundary sets in Experiment 6. Bold bar in centre gives ±1 within-subjects 95% confidence interval (Loftus & Masson, 1994, 2003).

An analysis of variance examining the effects of category (Within-category and Category boundary) and language (Himba and English) with category as a repeated measure, confirmed these impressions.

There was a significant effect of category, (mean, (S.D.): Within-category = 6.40 (1.73), Category boundary = 4.33 (2.31)), $F(1, 57) = 54.43, MSE = 2.50, p < 0.001$, indicating JNDs were significantly smaller at the category boundary than within-category. The language by category interaction was also significant, $F(1, 57) = 10.91, MSE = 2.50, p < 0.005$. Paired samples t-tests showed that JNDs in both groups were lower for boundary than within-category sets, minimum $t(32) = 2.54$, maximum $p < 0.025$. However when a difference score was calculated by subtracting category boundary JNDs within-category JNDs and the resultant values compared, the category effect was found to be significantly greater for the English group, $t(52.06), p < 0.005$. 

76
3.5.4. Discussion

The results of Experiment 6 do not appear to be entirely consistent with the results of Experiments 3-5. Although the category effect is present in Experiment 6 in both language groups, it is larger in the English than in the Himba group. Since the green-blue boundary is marked in English, but not Himba, it is possible the difference in performance is attributable to an effect of language on perception. Although the floor effect made accurate assessment of the size of the category effect impossible, it is very unlikely that this affects the finding, because performance at floor was more pronounced amongst English than Himba participants in category boundary sets. However, the assertion that the differences between the two language groups found in this experiment are caused by differences in colour language should be treated cautiously because no difference was found between performance of two-to-four year old groups in the target detection task used in Experiment 4. So if there is a linguistic effect it appears not to influence performance of the target detection task. Furthermore, if the difference between the two groups in this experiment is due to a language effect, then it should be more pronounced when comparisons are made amongst adults whose knowledge of the colour terms in their language is more complete than that of two-to-four year olds.

3.6. Experiment 7: Just Noticeable Difference Task: English and Himba Adults

3.6.1. Introduction

As in Experiment 6, a paper based task was used to measure adult JND thresholds, with colorimetric properties defined in CIE (1976) L*a*b* space. However there were a number of differences between the two experiments. The task used in Experiment 7 was a 6-AFC task. Trials involved items made up of six rectangular stimuli, one of which contained a target (again the outline of a dog) defined by a colour difference. In order to avoid the floor effect found in the previous experiment and because adult chromatic thresholds are considerable lower than those of children (Knoblauch et al., 2001), stimuli capable of measuring JNDs down to 0.25ΔE units were produced.
3.6.2. Method

Participants
Sixteen English speaking (mean age = 23 years 5 months, S.D. = 6 years 3 months) and fourteen Himba speaking adults (estimated age range 16-40 years) took part in the study. Himba speakers were tested in villages in Namibia, English speakers at the University of Surrey. All participants had normal colour vision as assessed using the City Color Vision Test (Fletcher, 1980).

Apparatus
Apparatus was the same as that used in Experiment 6, with the addition of a grey painted plywood background (200mm x 290mm) on which stimuli were mounted.

Stimuli and Design
Stimuli were of two kinds. ‘No target’ stimuli were rectangular (80mm x 90mm) uniform coloured patches printed on high quality glossy photographic paper. Target stimuli differed only in containing the image of a dog (maximum length 66mm, maximum height 57mm) shown against the same uniform background. Experimental items consisted of six stimuli attached to the plywood background in a two by three array, as shown in Figure 3.9. On each trial five stimuli were ‘no target’ stimuli, the other was a target stimulus. Location of target stimuli varied randomly from trial to trial. Stimuli were created to make four sets that measured JNDs at within-category and category boundary locations in the blue to green region of colour space. Two sets measured JNDs at category boundary locations, differing in the direction of measurement, one set at within-category blue and one at within-category green locations, as shown in Figure 3.10. The colour of stimulus backgrounds was the same in each set. Target stimuli differed from ‘no target’ stimuli only in containing the target of a different colour. The colour of the target in target stimuli was varied such that in each of the four sets ΔE separation between the target and background could be varied between 8 levels of ΔE separation (0.25, 0.5, 1, 1.5, 2, 3, 4 and 5 ΔE units). This was achieved by varying the colour of the target. As separation sizes increased, the dog became increasingly noticeable against the background. ΔE separation differences were mostly due to a*b* variation. There were very minor L*a* differences, though overall these were equated for category boundary and within-category sets. On target stimuli, to minimize visibility of the border between target and background on
target stimuli, a five-pixel wide Gaussian blur was inserted between target and background, as in Experiment 6. Stimuli L*a*b* co-ordinates are given in Appendix 3. To prevent damage, stimuli were covered with a sheet of transparency before colorimetric measurements were taken and were stored in darkness when not in use. Subsequent remeasurement confirmed that stimuli colorimetric properties had remained constant.

Figure 3.9. (to left) Example of an item from Experiment 7. Target location can be either as shown or in any one of the other five squares.

Figure 3.10. JND measurement sets used in Experiment 7. Two sets measured within-category JNDs, two measured category boundary JNDs. In Experiment 7, target-background separation was varied by changing colorimetric properties of the target. Background colorimetric properties remained constant in each stimulus set.
**Procedure**

Items were displayed in the light-box described in Experiment 6. Himba speakers completed the experiment outside, in shaded conditions. English speakers were tested indoors with the light box again used to provide illumination. Both groups attempted all four test sets. Order was counterbalanced. Instructions were to identify which of the six stimuli contained the outline of dog. If unable to do so, participants were required to guess. Prior to attempting experimental trials, a set of practice stimuli was used to gain familiarity with the task.

The procedure used to determine the ΔE separation of the target stimuli used in each trial was almost identical to that used in Experiment 6, although separations were always decreased or increased by the smallest unit available. As before, three separate estimations of JND were made. Each was based on the mid-point of every run of incorrect responses.

### 3.6.3. Results

JNDs for each participant were calculated as in Experiment 6. JNDs from the two sets measuring JNDs at the blue-green boundary were combined to produce a single boundary JND estimate. This gave three JND estimates, one at the blue-green category boundary and one each at within-category green and blue locations. JNDs calculated in this way are shown in Table 3.4 and Figure 3.11. Overall, JNDs for Himba adults appear to be higher than for English adults. The pattern of performance appears similar in both groups, with JNDs smaller at the blue-green boundary than at either within-category location.

Data were analyzed using the same analysis of variance as was used in Experiments 4 and 5, examining effects of stimulus set, (Within-category green, Within-category blue, Boundary) and language group (English or Himba). Stimulus set was a repeated measure. Planned comparisons were used to compare performance at the three measurement locations.
Table 3.4. Mean JNDs in L*a*b* units (SD) for measurement locations used in Experiment 7.

<table>
<thead>
<tr>
<th>Measurement location</th>
<th>Within-green</th>
<th>Boundary</th>
<th>Within-blue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Himba</td>
<td>1.50</td>
<td>1.15</td>
<td>1.50</td>
</tr>
<tr>
<td></td>
<td>(0.43)</td>
<td>(0.36)</td>
<td>(0.74)</td>
</tr>
<tr>
<td>English</td>
<td>1.28</td>
<td>0.86</td>
<td>1.10</td>
</tr>
<tr>
<td></td>
<td>(0.22)</td>
<td>(0.18)</td>
<td>(0.31)</td>
</tr>
</tbody>
</table>

Figure 3.11. Mean JNDs (±1 SE) for Himba and English adults in Experiment 7. Bold bar on right of figure gives ±1 within-subjects 95% confidence interval (Loftus and Masson, 1994, 2003).

There was a significant main effect of stimulus set, (Means (S.D.): Within-category green = 1.38 (0.35), Boundary = 1.00 (0.31), Within-category blue = 1.29 (0.58)), $F(2, 56) = 7.23$, $MSE = 0.164, p < 0.005$, indicating that overall JNDs differed between stimulus sets. The interaction between stimulus set and language group did not approach significance, $F = 0.335, p = 0.717$. Overall JNDs were significantly smaller...
for English than for Himba participants (Means (S.D.): Himba = 1.39 (2330), English = 1.08 (1740)), $F (1, 28) = 12.57, MSE = 0.17, p < 0.005$. Planned comparisons showed JNDs at the blue-green boundary to be significantly lower than JNDs measured using either of the within-category stimulus sets, minimum $F (1, 28) = 6.38$, maximum $p < 0.025$. JNDs for the two within-category sets were not significantly different, $F (1, 28) = 0.47$, $p = 0.501$.

3.6.4. Discussion

The results of Experiment 7 indicate a category effect at the blue-green boundary. Although Himba adults’ JNDs were higher overall, no differences in the pattern of performance between the two language groups were found. The results are therefore similar to those of Experiments 4, which found evidence of CP in adults by measuring response latencies. Since a different measure (JNDs) was used in this experiment, this would appear to provide converging evidence that, in adults, colour CP does not depend on language. In addition, as no difference between groups was found in either Experiment 4 or 7, these experiments provide no evidence that colour language can modify colour perception. However in this respect, the results of this experiment differ from those of Experiment 6, which measured JNDs in two-to-four year olds and found that the size of the CP effect differed between the two language groups. This finding is discussed further in the next section.

3.7. General Discussion of Experiments 3 to 7

The experiments presented in Chapter 3 compared performance of Himba and English speakers on tasks using stimuli drawn from the blue-green region of colour space. Himba speakers lack separate terms for green and blue, so if linguistic colour categories can affect colour perception, differences between the two groups ought to be found. In general, the results of the experiments do not provide support for this idea.

Experiment 3 used an X-AB task to test two-to-four year olds in both language groups and found CP across the blue-green boundary.
Experiments 4 and 5 tested two-to-four year olds and adults from the two language groups using a target detection task. CP was found in all groups, with no differences between the different language groups.

Experiment 6 and 7 measured colour JNDs in the same groups. Again CP was found in all groups tested. In the case of two-to-four year olds, the categorical effect was larger in the English than in the Himba group. There was no evidence that the CP found in adults differed between the two language groups.

Overall, the most striking feature of the experimental results is the similarity in the pattern of performance found in the experiments, regardless of age, language or task. In particular, it appears that performance on all the tasks is improved in trials where stimuli come from the green-blue boundary region of colour space rather than blue or green regions.

These CP effects appear to be the result of perceptual processes because, as is argued elsewhere, the tasks used in Experiments 4-7 tap low-level perceptual processes. Experiment 3, on the other hand, used a task thought vulnerable to labelling strategies. However performance of Himba two-to-four year olds tested was not different from that of English two-to-four year olds and a significant CP effect was found. Franklin et al. (2005a) presents similar findings. The results of these experiments therefore appear to run counter to claims that colour language is responsible for the experience of perceptual colour categories. If this were the case, it is difficult to explain the similarities in performance found between the different language and age groups tested.

What is also notable is the different pattern of results found in the experiments presented here and those found in previous cross-cultural studies (e.g., Roberson et al., 2000; Roberson et al., 2005) which have provided evidence that differences in colour perception exist between speakers of different languages. These differences may arise because of differences in task demands. It is suggested in the introduction to this chapter that, as other authors have noted, X-AB (Roberson & Davidoff, 2000) and triadic judgement (Kay & Kempton, 1984) tasks may be vulnerable to labelling strategies. The tasks used in Experiments 4-7 here are either immune to such
strategies or at least considerably less vulnerable to them. The difference between the results of Experiments 4-7, which find CP in the absence of a linguistic boundary, and research that has found CP depends on the presence of a linguistic boundary, may then reflect these differences in task demands.

Because the tasks used here place more emphasis on perceptual processes than those used previously in cross-cultural studies of colour perception, it is suggested that the differences between different language groups found previously do not stem from large scale differences in colour perception. They may reflect differences in colour cognition. When perceptual processes are measured though, Himba speakers show CP across a boundary not marked in their language. So do young children and even infants, as numerous previous studies appear to show (e.g. Bornstein et al., 1976; Franklin et al., 2005b). The experiments presented here therefore provide converging evidence supporting the view that colour perception is marked more by universal tendencies than by cultural variation, as other work emphasising the evolutionary trajectory of colour term acquisition by languages (e.g., Berlin & Kay, 1969) and the apparent universal salience of particular areas of colour space (Regier et al., 2005) has suggested.

The results presented here also suggest that only the weakest versions of the linguistic relativity hypothesis can be supported. Numerous studies show that performance of tasks thought to involve higher cognitive functions, such as X-AB and triadic judgement tasks, do appear to be influenced by colour language. But this influence of language does not appear to influence low-level colour perception.

There are two caveats to these conclusions. First, the results of Experiment 6, which measured JNDs in two-to-four year olds, present a contrast to the results of the other experiments presented in this chapter. The significant difference in the pattern of performance between the two language groups found in this experiment was not mirrored in any of the other experiments. This difference may be caused by differences in colour language between the groups influencing colour perception by ‘warping’ perceptual space. However such an explanation is not consistent with the results of Experiments 4, 5 and 7. If learning colour categories could have this effect on perceptual processes, differences between the groups should also have been
apparent in these experiments. These differences should also be more pronounced in experiments involving adults (who have good knowledge of colour terms), than in young children. This was not the case. Performance patterns in both language groups in the experiments involving adults were similar. Performance of the two-to-four year olds from the different language groups in Experiment 4 was also very similar. These factors suggest that differences in colour language may not account for the significantly different pattern of results found between the groups tested in Experiment 6.

Second, even discounting the results of Experiment 6, the results of the other experiments presented in this chapter do not rule out the possibility that learning colour terms may have an influence on colour perception. What they do suggest is that any such influences are relatively minor in scale. The perceptual colour categories presumed to drive performance of the tasks used here may be universal, but may also be susceptible to language driven ‘fine tuning’. Comparing performance of speakers of different languages using tasks like those used in Chapter 2, may provide evidence demonstrating these hypothesised differences.
Chapter 4
Lens Brunescence and Colour Perception

4.1. Introduction

The previous chapter presented a series of experiments indicating that Himba speakers, whose language lacks separate blue and green colour terms, show a pattern of CP across the blue-green boundary similar to that found in English speakers. Whilst the possibility that linguistic colour category distinctions may serve to ‘fine-tune’ cannot be ruled out, colour language does not appear, on the basis of the experiments presented in Chapter 3, to cause colour CP. The evidence from these experiments does not support the ideas either that languages divide colour space in an arbitrary manner, or that the perceptual experience of colour categories is shaped by colour language.

Nevertheless, differences between the language groups were apparent in Chapter 2. In particular Himba adult JNDs (Experiment 7) were significantly higher than those of English adults. Similarly, response latencies on a target detection task were also higher for Himba than for English adults (Experiment 5). Although the same pattern was found in Experiment 6, which measured JNDs in Himba and English two-to-four year olds, the performance of two-to-four year olds on a target detection task (Experiment 4) was similar across languages. Whilst these differences in overall levels of performance may be caused by a number of factors, it is possible they reflect differences in perception related to factors other than colour language. Consequently this chapter considers a quite different explanation to account for differences in colour language. Rather than considering whether colour language differences alter colour perception, the theories considered in this chapter argue the possibility that differences in colour language may be influenced by differences in colour perception.

4.1.1. Changes in Colour Perception and Changes in Ocular Media Associated with Aging and Exposure to UV-B light

In the general introduction to this thesis, potential influences on colour perception besides colour language were discussed. The idea that differences in sensitivity to
short-wavelength light may help explain differences in colour language, in particular
the lack of separate blue-green terms in many languages, was considered. Rivers
(1901b), it will be recalled, was the first to draw attention to this possibility. It has
subsequently been shown that spectral lens density increases during life (see Pokorny,
Smith & Lutze, 1987) and that increasing exposure to UV-B hastens aging of the lens
(Javitt & Taylor, 1994), which causes it to turn a yellowish colour. This results in
increased absorption of short-wavelength light. Noting, first, that there are increasing
amount of UV-B light at the equator and, second, that languages spoken closer to the
equator are less likely to encode separate blue and green terms, Bornstein (1973)
suggested a connection between the two.

More recent studies have drawn attention to the apparent differences in tritan errors
made on colour vision tests between populations living at different latitudes. Tritan
errors are of a kind made by tritanopes, who lack an active S-cone system, or
tritanomalous subjects who, whilst not lacking an active S-cone system, make the
same kinds of confusions, involving yellow, through green to blue colours. Tritanopia
is a rare (<0.005%) inherited form of colour blindness, but other, acquired, factors can
also cause tritanomalous errors on colour vision tests. For example, diets low in some
B-vitamins, for example, can cause acquired tritan defects (Fletcher and Voke, 1985).
Relative increases in absorption of short-wavelength light would also increase tritan
errors. Laws, Davies, Corbett, Jerrett & Jerrett (1995) found that an unusually high
proportion of Setswana speaking Botswanans showed mild tritanomoly. A follow-up
study (Davies, Laws, Corbett and Jerrett, 1997) compared samples from the U.K.,
Ireland, Spain, Greece, Botswana, Malawi and South Africa. Whilst tritan defects
were absent in the Spanish, Irish and British samples, almost 20% of Greek and
African participants made colour vision errors, most of them tritan errors. This pattern
of errors was similar to that of a further British sample aged over 65 years. This led
the authors to conclude that the most likely cause of the number increased tritan errors
was pre-retinal filtering of short-wavelength light. A number of other studies have
also suggested a lowered sensitivity to short-wavelengths (equivalent to tritanomoly –
see below) corresponds to increased UV-B exposure (e.g., Ishak 1952a, 1952b;
Sperling & Hsia, 1957).
If the increased incidence of tritan errors in some populations found in the studies described above is caused by selective filtering of short-wavelength light there are several potential explanations. Bornstein (1973) suggested that, because high levels of UV-B damage the retina, selective filtering of short-wavelength would be useful in areas of high UV-B exposure and therefore may be an inherited characteristic. Besides offering protection, such an adaptation would also enhance visual acuity (Foley-Fisher, 1968). Like Rivers seventy years earlier, Bornstein felt an increase (relative to other populations) in macula pigmentation the most likely candidate. However, as is stated in the general introduction, no evidence exists that macula lutea density increases amongst people living closer to the equator. Nor is there much evidence of inherited differences in short-wavelength light filtering. Gaines and Powell (1981) compared populations of Negroid and Caucasian children, but found no evidence to support the idea. On the other hand, there is evidence (Javitt & Taylor, 1994) indicating that increased exposure to UV-B turns the lens yellow. It follows then that the cause of increased incidence of tritan errors found by Davies et al., (1997) and Laws et al., (1995) can, at present, be satisfactorily accounted for by premature lens aging.

It is worth making clear the links being proposed between lens aging and colour language. Lens aging produces yellow ocular pigmentation and this in turn has effects on vision that are identical to tritanomoly or tritanopia (Judd, Plaza & Farnsworth, 1950). The pattern of confusions produced, rendering blues and greens less discriminable, blues blacker and purples redder has parallels with the linguistic colour categories frequently found in equatorial regions (Lindsey and Brown, 2002). That is to say the colour category structures found in many equatorial languages seem to reflect tritanomalous colour deficiencies.

Lindsey and Brown’s (2002) study, also mentioned in the general introduction, provides evidence that colour naming is affected by absorption of short-wavelength light. In their experiment, sixteen to twenty-seven year old American English speakers named colours as viewed through clear and yellow lenses. As the simulated lenses became older and more brunescent (yellower), participants named fewer and fewer stimuli ‘blue’ or ‘purple’. The authors claimed, The qualitative similarity between the colour naming at the oldest, darkest lens values in our experiment and
colour naming by the speakers of ‘grue’ and ‘dark’ languages is striking. The claim made then is that the lack of separate blue and green terms in many languages may be explicable in terms of premature lens aging caused by exposure to high levels of UV-B light. Such a claim is attractive because it would represent a universal causative factor and so tie in both with Berlin and Kay’s (1969) theory of universal colour categories and the known prevalence of ‘grue’ languages (those lacking separate blue and green colour terms) around the equator.

There are however several problems with Lindsey and Brown’s (2002) theory. First, the lens yellowing mechanism suggested seems to be acquired rather than inherited. If the eleven basic categories are assumed to be universal (through the young lens), then the blue and green categories ought to appear distinct for young speakers of ‘grue’ languages. This raises the issue of why, even given premature lens aging, separate ‘blue’ and ‘green’ terms are not encoded. In response, Lindsey and Brown argue that even if tritan colour anomalies only affect a proportion of a given population, then a significantly larger proportion of all conversations would involve participants with colour vocabularies that are affected. Thus, it is argued, a ‘grue’ term could become accepted in a language even if many of the speakers of that language were able to adequately distinguish green from blue. Steel and Belpaeme (2005) make a similar point.

A second objection to Lindsey and Brown’s (2002) theory comes from a similar study carried out by Hardy et al. (2005). Lindsey and Brown’s (2002) original findings were replicated but the stimuli were also named by a group of older observers (average age was 74 years). Despite their age this group, amongst whom prior ocular media measurement confirmed lens brunescence, nevertheless named stimuli in much the same way as younger observers named unfiltered stimuli. This suggests that people are able to preserve the location of category boundaries in colour space even in the face of the changes brought about by the aging lens. For this to be the case, the visual system must, as Hardy et al. (2005) point out make compensations to account for the

---

19 ‘Grue’ languages refers to those languages that do not encode separate blue and green colour terms.
20 Lindsey and Brown (2002), pp 509.
21 For example, in a conversation involving two speakers, if for one speaker the terms blue and green do not signify a meaningful perceptual difference, then that person’s colour vocabulary would be affected.

89
slow changes over time in the spectral composition of light entering the eye. This appears problematic for Lindsey and Brown’s (2002) theory. The reason is that once, say, the blue-green distinction is encoded in a language, the situation where it has no perceptual validity for a proportion of that languages speakers would appear not to arise. Then again, it is not known how the level of lens brunececence in the aged sample tested by Hardy et al., (2005) compares to that of speakers of pre-industrialised languages in equatorial regions. No-one has yet made the necessary measurements. As Lindsey and Brown (2002) point out, if lens aging is related to lifetime UV-B exposure, then a thirty-year-old Southern African may have a lens similar to that of a hundred year old northern European. The contrast would be even greater if the typical Southern African were to spend more time outside. Although Hardy et al. (2005) demonstrate that colour naming can be preserved, it is not clear whether this preservation of blue-green distinctiveness survives in situations of greater lens brunececence than those of the participants they tested. It is not clear either how the profile of lens bruneceence in an equatorial population compares to that of the participants in Hardy et al’s (2005) study. Consequently, it is not possible to discount Lindsey and Brown’s (2002) hypothesis on the basis of Hardy et al’s (2005) study. Furthermore, absorption of short-wavelength may have the effect of compressing blue and green regions of colour space, relative to other regions, so that even if it remains possible to distinguish between green and blue stimuli, such considerations apply to a relatively smaller portion of overall colour space.

A final criticism that applies to both Lindsey and Brown’s (2002) and Hardy et al’s (2005) studies is that stimuli were all highly saturated and of equivalent brightness levels. Given that changing brightness levels influence colour naming (shown for example in the Bezold-Brücke effect), neither of the two studies discussed above catalogue how lens bruneceence might effect naming of stimuli at differing brightness levels.

In summary, there appears to be a higher than expected incidence of apparent tritanopia in speakers of languages that do not use separate blue and green terms. Underlying this tritanomaly is thought to be increased absorption of short-wavelength light prior to its reaching the retina, probably caused by premature aging of the lens.
This may result in less ‘perceptual-pressure’ to split blue and green categories (Davies et al., 1997) and hence may explain at least some proportion of inter-language colour category differences.

4.1.2. Overview of Experiments
This chapter presents five linked experiments examining the role increased absorption of short-wavelength light has on colour perception in the blue-green region of colour space. Two groups of English speakers were compared: one group comprised participants aged fifty years and over, the other of participants aged under thirty.

As described above, there is evidence that lens brunescence increases tritan type errors on colour vision tests (e.g., Davies et al., 1997) but, when it occurs naturally over lifespan, does not diminish the ability of participants to accurately name blue and green stimuli (Hardy et al., 2005). The experiments presented in this chapter aim to extend previous work by comparing performance using in addition to a naming task, JND and similarity judgement tasks. An additional experiment was included to provide some indication of the relative absorption of short-wavelength light. Differences in performance between the two groups can be used to assess the idea that differences in the colour terms languages encode might, to some extent, reflect physiological differences in colour vision amongst their speakers. In the case of the experiments presented here, the issue is whether lens aging, believed to be accelerated amongst speakers of ‘grue’ languages, could help to explain the lack of separate blue and green terms in those languages.

Large scale differences between the two groups were not expected for several reasons. Most important was that all participants who took part in the experiment passed the City University Colour Vision Test (Fletcher, 1980). This test includes plates designed to detect tritanopia and it is presumed that those with advanced brunescence would fail those plates. The two groups of participants were also relatively close in age (average ages were twenty-one and fifty-eight years). Participants were resident of a region (UK) of relatively low UV-B intensity. Similarly, it is likely that participants spent large amounts of time indoors and wore sunglasses when outside in sunlight. These latter factors are likely to limit UV-B damage to the eye.
Experiment 8a measured pre-retinal filtering of light to establish that significant differences in short-wavelength light absorption exist between the two age groups.

Experiment 8b measured JNDs in the blue to green region of colour space to examine first whether age related differences in patterns of JNDs exist and second whether any such differences reflect pre-retinal filtering of light.

Experiment 8c was a dyadic judgements task exploring whether there are differences in perceived similarity of stimuli between the two age groups. Again the findings were linked to pre-retinal filtering.

Experiment 8d was a colour naming experiment to examine whether differences in naming exist between the two groups tested.

Experiment 8e was a colour naming experiment using the same stimuli as in Experiment 8d, though with simulated lens brunescence.

Experiments 8a-e used stimuli specified in a colour space derived from CIE (1931) colour space. Because equal perceptual spacing between stimuli was an important consideration, stimuli were specified in perceptually uniform derivation, CIE (1976) L*u*v* colour space. Munsell colour space was not used because it retains perceptually uniformity only when changes affect just one of its three dimensions. When equiluminant stimuli are specified in straight lines, as was the case here, changes in both the hue and saturation of stimuli specified in Munsell space occur. Although CIE spaces are suited to the purposes for which they were employed here, the appearance of stimuli thus defined is not readily apparent from the notation used. In consequence where it is useful the appearance of the stimuli used in the experiments is specified.

The relation between stimuli used in Experiments 8b-e were also arranged to highlight any differences in tritan vision between the groups. Stimuli were placed along tritan axes. These are represented by straight lines drawn through the colour spaces used. At constant brightness levels, a set of stimuli lying along any one of these lines should
appear identical to a tritanope. Changes in colour vision caused by absorption of short-wavelength light should be particularly salient using such stimuli.

4.2. Experiments 8a-e

Participants
There were thirty-six participants, divided into two groups. One group (under 30 group) consisted of twenty participants under thirty years of age (ten males, ten females: mean age = 21 years 9 months, SD = 3 years 9 months). The second group (over 50 group) included sixteen participants over fifty years of age (six males, ten females: mean age = 58 years 6 months, SD = 4 years 10 months). All participants had normal colour vision as assessed by the City University Colour Vision Test (Fletcher, 1980).

Apparatus
The experiments were run using a personal computer and a calibrated AOC sixteen-inch monitor\textsuperscript{22}. Colour readings were made using a Cambridge research systems colourCAL colourimeter.

4.2.1. Experiment 8a: Temporal Achromatic Measurement Luminance Adjustment Task (AMLA)

4.2.1.1. Introduction
Despite being well established that lens brunsecence increases with age, all the participants in the experiments presented here passed the City University Colour Vision Test (Fletcher, 1980). Experiment 8a therefore investigated whether measurable differences in short-wavelength light absorption existed between the two age groups.

The method used was an adaptation of flicker photometry\textsuperscript{23} devised by Lillo and Moreira (2005). The task presents, on a display, a square alternating rapidly between chromatic and achromatic. The square appears to flicker rapidly. The achromatic

\textsuperscript{22} CIE (1931) \(x, y, Y\) values of the white point was, \(x = 0.310, y = 0.316, Y = 100.0\).
\textsuperscript{23} Kaiser and Boynton (1996) Chapter 9 contains describes the theoretical basis of flicker photometry.
phase can be adjusted by observers between black and white. By doing this the amount of perceived flicker can be minimised. At fast flicker rates, only the achromatic mechanism is responsive and flicker is perceived to be at a minimum when the stimuli appear equally bright. In a situation where the chromatic phase of the flicker is dominated by short-wavelength light, an observer with a brunescent lens will tend to perceive minimum flicker when the achromatic phase is less luminant than is the case for an observer with a less brunescent lens. The reason lies in the increased absorption of short-wavelength light, which makes the chromatic phase appear dimmer and thus require a less luminant achromatic accompaniment to minimise flicker.

S-cones are not thought to contribute to brightness perception, but both L- and M-cones are sensitive to short-wavelength light (see Kaiser and Boynton, Chapter 5). Consequently, although the task used in this experiment does not involve S-cones, if results indicate that absorption of short wavelength light by L- and M-cones is affected in the over 50 group, it seems reasonable to assume both that the cause is lens brunescence and that the quantal catch of S-cones is also reduced.

4.2.1.2. Method

Stimuli and Design

Stimuli were coloured squares (155 x 155 mm) displayed individually at the centre of the display against a grey background. The central portion of each square was covered with a fixation disc (55 mm diameter) to ensure that performance of the task was not affected by the macula, located at the centre of the retina and approximately 1.5 mm in diameter. There were thirteen stimuli in total, differing in colour. The range of colours represented by the stimuli was a sample of the most saturated stimuli obtainable on the monitor. Figure 4.1. shows the location of the thirteen stimuli on a CIE (1976) L* u* v* diagram. As can be seen, the stimuli lie along the monitor gamut line defining the most saturated stimuli that could be produced using the equipment available. Stimuli were broadband, but dominant wavelength of stimuli 1-13 increases from approximately 460 nm (stimulus 1) to 600 nm (stimulus 13). Stimuli CIE (1976) L* u* v* values along with hue angle (huv) are given in Table 4.1. In each trial, the fill colour of the 155 mm x 155 mm square alternated between one of the thirteen chromaticities defined above and achromatic, at a rate of 30 Hz.
achromatic phase could be adjusted by the participant between black and white in forty-one steps corresponding to equal changes in luminance, calculated using a gamma function. Stimuli were viewed singly in a random order, against a uniform grey background. The design of a single trial is shown in Figure 4.2.

Figure 4.1. Stimuli used in Experiment 8a plotted on CIE (1976) u* v* diagram.
Table 4.1. CIE (1976) $u^* v^*$ values of stimuli used in Experiment 8a, together with hue angle and $L^*$ value.

<table>
<thead>
<tr>
<th>stimulus</th>
<th>$u^*$</th>
<th>$v^*$</th>
<th>$L^*$</th>
<th>Hue angle° (huv)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-15.2</td>
<td>-137.0</td>
<td>37.3</td>
<td>263</td>
</tr>
<tr>
<td>2</td>
<td>-44.7</td>
<td>-92.8</td>
<td>63.0</td>
<td>244</td>
</tr>
<tr>
<td>3</td>
<td>-66.4</td>
<td>-46.6</td>
<td>79.9</td>
<td>215</td>
</tr>
<tr>
<td>4</td>
<td>-74.1</td>
<td>-12.8</td>
<td>90.0</td>
<td>190</td>
</tr>
<tr>
<td>5</td>
<td>-77.0</td>
<td>13.1</td>
<td>88.7</td>
<td>170</td>
</tr>
<tr>
<td>6</td>
<td>-86.2</td>
<td>62.5</td>
<td>87.0</td>
<td>144</td>
</tr>
<tr>
<td>7</td>
<td>-89.4</td>
<td>111.7</td>
<td>85.6</td>
<td>129</td>
</tr>
<tr>
<td>8</td>
<td>-45.2</td>
<td>112.7</td>
<td>89.8</td>
<td>112</td>
</tr>
<tr>
<td>9</td>
<td>-10.5</td>
<td>110.7</td>
<td>93.8</td>
<td>95</td>
</tr>
<tr>
<td>10</td>
<td>17.1</td>
<td>112.0</td>
<td>96.9</td>
<td>81</td>
</tr>
<tr>
<td>11</td>
<td>42.9</td>
<td>99.6</td>
<td>88.1</td>
<td>66</td>
</tr>
<tr>
<td>12</td>
<td>99.6</td>
<td>78.1</td>
<td>74.5</td>
<td>38</td>
</tr>
<tr>
<td>13</td>
<td>187.6</td>
<td>46.5</td>
<td>57.8</td>
<td>13</td>
</tr>
</tbody>
</table>
Figure 4.2. Trial design in Experiment 8a. The chromatic stimulus alternates with an achromatic background at 30 Hz. Participants adjusted achromatic luminance to minimise flicker.

Procedure
The display was viewed in a dark room from a distance of 50 cm. Participants were instructed to fixate on the central disk, and then to adjust the achromatic stimulus luminance until flicker was perceived to be at a minimum. Adjustments were made using keyboard arrow keys. When each adjustment was complete, the space bar was used to initiate the next trial. Two practice trials, using achromatic stimuli were used.
Following completion of the practice trials, experimental trials were attempted. There were two blocks of trials. In each block the thirteen stimuli were presented in a random order. The achromatic adjustment settings selected to minimize perceived flicker in each trial were recorded.

4.2.1.3. Results

Examination of practice trial settings confirmed that participants in both age groups were able to make achromatic matches accurately. To compare the relative luminance settings made by each group, data was transformed as follows. Individual’s settings for the two trials involving each stimulus were first averaged. These values were then expressed as a proportion of the average setting, for that stimulus, made by participants in the under 30 group. This made the new mean score for the under 30 group equal to zero for each of the thirteen stimuli. Average scores in the over 50 group then reflected, for each stimulus, the proportional difference in luminance settings between the two groups. Figure 4.3. shows the data plotted in this way. As can be seen, for stimuli with shorter dominant wavelengths, the average luminance setting made by participants in the over 50 group were lower than in the under 30 group. As dominant wavelength increases, the opposite pattern emerges – participants in the over 50 group required relatively more luminance to make matches.

Data transformed as described in the previous paragraph was examined using an analysis of variance comparing the effects of measurement location (stimuli 1-13) and age group (over 50, under 30), with measurement location a repeated measures factor. This showed there were significant differences in the settings made by the two groups, $F(12, 408) = 8.49, \text{MSE } = 0.003, p < 0.001$. Planned comparisons confirmed that relative to participants in the under 30 group, settings made by participants in the over 50 group were less luminant for the three stimuli with the shortest dominant wavelengths, minimum $F = 4.339$, maximum $p < 0.05$. For all but one of the stimuli with the six longest dominant wavelengths, participants in the over 50 group made significantly more luminant matches than those in the under 30 group, minimum $F = 6.442$, maximum $p < 0.025$. The exception was stimulus 12 ($\text{huv } = 66$) where there was no significant difference between performance of the two age groups, $F = 2.624$, $p = 0.114$. 
Figure 4.3. Results of temporal AMLA task used in Experiment 8a showing relative differences in performance between the two age groups, with mean settings for participants in the under 30 age group rescaled to zero. The x-axis scale gives hue angle and indication of change in dominant wavelength. Error bars indicate ±1 SE. Bold error bar at top left of figure shows ±1 within-subjects 95% confidence interval (Loftus and Masson, 1994, 2003).

4.2.1.4.-Discussion

The results indicate that for the three stimuli with the shortest dominant wavelength (or the three largest hue angles) participants in the over 50 group required significantly less luminance to minimise perceived flicker relative to those in the under 30 group. It is likely that this indicates increased absorption of short-wavelength light by the lens. Moreover, the relatively more luminant achromatic matches made by participants in the over 50 group to stimuli of the longest dominant wavelengths, (or smallest hue angles) is also consistent with this in interpretation. This is because the brightness of the achromatic phase is influenced by the presence...
of a short-wavelength component in the achromatic light. If this short-wavelength component is absorbed to some extent by the lens, its relative contribution to brightness is reduced. The chromatic stimuli of longer dominant wavelengths contain relatively less, if any, short wavelength light to be absorbed by the brunescent lens. Thus a relatively brighter achromatic adjustment is required to match the brightness of a stimulus of longer dominant wavelength.

Experiment 8a therefore appears to confirm that in the over 50 group there is a significant increase in the relative absorption of short-wavelength light. Such differences between the two groups exist despite no differences in performance of the City University Colour Vision Test (Fletcher, 1980).

4.2.2. Experiment 8b: Temporal 2-AFC Detection Task

4.2.2.1. Introduction
In Experiment 8b JNDs were measured to examine whether the differences in short-wavelength light absorption noted in the previous experiment correspond to differences in JNDs. Rivers (1901b) suggests that a population he studied in Murray Island (close to the equator) showed, "... a relative insensitivity to blue... as compared to Europeans."^24

Absorption of short-wavelength light should primarily affect perception of blue stimuli. Predicting the consequences of short wavelength light absorption on JNDs, which were measured in Experiment 8b is made difficult by a number of factors. For example, it may be that, if luminance is kept constant changes in the appearance of blue stimuli will be relatively less noticeable through a brunescent lens. Given this the difference in JNDs between the two groups might be expected to increase as the point at which JNDs are measured becomes bluer (shorter dominant wavelength). By this account (and given that changes in chromatic appearance in blue and green regions of colour space consist primarily of differences in short-wavelength light) JNDs for the over 50 group should become relatively larger (in relation to JNDs for the under 30

---

^24 Rivers (1901b), pp 51.
group) as the JND measurement location shifts from green (longer dominant wavelength) to blue (shorter dominant wavelength).

On the other hand, Weber's Law \((\Delta I/I = \text{constant})\) suggests that the reduction of the overall level of S-cone activation (caused by lens absorption resulting in a lower photon capture by the S-cones), should result in smaller changes in chromaticity being perceptible. By this account, JNDs should be relatively lower for the over 50 group as stimuli become bluer (assuming there is no influence of noise). However a point must be reached when Weber's law breaks down because of the effects of noise. Schefrin, Shinomori and Werner (1995) measured JNDs along tritan axes, finding that age related increases in JND were accentuated at low levels of cone stimulation but that there were no apparent age related differences in Weber fractions. They concluded that (besides increasing absorption of short-wavelength light and lower overall levels of retinal illumination) age related increases in JNDs are caused either by increased neural noise, or by S-cones becoming inefficient at capturing photons. Although exposure to UV-B and sunlight is known to selectively damage S-cones (e.g. Harwerth & Sperling, 1971), it is not clear whether exposure to UV-B could selectively damage S-cones in such a way as to reduce efficiency of photon capture. The effect though should be the same as that of increased lens brunescence. If age related differences in JNDs are accentuated at lower levels of S-cone stimulation, as Schefrin et al.'s (1995) study implies, differences between the two groups JNDs, measured along a tritan axis, may be relatively higher in green than in blue regions of colour space (with JNDs larger in the over 50 group) because levels of cone activation are higher in blue relative to green regions of colour space. A subsequent study by Shinomori, Schefrin and Werner (2001) which measured wavelength JNDs reached similar conclusions and also found that JNDs between 450-510 nm are mediated by S-cones.

However these findings should be used cautiously in making predictions about the performance of participants in Experiment 8c. Both Schefrin et al. (1995) and Shinomori et al. (2001) used stimuli standardised in Trolands (a measure of retinal illuminance) which nullifies the effects of lens brunescence. Stimuli in Experiment 8b were of constant \(L^*\) (a measure of brightness) value and hence retinal illuminance was not controlled across participants. Retinal illuminance should though be generally
lower in the over 50 group because in addition to brunescence, a range of factors cause an overall lowering of the amount of light reaching the retina (see Schefrin, Werner, Plach, Utlaut & Switkes (1993) for a brief outline of these factors.) Since no measures were taken enabling retinal illuminance levels to be calculated, it is not possible to know whether the levels of S-cone stimulation for the stimuli used in Experiment 8b are low enough for the age related differences in performance proposed by Schefrin et al. (1993) to take effect.

A further factor that needs consideration is the possible existence of mechanisms that compensate for the increased absorption of short-wavelength light to preserve the brightness of different coloured objects. Kraft and Werner (1994) suggest, on the basis of heterochromatic flicker photometry and heterochromatic brightness matching studies, that such a mechanism exists. They suggest it is able to compensate adequately to increases in lens absorption at relatively longer wavelengths, but cannot compensate adequately at shorter wavelengths (420-480 nm) as age and lens density increase. This implies that relative differences in JNDs between older and younger observers ought to increase when measured in regions of colour space where the dominant wavelength of stimuli is less than 480 nm.

Finally, previous studies investigating the effects of aging on JNDs have not considered CP. If it is proposed that lack of separate green and blue terms in many languages is related to lens brunescence, then it might be expected that effects consistent with CP (in this case lower JNDs at the blue-green category boundary) would diminish as brunescence increases.

To summarise, chromatic JNDs in the blue-green region of colour space appear to be mediated by S-cone activity. Increasing levels of S-cone activation lead to relatively higher JNDs. However at relatively low levels of retinal illumination, JNDs for older participants are relatively higher than for younger participants, an increase that may be caused by an increase in noise. Since the stimuli in Experiment 8b, here, are equiluminant, it is difficult to predict how the factors thought to influence S-cone JNDs (brunescence, noise, level of cone activation) will affect the pattern of JND found in the two age groups tested. It is also the case that the studies mentioned in the previous three paragraphs all used single wavelength stimuli. The stimuli used in
Experiment 8b were broadband and the results should therefore have an ecological validity lacking in the studies mentioned above.

4.2.2.2. Method

Stimuli and Design

The same temporal 2-AFC detection task used to measure hue JNDs in Experiment 2 (section 2.3.2.) was used to measure chromatic JNDs in this experiment. The following alterations were made. Stimuli were equiluminant ($L^* = 61.7$) colours specified in CIE (1976) $L^* u^* v^*$ colour space. JNDs were measured in $\Delta E$, rather than Munsell hue, units. Stimuli were displayed using smaller patches than those used in Experiment 2 and viewing distance was increased\(^2\). Viewing angles were $2^\circ \times 4^\circ$ for the patches. Colour reproduction was carried out using the phosphor values of red, green and blue guns and a look-up table detailing luminance values for each gun.

Procedure

The procedure was the same as that used in Experiment 2, though with the following alterations. JNDs were measured at three locations, shown in Figure 4.4. These locations had equal intervals between adjacent pairs ($\Delta E = 35.2$ units) along a tritan confusion line defined by the fundamentals identified by Smith & Polkomy (1972)\(^3\). The specified line was chosen to maximise the luminance and saturation of stimuli displayed in the experiment, within the limits of the display gamut. JNDs at each of the three measurement locations were measured along this tritan confusion line. Table 4.2. gives the $L^* u^* v^*$ values for the measurement locations and colour appearance.

JNDs were measured in three blocks of ninety-six trials, as in Experiment 2, which were taken in a random order. After each trial, a space bar response was required to initiate the following trial. After incorrect trials, feedback was given in the form a message, ‘incorrect’ briefly flashed on the screen. The maximum separation range between stimuli displayed in any trial was twenty $\Delta E$ units. Participants failing to correctly perform the task on any trial at this separation were required to attempt the block of trials again.

---

\(^2\) Rectangular single stimulus patches were 105 x 52.5 mm, abutting pairs were 52.5 x 52.5 mm. Viewing distance was 150 cm.

\(^3\) These fundamentals define imaginary lines in colour space. For observers lacking short-wavelength cones (tritanopia), stimuli lying along these lines are indistinguishable.
Figure 4.4. JND measurement locations used in Experiment 8b. JNDS were measured along a tritan axis around each of the three locations.

<table>
<thead>
<tr>
<th>stimulus</th>
<th>$u^*$</th>
<th>$v^*$</th>
<th>Hue angle* ($h_{uv}$)</th>
<th>Appearance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-48.0</td>
<td>25.57</td>
<td>153.0</td>
<td>green</td>
</tr>
<tr>
<td>2</td>
<td>-39.91</td>
<td>-8.71</td>
<td>192.3</td>
<td>blue-green</td>
</tr>
<tr>
<td>3</td>
<td>-31.82</td>
<td>-43.0</td>
<td>233.5</td>
<td>blue</td>
</tr>
</tbody>
</table>

4.2.2.3. Results

One participant in the over 50 group was unable to provide a JND at one of the measurement locations (stimulus 1, green)\(^{27}\), despite repeated attempts to do so. This

\(^{27}\) Because the participant made repeated attempts to provide a JND and had normal colour vision as assessed by the City University Colour Vision Test (Fletcher, 1980), it is likely that their JND for this point exceeded that of the maximum separation range. Consequently, for this point, the participant was assigned a JND value corresponding to the maximum (twenty $\Delta E$ units) although it was likely to be an underestimate of actual JND.
participant was assigned the maximum possible JND value for that location (20 ΔE units). JND estimates were calculated using median values, as in Experiments 1 and 2. Table 4.3. gives mean JNDs (in ΔE units) across participants in each group for the three measurement locations. Figure 4.5. shows the same information graphically. JNDs for participants in the over 50 group appear to be larger overall than in the under 30 group. JNDs for both groups appear lower at the blue-green boundary than green or blue measurement locations. The pattern of performance appears to differ somewhat between the groups. In the over 50 group the relative increase in JNDs at the green location appears larger than for the under 30 group.

<table>
<thead>
<tr>
<th>Measurement location</th>
<th>Table 4.3. Mean ΔE JNDs for over 50 and under 30 groups in Experiment 8b. Figures in brackets are standard errors.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3</td>
<td>Over 50 7.76 5.13 8.14 (0.96) (0.43) (0.62)</td>
</tr>
<tr>
<td>green blue- blue green</td>
<td>Under 4.78 4.29 6.97 (0.25) (0.29) (0.53)</td>
</tr>
</tbody>
</table>
JNDs for participants in the under 30 group were lower overall than those of participants in the over 50 group, $F(1, 34) = 7.43, \text{MSE} = 3.30, p < 0.025$. There was a significant main effect of measurement location, (Means, (S.D.): Green $= 6.11$ (3.04); Blue-green $= 4.67$ (1.56); Blue $= 7.49$ (2.45)), $F(2, 68) = 26.71, \text{MSE} = 2.70, p < 0.001$. Planned comparisons showed that JNDs at the blue-green measurement location were smaller than at either green or blue locations, minimum $F(1, 34) = 19.86$, maximum $p < 0.001$. The interaction between group and measurement location was also significant, $F(2, 68) = 4.34, p < 0.025$. Planned comparisons confirmed that the significant interaction was due to JNDs at the green measurement location were relatively higher than those obtained at the blue-green location for the over 50 group compared to the under 30 group, $F(1, 34) = 9.22, p < 0.01$. The difference between JNDs at the blue relative to the blue-green measurement locations was not significantly different between the two groups $F(1, 34) = 0.21, p = 0.65$. 

Figure 4.5. Mean JNDs ($\pm 1 \text{ SE}$) for over 50 and under 30 groups in Experiment 8b. Bold error bar to right gives $\pm 1$ within-subjects 95% confidence interval (Loftus & Masson, 1994, 2003).
The analysis was repeated with the data from the participant to whom an artificial JND value was assigned removed. The pattern of significance in the results remained the same.

4.2.2.4. Discussion

Although the previous experiment found measurable differences in absorption of short-wavelength light between the two groups, no corresponding difference in JNDs were found in Experiment 8b. The increase in JNDs at the blue relative to the blue-green measurement locations were no different in the two age groups, despite the fact that more short wavelength light appears to be absorbed by the lens of older participants. The dominant wavelength of stimuli at the blue measurement location ($u^* = -31.82, v^* = -43.0$) was approximately 480 nm. The mechanism proposed by Kraft and Werner (1994) to compensate for increased absorption of short-wavelength light is suggested to operate efficiently when wavelength exceeds 480 nm. This may explain why the relative increase in JNDs between the blue-green and blue measurement locations were similar between the two groups.

A second notable factor was that JNDs in the over 50 group were relatively higher at the green measurement location than for the under 30 group. In the previous experiment, relative adjusted luminance settings for stimuli of similar hue angle (and dominant wavelength) to the green measurement location were similar for the two groups, as can be seen by comparison of relative adjusted luminance settings in Figure 4.3. with hue angles for JND stimuli in Table 4.2. This suggests the difference in performance between the two groups is not related to relative differences in absorption of light by the lens. It is possible that the relative difference in JNDs relates to the lower level of S-cone activation for stimuli at the green measurement location relative to the other two measurement locations. As noted in the introduction to this section, Schefrin et al. (1995) found that at low levels of activation, S-cone JNDs were elevated in older observers. Whilst levels of S-cone activation may explain the differences noted here, there is no way of confirming this possibility.

None of the measures made in the experiments reported here allow retinal illuminance to be established accurately. Hence it is not possible to establish whether levels of S-
cone stimulation in the over 50 group are sufficiently low to account for the differences in JNDs in the manner Schefrin et al. (1995) propose.

A third important finding is that the pattern of performance is similar to the patterns of performance found in JND experiments carried out in Chapters 1 and 2 (Experiments 1, 2, 6 & 7). In those experiments, as in this one, JNDs at the blue-green boundary were found to be lower than at within-category green and blue measurement locations, providing evidence for colour CP. In this experiment, a third colour metric was used, CIE (1976) u* v* space, and the general pattern of findings was replicated, as evidenced by lower JNDs at the category boundary. If increased absorption of short-wavelength light by the lens is connected to the lack of separate blue and green colour terms in many languages, then it might be expected that CP effects would be reduced in aged populations, regardless of colour language. However, examination of the JNDs of the over 50 group in this experiment, in Figure 4.5, shows that the CP effect appears to be robust in the face of changes in colour vision brought about by the aging process.

4.2.3. Experiment 8c: Diadic Judgements Task

4.2.3.1. Introduction
Hardy et al. (2005) found no difference in the colour naming of stimuli in the blue to green regions of colour space amongst older and younger populations of English speakers. However similar naming patterns cannot be taken to indicate that the relative appearance of stimuli (to observers in different age groups) is similar. By the same token, although Experiments 8a and 8b suggest that group differences in the absorption of short-wavelength light do not appear to correspond to differences in JNDs, it is not clear that performance at threshold mirrors performance of a supra-threshold task.

Because absorption of short-wavelength light produces effects on colour vision very similar to tritanomoly (Judd, Plaza & Farnsworth, 1950), similarity judgements might be affected. Specifically, stimuli that lie along the same tritan axis may be judged as relatively more similar by older observers than by younger observers. To investigate
this issue, Experiment 8c used a diadic judgement task in which participants were required to rate how similar/dissimilar pairs of stimuli appeared. The stimuli used in the experiment varied in brightness as well as in chromaticity. Consequently ratings of similarity between stimulus pairs differing in brightness could be compared to ratings between stimulus pairs differing in chromaticity along a given tritan axis. Good evidence of tritan axis compression would be provided if relative to the under 30 group, participants in the over 50 group place more emphasis on differences in brightness than on differences in tritan axis chromaticity.

In the INDividual Differences SCALing (INDSCAL) analysis used in Experiment 8d, each participant's data was used to construct a two-dimensional plot to represent the level of similarity between stimuli. In this plot, stimuli judged similar are located close together, stimuli judged less similar are further apart. Two measures of fit, stress and average squared correlations (RSQ) can then be used to assess how well the derived plot fits the data fits, with higher values indicating a better fit.

INDSCAL also produces an index of subject weights that shows how much relative importance each participant assigns to each of the (in this case two) dimensions. These represent the end point of a vector coming from the origin of the weight space. In the two-dimensional weight space used here, vector angles of less than 45° indicate more weight being assigned to one dimension, more than 45° that more weight is assigned to the other dimension. An angle of 0° or 90° suggests the participants' judgements are based on one dimension at the exclusion of the other. Figure 4.6 gives a hypothetical example of such a plot derived from two participants' subject weights for a two dimensional solution. In making judgements, participant 2 places more emphasis on dimension 2, whilst participant 1 places more emphasis on dimension 1. The relative importance of dimensions to groups of participants can be compared by comparing the angles of each participant's vector (A more detailed description of INDSCAL can be found in Norušis (1994), Chapter 7.)

Stress values indicate the extent of the solution departure from the data (Kruskal & Wish, 1978, give a detailed interpretation of stress levels).

RSQ represent the proportion of variance of the scaled data accounted for by their corresponding distances.
Figure 4.6. Hypothetical example of subject weight plot for two-dimensional INDSCAL solution. The bold dashed line is at 45° to the axes. The participant above this line places more slightly emphasis in making judgments on dimension 2 than on dimension 1. The participant below the line places more emphasis on dimension 1.

4.2.3.2. Method

Stimuli and Design

Stimuli consisted of coloured squares (51 mm x 51 mm) displayed in pairs separated by 51 mm around the centre of the monitor screen against a neutral grey background (see Figure 4.8.). There were nine stimuli in total at three at three brightness ($L^*$) levels; 61.7, 41.3, 20.44, with three stimuli at each level. Location of stimuli in CIE (1976) $L^* u^* v^*$ space is shown in Figure 4.7, together with nominal hue. Appendix 4 shows a similar diagram with numerical values of stimuli marked. Stimuli at $L^* = 61.7$ were the same as the measurement locations around which JNDs were measured in Experiment 8b. Stimuli lay along three parallel tritan axes, one for each $L^*$ level.

$L^* \text{ value} = 61.7, x = 0.310, y = 0.316.$
Stimuli on each of the tritan axis were equally spaced, with the central stimulus on each axis being located at a Munsell hue of approximately 7.5BG, the nominal blue-green boundary. Table 4.4 gives the ΔE distances between adjacent stimulus pairs. As is shown, ΔE differences between stimuli of the same nominal hue at different value levels were similar (between 24 and 28 ΔE units.) Differences between pairs of stimuli at the same L* level were also similar. Distances between pairs of stimuli differing in chromaticity differed considerably depending on L* level. Separations between adjacent stimuli at the lowest L* level were less than 12 units. At the highest L* level, equivalent separations were approximately 35 ΔE units. These changes in ΔE distance between stimuli at different L* levels could not be avoided whilst keeping hue angle constant.

Figure 4.7. Location of stimuli used in Experiment 8c in CIE (1976) u*v* space at three different L* levels. Nominal hue of stimuli is also shown (G = green, BG = blue-green, B = blue).
Table 4.4. ΔE distances (L*u*v*) between adjacent stimulus pairs used in Experiment 8c.

<table>
<thead>
<tr>
<th>L*</th>
<th>G</th>
<th>BG</th>
<th>degradation</th>
</tr>
</thead>
<tbody>
<tr>
<td>61.7</td>
<td>35</td>
<td>35</td>
<td>↑</td>
</tr>
<tr>
<td>27.2</td>
<td>24.5</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>41.4</td>
<td>23.6</td>
<td>23.6</td>
<td>↑</td>
</tr>
<tr>
<td>27.8</td>
<td>25</td>
<td>27.6</td>
<td></td>
</tr>
<tr>
<td>20.44</td>
<td>11.7</td>
<td>11.7</td>
<td>↓</td>
</tr>
</tbody>
</table>

Procedure
All possible pairs of different stimuli (thirty-six), were displayed. Location of stimuli, in either right or left hand square was randomized. The test display remained until a response was made. The task was to judge the similarity between pairs using a seven-point Likert scale. There were two blocks of trials, taken consecutively in a random order. Each block consisted of the thirty-six combinations, making seventy-two trials in total. In one block, a value of 1 was used to indicate maximum similarity and 7 to indicate maximum dissimilarity. In the other block the scale was reversed. An appropriately labelled scale appeared on the computer screen during each block of trials. Figure 4.8. shows the display used for during experimental trials.
Figure 4.8. Example of display used to obtain similarity judgement data in Experiment 8c. Participants took two blocks of trials, with scales reversed in one of the blocks.

Stimuli were viewed in a dark room from a distance of 50 cm. The task was explained and responses were made using a computer keyboard. A mouse response was used to initiate successive trials. Following each trial there was a two-second interval, during which the display was filled with a grey background.

4.2.3.3. Results
Similarity rating data from blocks of trials where ratings of 1 represented maximum dissimilarity and 7 maximum similarity were rescaled so that for all trials a rating of 7 represented maximum dissimilarity. Each participant's responses from the two trials involving the same stimulus pair were than averaged and formed into individual distance matrices. These matrices were then analysed using INDSCAL. This computed a two-dimensional group space solution from the thirty-six input matrices (one for each participant) representing a common structure. The solution fitted the
data adequately: average stress values across individual matrices = 0.178, average squared correlations\(^3\) (RSQ) = 0.758. Figure 4.9. shows the derived solution. As can be seen in the figure, most stimuli fall in a circular order around two dimensions. Stimulus, BG 41.3, intermediate in hue and L* value, at the plot centre, whilst it is also notable that at L* = 20.44, the blue-green stimulus is judged as being more similar to the green than to the blue stimulus of equivalent brightness. The location of the stimuli on the plot suggests that dimension 1 corresponds to L* value (brightness), with darker stimuli at the right of the plot. Dimension 2 appears to correspond to hue.

Figure 4.9. Stimulus space derived from INDSCAL analysis of participant similarity matrices in Experiment 8c. Letters in labels indicate nominal colour of stimuli, (B = blue, BG = blue-green, G = green), numbers in labels indicate stimuli L* value.

Figure 4.10. is a plot showing individual participant weights for each of the two dimensions derived from the analysis. Mean weirdness\(^2\) ratings across participants in the two groups were similar (mean weirdness: over 50 = 0.14, under 30 = 0.18).

\(^3\) RSQ represent the proportion of variance of the scaled data accounted for by their corresponding distances.

\(^2\) Weirdness indicates how far a participant’s (or group’s) weight vector is from the average weight vector, with larger values indicating larger deviation. (See Norušis (1994), Chapter 7 for detailed description)
Inspection of Figure 4.10 suggests that overall both groups place more emphasis in making judgements on dimension 1 (corresponding to brightness) relative to dimension 2 (corresponding to chromaticity). However, it also appears that participants in the under 30 group give relatively more weight to dimension 2 in making judgements than those in the over 50 group. Whilst the vector angle of 50% of those in the under 30 group are less than 50° (relative to the vertical axis) this is the case for only two (12.5%) of over 50 participants. Empirical support for this impression was provided by comparison of the angle (in degrees) of each participant’s vector using an independent samples t-test. This found a significant difference between the two age groups, (means (SD): over 50 = 51.49 (8.87), under 30 = 56.31 (7.48)), $t(34) = 1.74, p < 0.05$ (one-tailed).^{33}

![Figure 4.10. Plot of participant weights from INDSCAL analysis of similarity matrices: Experiment 8c. Bold dashed line indicates 45° vector angle. Other dashed lines indicate 40° and 50° vector angles.](image)

^{33} The under 30 group also appears to consist of two subgroups when viewed in Figure 4.10. However, it was not clear on what this sub-division could be based.
4.2.3.4. Discussion

Although the INDSCAL solution derived from all thirty-six participant matrices fits the data relatively well, the results nevertheless indicate that differences in the relative appearance of stimuli exist between the two groups. As comparison of vector angles shows, those in the over 50 group tended to weight their judgments significantly more on the dimension corresponding to brightness than on that corresponding to chromaticity. In other words, participants in the over 50 group appear to place less emphasis on chromatic differences (relative to brightness differences) between stimuli aligned along tritan axes, than participants in the younger group.

Given first that the consequences of increasing lens brunescence are similar to tritanomoly and second that the results of Experiment 8a appear to indicate increased brunescence in the over 50 group, these results are not surprising. What they imply is that changes in colour appearance mediated by the S-cone pathway are relatively less salient in the over 50 group. The region of colour space in which S-cones mediate discrimination is for stimuli with dominant wavelength between 450-610 nm (see Shinomori et al., 2001), that is the blue-green region of colour space.

It seems possible that differences in the perceived similarity of stimuli found in this experiment may be more pronounced in other populations. Using the same colour vision test as was used in this study, Davies et al. (1998) found a high incidence of tritan errors. None were made by participants in Experiments 8a-e. If the tritan errors made by participants in Davies et al.'s (1998) study errors indicate increased lens brunescence, it follows that such groups should judge blue-green stimuli to be more similar than they were judged by the over 50 group examined here. The results therefore appear to support the idea that lens brunescence, by compressing the relative difference in appearance between blue and green stimuli, may result in less 'perceptual pressure' for separate linguistic 'blue' and 'green' colour categories.
4.2.4. Experiment 8d: Colour Naming Task

4.2.4.1. Introduction

Experiment 8d was a naming experiment in which participants were required to name a similar, but slightly extended range of stimuli to those used in Experiment 8c. Lindsey and Brown (2002) found that simulated brunescence resulted in stimuli being named blue less frequently. However, Hardy et al. (2005) demonstrate that aged observers in fact name stimuli in a similar manner to younger observers. One shortcoming of the two studies mentioned above is that stimuli were all at the same luminance level (Lindsey and Brown (2002) do not report the luminance of the stimuli used, only Munsell Value (equivalent to L* = 61.7). The stimuli used by Hardy et al.'s (2005) were of the same Munsell value as those used by Lindsey and Brown (2002) and had a luminance of 20 cd/m^2.) Consequently, neither study addresses the issue of whether changes in luminance or brightness influence naming. Such changes might be expected to differ between groups with differing levels of lens brunescence, particularly at lower luminance levels, because of the effect of lens brunescence on the activity of short-wavelength cones. An additional difficulty with the two studies under consideration is that named stimuli were all of high saturation and thus may not be representative of colour samples typically encountered in the environment.

Although the range of stimuli used in Experiment 8d is far more restricted than those used in Lindsey and Brown (2002) and Hardy et al.'s (2005) studies, the stimuli used in Experiment 8d were at three different luminance levels and of relatively lower saturation than those used by Lindsey and Brown (2002) or Hardy et al.'s (2005). Naming responses were not restricted to ‘blue’ and ‘green', but were constrained to the eleven basic colour terms, because at the lowest luminance levels, achromatic responses were possible. In addition, absorption of short wavelength light makes blue stimuli appear redder, so it appeared possible that participants in the over 50 group might name some blue stimuli purple.

^ Stimuli named in Lindsey and Brown (2002) and Hardy et al. (2005) were a subset of the Munsell chips used in the World Colour Survey.
4.2.4.2. Method

**Stimuli and Design**

Stimuli consisted of coloured squares (51mm x 51mm) presented singly against the same neutral grey background used in Experiment 8c. There were fourteen stimuli. Nine of the stimuli were identical to those used in Experiment 8c. The additional five stimuli were added to extend the three tritan axis, defined by L* level, along which stimuli lay. Stimuli were added to each end of two of the tritan axis (L* = 41.3 and 20.44), but only at one end (into the green region of colour space) of the third tritan axis (L* = 61.7). Figure 4.11. shows the location, in CIE (1976) u*v* space of the fourteen stimuli used in the experiment. Appendix 4 shows a similar diagram with numerical values of stimuli marked. Luminance values of the stimuli were, for each of the three L* levels, 30.05, 12.05 and 3.10 cd/m².

![Figure 4.11. Location of stimuli used in Experiment 8d in CIE (1976) u*v* space at three different L* levels.](image)

35 The gamut of the monitor used to display stimuli did not permit extending this axis further in both directions.
Procedure
Stimuli were presented in a darkened room at a viewing distance of 50 cm. Stimuli were viewed once each, in a random order, remaining on display until a naming response was made. The eleven basic colour terms appeared in a typed list at the top of the display. Responses were made verbally and recorded by the experimenter. Successive trials were initiated by a mouse pad response which initiated a two-second interval, during which the display was filled with the uniform grey background prior to display of the next stimulus.

4.2.4.3. Results
Stimuli were named either ‘blue’ or ‘green’ in over 98% of trials (over 50 group = 97.8%, under 30 group = 98.2%). Table 4.5. gives the mean proportion of blue responses made for stimuli at each L* level by the two age groups, as well as the overall mean proportional use of the term ‘blue’. Although the proportion of ‘blue’ responses to stimuli at the two highest L* levels is similar in both groups, it appears that stimuli at the lowest L* level are named ‘blue’ less frequently by participants in the over 50 age group.

<table>
<thead>
<tr>
<th>Stimulus L* value</th>
<th>Proportion ‘blue’ responses</th>
<th>Table 4.5. Mean proportional use of term blue to name stimuli in Experiment 8d, for over 50 and under 30 age groups, for stimuli at each of the three L* levels and overall. Figures in brackets are standard deviations.</th>
</tr>
</thead>
<tbody>
<tr>
<td>61.70</td>
<td>0.44 (0.11)</td>
<td>0.49 (0.13)</td>
</tr>
<tr>
<td>41.36</td>
<td>0.46 (0.10)</td>
<td>0.44 (0.08)</td>
</tr>
<tr>
<td>20.44</td>
<td>0.25 (0.09)</td>
<td>0.34 (0.09)</td>
</tr>
<tr>
<td>Overall</td>
<td>0.38 (0.09)</td>
<td>0.42 (0.07)</td>
</tr>
</tbody>
</table>

These impressions were confirmed by an analysis of variance examining the effects of brightness (L* = 61.7, 41.3, 20.44) and group (Over 50, Under 30), with brightness a repeated measures factor, on the use of the term ‘blue’ to name stimuli.
There were significant main effect of brightness level, (Means, (S.D.): \(L^* 61.7 = 0.46\) (0.08); \(L^* 41.3 = 0.45\) (0.09); \(L^* 20.44 = 0.30\) (0.10)), \(F(2, 68) = 35.33, MSE = 0.008, p < 0.001\), indicating that the overall proportion of 'blue' responses was different across brightness levels. The interaction between brightness level and group was also significant, \(F(2, 68) = 3.33, p < 0.05\). Planned comparisons indicated that relative to use of 'blue' to name stimuli at \(L^* = 41.3\), participants in the over 50 group used the term less frequently than those in the under 30 group to name stimuli at \(L^* = 20.44, F(1, 34) = 8.01, MSE = 0.11, p < 0.01\).

Because four stimuli were presented for naming at the highest \(L^*\) level and five were presented at each of the other two levels, analysis of data using of proportional scores may be misleading.\(^{36}\) To investigate this, the data was reanalysed by assuming that a fifth stimulus at the highest \(L^*\) level was present and was named 'blue' by all participants. This assumption seems reasonable for two reasons. First, the absent stimuli at the highest \(L^*\) level, which was out of gamut, would have been a saturated blue. Second, the bluest stimulus actually presented at the highest \(L^*\) level was named 'blue' by all participants. Reanalysed in this way, the pattern of significant findings in the results remained the same as that found in the previous analysis.

4.2.4.4. Discussion

Experiment 8d found differences between the two age groups in naming stimuli in the blue-green region of colour space. In particular, participants in the over 50 group use the term 'blue' to name stimuli at the lowest \(L^*\) level less frequently than participants in the under 30 group. These results differ from those of Hardy et al. (2005), who found no differences in naming between different age groups and it seems probable the reason for these differences lie in the varying luminance levels of the stimuli used in Experiment 8d (the stimuli named in Hardy et al's (2005) study were of equivalent luminance levels.) The results of the experiment therefore appear to provide at least limited support for Lindsey and Brown's (2002) suggestion that lens brunescence may alter perception and naming of green and blue stimuli. Indeed, one of the proposed effects of lens brunescence is to reduce the area of colour space.

\(^{36}\) For stimuli at the highest luminance level, proportional scores varied between 0 and 1 in increments of 0.25. At the other two levels the increments were 0.2.
named ‘blue’. For stimuli at the lowest L* level participants in the over 50 group do use ‘blue’ to name stimuli relatively less frequently than those in the under 30 group.

However the fact that differences in use of the term ‘blue’ between the groups examined here appear confined to stimuli of the lowest brightness suggests the differences may relate to the Bezold-Brücke effect rather than lens brunescence. The Bezold-Brücke effect refers to changes in the perception of hue as brightness increases (See Lillo, Aguado, Moreira and Davies (2004) for a detailed description). Stimuli with dominant wavelength below 500 nm appear bluer as brightness increases. Conversely, lowering the brightness of a blue stimulus makes it appear greener. In addition to the selective absorption of short-wavelength light of which much has been made here, the aging process causes a general reduction in the amount of light reaching the retina. A major factor is that, under given conditions pupil size tends to decrease as age increases (see Wyszecki & Stiles, 1982, Chapter 2). This decrease in the overall level of retinal illumination in older observers would appear to predict that, following the Bezold-Brücke effect and in the absence of any compensatory mechanism, blue stimuli should appear greener. This was the case at the lowest L* level in Experiment 8d. However, it is reasonable to ask why the same pattern is not evident at increased L* levels. Hardy et al. (2005) suggest the visual system must compensate for changes in the spectral composition of light, via the process of colour constancy. One possibility is that this compensation system is unable to cope when stimuli are sufficiently dark. A second is that the same effect would have been noted had the stimuli at the two highest L* levels been more closely spaced.

4.2.5. Experiment 8e: Colour Naming Task with Simulated Lens Brunescence

4.2.5.1. Introduction
To further examine the effects of short-wavelength light absorption on naming of blue and green stimuli an additional experiment was carried out in which young participants with simulated lens brunescence named the same set of stimuli used in the previous experiment. Simulation of lens brunescence was achieved by the use of a
physical filter incorporated into a set of goggles, a method similar to that used by Zlatkova, Coulter and Anderson (2006) to investigate the effects of brunescence on contrast sensitivity and blue on yellow acuity. Both Lindsey and Brown (2002) and Hardy et al. (2005) found that simulated lens brunescence in young observers produced differences in colour naming, most notably a marked reduction in use of the terms ‘blue’ to name stimuli. However naming performance of an older and younger group (without simulated brunescence) was similar in Hardy et al.’s (2005) study. As is pointed out in section 4.2.4.1., both the colour naming studies mentioned above used equiluminant (prior to filtering) stimuli. Consequently the effects of simulated brunescence on the naming of stimuli of varying luminance are not clear.

4.2.5.2. Method

Participants
There were fourteen participants, (seven males, seven females: mean age = 27 years 1 months, SD = 3 years 5 months). All participants had normal colour vision as assessed by the City University Colour Vision Test (Fletcher, 1980).

Stimuli Apparatus, Design and Procedure
The experiment was identical to Experiment 8d except that participants named stimuli viewed through a yellow filter built into a pair of goggles. The goggles were constructed from Lee filters (CP10Y and CP20Y) and had the effect of reducing the amount of short wavelength light reaching the eye. Figure 4.12. shows estimated optical density for a twenty-seven year old (the average participant age) wearing the goggles and for a fifty-nine year old (average age of participants in over 50 group in Experiment 8d). For comparison a twenty-seven year old’s estimated optical density whilst not wearing the goggles is also shown. As can be seen, the effect of wearing the goggles closely simulates the optical density estimated for over 50 participants in Experiment 8d.

---

37 Lens densities were calculated on the basis of data presented in Pokorny et al. (1987). Spectral transmission data for the filters was converted to optical density difference data using formula described in Wyszecki and Stiles (1982), section 2.4.6.
Figure 4.12. Estimated optical density for a twenty-seven year old in simulated lens brunescence colour naming experiment (8e) with and without filter. Also shown is estimated optical density of a fifty-nine year old. Optical density difference refers to difference of lens density at $\lambda = 700$ and $\lambda$ values on x-axis.

The next figure, Figure 4.13, shows how the location of the stimuli in $u^*v^*$ colour space alters when measured through the yellow lens. As can be seen in the figure, changes in chromaticity of stimuli occur primarily on the $v^*$ axis, whilst the changes are greater along the $v^*$ axis for stimuli that appear bluer. Chromaticity values of the achromatic point are also altered in the same way, such the $u^*v^*$ values of the grey background are altered to those of a hue that would normally appear greenish-yellow.
4.2.5.3. Results

Stimuli were named either 'blue' or 'green' in over 98% of trials. Figure 4.14 shows the relative proportion of 'blue' responses made to stimuli at each of the three different L* levels. Results from the two groups who took part in Experiment 8d are also included. As can be seen, participants in Experiment 8e show a pattern of naming...
that resembles that of participants in the over 50 group. However it is notable that stimuli at the highest $L^*$ level are named 'blue' less frequently than by either of the other two groups.

![Figure 4.14](image)

**Figure 4.14.** Proportional use of 'blue' term to name stimuli in Experiment 8e ($\pm 1$ S.E.), together with same data for participant groups in Experiment 8d ($\pm 1$ within-subjects 95% confidence interval (Loftus & Masson, 1994, 2003)).

Data was analysed with an analysis of variance examining the effects of brightness ($L^* = 61.7, 41.3, 20.44$) and group (Over 50, Under 30, Simulated brunescence), with brightness a repeated measures factor, on the use of the term 'blue' to name stimuli.

As in Experiment 8d, there were significant main effect of brightness level, (Means, (SD): $L^* 61.7 = 0.43 (0.12)$, $L^* 41.3 = 0.44 (0.08)$, $L^* 20.44 = 0.29 (0.10)$), $F (2, 79.54) = 36.58, MSE = 0.011, p < 0.001$, indicating that the overall proportion of 'blue' responses was different across brightness levels. The interaction between brightness level and group was also significant, $F (4, 79.54) = 3.09, p < 0.05$. Post hoc Dunnett's T3 tests indicated that whilst there were no significant differences between
the simulated brunescence and over 50 groups \( (p = 0.132) \), a significant difference in naming did exist between the simulated brunescence group and the under group \( (p < 0.005) \).

As in the previous experiment, data was reanalysed on the assumption that a fifth stimulus at the highest \( L^* \) level had been named ‘blue’. The pattern of significant findings remained the same as in the previous analysis.

4.2.5.4. Discussion

Experiment 8e provides further evidence that naming stimuli at different brightness and luminance levels produces differences in the pattern of naming. The most notable finding is that simulating lens brunescence produces a pattern of naming that differs significantly from that of participants in the under 30 group, in Experiment 8d, but does not differ significantly from that of the over 50 group. As with Experiment 8d, the findings would not be predicted on the basis of Hardy et al.’s (2005) study.

4.3. General Discussion of Experiments 8a-e

The results of Experiments 8a-e may be summarised as follows. Experiment 8a found significant differences between the two groups on a flicker photometry task indicative of the predicted differences in absorption of short-wavelength light. Differences in performance of the two groups on a JND task (Experiment 8b) were also found. However there was little evidence suggesting that the difference in short-wavelength light absorption between the two groups caused the differences in JNDs found. As in experiments reported earlier in this thesis, Experiment 8b provided evidence for CP, with JNDs lower around the blue-green category boundary than at either of the within-category blue or green locations tested. Nor did the size of the CP effect appear diminished in the over 50 group. In Experiment 8c, a dyadic judgement task, small but significant differences in performance of the two age groups were found. These suggest that differences between blue and green stimuli appear relatively more similar to older than to younger observers. Such an effect is consistent with increased short-wavelength light absorption at the lens. Experiment 8d, found differences in naming between the two groups, but only for stimuli at the lowest brightness level. A
final Experiment 8e, in which lens brunescence was simulated, found further evidence of differences in naming patterns, but in a direction not predicted by previous studies.

Overall, the results of Experiments 8a-e provide mixed support for the idea that the division of colour space into separate blue and green named regions by languages is influenced by lens brunescence, as suggested by Lindsey and Brown (2002). If it is assumed that signs of tritanomaly found in equatorial populations, by for example, Davies et al. (1997), reflect lens brunescence then the effect on this process on colour appearance may be greater for those populations than that found in the over 50 group tested here. In terms of colour appearance this may equate to differences between blue and green stimuli appearing less salient than to differences between stimuli in other regions of colour space, separated for example by differences based on brightness. However, it is important to add that such differences do not appear to be associated with a loss of CP across the blue-green category boundary. As stated above, no evidence of such an effect was found in Experiment 8b, nor in the experiments reported in Chapter 3 which compared performance of Himba and British adults (Experiments 5 and 7). Indeed the Himba adult groups and the over 50 group of British speakers tested show considerable similarities in their performance compared to that of young British adults: that is although measures (JNDs, response latencies) are elevated, this elevation does not appear to result in diminished CP. The elevation of these experimental measures is however not inconsistent with the compression of perceptual space along a blue-green axis.

It should however be noted that considerable caution is required in attempting to draw inferences about the vision of people living in equatorial regions on the basis of either an elderly sample of English speakers living in the northern hemisphere, (as is done here and by Hardy et al, 2005), or on the basis of the simulated lens aging (Lindsey & Brown 2002). Although such studies can provide useful insights, it is not altogether clear how the results can be applied to those living in equatorial regions. For example, although it appears likely that premature lens aging may occur amongst inhabitants of equatorial regions, it is also likely that other age related influences on vision (for

---

38 None of the participants who took part in Experiments 8a-e showed colour vision defects on the screening test used, whereas work such as Davies et al. (1997) suggests that tritanomalous vision defects found on screening test reflect lens brunescence.
example, senile miosis) are not related to UV exposure. Consequently the performance of an elderly English speaking group may differ from that of a younger group from around the equator, because performance is affected by factors other than the shared lens brunescence. Experiment 8b found differences between over 50 and under 30 groups that appear to reflect differences between the groups caused by factors other than levels of lens brunescence. Similarly, Hardy et al. (2005) criticise Lindsey and Brown’s (2002) study for a failure to take account of long-term adaptation effects. However, applying Hardy et al.’s. (2005) results to the issue of why some languages encode separate blue-green terms and others do not is also problematic. What their results suggest is that in certain populations and certain circumstances the blue-green distinction may be preserved at increased levels of lens brunescence. As Experiments 8d and 8e imply, differing brightness and/or luminance levels may in fact produce differential patterns of naming in the over 50 and under 30 age groups tested. In addition it is not known what the effect on naming would be for younger people with naturally accelerated lens aging, nor how the level of lens brunescence found in elderly populations in Britain and the USA compares to that of populations living in regions of high UV exposure.

What is required is further work with groups living in regions of high UV exposure. At present the only data comes from the studies using colour vision tests reported earlier. Ideally, future work would measure the range of lens brunescence found in populations exposed to high levels of UV and use similarity judgement tasks to assess the relative weight given to judgements along the blue-green axis in such populations. Additional work could assess what level of lens brunescence is required to abolish the CP effect around the green-blue boundary. It may be, for example, that in some societies significant minorities exist who, due to severe lens brunescence, do not perceive greens and blues hues as distinctly different. In such circumstances, sociological factors may play a role in whether separate green and blue terms become encoded in the language as Lindsey and Brown (2002) suggest.
Chapter 5

General Discussion

Different languages divide colour space in different ways using different numbers of colour terms. This thesis builds on a long tradition of research that has attempted, from a range of differing perspectives, both to explain this variation and draw wider inferences from it. Three questions were posed in the introduction. The first related to the influence of colour category membership on colour perception, the second to potential influences of colour language on colour perception. The third question reversed the premise of the second and asks whether differences in colour perception may influence differences in colour language. The extent to which the experiments presented in this thesis assist in resolving these questions are discussed below.

5.1. Summary of Research Questions and Findings

5.1.1. The influence of colour category membership on colour perception

Two experiments presented in Chapter 2 appear to show that colour category membership does influence colour perception. The tasks, which measured hue JNDs, showed that JNDs are lower at the blue-green boundary than are JNDs measured at within-category green or blue locations. This implies that perceptual space may be "warped" across the green-blue category boundary. In turn, this implies that colour category membership influences colour perception, a finding in agreement with previous studies (e.g. Roberson & Davidoff, 2000) which provide evidence for colour CP. The experiments presented in Chapter 2 were, however, more exclusively perceptual in their task demands than those used in previous studies and thus show the influence of colour categories on perceptual performance more clearly than previous studies. However the experiments do not reveal anything about the origin of colour categories themselves, for the following reason. Traditionally, two opposed views have sought to explain the origin of colour categories. One view suggests that colour categories are universal and perceptible regardless of whether or not a given individual's language encodes them. An opposing view supposes that colour categories are derived from language. In its strongest form, this relativist view claims that the continuum of electromagnetic radiation is cut up into categories by language
on a more or less arbitrary basis. For the English speakers who took part in the experiments presented in Chapter 2, the proposed universal colour categories coincide with the proposed relativist categories. Thus there is no way of knowing, on the basis of the experiments presented in Chapter 2, whether the categories which appear to influence performance are linguistic or universal in origin.

5.1.2. Potential influences of colour language on colour perception

To examine the origin of the colour categories that appear to underlie performance of the experiments presented in Chapter 2, additional experiments were carried out. These experiments, presented in Chapter 3, compared performance of English and Himba speakers in two different age groups. Himba speakers lack separate terms for blue and green and thus, if linguistic colour categories exert an influence on colour perception, might be expected to perform differently from English speakers. Similarly, the influence of colour categories, if linguistic in origin, might be expected to be diminished in young children, whose knowledge of colour terms is incomplete.

The experiments used in Chapter 3 were similar to those used in Chapter 2 in that they attempted to measure performance of low-level perceptual mechanisms. What the experiments demonstrated was consistent regardless of age or language group—namely that performance of trials involving stimuli from the blue-green boundary was superior to performance of trials involving stimuli drawn from neighbouring within-category regions of colour space. These findings differ significantly from recent cross-cultural studies of colour perception (e.g. Roberson et al., 2000, 2004) which have tended to provide evidence of an influence of linguistic colour categories on performance. An explanation that may account for these differences is that the tasks used in Chapter 3 measured perceptual processes more directly than those used by for example Roberson et al., 2000. It is argued that tasks used in previous studies are vulnerable to effects of memory and/or language and that consequently it is erroneous to draw conclusions about perceptual processes from tasks that do not measure them directly. Furthermore, it is argued that the results of the experiments presented in Chapter 3 do not coincide with ideas that colour categories are formed by learning.
names for particular regions of colour space. Rather, the results of these experiments are interpreted as suggesting the existence of universal perceptual colour categories that languages may or may not encode. However, it should be pointed out that results of the experiments presented in Chapters 1 and 2 do not rule out the possibility that learning linguistic colour terms may have measurable influences on colour perception. The scope of such influences though is supposed to be relatively limited and does not appear to underpin the CP effects observed.

5.1.3. Do differences in colour perception influence differences in colour language?

If the conclusions drawn from the cross-cultural work presented in Chapter 3 are considered valid, it becomes reasonable to ask on what basis languages differ in their colour lexicons. The idea that differences in colour language reflect differences in colour perception is one that antecedes the relativist-universalist debate. One such proposal, which is considered in Chapter 4, is that differences in the relative absorption of short wavelength light caused by lens brunescence may cause differences in colour perception that could partially account for the occurrence of 'grue' languages. This idea was examined by comparing English speakers of differing age groups (such an idea is feasible because lens brunescence increases with age). A series of experiments found that the CP effect, assessed by measuring JNDs, did not appear to be diminished in the older group tested, but that differences in colour naming in the blue-green region were apparent, as were differences in the relative salience of changes in appearance of stimuli along a blue-green axis. Whilst the apparent preservation of the CP effect counts against the idea that lens brunescence may play a role in some languages lacking separate blue and green colour terms, differences in colour naming and particularly the apparent compression of the blue-green axis are consistent with this idea. Caution though is needed in relating the results of the experiments presented in Chapter 4 to the wider debate concerning 'grue' languages because participants were not speakers of 'grue' languages. Moreover it is not certain to what extent the results should be interpreted as reflecting the influence of lens brunescence rather than other factors (besides lens brunescence) which may be

---

39 Though not impossible. It could, for example, be argued that measures of JNDs and response latency are not influenced at all by colour category membership and that the correspondence of improved JNDs etc with the blue-green category boundary is coincidental. By this account though claims for colour CP would be invalid.
associated with the aging process. Overall though the experiments in Chapter 4 do offer tentative support for the idea that differences in short wavelength light absorption may play a role in the lack of separate blue-green terms in many languages.

5.2. Implications for Differing Theoretical Perspectives

One major theoretical reference points guiding the work presented in this thesis has been the contrasting relativist and universalist perspectives on colour categorisation. Another has been the extent to which evidence for colour CP can be provided by tasks that measure low-level perceptual processes, associated with which are theories of perceptual learning. The implications of the findings presented here on theories of CP and perceptual learning are considered in the following section before some comments are made regarding the relativist-universalist debate.

5.2.1. Categorical perception and perceptual learning

Different theories of categorical perception mentioned briefly in Chapter 1 include those that suggest the phenomenon is innate (e.g. Bornstein, Kessen & Weiskopf, 1976), those that suggest it is caused by the warping of perceptual space (Harnad, 1987) and those that suggest CP may be a result of labelling (e.g., Kay and Kempton, 1984). What the experiments presented here suggest is that in different circumstances CP may be driven by different factors. For example, labelling of stimuli may produce CP in experiments that involve a delay between target and test phase. Such an effect is proposed here to account for the results of cross-cultural studies that find a link between colour CP and colour language (e.g. Roberson et al., 2000). In circumstances where the use of labels is discouraged, such as the adapted triadic judgement task used by Kay and Kempton (1984) a necessary link between linguistic colour categories and categorical perception appears absent. But as the experiments in Chapters 2 show, evidence for colour CP can be found on tasks that measure low-level perceptual processes. This implies that ‘warping’ of perceptual space may be involved in CP, or that CP may be innate. Since Chapter 3 provides evidence that CP does not depend on colour language the findings appear to refute the notion that CP is caused by a ‘warping’ of perceptual colour space contingent upon learning colour.
terms. In terms of perceptual learning, as is stated elsewhere, the findings presented here do not discount the possibility that colour term learning may cause a degree of perceptual 'warping' and that further testing of groups speaking different languages would find differences in the extent of CP congruent with differences in colour language. In general, the results of the experiments in Chapters 2 and 3 are neutral with regard to perceptual learning. They do though suggest that it would be a mistake to assume that evidence of colour CP provides evidence for perceptual learning unless clear evidence can be found indicating a tight link between variations in colour CP corresponding to variations in colour language on tasks that measure perception directly. Language specific differences on tasks susceptible to influences of memory and labelling are difficult to accept as evidence of perceptual learning. Similarly, it is hard to accept evidence of colour CP based on tasks susceptible to the same influence. The reason in both cases is that it may be doubted whether the tasks used measure perceptual processes directly.

5.2.2. Colour categories and universalist-relativist debate
With reference to colour language, the controversy between universalist and relativist standpoints has focused on the degree to which the division of colour space by language is constrained. Besides numerous studies that have focused on patterns of naming, the location of category foci, boundaries and centres, previous findings of CP have tended to be language specific and to provide evidence supporting relativist interpretations. The cross-cultural experiments in Chapter 4 also find CP, but this time corresponding to the proposed universal colour categories. Such findings provide strong evidence that perceptual colour categories are universally shared, particularly if criticisms of cross-cultural demonstrations of colour CP made previously are accepted. Whilst the findings do not rule out an influence of language on colour perception, they do not fit with extreme relativist views that languages divide colour space arbitrarily. As such the cross-cultural experiments reported here add to a body of evidence from a range of other sources, for example the emergence of colour terms in a hierarchical order in languages (Berlin & Kay, 1969), the location of category centroids (Kay & Regier, 2003) and colour CP in infants (Franklin et al., 2005b) which suggest that colour categories are not arbitrary.
5.3. Suggestions for Future Research

A limitation of all the studies presented in this thesis is that stimuli were confined to blue and green regions of colour space. Whilst this has permitted a rigorous examination of factors that may influence the colour categories used to name this region and conversely of the influence linguistic colour categories may have on colour perception in this region, generalisation of any conclusions drawn must be tentative.

With regard to the issues addressed in Chapter 4, that is the possible influence of lens brunescence on colour perception and naming, this limitation is less serious, because the aims of Chapter 4 were narrowly defined. However Chapter 2 and 3 addressed rather general issues, namely whether colour CP is perceptual and whether colour CP is influenced by colour language. To extend the validity of the conclusions drawn in these chapters, further studies should test for CP across other colour category boundaries. Such studies would serve not only to test theories of colour CP, but may also help to refine them. Assuming evidence of CP across other boundaries is found (some such evidence already exists: see for example, Franklin and Davies, 2004; Franklin et al., 2005a) an issue of particular interest is whether there are differences in CP effects across different chromatic colour category boundaries. It seems possible that CP effects across a boundary dividing primary and secondary categories (such as blue and purple) between two primary categories (blue and green) or two secondary categories (purple and pink) may differ. Similarly, changes between colour categories need not be defined solely in terms of hue differences, but may be defined by changes in brightness (brown and yellow) or combinations of hue and brightness (pink and red). Tests of colour CP across such boundaries would not merely be 'stamp-collecting' but are likely to provide valuable insights into the nature of perceptual colour categories. Conducting such experiments would present various technical challenges – few colour categories (that is either in the English language or the proposed universals) occupy as much of colour space as blue and green. At present the gamut of most displays prohibits reproduction of many regions of colour space. Consequently practical difficulties of stimulus specification exist.

A second area in which further work may prove fruitful relates to more detailed examination of whether learning colour terms can produce difference in colour
perception. The experiments presented here find no evidence of such effects and would appear to refute the idea that colour CP is caused by learning colour terms. However it may be that learning colour terms does produce measureable differences in perception. Therefore, further comparisons of different language groups might show differences in the overall size of the CP effect that are language dependant. In particular cross-linguistic comparisons of JND measurements collected using a task like the temporal 2-AFC detection task, which is more sensitive than the tasks used in Chapter 3, may reveal such differences. Results of such studies would be of interest primarily to those concerned with theories of perceptual learning. With regard to theories of colour term distribution in languages, they are likely to be of less interest—unless they contradict the results of the experiments presented here and provide evidence that colour CP is language dependant. This however would appear an unlikely prospect.

Finally, there seems to be considerable scope for further studies investigating whether differences in colour perception can be found amongst different groups and whether such differences, if they exist, covary with differences in language. That differences in rates of colour blindness exist between different groups appears certain; for example, 8% of Caucasian, 5% of Asian and 4% of African males suffer from Daltonism (Wolfmaier, 1999). Rates of acquired tritanopia appear higher in locales exposed to high levels of UV-B (Davies et al., 1998). Systematic comparisons of colour perception amongst different groups may reveal further differences. More specifically, investigating differences in the perception of the blue-green region of colour space may show a link with ‘grue’ languages. At present and although such theories have a long history, they lack corroborating evidence. There do appear to be grounds for supposing differences in blue-green perception exist: lens aging may occur at an accelerated rate among speakers of ‘grue’ languages, diets low in vitamins B11 and B12 can also produce acquired tritanopia and such diets are common in, for example, rural areas of southern Africa (Davies et al., 1998) where ‘grue’ languages predominate. However current evidence is somewhat indirect. Future studies should establish levels of brunescence in groups living around the equator and also establish how this effects colour appearance and colour naming in such groups. A particularly interesting experiment would be to test participants from regions of high UV-B exposure who speak languages that nevertheless encode separate green and blue
terms. Increased green-blue confusions amongst such groups, compared to say equivalent aged natives of the UK, would provide firmer evidence than currently exists that at least some differences in colour naming are influenced by differences in colour perception. If it is assumed that green-blue confusions only affect a proportion of members of any group then it would be important to establish how increases in the proportion of the group affected impact upon the linguistic utility of separate blue and green terms. Similar studies have already been carried out using computational modelling (e.g. Steels & Belpaeme, 2005).

5.4. Conclusions

The work presented in this thesis finds evidence of colour CP on tasks that measure low-level perceptual processes. Such effects do not appear to depend on linguistic colour categories. This suggests that colour categories in language may reflect a set of universal colour categories and that perceptual colour categories are not a consequence of learning colour words. Differences in colour language may be influenced by language independent differences in colour perception, such as systematic variations in lens brunescence.

---

40 This appears on the basis of Davies et al.'s. (1998) studies of tritanomolous colour vision defects to be the case.
References


Bornstein, M.H. & Korda, N. (1984). Discrimination and matching within and between hues measured by reaction times: some implications for categorical
perception and levels of information processing. *Psychological Research, 46*, 207-222.


Press.

Research, 33(8), 1053-1065.

Psychological Review, 96, 433-358.


Cognition. 57(3), 217-239.

Ltd.

and Management. Bristol: Adam Hilger Ltd.

Vision Research, 8, 1055-1065.


knowledge does not affect categorical perception of color in toddlers. Journal of 
Experimental Child Psychology, 90, 114 -141.


Appendix 1
Munsell and CIE Colour Spaces

1. The Munsell Colour System

Munsell colour space is three dimensional, specifying stimuli in terms of hue (corresponding to colour), value (lightness) and chroma (similar to saturation), as shown in figure A1. Hue is angular and specified by abbreviations of 5 main hues; red (R), yellow (Y), green (G), blue (B) and purple (P), and 5 intermediate hues; YR, GY, BG, PB, and RP. Each of these 10 lettered locations cover 36° arcs. Subdivisions are noted numerically on a scale of 1-10, giving hue values such as 5R, 7.5BG and 4PB etc. Value is specified using linear values. Values range between 0 (black) and 10 (white). Chroma values are radial. The maximum realisable chroma varies depending of hue and value. Stimuli are specified by giving the hue, value and chroma coordinates, in that order. Thus 5B 3/3 equates to a dark desaturated blue, whilst 5YR 8/9, would represent a light, saturated orange.

Figure A1. Schematic representation of Munsell colour space.
Each of the three Munsell dimensions is standardised to be perceptually uniform (Newhall, Nickerson and Judd, 1943). This means that if values in two dimensions are kept constant, equal numerical changes in the third dimension represent equal perceptual differences. For example the difference between a stimulus with Munsell notation 5B 6/8 and another of notation 5B 8/8 is perceptually equal to the difference between 5B 6/8 and 5B 4/8. In this case, the only changes are in value, and these are the same for both pairs of stimuli. Thus the perceptual difference between both pairs of stimuli is the same. Another two stimulus pairs, 2.5Y 2/4, 5Y 2/4 and 5Y 2/4 and 7.5Y 2/4 are also perceptually equal, though in this instance the difference is one of hue. Changing more than one dimension at a time results in a breakdown of perceptual uniformity. Thus stimulus pairs 4P 5/6, 3P 4/5 and 4P 5/6, 5P 6/7 cannot be said to be equally perceptually different.

Munsell stimuli should be viewed under conditions of controlled illumination, preferably illuminant C, which appears white and has a colour temperature of 6700°. This approximates daylight. If Munsell colours are viewed on a monitor, a white point of x = 0.31, y = 0.316 approximates illuminant C.


CIE colour spaces are based on data from colour matching functions made by several researchers, (e.g. Guild, 1931; Burch, 1959) representing the CIE 1931 Standard Colorimetric Observer and the CIE 1964 Supplementry Standard Colorimetric Observer and defined ultimately from three primaries, X,Y and Z, which are theoretical super-saturated colours lying outside the bounds of the spectral locus. Equal changes in the 3 variables do not necessarily produce equal perceptual distances. Smaller changes in the Z primary for example can be detected than in the Y primary. CIELAB and LIELUV are two attempts to transform XYZ space into perceptually uniform colour spaces, such that a minimum perceived colour change in any direction from any point in the space is numerically equivalent to a minimum perceived colour change in any direction from another point in colour space.
In both CIELAB and CIELUV, the chromatic dimensions, $a^*b^*$ and $u^*v^*$ do not map directly onto the perceived colour dimensions. Lines along a constant dimension do not equate with equal equivalent hues. For example, the Bezold-Brücke hue shift illustrates how changes in retinal illumination can cause changes in perceived hue. (Wyszecki and Styles, 1982, pp. 420-424). At lower illumination levels stimuli look more red or green, at higher levels blue or yellow. For colours of a particular hue therefore, $u^*v^*$ and $a^*b^*$ values change as lightness changes. Another feature of CIELAB and CIELUV is that colour information is centred on the colour of the white point used. In the case of stimuli generated on a monitor, it is usual to use the white point of that monitor.

2.1. CIELAB

The formula used to derive CIELAB is related to the Adams-Nickerson colour difference formula. There are three Cartesian axes: $L^*$ represents luminance with values of 100 representing white and 0 black, $a^*$ approximates the red-green axis and $b^*$ the yellow-blue axis posited by opponent process theories of colour perception, as shown in figure A2.

![Figure A2. Schematic representation of $L^*a^*b^*$ colour space.](image-url)
Perceptual distances are calculated using the following formula, where $\Delta E$ denotes perceptual distance and the other $\Delta$ quantities the differences between corresponding co-ordinates of the two stimuli.

$$\Delta E = \sqrt{(\Delta L)^2 + (\Delta a^*)^2 + (\Delta b^*)^2}.$$  

2.2. CIELUV

CIELUV was introduced at the same time as and is similar to CIELAB. The co-ordinates used are $L^*u^*v^*$ co-ordinates. $L^*$ values are the same as in $L^*a^*b^*$ colour space. Values for $u^*$ and $v^*$ are calculated using a different formula. The formulae for computing these co-ordinates may be found in Wyszecki and Styles (1982, p. 165).

CIELAB and LUV systems each work better on different validation data sets (Shevell – The Science of Colour, p. 206), making it difficult to choose between the two colour spaces.
Appendix 2

L*a*b* co-ordinates for stimuli used in Experiment 6

### CATEGORY BOUNDARY

<table>
<thead>
<tr>
<th>Target</th>
<th>L*</th>
<th>a*</th>
<th>b*</th>
<th>L*</th>
<th>a*</th>
<th>b*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue-Green</td>
<td></td>
<td></td>
<td></td>
<td>Green-Blue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>49.94</td>
<td>-48.81</td>
<td>-25.17</td>
<td>2.5</td>
<td>48.69</td>
<td>-47.07</td>
</tr>
<tr>
<td>5</td>
<td>48.62</td>
<td>-51.04</td>
<td>-23.31</td>
<td>5</td>
<td>48.94</td>
<td>-45.04</td>
</tr>
<tr>
<td>7.5</td>
<td>48.71</td>
<td>-51.11</td>
<td>-21.82</td>
<td>7.5</td>
<td>49.69</td>
<td>-44.34</td>
</tr>
<tr>
<td>10</td>
<td>48.04</td>
<td>-53.57</td>
<td>-19.87</td>
<td>10</td>
<td>49.78</td>
<td>-42.80</td>
</tr>
<tr>
<td>12.5</td>
<td>48.34</td>
<td>-54.68</td>
<td>-19.41</td>
<td>12.5</td>
<td>49.44</td>
<td>-41.45</td>
</tr>
<tr>
<td>15</td>
<td>48.01</td>
<td>-55.69</td>
<td>-17.65</td>
<td>15</td>
<td>49.83</td>
<td>-39.40</td>
</tr>
<tr>
<td>20</td>
<td>47.65</td>
<td>-57.67</td>
<td>-15.17</td>
<td>20</td>
<td>49.98</td>
<td>-38.47</td>
</tr>
<tr>
<td>25</td>
<td>47.55</td>
<td>-63.39</td>
<td>-11.61</td>
<td>25</td>
<td>50.19</td>
<td>-34.74</td>
</tr>
<tr>
<td>30</td>
<td>47.77</td>
<td>-66.05</td>
<td>-10.31</td>
<td>30</td>
<td>50.71</td>
<td>-30.86</td>
</tr>
<tr>
<td>35</td>
<td>48.98</td>
<td>-67.63</td>
<td>-9.65</td>
<td>35</td>
<td>50.84</td>
<td>-27.74</td>
</tr>
<tr>
<td>40</td>
<td>47.24</td>
<td>-69.64</td>
<td>-3.14</td>
<td>40</td>
<td>49.20</td>
<td>-22.57</td>
</tr>
</tbody>
</table>

### CATEGORY CENTRE

<table>
<thead>
<tr>
<th>Target</th>
<th>L*</th>
<th>a*</th>
<th>b*</th>
<th>L*</th>
<th>a*</th>
<th>b*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Green</td>
<td></td>
<td></td>
<td></td>
<td>Blue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>44.44</td>
<td>-76.27</td>
<td>26.84</td>
<td>2.5</td>
<td>49.98</td>
<td>-36.47</td>
</tr>
<tr>
<td>5</td>
<td>44.30</td>
<td>-81.05</td>
<td>23.18</td>
<td>5</td>
<td>50.24</td>
<td>-33.18</td>
</tr>
<tr>
<td>7.5</td>
<td>44.71</td>
<td>-81.73</td>
<td>21.16</td>
<td>7.5</td>
<td>50.48</td>
<td>-32.04</td>
</tr>
<tr>
<td>10</td>
<td>44.76</td>
<td>-80.12</td>
<td>19.34</td>
<td>10</td>
<td>50.72</td>
<td>-30.86</td>
</tr>
<tr>
<td>12.5</td>
<td>45.46</td>
<td>-81.05</td>
<td>18.58</td>
<td>12.5</td>
<td>50.72</td>
<td>-29.05</td>
</tr>
<tr>
<td>15</td>
<td>45.76</td>
<td>-81</td>
<td>14.90</td>
<td>15</td>
<td>50.84</td>
<td>-27.74</td>
</tr>
<tr>
<td>20</td>
<td>46.55</td>
<td>-76.59</td>
<td>10.50</td>
<td>20</td>
<td>49.30</td>
<td>-22.57</td>
</tr>
<tr>
<td>25</td>
<td>48.68</td>
<td>-76.68</td>
<td>7.30</td>
<td>25</td>
<td>56.61</td>
<td>-19.67</td>
</tr>
<tr>
<td>30</td>
<td>56.23</td>
<td>-73.88</td>
<td>3.46</td>
<td>30</td>
<td>54.88</td>
<td>-10.81</td>
</tr>
<tr>
<td>35</td>
<td>46.00</td>
<td>-72.38</td>
<td>0.68</td>
<td>35</td>
<td>54.56</td>
<td>-5.78</td>
</tr>
<tr>
<td>40</td>
<td>47.27</td>
<td>-60.54</td>
<td>-3.14</td>
<td>40</td>
<td>53.04</td>
<td>3.13</td>
</tr>
</tbody>
</table>
Appendix 3

L*a*b* co-ordinates for stimuli used in Experiment 7

### CATEGORY BOUNDARY

<table>
<thead>
<tr>
<th></th>
<th>Blue-Green</th>
<th>Green-Blue</th>
</tr>
</thead>
<tbody>
<tr>
<td>L*</td>
<td>a*</td>
<td>b*</td>
</tr>
<tr>
<td>Background</td>
<td>43.37</td>
<td>-21.48</td>
</tr>
<tr>
<td>separation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>43.36</td>
<td>-21.40</td>
</tr>
<tr>
<td>0.5</td>
<td>43.19</td>
<td>-21.31</td>
</tr>
<tr>
<td>1</td>
<td>42.92</td>
<td>-20.96</td>
</tr>
<tr>
<td>1.5</td>
<td>42.87</td>
<td>-20.70</td>
</tr>
<tr>
<td>2</td>
<td>43.26</td>
<td>-20.61</td>
</tr>
<tr>
<td>3</td>
<td>43.51</td>
<td>-19.98</td>
</tr>
<tr>
<td>4</td>
<td>43.36</td>
<td>-18.92</td>
</tr>
<tr>
<td>5</td>
<td>44.11</td>
<td>-18.08</td>
</tr>
</tbody>
</table>

### CATEGORY CENTRE

<table>
<thead>
<tr>
<th></th>
<th>Green</th>
<th>Blue</th>
</tr>
</thead>
<tbody>
<tr>
<td>L*</td>
<td>a*</td>
<td>b*</td>
</tr>
<tr>
<td>Background</td>
<td>43.31</td>
<td>-24.73</td>
</tr>
<tr>
<td>separation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>43.24</td>
<td>-24.89</td>
</tr>
<tr>
<td>0.5</td>
<td>43.30</td>
<td>-25.16</td>
</tr>
<tr>
<td>1</td>
<td>43.53</td>
<td>-25.60</td>
</tr>
<tr>
<td>1.5</td>
<td>43.94</td>
<td>-25.03</td>
</tr>
<tr>
<td>2</td>
<td>43.71</td>
<td>-25.92</td>
</tr>
<tr>
<td>3</td>
<td>44.01</td>
<td>-20.13</td>
</tr>
<tr>
<td>4</td>
<td>43.26</td>
<td>-28.29</td>
</tr>
<tr>
<td>5</td>
<td>43.76</td>
<td>-29.40</td>
</tr>
</tbody>
</table>
Appendix 4

Location of stimuli used in Experiments 8a and 8b with L* u*v* co-ordinates marked. Filled symbols indicate stimuli used in both experiments. Unfilled symbols indicate stimuli used only in Experiment 8b.
Appendix 5

Location of stimuli used in Experiments 8e as viewed through yellow filter, with u*v* co-ordinates marked. Values of achromatic point are also shown.

L* = ◇ filter 61.7  □ filter 41.3  △ filter 20.44