The nature and origin of categorical colour perception: cross-cultural and interference task approaches

Alison Wiggett

Thesis submitted as part-fulfilment for the requirements of the degree of Doctor of Philosophy in Psychology.

Department of Psychology
University of Surrey
March 2004
The current thesis presents a series of experiments examining the relationship between language and colour cognition. The main focus is on the categorical perception (CP) of colour, the finding that discrimination of colours that straddle a category boundary is more accurate than within category colour discrimination. It has recently been argued that CP is not a perceptual effect, but rather a direct effect of language due to the comparison of stimulus labels (Roberson & Davidoff, 2000). The current thesis tests this account in a series of colour discrimination tasks with verbal and visual interference, a visual search task and a cross-cultural comparison of speakers of languages that make different categorical colour distinctions. The results of the colour discrimination tasks suggest that the effect of verbal interference on CP is dependent on task design. CP was found to survive verbal interference when the type of interference was not predictable, suggesting that the task was open to encoding strategies. CP did not survive when the colour target was presented with incongruent Stroop interference. An account of CP is proposed in which target name generation is necessary for CP. However, incongruent Stroop interference at test stimuli presentation did not selectively affect CP, suggesting that the process leading to CP is more complex than a simple matching-to-labels account would suggest. Target name generation may activate or reinforce a category code, which in turn facilitates cross-category discrimination. However, as evidence for CP was also found in the visual search task - suggesting that CP can be a perceptual effect - the account of CP proposed here may be specific to certain types of tasks only. The possibility that CP is a memory effect due to a shift towards prototype (Huttenlocher, Hedges & Vevea, 2000) was also considered. The findings presented here suggest that CP is dissociated from a shift towards prototype. The cross-cultural comparison of English and Owambo speakers on a triads and a visual search task allowed a further test of the role of language in colour CP. Language effects were found on both tasks, however, the differences were not all predicted by differences in colour naming. The results suggest that the representation of colour may also differ cross-culturally. Overall, the results presented here suggest that colour perception is not independent of colour language.
Acknowledgements

I would like to thank Ian for supervising my PhD and for being so supportive and patient throughout the last 3 years. Thanks also to Emre and Mike for your help with programming and thanks to Ally and Amy for all your help collecting and analysing data.

I’d like to thank Marieta Bester for her help with collecting the Namibian data and Vei, Erica and Anton for letting us stay. Thanks to Anna for making the trip so special and fun and thanks for the exquisite musical entertainment you provided.

Thank you Anna, Maya, Chris, Tanika, Alex, Spyri, Emily and Adrian for being great friends/housemates/lunch buddies. It’s been a pleasure.

Special thanks to Helen, Richard, John and my Grandparents who have all supported me throughout my PhD. Thank you for your love and support.

I would especially like to thank Si who, despite being thousands of miles away, was always there for me. Thank you for your encouragement, support, patience and love.
# Table of contents

Abstract ................................................................. i  
Acknowledgements .................................................. ii  
List of Figures ......................................................... viii  
List of Tables ....................................................... xiv  

## Chapter One: Colour and Language
1.1. General Introduction ............................................. p.1  
1.2. The nature of categories ....................................... p.3  
   1.2.1. Classical theory of categorisation .................. p.3  
   1.2.2. Prototype Theory ....................................... p.5  
1.3. Categorical Perception .......................................... p.7  
   1.3.1. Naturalistic theories ................................... p.8  
   1.3.2. Perceptual change theories ......................... p.10  
   1.3.3. Language theories ..................................... p.11  
1.4. Colour categories .............................................. p.12  
   1.4.1. The physiological basis of colour perception .... p.13  
   1.4.2. The influence of language on colour categorisation...... p.16  
      1.4.2.1. Linguistic relativity .......................... p.16  
      1.4.2.2. Cognitive universals ...................... p.18  
1.5. Summary ......................................................... p.23  
1.6. Overview of the chapters ..................................... p.24  

## Chapter Two: The effect of verbal and visual interference on the categorical perception of colour
2.1. General Introduction ............................................. p.26  
2.1.1. Working Memory ......................................... p.29  
2.1.2. Verbal and perceptual codes in colour memory ....... p.31  
2.1.3. Verbal Overshadowing ................................... p.33
3.2. Experiment 5: Stroop interference at target presentation

3.2.1. Introduction ................................................................. p.85
3.2.2. Method ................................................................................. p.87
3.2.3. Results ................................................................................ p.90
3.2.4. Discussion ......................................................................... p.96

3.3. Experiment 6: Stroop interference at test presentation

3.3.1. Introduction ..................................................................... p.99
3.3.2. Method ............................................................................... p.101
3.3.3. Results .............................................................................. p.103
3.3.4. Discussion ......................................................................... p.107

3.4. Experiment 7: Stroop interference in a quick discrimination task

3.4.1. Introduction ...................................................................... p.110
3.4.2. Method ............................................................................... p.111
3.4.3. Results .............................................................................. p.112
3.4.4. Discussion ......................................................................... p.115

3.5. Experiment 8: Stroop at test in a quick discrimination task

3.5.1. Introduction ...................................................................... p.117
3.5.2. Method ............................................................................... p.118
3.5.3. Results .............................................................................. p.119
3.5.4. Discussion ......................................................................... p.122

3.6. General Discussion .............................................................. p.124

Chapter Four: Is categorical perception due to a shift towards prototype?

4.1. General Introduction .......................................................... p.128
4.1.1. Perceptual magnet effect ......................................................... p.129
4.1.2. Shift towards prototype .......................................................... p.129
4.1.3. CP, perceptual magnet effect and shift towards prototype ...... p.131
4.1.4. Categorical perception and the shift towards prototype........ p.135

4.2. Re-analysis of 2AFC and same-different colour discrimination tasks p.136
4.2.1. Re-analysis of Experiment 1: 2AFC task ................................. p.136
4.2.2. Re-analysis of Experiment 2: same-different task ................. p.137
Appendices

Appendix 1  Perceptually uniform colour spaces .............................................. p.242
Appendix 2  List of word pairs used in the verbal interference tasks in
            Experiments 1-3 .............................................................. p.245
Appendix 3  Munsell codes, CIE Yxy and L*u*v* co-ordinates of colour
            stimuli used in Experiments 1-8 ...................................... p.246
Appendix 4  Triad choice by individual cross-category and within category
            classification ................................................................. p.247
List of Figures

Chapter Two

Figure 2.1. Diagram illustrating the task used by Roberson and Davidoff (2000) ............................................................... p.27

Figure 2.2. Discrimination accuracy for within and cross-category comparisons in blocked interference conditions (Taken from Pilling et al., 2003, Experiment 1) ...................... p.37

Figure 2.3. Schematic representation of the stimuli matrix showing Munsell hue (horizontal axis) and value (vertical axis) at constant chroma 7.4 .............................................................. p.40

Figure 2.4. Examples shape pairs used in the visual interference task .. p.44

Figure 2.5. Mean accuracy for within category (A-a) and cross-category (A-B pairs) .............................................................. p.47

Figure 2.6. Mean A' for within (A-a) and cross-category (A-B) pairs .... p.52

Figure 2.7. Mean accuracy for focal and boundary stimuli for each interference condition (blocked) ........................................ p.56

Figure 2.8. Mean accuracy for focal and boundary stimuli for each interference condition (unblocked) ...................................... p.57

Figure 2.9. Illustration of search conditions in Experiment 4 ............. p.63

Figure 2.10. Mean reaction times in milliseconds for within-category and cross-category search plotted against display size (4, 16, 36) for target present trials only ........................................ p.65

viii
Figure 2.11. Mean reaction times in milliseconds for within-category and cross-category search plotted against display size (4, 16, 36) for target absent trials only .................................................... p.66

Figure 2.12. Mean percent accuracy for within-category and cross-category search plotted against display size (4, 16, 36) for target present trials only ........................................................... p.67

Figure 2.13. Mean percent accuracy for within-category and cross-category search plotted against display size (4, 16, 36) for target absent trials only ............................................................ p.68

Chapter Three

Figure 3.1. Schematic representations of neutral (a), congruent (b) and incongruent (c) cross-category trials in 5a ............................................ p.88

Figure 3.2. Examples of a congruent trial in Experiment 5b. Stimuli were presented as colour-in-word stimuli to ensure concurrent processing of the two dimensions ......................... p.89

Figure 3.3. Mean discrimination accuracy for within (A-a) and cross-category (A-B) pairs in Experiment 5a ................................................. p.91

Figure 3.4. Mean discrimination accuracy or cross-category (A-B) and within (A-a) pairs in Experiment 5b ......................................................... p.94

Figure 3.5. Schematic representations of neutral (a), congruent (b), incongruent (c), and mixed (d,e) cross-category trials presented in Experiment 6 ......................................................... p.103
Figure 3.6. Mean discrimination accuracy for cross-category (A-B) and within category (A-a) pairs in Experiment 7 ...................... p.106

Figure 3.7. Example of a congruent cross-category trial presented in Experiment 7 ............................................................. p.112

Figure 3.8. Mean discrimination accuracy for cross-category (A-B) and within (A-a) pairs in Experiment 7 ............................ p.114

Figure 3.9. Example of an incongruent cross-category trial used in Experiment 8 ............................................................. p.119

Figure 3.10. Mean discrimination accuracy for cross-category (A-B) and within category (A-a) pairs in Experiment 8 ..................... p.121

Chapter Four

Figure 4.1. Warping of perceptual colour space ......................... p.131

Figure 4.2. Accuracy (A') around the prototypes compared to either side of the prototypes (reproduced from Davies, Özgen, Pilling & Wiggett, 2003) ................................................................. p.132

Figure 4.3. The effect of a shift towards prototype ..................... p.133

Figure 4.4. Same-different accuracy (A') for peripheral pairs. (Reproduced from Davies et al., 2003) ............................................ p.134

Figure 4.5. Schematic representation of stimuli used in the categorical perception experiments in Chapter 2 and 3. Stimuli representations shift towards the prototype. Shift towards prototype arguments predict CP .............................................. p.135
Figure 4.6. Mean accuracy for within category (analysed by order of presentation) and cross-category discrimination (Re-analysis of Experiment 1) .............................................................. p.137

Figure 4.7. Mean accuracy (A') for within category (analysed by order of presentation) and cross-category discrimination (Re-analysis of Experiment 2) .............................................................. p.138

Figure 4.8. Schematic representation of order effects found in Experiment 2. A' for within category pairs as a function of stimulus order .............................................................. p.139

Figure 4.9. Diagram showing two possible directions of the shift towards prototype. The direction of the shift is dependent on whether the prototype is defined as the average of all stimuli (P\text{t}) or the pre-existing prototypes (P\text{p}, P\text{g}) .............................................................. p.140

Figure 4.10. Possible effects of incongruency on within category discrimination .............................................................. p.143

Figure 4.11. Mean accuracy for within category discrimination (analysed by order of stimulus presentation) and cross-category discrimination (Re-analysis of Experiment 5a) ..................... p.144

Figure 4.12. Mean Accuracy scores for within category (analysed by order of presentation) and cross-category discrimination (Re-analysis of Experiment 5b) .............................................................. p.146

Figure 4.13. Mean accuracy for within category discriminations (analysed by order of presentation) and cross-category presentation (Re-analysis of Experiment 6) .............................................................. p.148
Figure 4.14. Mean accuracy for within category discrimination (analysed by order of presentation) and cross-category discrimination (Re-analysis of Experiment 7) ........................................... p.151

Figure 4.15. Mean accuracy for within category discrimination (analysed by order of presentation) and cross-category discrimination (Re-analysis of Experiment 8) ........................................... p.153

Chapter Five

Figure 5.1. Schematic representation of the lists used in the visual search task ............................................................ p.170

Figure 5.2. Location and distances of colours in CIE L*u*v* space for all four search conditions ........................................... p.171

Figure 5.3. Search time in seconds for one-target conditions (1T-pure and 1T-mixed, Experiment 9) ........................................... p.173

Figure 5.4. Search time in seconds for the two target conditions (2T-pure and 2T-mixed, Experiment 9) ........................................... p.174

Figure 5.5. Schematic representation of the stimuli used for the Red/Pink 2 triads in Experiment 10 ........................................... p.183

Figure 5.6. Schematic representation of the triads in Experiment 10 ...... p.184

Figure 5.7. Mean category scores for each cross-category triad set type for English and Balantu speakers ........................................... p.192
Figure 5.8. Mean Value and hue scores for each within category triad set for English and Balantu speakers ............................................ p.194

Figure 5.9. Mean category/hue scores for cross and within category triads for English and Balantu ................................................. p.196
List of Tables

Chapter Two

Table 2.1. Naming frequencies (%) across participants for forced choice (blue or green) and free response conditions ....................... p.41

Table 2.2. Examples of word pairs used in the verbal interference task... p.44

Chapter Three

Table 3.1. Mean reaction time in neutral, congruent, and incongruent Stroop conditions for cross-category (A-B) and within category (A-a) discriminations in Experiment 5a ..................... p.87

Table 3.2. Mean reaction time in neutral, congruent, and incongruent Stroop conditions for cross-category (A-B) and within category (A-a) discriminations in Experiment 5b ..................... p.90

Table 3.3. Mean reaction time in neutral, congruent, incongruent and mixed Stroop conditions for cross-category (A-B) and within category (A-a) discriminations in Experiment 6 ...................... p.101

Table 3.4. Mean reaction time ms in neutral, congruent, and incongruent Stroop conditions for cross-category (A-B) and within-category (A-a) discriminations in Experiment 7 ............. p.110

Table 3.5. Mean reaction time in neutral, congruent, incongruent, and mixed Stroop conditions for cross-category (A-B) and within-category (A-a) discriminations ......................... p.116
Chapter Four

Table 4.1. Summary of the re-analysis of Experiments 1 and 2 and 5-8 for within category order effects in Chapter 4 ......................... p.151

Chapter Five

Table 5.1. Colours used in the visual search task, dominant term given by English and Kwanyama speakers and CIE L*u*v* co­ordinates for each colour .......................................................... p.167

Table 5.2. Triad sets used in Experiment 10 ............................................. p.182

Table 5.3. Perceptual hue and value CIE distances (ΔE) for each triad set ................................................................. p.183

Table 5.4. English Naming Data (Experiment 10) .................................... p.186

Table 5.5. Balantu Naming Data (Experiment 10) .................................... p.187

Table 5.6. Frequency of choice for each tile (with position of English category boundary) for English and Balantu ...................... p.189

Table 5.7. Mean Category Scores for each cross-category triad set type for English and Balantu ......................................................... p.192

Table 5.8. Mean Value and Hue scores for each triad set type for English and Balantu ................................................................. p.194
Chapter One

Colour and Language

1.1. General Introduction

"Imagine a tribe of colour-blind people, and there could easily be one. They would not have the same colour concepts as we do. For even assuming they speak, e.g. English, and thus have all the English words, they would still use them differently than we do and would learn their use differently. Or if they have a foreign language, it would be difficult to translate their colour words into ours."

Ludwig Wittgenstein

Colour categories are lexical categories (for example red, blue, and green) as well as perceptual categories (the experience of red, blue, or green) and the experience of colour is on some level initiated by neurophysiological processes in the eye and brain. A fundamental question is how are these levels of colour perception and categorisation related? Do the semantic categories map directly on to the perceptual categories? Or are the perceptual categories formed by our linguistic categories?

Wittgenstein's remark on colour, quoted above, highlights the difficulties involved in studying the relationship between colour language and colour perception: even if the tribe of colour blind people used the same English colour words, they clearly could not have the same colour concepts since they would not have perceived colour the way we do. In other words their internal representations of colour space would not be the same as ours. Secondly, if the tribe spoke a different language to ours, we would not know which of their terms, if any, corresponded to our own terms. Thus, we are faced with the problem of not knowing how colour space is represented in speakers of different languages.
This points to the question addressed by the current series of experiments: how are colour categories encoded and represented, both linguistically and perceptually? Thus, the main focus is on the relationship between colour perception, colour categorisation and language. Experimentally, this can be addressed by testing what types of colour codes are used when colours are encoded, stored and recalled. A core question is whether these representations are visual or verbal in nature.

The thesis also addresses the relationship between colour categorisation and language cross-culturally. The fundamental question is whether the language we speak influences the way we perceive the world. The colour domain is ideal for testing this hypothesis as colour vocabularies vary considerably cross-culturally which allows a test of whether linguistic differences are paralleled by non-linguistic cognitive differences. Relating to the point made above, that colours can be represented verbally and perceptually, the comparison of speakers of different languages should shed light on the processes involved. If our colour categories are hard-wired, speakers of different languages should nonetheless have the same representations of colour. Thus, the main question is: are the variations in colour categorisation merely on a lexical, surface level, or do language and cultural influences shape the internal representations and possibly the perception of colour?

The focus in this thesis is on the categorical perception of colour (hereafter CP). Colour space is a continuum, yet we perceive qualitative changes at boundaries and hence perceive colours categorically. An important question is whether these qualitative changes are hardwired in the visual system or whether they are acquired in accordance with the language we speak. The number of colour terms used to describe colour space varies across languages which means that the distribution of category boundaries also varies. If language is important in colour categorisation, CP should only be found across boundaries of existing linguistic categories.

1.2. The nature of categories

The question of how we form ideas and concepts from raw sensory experience has a long history in psychology and philosophy. Aristotle stated that “as a result of seeing the same thing happen many times we look for the universal and have proof; the universal becomes clear from a number of particular instances” (from Homa, 1984, p. 53). Aristotle proposed the existence of ten categories that were mutually exclusive and jointly exhaustive of reality. The British Empiricist John Locke also puzzled over the question of how categories emerge. He argued that the mind computed general ideas from directly experiencing the world and noted that higher level processes such as reasoning and communicating could only be done if sensations and experiences were grouped into general concepts and categories (Homa, 1984).

Today the same question is one of the fundamental questions of cognitive science (Harnad, 1987). If we did not categorise the world around us, each instance would have to be treated as new and unique. The task of dealing with all inputs would be much harder, if not impossible. Rosch and Lloyd (1978) stated that categorisation is important since “the world consists of an infinite number of potentially different stimuli” (p.1). Therefore, it is an important task for all organisms to segment the world into groups or categories so that for the purpose of cognitive economy (Rosch, 1978) non-identical stimuli can be treated as equivalent. The question is how do these categories emerge, how do we learn them and how are categories and instances of categories represented. Two of the main theories of categorisation will be briefly outlined here: classical theory and prototype theory.

1.2.1. Classical theory of categorisation

The classical theory of categorisation proposes that categories are formed on the basis of defining attributes. Stimuli that have a certain set of attributes are regarded as members of the category; the attributes are both necessary and sufficient for category membership. This means that if an item, object, or instance does not have all the attributes, it is not a member of the category. The implications of this rule are firstly,
that category membership is all-or-none. Secondly, all members of a category are equally representative of that category. Furthermore, categories are thought to be ordered in strict hierarchies. An example of work done in this tradition is that of Collins and Quillian (1970). They asked simple yes-no questions such as "is a collie a dog" or "is a collie an animal". Response times to the former question were shorter compared to response times to more general questions. The authors argued that, when asked "is a collie a dog", people had to infer that collies were dogs and that dogs were animals. The model of categorisation constructed from this is ordered in strict hierarchies. At the first level there are general categories such as 'animal'. At the next level the categories are more specific such as 'dog' or 'bird', and at most specific level instances such as 'collie' or 'canary' are encoded. Thus, categories at lower levels are automatically also members of the higher level category and the decision as to whether something is a member of a category is done by inferring from one level of the hierarchy to the next.

The classical theory of categorisation has several shortcomings. The first problem is that the theory does not consider the notion of typicality. Rosch and Mervis (1975) found that not all members of categories are judged as equally typical. For example, a sparrow tends to be rated as a more typical bird than an ostrich. Thus, membership is graded rather than all-or-none. Furthermore, Rosch found that items that were judged to be more typical were also categorised faster than less typical examples (Rosch, 1978).

A further criticism of the classical theory is that the boundaries of categories are not always well defined. McCloskey and Glucksberg (1978) reported category membership decisions made by the same participants on two separate occasions. The results showed that consistency between subjects and across sessions was high for highly typical category members, but much less consistent for intermediate typicality. This suggests that category boundaries tend to be fuzzy rather than clearly defined.

Thirdly, not all categories have clearly defined attributes or features. The example famously advanced by Wittgenstein (1953) was that of the category 'game'. Everybody knows what a game is, yet it seems impossible to provide a clear set of defining attributes that need to be fulfilled for something to qualify as a game. Thus,
even when a category may have a defining set of attributes, people are often poor at defining them (Medin, 1989). The idea that people know what defines a category and that this knowledge is consciously used to determine category membership is hard to substantiate as the above example suggests that the knowledge may not always exist. Furthermore, it has been found that even when the knowledge does exist it may be wrong (McNamara & Sternberg, 1983).

1.2.2. Prototype theory

The prototype theory of categorisation opposes the idea that we learn what is necessary and sufficient to determine category membership by feature analysis. The prototype theory postulates that we categorise not by identifying critical properties, but by learning a prototype, an ideal member of the category, against which other members are matched. Thus, there is no uniform structure to categories, rather the boundaries are fuzzy and category membership is graded. The abstracted prototype functions are used in subsequent classification or recognition instances more than individual exemplars.

Recent studies on the nature of categorisation of artificial perceptual categories such as polymorphous stimuli, for which no single feature is either necessary or sufficient for category membership, support the prototype theory (Depy, Fagot, Vanclain, 1997). In these studies people usually learn to classify sets of stimuli such as dot-patterns or shapes into two categories. The stimuli shown are variants of the two category prototypes. It has been shown that classification of the unseen prototypes is more accurate than that of new exemplars (Posner & Keele, 1968). There is evidence to suggest that the prototype is classified more accurately than the trained stimuli (Breen & Schvaneveldt, 1986; Homa, Sterling, Trepal, 1981; McLaren, Bennett, Guttman-Nahir & Mackintosh, 1995). Under certain conditions the prototype has been shown to receive higher oldness ratings than the actual training instances even though the prototype was novel (Homa, Goldhardt, Burrue-Homa & Smith, 1993). Furthermore, for estimates of spatial location, Huttenlocher, Hedges & Duncan (1991) found that dots were misplaced towards the prototypical location. Overall, these studies support the idea that prototypes are the central, ideal, and most salient category instance.
Medin & Barsalou (1987) have pointed out that a distinction needs to be made between perceptual and conceptual categories. Perceptual categories are for example colours, shapes, speech sounds, and spatial frequency patterns. Conceptual categories are knowledge based categories, either natural taxonomic categories such as animal or tree, or categories of artefacts such as furniture and car. However, Rosch, Mervis, Gray, Johnson and Boyes-Braem (1976) have argued that many abstract concepts and categories are to some extent organised around perceptual similarities, an idea also advanced by Goldstone and Barsalou (1998). Thus, Rosch’s prototype theory has been applied to both conceptual (Rosch & Mervis, 1975) and perceptual (i.e. colour, Rosch, 1973) categories.

Categories, according to Rosch (Rosch, 1978; Rosch & Mervis, 1975), are defined on three levels: a basic level, a super-ordinate level and a sub-ordinate level. An example of a basic level category is ‘table’, the corresponding super-ordinate category is ‘furniture’, the sub-ordinate category ‘kitchen table’. Rosch argues that categorisation first takes place at the basic level, at this level the features are common to most members of the category (in the example given above ‘table’). At the super-ordinate level, however, the instances or examples have fewer common features (furniture). Lastly, at the sub-ordinate level there is considerable overlap between different categories (for example a kitchen table and a living-room table) and between categories at the basic level. Crucially, categories are most distinct at the basic level of comparison, it is here that categories have the highest ‘cue validity’ as categories on this level are perceptually similar.

Furthermore, Rosch argued that the basic level categories were psychologically the most salient. There is evidence to suggest that people usually use the basic category names to describe objects and that children generally acquire basic category names first in language acquisition (Rosch, 1978). The notion of saliency of basic categories is particularly relevant to Rosch’s work on colour categories (see § 1.4.2.2.).

Harnad (1987) points out that the prototype theory of categorisation holds well for some categories, for example facial expressions. However, the problem is that for many categories there is no prototype, for example there is no ‘ideal’ chair. Chairs
must share certain features to be classified as chairs. Thus, for some categories prototype theory seems to fall back on ideas of categorisation it set out criticising. Or as Margolis (1994) argued “the problems that infect earlier theories crop up in their successors” (p.74). Van Brakel (1991) is also strongly critical of prototype theory, especially the assumption that “there are pan-human psychological essences or core meanings that refer to basic level natural kinds” (p.233), an assumption, which van Brakel argues, is based on ethnocentric ideas. However, different theories may be appropriate to different kinds of categories. Homa (1984) argues that explanations may have been inappropriately extended to categories that may be qualitatively different or may possess different types of structures, hence “the premise that all categories are processed in fundamentally similar ways is almost certainly false” (p. 51).

1.3. Categorical Perception

As the process of grouping instances into categories is said to be fundamental to all organisms (Rosch & Lloyd, 1978), the way categories are subsequently perceived is in turn argued to be fundamental to cognition (Harnad, 1987). As outlined earlier, for the purpose of cognitive economy we categorise the world around us as it allows us to treat different instances as equivalent. The cognitive result of this process is said to manifest itself in categorical perception (CP): differences between instances within categories appear smaller, whereas across category differences are perceptually stretched. In experimental terms, CP is defined by faster and more reliable across-category discrimination compared to within-category discrimination. Crucially, this is the case even when distances between stimuli within a category and across a category boundary are held equal suggesting that a qualitative change occurs across the category boundary facilitating discrimination.

The first reports of categorical perception were of speech perception (Liberman, Harris, Hoffman & Griffith, 1957). A continuum of speech sounds (e.g. from /ba/ to /da/) varying in equal steps between phonemes, was shown to be more discriminable for sounds straddling the category boundary than phonemes belonging to the same category (Liberman et al., 1957; Pisoni, 1973). Early theories of CP in this domain
emphasised the uniqueness of speech perception for CP. Speech perception was argued to be affected by speech production, thus how something is heard is influenced by the way it is produced. The discontinuities in /pa/, /da/, /ba/, /ga/ are thought to be due to the discontinuities required to produce them (Harnad, 1987). Liberman et al. suggested that sounds are perceived by inferring how they would have had to be pronounced: the motor theory of perception.

Later research found evidence for CP in other domains of perception, including visual perception. CP has been found in the perception of line length (Tajfel & Wlikes, 1963), facial expressions (Calder, Young, Perret, Etcoff, Seth & Rowland, 1996; Young, Rowland, Calder, Etcoff, Seth & Perret, 1997), facial actions (Campbell, Woll, Benson & Wallace, 1999) and faces of different species (Campbell, Pascalis, Coleman, Wallace & Benson, 1997). Furthermore, using morphed images Newell and Bülthoff (2002) have presented evidence for CP in the perception of familiar objects such as bottles and glasses. CP has also been found across a range of colour perception tasks (Bornstein & Korda, 1984; Boynton, Fargo, Olson & Smallman, 1989; Özgen & Davies, 2002; Pilling, Wiggett, Özgen & Davies, 2003; Roberson & Davidoff, 2000; Uchikawa & Shinoda, 1996). The results of these studies support the idea that the brain somehow divides perceptually similar stimuli into qualitatively different categories to allow for more efficient processing (see Harnad, 1987 for a review). In addition to the early motor theory of CP, three main classes of theories are found in the literature: naturalistic theories, perceptual change theories, and labelling theories (Pilling, 2001). These are discussed below.

1.3.1. Naturalistic theories

Naturalistic theories presume CP to be the result of innate, physiological processes that produce discontinuities in the perceptual representation of stimulus dimensions. According to this theory, uniform physical stimulus dimensions are mapped onto perceptual representations, and this results in local discrimination maxima around which the category boundaries form. Naturalistic theories, therefore, propose that CP is both inborn and perceptually mediated. Support for innate CP comes from studies on pre-linguistic infants and animals (Bornstein & Marks, 1982). Bornstein, Kessen &
Weiskopf (1976) have shown that 4-month old infants perceive colours in a categorical fashion. After habituating the infants to a certain colour, Bomstein et al. found that the babies looked more at the novel stimulus if it came from a different adult category than if it was taken from the same adult category. There are, however, some problems with Bomstein et al.'s study. The stimulus differences, for example, were equated in wavelength and therefore variations in discrimination thresholds were not taken into account. Franklin and Davies (2003) have recently addressed these issues and, using reflective colours instead of lights, replicated Bomstein et al.'s study. Franklin and Davies' results are fully consistent with Bomstein et al.'s findings and could therefore be taken to support a naturalistic account of CP.

These findings suggest that colour categorisation is primarily driven by biological processes and that language is not essential for dividing up the spectrum. Naturalistic theories of CP are in line with universality accounts of colour categorisation. Studies showing that animals and infants respond in a categorical fashion to colours have been taken to support the claim that there are universally salient colour categories even for speakers that do not encode all basic colour categories in their language. As Bornstein and Marks (1982) argue, “if infants are born with a capacity to see different colors, it should not matter that the baby is a Kwakiutl of the Pacific Northwest, a Hanunoo in the Philippines, or an American Indian: all should see red as distinct from blue, blue as separate from yellow, and so forth” (p.69). Franklin and Davies, however, argue that it is possible that the visual system is initially tuned to certain points in colour space at which boundaries are perceived, but that later on language re-shapes the perceived colour space by emphasising and drawing attention to certain colour discrimination and not others.

Naturalistic theories cannot give a full account of CP. The theory cannot explain the flexibility and variability of CP. CP for musical intervals has been shown to be greater for trained musicians than for non-musicians (Zatorre & Halpern, 1979). Furthermore, CP can emerge through learning new categories (Goldstone, 1994, Linvingston, Andrews and Harnad, 1998). The results from CP for familiar objects mentioned above (Newell & Bülthoff, 2002) can also not be explained by naturalistic theories of CP as glasses, bottles and vases are clearly not naturally occurring categories.
In colour perception, CP differs for speakers of different languages (Davies & Corbett, 1997; Kay & Kempton, 1984) and Özgen and Davies (2002) have recently shown that colour CP can emerge by training people to split a pre-existing category (e.g. blue) into two new categories (light and dark blue). These findings suggest that even if colour categories are to a certain extent hardwired, they are mutable.

1.3.2. Perceptual change theories

Goldstone, Lippa and Shiffrin (2001) have shown that learning new categories of faces can lead to changes in the internal representation of similarity space. Livingston, Andrews and Harnad (1998) report similar findings in a task using artificial categories (figures resembling micro-organisms), similarity ratings after training showed within category stimuli were rated as more similar. Perceptual change theories postulate that learning a new category either through massed practice or in learning to distinguish by name, results in a representative change of the stimulus dimension. Hence, cross-category differences appear to be more different, within category differences appear more similar. Özgen and Davies (2002) used this method to study categorical colour perception. After training discriminations of colours that straddled the new boundary were more accurate than discriminations within the new category.

Perceptual change theories are similar to naturalistic theories as both propose CP to be due to perceptual discontinuities of similarity space. However, the theories differ in that the perceptual change theories assume CP to be acquired rather than innate. Evidence from the perceptual learning literature supports this position (see Goldstone, 1998 for a review). The problem with perceptual change theories is that, although acquired CP effects have been found in the short term (Goldstone, 1994; Özgen & Davies, 2002), no attempt has been made to test for evidence of long term changes in the representation of similarity space.
1.3.3. Language theories of CP

Labelling, or direct language, theories of CP (Kay & Kempton, 1984; Roberson & Davidoff, 2000; Rosen & Howell, 1987) suggest that discriminations between stimuli are made, at least in part, by comparisons of verbal labels. The discrimination of stimuli that straddle a category boundary is advantaged since stimuli belonging to separate categories are given different labels (e.g. blue-green). Within category discriminations, however, would be less accurate since stimuli have the same label (e.g. blue-blue). Goldstone et al. (2001) call this process a 'strategic judgement bias'; category names are used as a cue to the similarity of stimuli.

This idea is consistent with dual-code theories of category representations. For instance, according to Paivio (1978, 1986) stimuli can be represented by two independent codes: a visual code and a verbal code. Furthermore, recognition memory is improved if the stimulus is encoded at both levels. Also, in colour perception, studies have shown that colours can be encoded both verbally and visually (Schooler & Engstler-Schooler, 1990, Schooler, Fiore & Brandimonte, 1997), however Schooler and Engstler-Schooler also showed that verbal codes did not always improve memory accuracy. Pilling et al. (2003) have suggested that for colour, within category judgements may be primarily based on the comparison of perceptual codes, whereas cross-category comparisons are likely to be based on the comparison of visual and verbal codes. If all judgements were based entirely on the comparison of verbal labels, within category discriminations should be at chance level. This is not the case (Pilling et al., 2003; Roberson & Davidoff, 2000). This is in line with Goldstone et al.'s (2001) findings that strategy judgement biases (using category name codes) affect mostly cross-category comparisons, whereas representational changes were found to affect within category judgements. Schooler and colleagues’ argument that verbal labels do not necessarily facilitate memory also supports this as their findings were based on only within category colour discriminations.

There are also problems with language accounts of CP. Discrimination of stimuli is not always best at identification boundaries, as Calder et al. (1996) have shown for facial expressions. Also, differences in naming do not always lead to differences in
CP (Malt, Sloman, Gennari, Shi & Wang, 1999). Furthermore, no evidence for language effects on CP were found in a comparison of English and Russian speakers (Davies, Corbett, Laws, McGurk, Moss & Smith, 1991; Laws, Davies & Andrews, 1995). The Russian language has two separate terms for blue. However, this did not affect similarity judgements. Labelling theories can also not account for category effects found in animal and infant perception (Bornstein et al., 1976; Bornstein & Marks, 1982). Goldstone et al.’s (2001) experiment also provides evidence against a direct labelling account of CP. The perceived similarity between faces was tested with two stimuli from a trained category and a third neutral (i.e. untrained) stimulus. The stimuli were perceived as more similar after category training. This, so argued Goldstone et al., could not be due to labelling since the neutral stimulus was never presented during training and could thus have not been labelled. However, in regard to colour CP, Roberson and Davidoff (2000) and Winaver, Witthoft, Wu and Boroditsky (2003) have presented strong evidence in favour of a direct language account of CP (these will be discussed in more detail in Chapter 2).

The focus of the experiments presented in this thesis is on CP since testing categorical colour perception should allow a test of the nature and representation of colour categories in the information processing system. Depending on whether CP is truly perceptual or due to higher level, cognitive or linguistic processes, certain experimental manipulations should affect CP and so help determine the nature and locus of CP.

1.4. Colour categorisation

Terms such as red, green, or pink refer to the colour appearance of objects. Thus, in the sense that it refers to perceptual properties, colour is a perceptual category. According to Rosch (1973), ‘colour’ is the super-ordinate category; ‘red’, ‘green’, ‘pink’ etc. are basic level categories defined by a prototype structure. The prototypical colours are the perceptually most salient examples of each colour category. A fundamental question is where in the perceptual system and how do these categories emerge?
1.4.1. The physiological basis of colour perception

Colour vision is dependent on light reaching the retina in the eye. The human eye is sensitive to electromagnetic wavelengths in the range of about 380 to 750nm (Boynton, 1979). The shortest wavelength that humans can perceive appears violet. At about 470nm the reddish component disappears and the light is perceived as unique blue. As the wavelengths increase, the green component becomes more visible until unique green is perceived at about 510nm. Unique yellow is perceived at about 575nm; the longest visible wavelength of the spectrum is perceived as red. Light has certain properties which partly determine the three dimensions of colour: hue, brightness and saturation. Hue is a function of the dominant wavelength reaching the eye, brightness refers to the intensity or amplitude of the wavelength and saturation refers to the purity of the dominant wavelength.

The first to propose a trichromatic theory of colour vision was Young in 1802 (Wasserman, 1978). Young thought it unlikely that the retina contained an infinite number of colour sensitive particles and hence the number had to be limited. He suggested there may be three. Modern experimental methods have verified Young’s line of reasoning (later known as the Young-Helmholtz Theory) showing that the human retina contains three different photopigments that are the basis of colour vision (Boynton, 1979; Wooten & Miller, 1997). The cone receptors in the retina are the first stage in colour processing; each type of cone has a different peak spectral sensitivity. They either peak in the short-, middle-, or long-wavelength regions of the visible spectrum (Boynton, 1988).

The nineteenth century physiologist Hering challenged the trichromatic theory as an adequate explanation of colour perception on the grounds that it did not explain why, when mixing a yellowish red and a yellowish green, the resulting sensation was unique yellow (Wooten & Miller, 1997). Hering’s ideas were largely born out of the subjective appearance of the spectrum’s four primary hues: blue, yellow, red, and green and the notion that it does not seem possible to see or imagine a yellowish blue or a greenish red. Thus, he argued these hues must be primary, indivisible hue sensations which are known to him and everyone else by common experience. He
speculated that this was due to an ‘opponent colour’ process, yellow-blue and green-red being manifestations of single physiological processes.

James and Hurvich (1955, in Wasserman, 1978) quantified the opponent process theory psychophysically. Using lights of the four unique hues, they measured the amounts of certain lights needed to cancel out opposing hues. For example an orange light at 600nm appears to contain both red and yellow components. It is possible to cancel the red by using a green cancellation stimulus, thus leaving a pure yellow light.

Following on from hue cancellation studies, quantitative methods for describing colour appearance were used. Boynton and Gordon (1965) presented monochromatic lights and subjects had to judge the appearance using only ‘blue’, ‘yellow’, ‘red’, or ‘green’. Responses could either be given with just one colour word or a combination of two. Thus, if unique blue was presented, the response was blue, if a greenish blue was presented, the response was blue, green (in that order). The results support the opponent colour model in that responses were never blue, yellow or red, green (or vice versa). Stemheim and Boynton (1966) reported a similar experiment in which observers also assigned relative proportions of hues. Thus, when presented with a 490nm stimulus, it might have been described as 50% blue, 50% green (Wooten & Miller, 1997). Stemheim and Boynton’s aim was to assess how elemental certain hues were. For a colour term and hence a hue to be denoted elemental, the term had to be both sufficient and necessary to describe the colour.

DeValois and Jacobs (1968) and DeValois and DeValois (1975) presented physiological evidence in support of opponent processes in colour vision. They identified cells in the lateral geniculate nuclei of the macaque monkey that responded in an opponent manner to stimulation by light in the receptive field. They proposed that colour information was transformed from the responses at the level of the cones, where responses increased with increasing intensity of the stimulation, to an ‘opponent system’. In the opponent stage, some wavelengths were found to cause an increase in a given retinal signal and other wavelengths a decrease. The cells isolated showed excitation to red wavelengths and inhibition to green (+R-G) and excitation to yellow and inhibition to blue (+Y-B), as well as +G-R and +B-Y cells. A revision of the opponency theory was later published by DeValois and DeValois (1993).
acknowledging that the proposed axes were incorrect (see also Jameson and D’Andrade, 1997).

It is nowadays generally accepted that both trichromacy and opponent processes are needed to account for the different stages of colour vision. Indeed, Helmholtz himself had proven the compatibility of the two theories by showing that a simple transformation could change the three receptor outputs into difference signals and one additive signal (Wasserman, 1978, pp.91-93). It has been suggested that opponent processes are based on the relationship between excitatory and inhibitory signals which are in turn initiated by the three types of receptors (Wooten & Miller, 1997).

At the cortical level of analysis, the question is to what extent is colour analysed separately from other aspects of the visual information such as form and motion. Zeki (1980; 1983) identified two adjacent regions in the prestriate cortex where colour specific cells were more often found than in other areas. Thus, Zeki argued that these areas (V4 and V4A) were specific to the analysis of colour and stated that there was “little difficulty in equating the perception of colour with the response of the individual colour coded cells” (1980, p.418).

However, as Boynton (1988) has pointed out there are some problems with drawing conclusions about human colour perception from Zeki’s study. As the monkeys were anesthetized they were not only feeling no pain but the monkeys were also seeing no colours. Saunders and van Brakel (1997) point out that this highlights the importance of making the distinction between passive wavelength responders and active colour see-ers. It has also been argued that the recordings of opponent cells in the LGN do not support the yellow/blue – red/green opponency hypothesis (Jameson and D’Andrade, 1997). Jameson and D’Andrade argue that the results are in fact in line with the axis being cherry/teal and chartreuse/violet, a point consistent with DeValois and DeValois’ (1993) revision of opponency theory. They further make the claim that the principle justification for the yellow/blue and red/green opponent process model is that we find it “virtually impossible to think of cancelling or scaling other opposite colours”, thus the main support for the four-basic-hue model is its “intuitive appeal and historical continuity” (p.311).
Although the exact processes are as yet not fully understood, the physiology of the perceptual system is no doubt the basis of colour vision. Without these processes our world would appear colourless. However, trichromatic processes and opponent processes are located very early on in visual processing, they are both retinal processes. Between the retina and the final experience of colour, colour categories appear, but the mechanism is not yet known. Abramov (1997) has pointed out that physiology in itself cannot tell us how visual sensory information is transformed into a categorical percept. Higher level processes are also likely be involved in colour categorisation.

1.4.2. The influence of language on colour categorisation

Colour categories are lexical as well as perceptual categories; we use a set of labels to designate certain colour experiences. The relationship between colour categories and colour language has been studied extensively and has been the principal testing ground in the linguistic relativity versus cognitive universals debate. The visible spectrum of wavelength is a continuum, the way we ‘cut up’ the spectrum into categories, however, varies significantly across languages. These differences in colour language are seen as affording a test of whether language affects colour perception. Thus, the question at the heart of this debate is whether our lexical colour categories merely describe innate, biological colour sensations, or whether language plays an important role by driving the segmentation of the colour spectrum into categories.

1.4.2.1. Linguistic relativity

Does the language we speak influence the way we see the world? The linguistic relativity, or Whorfian, hypothesis suggests it does. Whorf is commonly credited with the formal expression of the hypothesis (see Roberson, 1998 for a review). Whorf famously stated that:
“We dissect nature along the lines laid down by our native language. The categories and types we isolate from the world of phenomena we do not find there because they stare every observer in the face, on the contrary, the world is in a kaleidoscopic flux of impressions which has to be organised by our minds – and this means by the linguistic systems of our minds” (Whorf, 1956, pp. 212-213).

Whorf argued that our categories and concepts are not out there in the world, they are not pre-given. Rather, we use language to structure and make sense of the world. The structure and categories will depend on the language we speak. On this idea, Whorf’s writings contain two propositions: one of linguistic relativity and one of linguistic determinism. These propositions or interpretations differ in the role they ascribe to language in determining our reality, as summarised by Brown (1976). Linguistic relativity is defined as the idea that structural differences between language systems are paralleled by non-linguistic cognitive differences in native speakers of the languages in question. Linguistic determinism makes a stronger claim in that the structure of a language is proposed to strongly “influence or fully determine the world-view” (p.128) of the speaker of that particular language. The widespread variation in how languages divide up the colour spectrum means that the colour domain is seen as an ideal testing ground of the linguistic relativity hypothesis. The rationale is that, if language influences thought, the seemingly arbitrary linguistic divisions of the colour spectrum should influence the way colours are perceived.

Amongst the earliest investigations of the relationship between colour language and the perception of colour was that conducted by Brown and Lennenberg (1954). They proposed that if language was paralleled by non-linguistic cognitive differences, the availability of verbal labels should affect the ease with which the colour of a stimulus could be remembered. For English speakers, colours that were easily named (high codability) were more likely to be remembered correctly than colours that were less codable. However, a later cross-cultural comparison of English and Zuni (a native American language) produced ambiguous results (see Pilling, 2001).

Lantz and Stefflre (1964) proposed a refined method for measuring the effect of language on colour cognition. They used communication accuracy as a measure of
nameability which took the context in which colours were used in a language into account. Communication accuracy was defined as the ease with which a target colour could be described to another person so that he or she could select it from an array of colours. Lantz and Stefflre showed this to be a better predictor of English memory than codability. Furthermore, in a cross-cultural study comparing native Spanish and Mayan speakers, communication accuracy correlated highest with colour memory within a language (Stefflre, Castillo & Morley, 1966).

1.4.2.2. Cognitive universals

The 1960s saw the emergence of a strong universalist trend in psychology, primarily driven by the rise of cognitive psychology and cognitive science (Hunt & Agnoli, 1991). For example in the field of language acquisition, Chomsky (1968) argued that the grammar of languages did not differ arbitrarily and hence proposed that universal principles governed the grammatical structures of all language. In 1969 Berlin and Kay published ‘Basic Color Terms: Their universality and evolution’. The publication of this seminal work is generally regarded as signifying a shift from a belief in relativity to universals in colour categorisation. The main argument in Berlin and Kay’s work was that the emergence of colour categories across languages was governed by universal principles.

Berlin and Kay

Berlin and Kay studied speakers of twenty languages and searched the anthropological literature of a further 78 languages. The colour tasks conducted on the speakers of twenty languages included choices of best examples, or foci, of colour terms across languages. Using an array of Munsell colours, respondents chose the best examples of their colour categories and also determined where the boundaries between categories were. The results showed that, although the distribution of boundaries differed, there was widespread agreement on the position of focal colours. A model was proposed in which basic colour terms emerged and entered languages in a fixed evolutionary order.
The Berlin and Kay model is based on eleven basic colour terms and comprises seven stages of colour language development. According to the model, languages with only two terms will divide the colour spectrum roughly into black and white (or dark and light) (Stage I); language with three colour terms will include the term red (Stage II); languages with four terms should add either yellow or green (Stage III); languages with five terms will add whichever term was not added at the previous stage (Stage IV). At the next stage of the sequence blue is added (Stage V); followed by brown (Stage VI). At the final stage, orange, pink and purple are added in no particular order (Stage VII). The term ‘grey’ is a wildcard term and can appear at any of the stages.

Berlin and Kay argued that these 11 colour terms were basic terms and that the categories they referred to were basic categories. Basicness was defined by following criteria: basic terms were those colour terms that were general and salient. A term was classed as general if it could be applied to a diverse range of objects and its meaning was not included in that of another term. A term was considered salient if it was easily elicitable and occurred in the idiolect of most speakers of a language. Furthermore, it was used consistently and with a high consensus.

Berlin and Kay’s original formulations and results were hugely influential throughout psychology, linguistic and anthropology. However, the theory has also been met with much criticism over the years (see Simpson, 1991). One of the main criticisms has been that the data was collected not on monolingual speakers of the languages tested but on bilingual speakers living in the San Francisco Bay Area in California. Furthermore, for some languages there was only one respondent, tested three times (Saunders & van Brakel, 1988). A second line of criticism has focused on the problem of defining basic colour terms (Crawford, 1982; Dedrick, 1996, 1998).

Kay and MacDaniel

The theory was later revised by Kay and MacDaniel (1978); using fuzzy set theory to model colour categorisation they addressed some of the shortcomings of Berlin and Kay’s arguments. Kay and MacDaniel’s theory linked basic linguistic categories to fundamental neural responses (FNR). The proposition was made that colour categories were directly based on FNRs which in turn were closely defined by the physiology of the visual system. The basic colour categories defined by Berlin and
Kay were subdivided into three types of categories. Primary colour categories (red, green, blue, yellow, black and white) were said to emerge directly from FNRs of opponent mechanisms (DeValois & Jacobs, 1968; DeValois, 1975). The second type of fuzzy set (fuzzy union) was argued to be formed by the combination of the six primary colour, e.g. blue and green. These categories were labelled composite categories and are found in smaller colour lexica with less than six terms (Wooten & Miller, 1997). The third type of category (fuzzy intersection) were so-called derived categories that were compositions between elemental colours, e.g. red AND yellow.

Kay and MacDaniel’s theory has also not gone without criticism, foremost it has been argued that there is no justification for the direct linking of primary colours to neurophysiological responses as the response of the opponent cells in the LGN are not direct neural correlates of psychological colour experience. Jameson and D’Andrade (1997) argued that the fundamental chromatic axis (here the maximum excitation of cells) are not yellow-blue and red-green, but rather intermediate colours (see § 1.4.1). Abramov (1997) as well as Mollon (1992) have pointed out that there is as yet no direct evidence that the hue sensations of red, green, blue and yellow correspond directly to colour coded cells in the visual systems. In fact, it has been argued that no such colour-coded cells exist (Saunders & van Brakel, 1997). Hence, Kay and MacDaniel’s claims about the underlying physiology of colour categories remains speculative. Despite these criticisms, their work is often cited as proving a strong correlation between physiology and colour naming (see Saunders & van Brakel, 1997 for a critique of the basic colour term tradition).

Rosch
Support for the proposition that colour experience is governed by cognitive universals also came from studies using psychological measures, foremost with the publication of Rosch’s pioneering cross-cultural investigations (Rosch, 1972; 1973). Rosch published a series of studies in which she compared the performance of American English speakers with the Dani of New Guinea (now Indonesia) on a number of colour tasks. The main findings and interpretations were strongly supportive of the notion of cognitive universals: the language spoken and the number of colour terms a language has, does not affect colour perception or colour cognition.
The Dani appeared to have only two colours terms that fulfilled the Berlin and Kay criteria of basicness: *mili* corresponding roughly to dark or cool colours, and *mola* corresponding to light or warm colours. On a recognition memory task in which a memory stimulus had to be identified in an array of Munsell colours, Rosch found that both English and Dani remembered focal colours better than non-focal colours. The focal colours in question corresponded to the best example of each basic colour category as determined by Berlin and Kay (1969). Focal colours were remembered better independent of whether the categories were encoded in a given language. Furthermore, the Dani found it easier to learn arbitrary names for focal than for non-focal colours (Rosch, 1972) and categories for which a focal colour defined the category centre were easier learned than categories with non-focal central stimuli (Rosch, 1973). Thus, Rosch argued that the focal colours, the prototype of each colour category, were universally salient, independently of the language spoken. The overriding claim was that language does not determine perception, but rather perception determines language.

Rosch's work has been taken by many as definite proof of the universality of basic colour categories and the psychological salience of focal colours (see Roberson, 1998). However, the issue of cognitive universals and/or language effects driving colour categories is still debated today. The universalist research tradition and the conclusions drawn have more recently faced strong criticisms (Lucy, 1992; 1997a; 1997b; Ratner, 1989; Saunders & van Brakel, 1988; 1989; 1997; Simpson, 1991). Critics of the theories have pointed out some potentially serious problems with the design and interpretation of Rosch's studies. Lucy and Shweder (1979) have argued that the array of Munsell colours used was biased towards focal colours. Rosch reported that the Dani would not learn nonsense words to describe the basic colour terms which led Saunders and van Brakel (1988) to ask "why wouldn't they learn words that name salient, natural, ideal-types of which prototypes or reference points were presented to them" (p.365). No appropriate controls were used on the memory tasks as no categories were tested that the Dani but not the English were familiar with (Roberson, Davies & Davidoff, 2000). It has also been pointed out that much of the data collected was discarded since it did not fit the universalist basic colour terms framework (Saunders & van Brakel, 1988) and that the data were coded to fit the
theoretical model (Saunders, 1995). Overall, as Davies (1997) has pointed out, “the inferential load that Rosch’s work is required to support is too heavy” (p. 186).

Despite Rosch’s work being both widely cited by ‘universalists’ and much criticised by ‘relativists’, no attempt had been made to replicate the findings until Roberson, Davies and Davidoff (2000) reported a comparison of English and Berinmo speakers. The Berinmo are a neighbouring group of the Dani and are reported to have five basic colour terms by the Berlin and Kay criteria. However, Roberson et al. found that applying the Berinmo colour terms to the Munsell array was not straightforward as there was much disagreement between participants in defining the best example of categories. Furthermore, the recognition memory task did not replicate Rosch’s finding. Each group was better at remembering the colours that were easily and consistently named in their own language. Also, the errors made were consistent with the naming patterns in each language. Thus, Roberson et al. presented evidence in support of a relationship between language and colour perception across languages and suggest that the inferences and conclusions widely made based on Berlin and Kay and Rosch’s work may not be entirely justified.

Whereas many of the studies in the field have focused on mapping the colour space of colour vocabularies (Greenfield, 1986; Iijima, Wenning & Zollinger, 1982; Kay, Berlin & Merrifield, 1991; Kuschel & Monberg, 1974; Lin, Luo, MacDonald & Tarrant, 2001; MacLaury, 1987; 1991; 1992; Zollinger, 1987), a number of studies have, like Roberson et al., tested the effect of language differences on various measures such as similarity judgements and recognition memory (Davies, Sowden, Jerrett & Corbett, 1998; Davies & Corbett, 1997; Kay & Kempton, 1984). The linguistic relativity hypothesis has also been tested in other areas. The way spatial relations are described has been found to vary cross-linguistically. Western languages tend to use an egocentric system in which the location of objects is described relative to the body (e.g. left/right and front/back). There are languages, however, that use absolute co-ordinates systems. Tzeltal, a Mayan language, for example describes the location of objects relative to the mountainous environment (uphill/downhill; see Brown & Levinson, 1993). A number of studies have emphasised the possibility of corresponding cognitive variability in the way space and spatial relations are represented (Bowerman & Choi, 2003; Brown & Levinson, 1993; Levinson, 1996;
The question of linguistic relativity has also been addressed by looking at variations in the grammatical structures of language. It has recently been argued that grammatical distinctions that have to be made in reporting events and objects in some languages but not in others, shape the way speakers of different languages internally represent the world (Boroditsky, Schmidt & Phillips, 2003).

Categorical colour perception is clearly relevant to this debate as CP has implications for linguistic relativity. If colour categories are based on low-level, hard-wired, innate processes (Bornstein & Marks, 1982), CP should be found for speakers of languages that do not designate all eleven categories. If, however, CP is driven by language, either through perceptual learning as argued by Özgen and Davies (2002) or if CP is based on direct language effects (Roberson & Davidoff, 2000), cross-cultural differences in CP should be found. CP should only be found across boundaries of existing linguistic categories. Note, however, that these two language accounts of CP have very different outcomes. In a perceptual learning account of CP, the effect is actually perceptual: language has warped perceptual colour space making cross-category discrimination easier than within category discrimination. In a direct language account of CP, the effect of language is an on-line effect of stimulus labelling during the task.

1.5. Summary

Categorisation is a fundamental cognitive process. The most prominent model of categorisation is prototype theory (Rosch, 1978) which suggests that category boundaries are fuzzy and hence category membership is graded rather than all-or-none. Categorical perception (CP) is an emergent property of categorisation. Given equally spaced stimulus dimensions, CP is defined by better cross-category discrimination compared to within category discrimination. CP could be due to a warping of perceptual space around category boundaries and could be either innate or emerge through perceptual learning. Alternatively, CP could be due to direct language effects.
Colour perception is based on physiological processes. However, both trichromacy and opponent processes happen very early on in visual processing. Between the retina and the final experience of colour, colour categories emerge. Thus, colour perception must involve higher level processes and it has been argued that language may shape our perceptual colour categories. Colour categories are, according to Rosch, basic level categories, defined by a prototype structure; the best example of each colour category is psychologically salient. Whereas many studies in the Berlin and Kay tradition have stressed the fact that there are considerable cross-cultural similarities in colour language, recent comparison of speakers of different languages have suggested that language is an important factor influencing colour cognition (Roberson, Davies & Davidoff, 2000).

Studies on pre-linguistic infants (Bornstein, et al., 1976; Franklin & Davies, 2003) have shown that 4-month old babies perceive colours categorically. This lends support to the idea that CP is hard-wired. However, there is also evidence to suggest that CP emerges when new colour categories are learned (Özgen & Davies, 2002). Furthermore, there is evidence for language effects on colour CP (Roberson & Davidoff, 2000). Hence, questions about the relationship between the biological and the linguistic levels of colour categorisation and the origin and nature of CP remain.

1.6. Overview of the chapters

Roberson and Davidoff (2000) argued that CP is a misnomer by showing that the effect was eliminated if a secondary verbal task was presented along with a delayed colour discrimination task. They proposed that colour CP was not a perceptual, but a direct language effect. Pilling et al. (2003) found further support for a language theory of CP. This account is further tested in Chapter 2. Thus, the first question addressed in this thesis is whether CP is a direct effect of stimulus naming.

The second question addresses the issue of the use of perceptual and name codes more directly by asking what happens when the use of colour name codes in colour discrimination is prevented. Introducing Stroop interference should have no effects if the tasks are done or can be done on the basis of perceptual codes alone. This will be tested in Chapter 3. A third question arises from work done on category prototype
effects (Huttenlocher, Hedges & Duncan, 1991; Huttenlocher, Hedges & Vevea, 2000) as it raises the possibility that CP is neither a perceptual nor a language, but rather a memory effect. This is relevant here since experiments testing CP often have a memory component. Chapter 4 tests the possibility that CP is a memory effect due to a shift towards prototype.

CP has implications for linguistic relativity. If CP emerges in accordance with linguistic category boundaries, either through perceptual learning or through direct influences of language such as naming, it should only be found for speakers of languages that make the categorical distinctions. If, however, our colour categories are innate and hardwired, CP should be found independent of linguistic category structures. The issue of cross-cultural relativity is addressed in Chapter 5.
Chapter Two

The effect of verbal and visual interference on the categorical perception of colour

2.1. General Introduction

Categorical perception is defined by better between category compared to within category discrimination. In the case of categorical colour perception this implies that even when perceptual distances in a colour space are held equal, discrimination of stimuli that straddle a category boundary (for example blue-green) will be superior to discrimination of within category stimuli (e.g. two blue stimuli). Colour CP has been found in tasks such as same-different judgements (Bornstein & Korda, 1984; Boynton, Fargo, Olson & Smallman, 1989), recognition memory (Uchikawa & Shinoda, 1996), similarity judgements (Laws, Davies & Andrews, 1995), and delayed colour discriminations (Roberson & Davidoff, 2000). Boynton et al. (1989), for example, conducted a same-different memory task in which the colour of a stimulus had to be remembered for 10 seconds and then compared to a second test colour. The results showed that the greater the degree of categorical difference between two colours, the less likely it was that participants incorrectly responded ‘same’.

The majority of studies mentioned above are taken to support the perceptual basis of categorical effects in colour perception. Uchikawa and Boynton (1987) for example suggest that even though the colour categories and hence CP found in conjunction with colour tasks could be influenced by the colour names, these category names describe fundamental colour sensations and are dependent upon an underlying physiology. Boynton et al. (1989) also concede that in a colour memory task “one may remember the name of a color category, rather than to retain an image of the color itself” (p. 230), however they conclude by stating that “categorization has reduced the size of perceived differences” (p.234). Thus, although it is often acknowledged that colour categorisation goes hand in hand with naming, the possibility that CP may be due to naming and matching labels, is not explicitly
considered by these authors: categorisation may be verbal; the resulting CP effect, however, is due to warping of similarity space.

Furthermore, in the literature colour perception is often given as an example of 'innate' categorical perception (e.g. Pevtzow & Harnad, 1997) and studies showing 4-month old infants perceive colours in a categorical way (Bornstein, 1987, Franklin & Davies, 2003) are taken to support this account.

The recent work by Roberson and Davidoff (2000), however, questions the perceptual basis of CP and raises the possibility that categorical perception effects could in fact be direct effects of language. Since most studies reporting colour CP include a memory component, it is conceivable that the effects found are not necessarily or exclusively located at the perceptual encoding stage. To study where CP is located within the information processing system, Roberson and Davidoff conducted a set of dual task experiments, combining delayed colour discrimination experiments with verbal and visual secondary tasks. Within and cross-category discrimination accuracy of stimulus pairs in the blue-green colour region was measured with either no interference, visual or verbal interference presented during the inter-stimulus interval. The design and results (Roberson & Davidoff, Experiment 2) are presented in Figure 2.1.

Fig. 2.1. Diagram illustrating the task used by Roberson and Davidoff (2000). The target colour was presented for 1 second. In the 5 second inter-stimulus interval either no interference, verbal interference or visual interference was presented. The two test colours (target + distractor) were presented for 1 second (or until response was made).
The visual interference was in form of a multicoloured dot pattern presented between target and test presentation; in the verbal interference condition a list of words was presented on screen. The instructions in the visual interference condition were to track the line in the dot pattern; in the verbal condition the list of words had to be read aloud. The results showed that verbal interference selectively eliminated CP and hence the advantage of cross-category over within category discriminability was said to be based not on cross-category pairs being perceptually more discriminable but cross-category stimuli being labelled differently. If the target stimulus is labelled e.g. blue on a cross-category trial, one of the test stimuli is accordingly labelled blue, the other green. On a within category trial both test stimuli are labelled blue. Thus, if the task is done by matching target and test name codes, cross-category pairs are advantaged over within category pairs. This would result in CP. Pilling, Wiggett, Özgen and Davies (2003) replicated Roberson and Davidoff’s findings with more carefully matched interference tasks for both same-different and 2AFC tasks.

The effects of language on colour discriminability recently reported by Winaver, Witthoft, Wu and Boroditsky’s (2003) strongly support a verbal account of CP. The authors report a series of experiments investigating whether the advantage of cross-category over within category discrimination is also found when the memory component is removed by presenting the target and test stimuli simultaneously. The task was a 2AFC task in which the stimuli were presented in a triangular fashion; the target was presented at the top, the two test stimuli (target and distractor) below. Winaver et al. tested whether CP is found in such tasks, and, if so, whether verbal interference reduced CP. Furthermore, the authors investigated whether CP is only found in speakers of languages that have a certain verbal colour boundary. The results showed that cross-category discrimination was significantly more accurate than within category discrimination even with no memory component. Furthermore, verbal but not visual interference significantly reduced this advantage, and the comparison of English and Russian speakers showed that the CP effect may be specific to the language spoken\(^2\). When the task was changed to a same-different task (which was in

\(^2\)The comparison of Russian and English speakers is of interest since the Russian language has two separate blue terms. Hence, Winaver et al. tested for CP effects across this category boundary and found that only Russian speakers showed the advantage of cross-category over within category discrimination.
effect an edge detection task), verbal interference did not selectively affect CP. Hence, the authors argued that there was an online effect of verbal coding on colour discrimination; if the task was made simple enough, however, the effect of language was reduced.

Roberson and Davidoff (2000) proposed that the effect of verbal interference found to eliminate CP in their experiments was on name retention during the inter-stimulus interval. The implication is that the colour stimuli were encoded verbally and that the task was in effect a verbal short-term memory task; hence having to carry out an additional verbal task interfered with the primary task, whereas having to carry out and additional visual task did not. This conclusion is, of course, based on the idea that there are independent verbal and visual short-term memory systems, as proposed in the working memory model (Baddeley, 1986; 1992).

Before reporting a series of experiments designed to test the direct language account, the current chapter will give a brief summary of the working memory model, followed by a review of studies looking specifically at colour memory and the question of how colours are encoded. The aim is to integrate these concepts and findings and use them to evaluate the account of CP offered by Roberson and Davidoff (2000).

### 2.1.1 Working memory

The dominant model of short-term memory has evolved from a unitary short term system to a multicomponent working memory model (Baddeley, 1986; 1992) which provides both temporary storage and manipulation of information in short term memory. Working memory comprises three subsystems:\(^3\) the central executive, the visuo-spatial sketchpad and the phonological loop. The central executive is thought of as an attentional control system, the visuo-spatial sketchpad is responsible for manipulating visual images and the phonological loop is assumed to store and rehearse speech-based information. Dual-task approaches have been used to test the separability and relative independence of these subsystems. The rationale being that if

---

\(^3\) Though it should be mentioned that Baddeley (2002) recently postulated a fourth system, the episodic buffer.
ways are found to disrupt visual, but not verbal processes (and vice versa), it is possible to explore the relative contribution of the different subsystems to complex cognitive tasks.

Dual task studies have shown that two tasks can be carried out simultaneously in short-term memory as long as the tasks do not require the same resources, i.e. as long as they are carried out in different subsystems. Logie, Zucco and Baddeley (1990) looked at recognition memory span for visual matrix patterns and visually presented letter sequences. The two tasks were combined with a concurrent verbal arithmetic task or a task requiring the manipulation of visuo-spatial material. Results showed that span for matrix patterns was disrupted by the visuo-spatial task but not the secondary verbal arithmetic task, whereas the reverse was true for letter span. Baddeley (1992) further reported studies showing that chess players who were required to memorise complex chess positions were able to do this despite having to carry out a verbal task at the same time. However, when required to carry out an additional visual task their performance deteriorated markedly.

Thus, the usefulness of the working memory model for studying the categorical perception of colour can be said to lie in the separability of the visuo-spatial and phonological subsystems: depending on whether CP is a truly perceptual effect or a language effect, selective interference should be found from either additional visual or verbal tasks. However, it is also important to note that studying CP within a working memory framework necessarily entails adding a memory component to the discrimination task. Any advantage of cross-category over within category discriminations could therefore be a memory effect during retention rather than an effect of language or a perceptual effect at encoding. However, if it was a memory effect, the size of the effect should increase when the memory load (for example length of the inter-stimulus interval) is increased.
2.1.2 Verbal and perceptual codes in colour memory

A problem with studying colour CP within a model of separate memory sub-systems is that there are multiple ways of doing a colour memory task. The colour could be labelled and the colour name remembered, or the visual image of a colour presented could be retained. Or, indeed, both codes could be used. Therefore, it is not possible to state a priori which subsystem will be used to carry out a colour memory task. A further complication is that colours (or for that matter all easy-to-name visually presented material; see Brandimonte, Schooler & Gabbino, 1997) may be automatically labelled when seen. Thus, even if participants are instructed to retain the visual image of a colour, the possibility of verbal codes affecting recall cannot be ruled out. Similarly, when verbal interference is found to disrupt discrimination accuracy (as in Roberson & Davidoff, 2000) it cannot be conclude that visual codes played no role in the task. Thus, there may be several different processes involved in the encoding, recall and discrimination of colours in memory tasks and their relationship may not be straightforward.

The idea that stimuli are coded both verbally and visually is not new and not limited to colours. Paivio (1986) has argued that it is precisely the fact that humans can deal with and process language and non-verbal objects and events simultaneously that makes human cognition unique. Paivio’s dual code theory suggests that when a stimulus is presented, it can be coded either in form of a visual or a verbal code or both visually and verbally. However, these codes are independent representations. Dual code theory also predicts that the recall or recognition of stimuli will be more accurate if a stimulus is represented in a visual and a verbal code. The independence of physical codes and name codes was demonstrated in a series of categorisation tasks by Posner and colleagues (Posner & Keele, 1967; Posner, Boies, Eichelman & Taylor, 1969). Using letters they tested reaction times to physically identical pairs (AA) and categorically identical pairs (Aa). The task was to respond ‘same’ if the letters had the same name. Hence, both AA and Aa pairs required a ‘same’ response. However, the results showed that response to physically identical pairs was faster than to categorically identical pairs but this advantage was lost with an inter-stimulus delay of 1500 ms; showing that the availability of the physical code of visual stimuli decays.
Therefore, whether verbal or visual codes are used in colour discrimination tasks is likely to depend on the demands of the task. This point becomes apparent when examining the evidence for the use of verbal and/or visual codes in the colour memory literature. Luck and Vogel (1997) have argued that colours are processed in visual short-term memory. They investigated the capacity of the visual working memory using an array of one to twelve colours and short stimulus presentation times of (100 ms with 900 ms ISI). The task was a same-different discrimination task in which either all colours in the array stayed the same or one stimulus changed colour; the task was to judge whether the arrays were same or different. Performance on the task was close to perfect on arrays consisting of one to three colours, then declined systematically from four to twelve colours present. To show that this effect was a pure visual memory effect, Luck and Vogel added a verbal memory load to the task in one of the experimental conditions. On half the trials, two digits were presented before the target colours and had to be retained and repeated out loud at the end of the trial. This manipulation was found to have no effect on overall performance and hence Luck and Vogel concluded that the colours were encoded perceptually and the task was carried out in visual short-term memory store.

The comparison of the study outlined above with Roberson and Davidoff’s work suggests that a verbal strategy is much more likely to have been adopted in Roberson and Davidoff experiments since both target presentation time and the ISI were much longer. The short presentation times and the presence of more than one target colour in Luck and Vogel’s experiment make it highly unlikely that participants attempted a naming strategy. Of course, the fact that verbal interference was shown to have no effect on accuracy supports this. The comparison of these two studies suggests that it is task dependent whether colours are encoded verbally or perceptually, or rather, which code is retained across the ISI. This idea is in line with Posner and Keele’s (1967) finding that the visual code of letters decays after 1500 ms. It is also noteworthy that Luck and Vogel used a set of highly discriminable colours (red, blue, violet, green, yellow, black and white) whereas Roberson and Davidoff stimuli were a set of blue and green stimuli from the boundary region. This again suggests that a purely perceptual strategy is likely to be easier in tasks such as Luck and Vogel’s since the colours were perceptually more dissimilar.
Clearly task designs and specifically presentation times are important as some task may invite a verbal strategy more than others. The shorter the presentation time and the more colours present, the less likely it is done on the basis of verbal labels. This suggests that, given sufficient time, colours are encoded both verbally and visually. Support for a dual-coding account of colours can be found in the work done on proactive inhibition – the finding that recall from a common category decreases over successive trials. Release from proactive inhibition is commonly found when new material is presented that requires new and different coding. Allen (1983, 1990) found that release from proactive inhibition occurred when switching from colour names to colour, but not vice versa. This asymmetry suggests that colours are encoded more elaborately than colour names, in other words both verbally and visually. That verbal codes are beneficial when cross-category colour discriminations are required is clear since these discriminations can be done successfully by simply matching colour labels. However, it could also be argued that this is only the case when there is no uncertainty of category membership, hence when naming is accurate. This raises the question of how beneficial verbal codes are in colour memory and discrimination tasks.

2.1.3 Verbal Overshadowing

It is generally assumed that verbal encoding of visually presented material helps recall (Atkinson & Shiffrin, 1971; Maki & Schuler, 1980). Schooler and Engstler-Schooler (1990), however, have argued that the tendency to rely on a verbal code when having to memorise colours can have a negative effect since the verbal code can ‘overshadow’ the original visual image of the colour. In a colour memory task, the participants were explicitly instructed to either verbalise the colour presented to them (by writing a description of the colour) or to visualise the colour, hence to attempt to keep a visual image of the colour. The target colour and five distractor colours taken from the same category as the test colour were then presented. The results showed that recognition accuracy in the verbalisation condition was significantly worse than in the visualisation condition. Schooler and Engstler-Schooler argued that the verbalisation of a visually presented stimulus can lead to “the formation of a nonveridical verbally biased representation” (p.62) which prevents the use of the intact visual code. This suggests that colours may well be encoded both verbally and
visually, but the more precise representation is, according to Schooler and Engstler-Schooler, visual.

Whether or not a verbal colour code is going to facilitate or impede discrimination depends, of course, on the type of discrimination. The results presented by Schooler and Engstler-Schooler are based on within category 5AFC tasks. Roberson and Davidoff's work on CP in colour discrimination tasks, however, was a comparison of within and between category discriminations. Verbal labels are clearly of more use for between category discriminations than within category discriminations. However, the interesting point is that participants in the visual encoding condition in Schooler and Engstler-Schooler were significantly better than in the verbal encoding condition. Also, in Roberson and Davidoff within category judgements were unaffected by verbal interference. If verbal coding and hence overshadowing were prevented by having to do an additional verbal task (as Schooler and Engstler-Schooler may predict), within category discriminations should have been more accurate in this condition. However, this did not happen. Again effects seem to be task dependent.

Schooler and Engstler-Schooler's work shows that instructions can affect whether verbal or visual codes are used. Interestingly, in their colour memory task, recognition was most accurate in the control condition, when neither verbal nor visual instructions were given. This could be taken to suggest that a combination of verbal and visual codes is most effective for assisting recall. This is supported by Garro (1986) who asked participants to report which strategy they used in a colour short-term memory task and found that a mixed strategy, using both verbal and visual codes, was the most common. Brandimonte et al. (1997) suggested that easy-to-name visually presented material might be labelled automatically. Whether colours fall into the category of easy-to-name stimuli probably depends on whether the colours are focal colours and how similar the colours in a stimulus set are. It has been found that focal colours are easier to name and remember than non-focal colours (see Garro, 1986; Rosch, 1972). Hence, it is possible that it depends on the nature of the colour to be remembered and the task design (presentation times, ISI, number of colour stimuli) whether i) a verbal code is used, ii) a verbal code is indeed useful and iii) whether a visual image of the colour is easily retained in short term memory.
In terms of investigating CP within a working memory framework, an important question is which of the two systems is being used when CP in colour memory tasks is found? The literature suggests that colours are encoded both verbally and visually when given sufficient time. Hence, both short-term memory subsystems should or could be involved in colour discrimination tasks with a memory component. Within a dual-task model this would imply that verbal and visual interference tasks should have similar effects on colour discrimination and therefore colour CP.

However, the fact that Roberson and Davidoff and Pilling et al. found only verbal interference tasks to eliminate CP and the work done by Schooler and Engstler-Schooler (1990), Brandimonte et al. (1997) and Garro (1986) on visual and verbal codes, suggest that these types of tasks may be open to encoding strategies. This idea was tested in Experiments 1 and 2. Experiment 3 tested the resilience of name codes to verbal and visual interference directly, whereas Experiment 4 tested CP in a more perceptual task, a visual search task.
2.2 Experiment One

Effect of unblocked interference on discrimination: 2AFC task

2.2.1. Introduction

The aim of the current experiment was to test the direct language account of colour CP as put forward by Roberson and Davidoff (2000). According to this account, cross-category pairs are more accurately discriminated than equally spaced within category pairs because the colours are labelled and hence the differences in discriminability arise from verbal coding processes (cross category pairs are given different labels which enables easy discrimination). Therefore, CP is a top-down activity driven by linguistic knowledge and not perceptual input.

It is important to note that in all of Roberson and Davidoff’s (2000) and Pilling et al.’s (2003) colour discrimination experiments, reliable CP effects were found when no interference was present. Hence, the finding that cross-category discrimination is superior to within category discrimination is a reliable effect. The CP effects reported by Bornstein and Korda (1984), Boynton et al. (1989) and Uchikawa and Shinoda (1996) for same-different judgements and recognition memory further support this point. However, to test where the effect is located in the information processing system, Roberson and Davidoff and Pilling et al. carried out experiments investigating the effects of different kinds of verbal and visual interference on CP. The verbal interference used in Roberson and Davidoff consisted of presenting words that had to be read out aloud; the visual interference was a multicoloured dot pattern presented during the ISI. In Pilling et al.’s experiments both verbal and visual interference required a response. In the verbal interference condition participants had to judge whether two sequentially presented words rhymed or not; in the visual interference condition the task was to judge whether a shape (a 2-D outline drawing) fitted into a second shape (also presented sequentially). In all of these colour discrimination experiments, both visual and verbal interference resulted in the reduction of overall accuracy, but CP only persisted in the visual interference condition; verbal interference eliminated CP. An example of the pattern of results found in these
experiments is given in Figure 2.2. These findings support an account of CP as a direct language effect rather than a perceptual effect as outlined above.

![Figure 2.2](image)

**Figure 2.2.** Discrimination accuracy for within and cross-category comparisons in blocked interference conditions (Reproduced from Pilling et al., 2003, Experiment 1)

Both Roberson and Davidoff (2000) and Pilling et al. (2003, Experiment 1) used blocked interference in their experiments, all trials within a block were either no-interference, visual or verbal interference, and therefore type of interference was known to the participants on each trial. It is possible that this may have encouraged the use of different coding strategies. The colour memory literature (see § 2.1.2) suggests that colours are coded both perceptually and verbally and that participants can be instructed to encode either perceptually or verbally (Schooler and Engstler-Schooler, 1997). Thus, it is possible verbal interference reduced CP in Roberson and Davidoff and Pilling et al.'s studies because participants, knowing that they would have to carry out an additional verbal task, did not attempt to use a verbal code in the verbal interference condition. This would suggest that the effect was not a direct effect on working memory but rather an effect of task strategy.
Experiment 1 was a delayed colour discrimination task with no, verbal and visual interference conditions, but conditions here were unblocked, making type of interference unpredictable. This could have different effects depending on where the locus of CP is. Firstly, the uncertainty could mean that participants would expect a naming strategy to be ineffective and hence adopt a perceptual coding strategy. This should eliminate CP in all three conditions if the direct language account is true. Secondly, participants may try to use a naming strategy on all trials and if a name could be retained over the ISI, CP should be found in all conditions. If, however, verbal interference blocked name retention, as argued by Roberson and Davidoff (2000), CP should be found only in the no interference and visual interference conditions, as in the experiments where interference conditions were blocked.

Colour discrimination accuracy was measured using a 2AFC task; after presentation of the target followed by either no, verbal or visual interference, the target colour was presented alongside a distractor and the task was to correctly identify the target. The stimulus set comprised within category pairs and cross-category pairs. Using Posner and Keele's (1967) terminology, within category pairs are represented as A-a pairs: although they are not physically identical, they have the same name and are hence categorically identical. Cross-category pairs are neither physically nor categorically the same, hence cross-category pairs are represented as A-B pairs. Crucially, the distances between stimuli were matched in CIE L*u*v* space; the perceptual distances between stimuli in within category pairs was equal to the distance of cross-category stimuli.

The current experiment tests the effect of verbal and visual interference on CP in unblocked interference conditions. This tests the possibility that the effects of verbal interference found to eliminate CP in Roberson and Davidoff and Pilling et al. were not directly on name retention but rather indirectly on choice of strategy.
2.2.2. Preliminary naming study

Subjects
Twenty subjects took part in the naming study, ten male and ten female (age range 18-30). All were students at the University of Surrey.

Stimuli
According to Bornstein and Monroe (1980), the blue-green category boundary for English speakers is around Munsell hue 7.5BG. Thus, for the following experiments stimuli were chosen so that half of the stimuli fell into the blue category and the other half in the green category. To check the suitability of these stimuli a preliminary naming study was conducted. The stimuli used in this experiment consisted of a 4x4 matrix of 16 stimuli. Chroma was held constant at Munsell chroma 7.4; hue as well as value was varied on four levels. The four levels of hue were 1.44B; 8.76BG; 6.06BG and 3.46BG; the four levels of value were 5.40, 5.80, 6.20 and 6.60. The stimulus matrix is shown schematically in Figure 2.3. All horizontally adjacent stimuli in the matrix had a perceptual distance of nine units in CIE L*u*v*; the vertically adjacent stimuli differed by four units. Although adjacent stimuli were equidistant in CIE L*u*v*, the difference in Munsell steps was slightly smaller for cross-category steps compared to within category steps (see Appendix 1 for a description of the Munsell and CIE systems). Each colour stimulus was 55mm², subtending 6.3° of visual arc at an approximate viewing distance of 500mm.

The stimuli were displayed on a 15" Sony Trinitron monitor controlled by a PC compatible with a Diamond stealth graphics card. Stimuli were equally spaced in the CIE (1976) L*u*v* system. In order to accurately display colours, it was necessary to calibrate the monitor by computing the gamma functions of the monitor’s three cathode ray guns. This was done according to the procedure outlined by Travis (1991). The chromaticity of all colour stimuli was checked using a Minolta CS-100 Colorimeter. Color Science Library (CSL v.2.0) functions were used to give the computer RGB values relative to the measured gamma function of each of the three

---

4 The Munsell system was standardised by comparing colours against a neutral grey background. It is possible that this method may have biased system against CP effects.
phosphors of the monitor to emulate colours of the specified CIE co-ordinates. A high-resolution timer DLL produced by ExacTics was used to ensure accurate event timing. The programme was written in Visual BASIC (version 6.0). Participants were screened for colour vision problems using the City University colour vision test (Fletcher, 1980).

![Stimuli Matrix](image)

**Fig 2.3.** Schematic representation of the stimuli matrix showing Munsell hue (horizontal axis) and value (vertical axis) at constant chroma 7.4

**Procedure**

There were two naming tasks: a forced choice (blue or green) and a free response task. Half the participants were in the forced choice condition, the other half completed the free response condition. Each colour was presented individually in the centre of the screen on a neutral grey background. The instructions were to name the colour; in the forced choice conditions the stimulus had to be categorised as either blue or green; in the free response condition any colour term could be used. In the forced choice condition the left mouse button was pressed to indicate ‘green’ and the right to indicate ‘blue’. In the free response condition, responses were spoken out loud and recorded by the experimenter. Each stimulus was presented until a response was made. The order of presentation was randomised across participants.
Results

Naming frequencies across participants for the four hues averaged across lightness are presented in Table 2.1. All stimuli with hue 1.44B and 8.76BG were predominantly labelled blue and all stimuli with hue 6.06BG and 3.46BG were labelled green. Naming consistency across participants was lower for the boundary stimuli (8.76BG and 6.06BG) than for the stimuli closer to the category centres (1.44B and 3.46BG). Furthermore, blue and green responses were less frequent in the free response condition because terms such as aqua and turquoise were permitted.

Table 2.1 Naming frequencies (%) across participants for forced choice (blue or green) and free response conditions.

<table>
<thead>
<tr>
<th></th>
<th>Blue</th>
<th>Green</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.44B</td>
<td>8.76BG</td>
</tr>
<tr>
<td>Forced choice</td>
<td>96.87</td>
<td>85.6</td>
</tr>
<tr>
<td>Free response</td>
<td>92.5</td>
<td>66.57</td>
</tr>
</tbody>
</table>

The results of the preliminary naming study confirmed that the blue-green category boundary for English speakers is around 7.5BG and that the current stimulus set straddles a category boundary: half the stimuli are blue and the other half are green.
2.2.3. Method

Participants
Twenty-eight participants took part (22 female, 6 male, age range 18 to 42). All were students at the University of Surrey; some received course credits for their participation.

Stimuli and Apparatus
The preliminary naming study showed that for English speakers the most frequent response to 1.44B and 8.76BG stimuli was blue, the most frequent response to stimuli at 6.06BG and 3.46BG stimuli was green. This was true at all four brightness levels. Hence, the blue-green category boundary in the matrix shown in Figure 2.3 can be set between Munsell hue 8.76BG and 6.06BG. The stimulus pairs used in this experiment consisted of horizontally adjacent pairs; all pairs were one step pairs. A total of 12 pairs were used, eight within category and four cross-category pairs. Stimuli from within category pairs were categorically identical as they were given the same name, hence these were A-a pairs. Cross-category pairs were physically as well as categorically distinct, hence A-B. The blue within category pairs were 1.44B – 8.76BG; the green within category pairs were 6.06BG – 3.46BG; the cross-category pairs were 8.76BG – 6.06BG. The colour stimuli were presented on screen as described in § 2.2.2.

Tasks
Primary Task
The primary task was a 2AFC discrimination task. Each experimental trial started with the presentation of a colour from one of the A-a or A-B stimulus pairs. This target stimulus was presented for 1000 ms, followed by a filled or unfilled interval of 5000 ms. The target stimulus was then presented together with a distractor stimulus, i.e. the colour of the pair that was not used as the target stimulus. The task was to choose the colour identical to the target. The colours were presented horizontally 3 mm apart; duration of presentation was 1000 ms. Presentations were balanced so that
for each colour pair both stimuli were presented as the target stimulus equally often. Furthermore, the allocation of the position of the target and distractor stimuli on-screen was balanced so that the target appeared in the left and right position with equal frequency. Responses were made by pressing the corresponding mouse button (left, right) and instructions emphasised accuracy. There were 216 trials and the ratio of A-a to A-B trials was 2:1.

Interference
The 216 trials were divided into three interference conditions: no interference, verbal interference and visual interference. In the first of these conditions, no-interference, participants were presented with just the colour discrimination task and a neutral grey screen for the duration of the ISI. In both the visual and the verbal interference condition participants were required to perform an additional task, presented on-screen.

Verbal Interference
In the verbal interference task, participants were required to carry out a rhyme judgement task. The stimuli used consisted of a list of word pairs as shown in Table 2.2 (for full list of word pairs see Appendix 2). Word pairs were presented successively in upper-case letters (Arial, font size 38). On each trial the first word was presented 300 ms after the offset of the colour stimulus and was presented for 1500 ms, followed by a 1000 ms blank interval. Then the second word of the pair was presented for 1500 ms. The task was to decide whether the two words rhymed or not. Responses were required before the offset of the second word, later responses were recorded as misses. Responses were made by pressing the appropriate arrow key on the computer keyboard, left indicating a positive response (the two words rhyme) and right a negative response (the words do not rhyme). Reaction times for the secondary task were measured from the onset of the second stimulus until a response was made.
Table 2.2. Examples of word pairs used in the verbal interference task.

<table>
<thead>
<tr>
<th>Rhyming Pairs</th>
<th>Non-rhyming pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chord</td>
<td>Broad</td>
</tr>
<tr>
<td>Theme</td>
<td>Seem</td>
</tr>
<tr>
<td>Quay</td>
<td>See</td>
</tr>
<tr>
<td>Aisle</td>
<td>Pile</td>
</tr>
<tr>
<td>Chord</td>
<td>Bread</td>
</tr>
<tr>
<td>Theme</td>
<td>Moose</td>
</tr>
<tr>
<td>Quay</td>
<td>Haste</td>
</tr>
<tr>
<td>Aisle</td>
<td>Lose</td>
</tr>
</tbody>
</table>

Visual Interference

The stimuli used in the visual interference task consisted of pairs of successively presented shapes. The shapes were two-dimensional outline drawings of different shapes, examples are given in Figure 2.4. Participants were required to decide whether the first shape could, in any rotation, fit inside the second shape. The visual-interference task had the same time course as the verbal-interference task in sequence of stimulus presentation, presentation time, interval between stimuli and type of response. A positive response indicating that the first shape fits inside the second shape was given by pressing the left arrow key; a negative response indicating ‘does not fit’ by pressing the right arrow key. The presentation of the second word or shape was followed by 700 ms unfilled interval after which the test colour was presented.

Figure 2.4. Examples of pairs of shapes used in the visual interference task. The first shape (A) fits into the second shape (B) in Example 1, but does not fit in Example 2.
On both the visual and the verbal interference tasks, half the trials required a positive and half a negative response. No shape or word pair was presented twice and the allocation of word and shape pairs to colour pairs was randomised. Participants were required to respond as quickly as possible, without compromising accuracy.

**Design**

In contrast to Roberson and Davidoff's (2000) and Pilling et al.'s (2003) studies, interference conditions in the current experiment were unblocked. Trials were presented in a randomised order across all six blocks, thus each block contained trials with each type of interference. Presentations were balanced so that equal numbers of each interference condition were presented in each block. Presentations of colour pairs were also balanced so that each colour-pair was presented equally often.

**Instructions**

Participants were informed that they would be presented with a target colour which had to be remembered. They would then be presented with two colours, one being identical to the target stimulus, the other a distractor. The task was decide which of the two colours was the target stimulus. Furthermore, they were told an additional task would be presented on some trials and that this would either be a verbal or a visual task. The task would be to judge whether two words rhymed or not or whether it was possible to fit a shape in a second shape. They were told that the order of trials was completely randomised and therefore type of interference task would not be known in advance. Before the experimental trials, there were three blocks of practice trials (30 trials in total, 10 trials of each type of interference). These trials were identical in nature to the experimental trials; however, auditory feedback was given indicating errors on practice trials for both the colour discrimination and the interference tasks.
2.2.4. Results

Interference task performance
Median correct reaction time (RT) scores on the verbal and visual interference tasks were calculated for each participant. Across participants, the mean of RTs was 957 ms for the rhyme judgement task and 992 ms for the shape judgement task. Accuracy on the interference tasks was calculated by adding the frequency of hits (correct response on positive trials) to the frequency of correct rejections (correct responses to negative trials). The resulting value was expressed as a proportion of the total number of trials in the interference condition. For the rhyming task the mean accuracy score was 78.24%, for the shape judgement task the mean accuracy score was 75.52%. The majority of errors on both the verbal and the visual interference task were not due to incorrect judgements, but rather failures to respond within the given time.

Discrimination Accuracy
The mean accuracy scores for within and cross-category pairs are presented in Figure 2.5. Results were analysed by a two-way ANOVA, the two factors being pair type (within versus across category) and interference (no interference, verbal interference and visual interference). Both factors were repeated measures. There was a highly significant effect of interference \(F[2, 54] = 26.28, \text{MSE}= 56.70, p<0.01\), and the main effect of pair type was also significant \(F[1, 27] = 16.14, \text{MSE}= 106.49, p<0.05\). The pair-type x interference interaction, however, was not significant \(F[2, 54] = 0.13, \text{MSE}= 56.02, p>0.05\).
Fig. 2.5. Mean accuracy (error bars show +/- 1 standard error) for within category (A-a) and cross-category (A-B pairs.)

Post-hoc testing by protected t-tests found that there was a statistically significant advantage for cross-category pairs over within category pairs in all conditions (minimum $t[26] = 2.12, p<0.05$).

Looking at just the cross-category pairs, it was found that accuracy was significantly better in the no-interference condition compared to both verbal ($t[25] = 4.73, p<0.001$) and visual ($t[25] = 3.99, p<0.01$) interference. There was no significant difference in accuracy for A-B pairs between the verbal and the visual interference conditions ($t[25] = 0.74, p>0.05$). For within category conditions the pattern of results was essentially the same. That is, there were no differences among interference conditions ($t[25] = 0.01, p>0.05$) while the no interference condition was significantly more accurate than either interference condition (minimum $t[25] = 4.55, p<0.001$).
2.2.5. Discussion

Consistent with previous research (Pilling et al., 2003; Roberson & Davidoff, 2000) a significant CP effect was found in the no interference condition, cross-category discrimination was more accurate than within category discrimination. In both the verbal and the visual interference conditions overall accuracy was significantly lower, however the CP effect persisted in both interference conditions. Thus, neither the verbal nor the visual task selectively reduced CP in unblocked conditions in a 2AFC task. It is important to note that the visual and verbal interference tasks used in the current experiment were identical to those used by Pilling et al (2003) for blocked conditions. Thus, any differences found between these experiments are likely to be due to the manipulation of interference conditions.

The finding that verbal interference reduces CP when interference conditions are blocked appears to be a robust effect; it has been reported by Özgen (unpublished manuscript), Pilling et al. (2003) and Roberson and Davidoff (2000). However, if verbal interference blocked name retention as argued by Roberson and Davidoff, CP should be found only in no interference and visual interference conditions whether interference conditions are blocked or unblocked. The results of the current experiment show this not to be the case. In the introduction the effects of strategic encoding choices were considered. It was argued that participants may expect a naming strategy to be ineffective in unblocked conditions and hence adopt a perceptual coding strategy on all trials. Again if the direct language account of CP is true, CP should have been eliminated in all three conditions. Again this clearly did not happen.

This leaves the possibility that participants, not knowing which interference task would be presented on any given trial, may have used a naming strategy on all trials. If a name code could be retained across the ISI, CP should be found in all three interference conditions. The results of the current experiment are consistent with this account. The question of whether the results can be generalised to other types of discrimination tasks will be addressed in the next experiment.
2.3. Experiment Two

Effect of unblocked interference on discrimination: Same-Different judgements

2.3.1. Introduction

The results of the previous experiment suggest that Roberson and Davidoff's account of CP as a direct language effect is incomplete. In an unblocked 2AFC task CP survived not only visual but also verbal interference. However, it is possible that this is an artefact of the method used in Experiment 1. Although Pilling et al. (2003) found the same pattern of results for 2AFC and same-different delayed discrimination tasks for blocked interference conditions, others have argued that the tasks are not necessarily equivalent. In a 2AFC task the two test stimuli are presented side by side, whereas in the same-different task only one stimulus at a time is present. Therefore, the discrimination judgement made in a 2AFC task is a relative judgement; the judgement required in a same-different task is an absolute judgement. Macmillan (1987) argues that the former may encourage a discriminative and the latter a categorical response.

Hautus and Lee (1998) compared four psychophysical procedures and found that 2AFC and same-different tasks differed in their sensitivities. Furthermore, Özgen (unpublished manuscript) found verbal interference to have no effect on CP in a blocked same-different colour discrimination task whereas CP was reduced in a 2AFC task. In the context of the categorical perception of speech, it has been argued that different processes or modes are involved in same-different and 2AFC tasks. Pisoni (1973) refers to them as auditory mode and phonetic mode, and Guenther, Husain, Cohen and Shinn-Cunningham (1999), following on from Durlach and Braida (1969), use the terms sensory-trace and context-coding. For colour discriminations it is possible that relative judgements, required when the target and the disctractor are presented side by side, invite a verbal strategy more than absolute judgements.

The results of Experiment 1 were interpreted as suggesting that a naming strategy had been adopted on all trials, and this code was found to survive both visual and verbal interference. If the relative judgements required in a 2AFC task invite a verbal
strategy more than a same-different task it is possible that a different strategy would be used for absolute judgements. Experiment 2 tested this possibility by repeating the task as a same-different colour discrimination task.

2.3.2. Method

Participants
There were twenty-six participants (21 female, 5 male; age range 18 to 39), all were students from the University of Surrey; some received course credits for their participation. None had taken part in the previous experiment.

Stimuli
The same set of colour stimuli as in Experiment 1 was used. Therefore, there were again four cross-category (A-B) and eight within category (A-a) pairs. However, as the current task was a same-different discrimination task, all sixteen identical pairs were also included. The stimuli in these pairs were physically as well as categorically identical, hence these are represented as A-A pairs.

Procedure
The primary task in this experiment was a same-different discrimination task. As in Experiment 1 each experimental trial started with the presentation of the target stimulus, followed by a filled or unfilled delay. Then the second colour stimulus was presented, the task was to judge whether the two stimuli were physically the same or different. The response was given by pressing the appropriate mouse button (left indicating ‘different’; right ‘the same’). Presentation times were identical to those in Experiment 1. Each participant completed 216 trials split into six blocks of thirty-six trials.

For the ‘different’ pairs (the A-a and A-B pairs), each stimulus within a pair was presented equally often as the target colour and the test colour and each pair was presented with equal frequency in all three conditions, as were the ‘same’ (A-A) pairs. In order to ensure the ratio of within and cross-category pairs was kept equal to that in Experiment 1, whilst also presenting all sixteen physically identical pairs with equal
frequency, it was necessary to present more same than different pairs across trials. The ratio of same to different trials was 2:1.

Again, the interference conditions in this experiment were unblocked, meaning that type of interference was randomised across all trials and therefore not predictable. The same interference tasks as in Experiment 1 were used. Three blocks of ten practice trials were given for the discrimination task and the interference tasks before the start of the experimental trials as described in Experiment 1.

2.3.3. Results

Interference task performance
As in Experiment 1, median correct RT scores on the verbal and the verbal interference task were calculated. The mean RT score was 1032ms for the verbal and 1045ms for the visual interference task. For the rhyming task the mean accuracy score was 75.12%, for the visual interference task 60.66%. Hence, more errors were made on the visual interference task in the same-different task than in the 2AFC task. However, as in Experiment 1 the majority of errors were not due to incorrect judgements but a failure to respond within the given time. It is unlikely that the poor performance on the visual task in the current experiment is a true effect of task difficulty. The same visual and verbal interference tasks produced more or less equal accuracy scores and reaction times in Experiment 1 and in Pilling et al.’s (2003) blocked 2AFC and same-different tasks.

Discrimination Accuracy
Since the number of same and different trials was not equal, a measure of accuracy that combined hits (correctly responding ‘same’ on same trials) and false alarms (responding ‘same’ on different trials) was computed to calculate accuracy for within and cross category pairs for each participant. Hence, A-prime (A’)^5 scores were

---

^5 A’ is the non-parametric equivalent of the d’ measure used in signal-detection (see Swets, 1996). The statistic produces a single score of signal detection independent of response bias. It is calculated from the proportion of hits (correct responses on signal trials) and the proportion of false alarms (incorrect responses on no-signal trials). A’ values vary between 1 (perfect performance) and 0.5 (chance level performance).
calculated. Mean accuracy for within and cross-category pairs in each interference condition are shown in Figure 2.6.

![Figure 2.6: Mean A' (error bars show +/- 1 standard error) for within (A-a) and cross-category (A-B) pairs.](image)

As the graph shows, the results essentially paralleled those found in the previous experiment. There was a CP effect in all conditions ($F[1,25] = 9.45$, MSE= 0.0089, $p<0.05$); the main effect of interference was also significant ($F[2,50] = 4.36$, MSE= 0.0049, $p<0.05$).

Post-hoc testing showed that cross-category discrimination was better than within category discrimination in the no and the verbal interference condition (minimum $t[25] = 2.35$, $p<0.05$). The difference did not reach significance in the visual interference condition ($t[25] = 1.40$, $p=0.17$). The comparisons of just cross-category and within category paralleled those of Experiment 1.
2.3.4. Discussion

In Experiment 1 the suggestion was made that single name retention should survive verbal interference. CP was found to survive in both the visual and the verbal interference conditions, it was argued, therefore, that a naming strategy may have been used on all trials. Since 2AFC tasks require a relative judgement whereas same-different tasks require an absolute judgement, a naming strategy may be more beneficial in a 2AFC task than a same-different task. However, the results of Experiment 2 show that repeating the delayed discrimination task as a same-different judgement task produced the same pattern of results as did the 2AFC task. Again, a significant CP effect was found in the no interference condition and in both the verbal and visual interference conditions. Hence, there is no suggestion that the fact that CP was found to survive verbal interference in unblocked experiments is specific to 2AFC delayed discrimination tasks. Therefore, in line with Pilling et al.’s (2003, Experiment 1) finding that blocked 2AFC and blocked same-different discrimination tasks produce highly similar patterns of results, Experiment 2 showed the same to be true for unblocked conditions: CP survived verbal interference.

When type of interference is unpredictable, the task is not open to strategic biases. Participants could either use a blanket perceptual or verbal code. The results of Experiments 1 and 2 are consistent with either account. However, as the task is open to strategic biases in blocked conditions, a perceptual strategy should have been adopted in the verbal interference conditions (as argued on page 36) in Pilling et al. (2003). However, CP was lost in the verbal blocks which suggest that a perceptual strategy in these types of tasks does not lead to CP. Thus, as CP survived in all conditions in the current experiments, it seems unlikely that a blanket perceptual strategy was used in Experiments 1 and 2. Therefore, assuming the blanket strategy used on all trials was a verbal strategy, the findings suggest that a colour name code can be retained at the same time as having to carry out an additional verbal task. Thus, a possible interpretation of the results of Experiments 1 and 2 is that name retention can survive verbal interference under certain conditions, i.e. when interference is unblocked. By using a colour-to-name matching task, this idea is tested directly in Experiment 3.
2.4. Experiment Three

Interference and name retention

2.4.1. Introduction

Roberson and Davidoff (2000) assumed that the effects of verbal interference found in colour discrimination tasks were on name retention during the ISI; the name code could not be retained because an additional verbal task had to be carried out. Interference presumably results from name retention and the verbal interference task competing for the use of the articulatory loop in working memory (Baddeley, 1986; Baddeley & Hitch, 1974). Verbal interference in the form of irrelevant speech and articulatory suppression has been found to lead to a reduction in recall of lists of items (Baddeley 2002; Baddeley, Lewis & Vallary, 1984; Neath, 2000; Salame & Baddeley, 1982). For example Salame and Baddeley (1982) tested serial recall of sequences of nine digits and found that concurrent speech significantly impaired performance. However, average memory scores showed some items were recalled correctly in all conditions. These studies show that verbal interference does affect the performance of verbal short-term memory. What these studies do not show, however, is that verbal interference reduces the number of items recalled to zero. Hence, they do not suggest that a single name code would be affected in a similar way. Within a working memory framework, it seems likely that a single colour name code could survive even when an additional verbal task was introduced. If this is true, the effects of verbal interference in Roberson and Davidoff and Pilling et al. are unlikely to have been on name retention per se.

Experiment 3 tested this idea directly by examining the resilience of verbal colour codes to verbal and visual interference. A 2AFC delayed colour-to-name matching task was used which required the name of the target colour to be selected after either no, verbal, or visual interference. If verbal interference directly affects name retention, as Roberson and Davidoff suggest, accuracy in the verbal interference condition should be less than in both the no interference and the visual interference condition.
The current experiment was run with blocked (3a) and unblocked (3b) interference conditions as this allowed a comparison with both Experiments 1 and 2 and Roberson and Davidoff and Pilling et al.'s blocked experiments. If verbal interference prevented name retention relative to the other types of interference, the unblocked design provided a check that this did not result from a choice of strategy.

2.4.2. Method

Participants
Twenty-two participants took part in each experiment (Experiment 3a: 3 men and 19 women; age range 18 to 32; Experiment 3b: 5 men and 17 women; age range 18-30). All were students at the University of Surrey. Some obtained course credits for participation. None had taken part in the previous experiments.

Stimuli
Four colour stimuli were used, two examples of blue and two examples of green. One stimulus in each colour category was a good example of the category (focal stimulus), and the other a less good example taken from closer to the blue-green category boundary (boundary stimulus). The focal stimuli were the most saturated example of that colour that could be displayed on the monitor used. The boundary stimuli were taken from the set of stimuli used in the previous experiments: blue 8.76BG 6.20/7.40; green 6.06BG 6.20/7.40. For the verbal and visual interference tasks, a subset of the word and shape pairs used in Experiments 1 and 2 were presented.

Procedure
The target colour was presented for 1000 ms, followed by a 5000 ms filled or unfilled ISI. During the ISI either no interference, verbal or visual interference was presented. The target was chosen randomly on each trial, each stimulus was shown with equal frequency. After the ISI, the two test stimuli were presented side by side for 1000ms. The test stimuli were black bordered squares with either “BLUE” or “GREEN” printed centrally in black uppercase font (size 38) on a neutral background. The position of the target name was chosen randomly on each trial but was presented on
the left and right with equal frequency. The task was to label the target, to retain this label and then to choose the correct target name at the test stage. The choice was made by pressing either the left or right mouse button, as appropriate. There were six blocks of 16 trials; in Experiment 3a two blocks in each interference condition in random order. In Experiment 3b, interference conditions were unblocked and hence randomised within each block of trials. The interference stimuli were the same as in Experiments 1 and 2, but only a subset of the stimuli was presented. Ten practice trials were completed before starting the experiment. Auditory feedback was given to indicate errors on the practice trials; no feedback was given on experimental trials.

2.4.3. Results

Experiment 3a

On the interference tasks, mean across participant accuracy and reaction time were 77.70%, 991 ms and 85.23%, 897 ms for visual and verbal interference, respectively. Mean percentage correct responses for the primary task for focal and boundary colours are presented in Figure 2.7.

![Figure 2.7](image-url)

Figure 2.7. Mean accuracy (error bars show +/- 1 standard error) for focal and boundary stimuli for blocked interference conditions.
As the graph shows, colour-to-name matching was more accurate for focal colours compared to boundary colours. However, interference conditions clearly did not affect accuracy. Results were analysed by a repeated measures two-way ANOVA, the factors were focality (focal, boundary) and interference (no, visual, verbal). Focal colours were matched more accurately than non-focal colours. There was no significant effect of interference \((F[2,42] = 0.90, \text{MSE}= 56.70, p>0.05)\); the focality x interference interaction was also non-significant \((F[2,42] = 1.53, \text{MSE}= 71.95, p>0.05)\). Post-hoc testing by protected t-tests showed that accuracy was significantly higher for focal colours than for boundary colours in all interference conditions \((t[21]= 3.06, p<0.01)\). There was no significant difference between the no interference condition and either the verbal or the visual interference condition in both the focal and the boundary condition.

**Experiment 3b**

On the interference tasks, mean across participant accuracy and reaction time were 81.53%, 961ms and 84.66%, 868ms for visual and verbal interference, respectively. Mean correct responses for the primary task for focal and boundary colours are presented in Figure 2.8.

<table>
<thead>
<tr>
<th>Interference condition</th>
<th>Focal</th>
<th>Boundary</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>100</td>
<td>70</td>
</tr>
<tr>
<td>Visual</td>
<td>90</td>
<td>60</td>
</tr>
<tr>
<td>Verbal</td>
<td>80</td>
<td>50</td>
</tr>
</tbody>
</table>

Figure 2.8. Mean accuracy (error bars show +/- 1 standard error) for focal and boundary stimuli for unblocked interference conditions.
As the graph shows, colour-to-name matching was again more accurate for focal colours compared to boundary colours and interference conditions did not seem to affect accuracy. Results were analysed by a repeated measures two-way ANOVA, the factors were focality (focal, boundary) and interference (no, visual, verbal). The results paralleled those of Experiment 3a. Focal colours again were matched more accurately than non-focal colours \((F[1,21] = 86.22, \text{MSE} = 253.93, p<0.001)\). There was no significant effect of interference \((F[2,42] = 2.59, \text{MSE} = 54.91, p>0.05)\); the interaction was also non-significant \((F[2,42] = 1.20, \text{MSE} = 55.93, p>0.05)\).

### 2.4.4. Discussion

Accuracy was greater for focal stimuli than boundary stimuli across all interference conditions. However, as naming is more reliable for non-boundary than for boundary stimuli (see § 2.2.2), the differences in accuracy probably reflect differences in reliability of naming focal and boundary stimuli, rather than differences in the resilience of the name codes. Boundary colours may sometimes have been labelled incorrectly and hence even though name retention was effective, a false response would have been recorded.

Verbal interference did not selectively affect accuracy of name matching, neither for focal nor boundary colours. This result challenges Roberson and Davidoff’s (2000) account of CP and the assumed effects of verbal interference. Their account suggests that the effect of verbal interference was on name retention. The current results, however, show that a colour name code can be retained whilst completing an additional verbal task. An interesting question then is why verbal interference eliminated CP in blocked conditions (Pilling et al., 2003; Roberson & Davidoff, 2000). The current results reinforce the strategic bias suggestion made earlier (Experiment 1 and 2). It was suggested that when interference is blocked and hence type of interference is known in advance, participants may choose not to use a naming strategy on the task. It is possible that having to complete an additional verbal task made retention of the name code more effortful, and an alternative strategy was used in the verbal interference blocks. For unblocked conditions, type of interference was
not known in advance and it is therefore likely that participants used a blanket (verbal) strategy. In the current task, since name recognition was explicitly required, no other strategies were available on the task.

The results of Experiments 1 to 3 support the idea that the advantage of cross-category over within category discrimination in delayed colour discrimination tasks is due to labelling the target and retaining the name code rather than the perceptual code. However, perceptual codes must also be used as within category discrimination were above chance in all conditions (Experiments 1 and 2). Furthermore, the finding that CP is a language effect may be highly task dependent. The relatively long (5 seconds) ISI may have invited or encouraged a verbal strategy on the task. The next experiment tests whether colour CP effects are found in more perceptually loaded tasks on which verbal strategies are unlikely to be useful. Experiment 4 compares within-category and cross-category visual search.
2.5. Experiment Four

CP and Visual Search

2.5.1. Introduction

It is possible that the language effects on CP reported in Experiments 1 and 2 (and in Pilling et al., 2003; Roberson & Davidoff, 2000) were task dependent. Having a memory component in a colour discrimination task may have the effect of inviting the use of a verbal strategy. Retaining a name code may be less effortful and easier than trying to remember a visual representation of the target across the inter-stimulus interval. One way of testing this is to use a purer perceptual task on which a verbal strategy is unlikely to be useful. Experiment 4 tests colour CP using a visual search task.

In a visual search task, a display of colours is presented which includes a set of distractors and a differently coloured target. The target may or may not be present on a given trial. The task is to detect whether a target is present or not. The decision about the presence of a target can be made very rapidly (Treisman & Gelade, 1980). However, search times also depend on the perceptual distance between target and distractor: the more similar they are, the more difficult search is. If the distance between target and distractor is over 20 units in CIE L*u*v (Bauer, Jolicoeur & Cowan, 1996), the target in the display tends to 'pop-out'. Increasing the number of distractors does not increase search time which suggests that the presence of targets that pop-out can be detected by parallel search.

Furthermore, it has been argued that target-distractor similarity is not the only factor determining pop-out. D'Zmura (1991) found that pop-out only occurred for targets that are linearly separable from distractors. Thus, when targets and distractors were colinear (i.e. fall on a straight line in colour space) and the target was between the two distractors, pop-out did not occur. Pilling (2001) pointed out that there is possibly a further factor influencing search: categorical identity of target and distractors. If differences across colour boundaries are more salient than differences within categories, search should be facilitated if the target is in a separate category from the distractors. Pilling found that the importance of linear separability decreased, and
hence pop-out was stronger, when the search straddled category boundaries. This suggests that categorical separations may operate over and above perceptual separations. Thus, even when distances are smaller than $\Delta E 20$, targets may be more likely to pop-out if the distractors are in separate categories.

The current experiment tests whether categorical identity influences search with homogenous distractors (all distractors the same colour). If CP is a labelling effect, no advantage of cross-category over within category discrimination should be found when verbal labels cannot be used. Visual search is a perceptually loaded task; decisions are typically made before verbal labels are available (Bauer et al., 1996). The task was an independent groups design, one group carried out search for blue targets, and the other searched for green targets. Within each group the target remained constant, the distractors changed depending on search condition (within or cross-category). The stimuli used were four equally spaced stimuli, two blue and two green. The stimuli were approximately 9 units (CIE $L^*u^*v^*$) apart. Before each search condition, participants learned to discriminate the target from the distractor in a training phase. As mentioned above, it is possible that categorical separations between targets and distractors shift the separation closer to pop-out and so facilitate the search decision.

If CP is a language effect, no difference between cross-category and within-category search should be found in the current experiment. However, a perceptual account of CP (Bornstein, 1987; Harnad, 1987) would predict cross-category search to be easier compared to within category search as distances across boundaries should be perceived as larger, hence cross-category targets should pop-out more and so facilitate search. On target absent trials, there should be no CP effect. This would be consistent with Pilling’s (2001) findings that categorical separations on target present trials influences search.
2.5.2. Method

Participants
Twenty participants (9 male, 11 female; age range 20–40) took part in this experiment. All were students at the University of Surrey and either received course credits or were paid for their participation. None of the subjects had taken part in the previous experiments.

Stimuli and apparatus
The stimuli used in this experiment were a subset of the stimuli used in Experiments 1 and 2. Only one lightness level was used (6.2). The four stimuli used were: 1.44B 6.2/7.4; 8.76BG 6.2/7.4; 6.06BG 6.2/7.4; 3.46BG 6.2/7.4. There were four search conditions: blue within (target 8.76BG, distractor 1.44B), blue across (target 8.76BG, distractor 6.06BG), green within (target 6.06BG, distractor 3.46BG), and green across (target 6.06BG, distractor 8.76BG).

Stimuli were presented in a notional 6x6 grid in the search display. Each colour was 17mm², subtending a visual angle of approximately 1.9° at an average viewing distance of 500mm. There were three set sizes: 4, 16 and 36. For set size 4 and 16, stimulus locations within the grid were randomly selected. For set size 36, all stimuli were present in the grid. On target-absent trials, all stimuli were the colour of the distractor. On target-present trials, one of the distractors was replaced with the target. The location of the target in the grid was random, with the exception that the target never appeared in any of the four corners. The target occurred equally often in the remaining 32 position within each condition.

Procedure
Participants were randomly allocated to either the blue or the green group. Each group completed one within-category and one cross-category search. The boundary stimulus (either 8.76BG (blue) or 6.06BG (green)) was always the target. In the within category conditions, the second within category stimulus was the distractor. In the cross-category search, the distractor was the boundary stimulus belonging to the other
category. This is illustrated in Figure 2.9. The order of conditions was randomised across participants.

![Diagram of search conditions](attachment:image.png)

**Fig. 2.9. Illustration of search conditions.** Each participant completed two search conditions, one within and one cross-category. The target was the same in both conditions. For example, the green group within category search was to search for a 6.06BG target amongst 3.46BG distractors, the cross-category search was to search for the same target amongst 8.76BG distractors. Within and cross-category distances ($\Delta E$) were equal.

Each participant completed a within category and a cross-category block (training phase and search task). Order of blocks was randomised across participants. In the training phase participants learned to discriminate the target from the distractor. The target and then the distractor were shown, followed by the presentation of 60 trials on which a single colour (either the target or the distractor) was presented in the centre of the screen. The colour had to be classified as either the target (by pressing the left mouse button) or the distractor (right mouse button). Auditory feedback was given to indicate errors. After completion of the 60 trials, feedback was given on the correct number of responses.
After the training phase, participants moved on to the search task. There were 192 trials, 64 for each of the set sizes (4, 16 or 36 stimuli) in each condition. In each condition the target was present on 50% of the trials and absent on the other 50% of trials. Order of trials was randomised across set size and target present/absent. Overall, each participant completed 192 within category search trials and 192 cross-category search trials.

Each experimental trial started with the presentation of a fixation cross for 250 ms, followed by a 400 ms interval after which the search array appeared on screen. Participants were instructed to press the left mouse button if a target was present and the right mouse button if no target was present. The stimuli remained on screen until a response was made. The inter-trial interval was 400 ms. The instructions emphasised both speed and accuracy. Auditory feedback was given to indicate errors. Each participant completed a training phase and 192 experimental trials once in the within category and once in the cross-category condition. There was a short break between blocks. Both accuracy and reaction times were recorded. On average the experiment took 45 minutes to complete.

2.5.3. Results

Very few errors were made during the training phase and all participants completed both search conditions. Mean reaction times and accuracy scores were calculated across participants for each set size for within-category and cross-category search. The comparison of the blue and the green group showed that the pattern of results was essentially the same, and hence the data from both groups were collapsed for the analysis. One participant in the blue group was excluded since the reaction times indicated that the task had been done on the basis of accuracy alone.

Reaction Times

Median reaction times were calculated for each participant for within-category and cross-category search. Mean reaction times across participants are presented in Figure 2.10 for target present trials and in Figure 2.11 for target absent trials.
As the graph shows, cross-category search was faster than within category search for all three set sizes. The results were analysed by a two-way ANOVA, the factors were category (within, cross) and set size (4, 16, 36). There was a significant main effect of category ($F[1,18] = 7.09$, MSE = 12027.78, $p<0.05$) and a significant effect of set size ($F[2,36] = 6.49$, MSE = 5132.42, $p<0.01$). There was no significant interaction of category and set size.

For both within and cross-category search, search in the 4 stimuli array was significantly faster than in the 16 (minimum $t[16] = 1.90, p<0.05$) and the 36 sets (minimum $t[16] = 1.92, p<0.05$). There was no significant difference between search in the 16 and the 36 stimuli conditions.
Fig. 2.11. Mean reaction times in milliseconds (error bars show +/- 1 standard error) for within-category and cross-category search plotted against display size (4, 16, 36) for target absent trials only.

The category effects found on target-present trials were not found for target-absent trials. Results were analysed by two way ANOVA (category, set size). There was a significant effect of set size ($F[2,36] = 5.31$, MSE = 23306.56, $p<0.05$); neither the effect of category nor the interaction were significant. Post-hoc tests for within category search showed that search in the 16 stimuli array was significantly slower than in both the 4 and the 36 stimuli condition ($t[16] = 2.09$, $p<0.05$). In the cross-category conditions, search in the 4 stimuli condition was significantly faster than search in the 16 stimuli condition ($t[16] = 1.99$, $p<0.05$).
Accuracy

Mean accuracy for cross-category and within category search are presented in Figure 2.12 for target present trials and in Figure 2.13 for target absent trials.

Fig. 2.12. Mean percent accuracy (error bars show +/- 1 standard error) for within-category and cross-category search plotted against display size (4, 16, 36) for target present trials only.

Cross-category search was more accurate than within category search for all three set sizes. Search accuracy was lowest in the 36 stimuli condition for both within and cross-category search. Results were analysed by a two-way ANOVA (category, set size). Both main effects were significant. Accuracy was higher when the search was cross-category compared to within category search ($F[1,18] = 6.76$, MSE = 224.25, $p<0.05$) and there was a significant effect of set size: $F[2,36] = 14.05$, MSE = 99.38, $p<0.001$). The category x set size interaction was not significant.
Search accuracy appeared to be higher in the cross-category compared to the within-category condition on target absent trials also. However, neither category ($F[1,18] = 2.18$, MSE = 239.24, $p>0.05$) nor set size ($F[3,36] = 0.32$, MSE = 16.97, $p>0.05$) were significant main effects. The interaction was also not significant ($F[3,36] = 1.31$, MSE = 19.64, $p>0.05$).

2.5.4. Discussion

For reaction times, there was a clear category effect on target-present trials. Thus, search was faster when the target and the distractor were from separate categories than when they were both from the same category. CP was also found for accuracy, cross-category search was significantly better than within-category search on target-
present trials. The results imply that the categorical difference improved performance by a constant amount irrespective of the number of distractors. As there was also a distractor effect one might have expected the effect of distractors to be less for the overall easier condition (i.e. the cross-category condition). Since this was not the case (there was no significant interaction) the interpretation of findings is somewhat less straightforward. However, as predicted, no category effects for reaction time or accuracy were found when no target was present.

Since perceptual distances (in CIE L*u*v*) in both conditions were equal, the results support the view that other factors than target-distractor similarity influence search. In line with Pilling’s (2001) findings, categorical membership was an important cue in deciding whether a target was present in the display. If perceptual space is warped across category boundaries, perceived distances in the cross-category condition were greater than in the within category conditions. As outlined in the introduction it is possible that this categorical separation of target and distractor in the cross-category condition meant that the search was influenced by pop-out. Although the visual search task was thought to be a relatively pure perceptual task, it is possible that pop-out may have lead to the target also being named and so name codes may have played a part by reinforcing the categorical separation and so facilitating search. Thus, ‘categorical’ pop-out would imply that not all stimuli in the array had to be named for name codes to be useful. It is possible that the target popped-out and was named.

It seems likely that the weight of language and perceptual factors in determining CP depends on the type of task. The findings of Experiment 4 support a more perceptual account of colour CP in visual search tasks and suggest that the language account of CP used to explain the effect in delayed colour discrimination tasks is not necessarily generalisable to other types of tasks.
2.6. General Discussion

The aim of Chapter 2 was to test the direct language account of colour CP. The results of both the 2AFC (Experiment 1) and the same-different task (Experiment 2) showed a significant CP effect across the blue-green category boundary in the no interference condition as predicted by previous research (Bornstein and Korda, 1984; Pilling et al., 2003; Roberson and Davidoff, 2000). The results in the interference conditions, however, differed markedly. Overall accuracy was lower in both the verbal and the visual interference condition compared to the no interference condition, but discrimination of stimuli that straddled the category boundary was still significantly better than within category discrimination in both the visual and the verbal interference condition. Thus, verbal interference did not have a selective effect on colour CP which suggests that Roberson and Davidoff's account of colour CP is incomplete.

Roberson and Davidoff argued that verbal interference disrupted colour name retention, hence CP found in the no interference and the visual interference condition did not arise from visual but verbal coding. The comparison of verbal labels only facilitates cross-category but not within category discrimination, hence CP is found. However, the current results show that CP can survive verbal interference and therefore suggest that the effect of verbal interference found by Roberson and Davidoff and Pilling may not have been an effect on name retention per se, but rather an indirect effect on the type of encoding strategy used. Knowing the type of interference on each given trial may have led participants to abandon a verbal strategy in the verbal interference condition. Having to carry out an additional verbal task could have meant that a verbal strategy was believed to be effortful or ineffective. This would suggest that CP is a language effect, but verbal interference (in blocked tasks) did not directly affect name retention, but rather choice of strategy.

*Unblocked interference*

In Experiments 1 and 2 type of interference was randomised across all trials. Since interference was not predictable, a blanket strategy was presumably adopted on all trials. Thus, there are two possible explanations for the results found. Firstly, it is
possible that naming was abandoned on all trials, performance was based on perceptual codes consistent with a perceptual account of CP. However, if this were the case, and if we assume that participants switched to perceptual encoding in the verbal interference conditions in Roberson and Davidoff’s and Pilling et al.’s blocked studies, CP should have been found in these conditions also. As this was not the case, a purely perceptual account seems unlikely. Hence, this leaves the possibility that not being able to predict type of interference resulted in the target being encoded verbally on all trials. If this was the case, it is evidence against the idea that the effect of verbal interference (in blocked conditions) is on name retention since, in the current experiment CP, survived both verbal and visual interference. This suggests that if a name was generated, it could be retained across all interference conditions, and again supports the idea that the effect in blocked experiments was on choice of strategy rather than on working memory per se.

**Name code retention**

From a working memory model (see § 2.1.1) of short term memory there are problems with Roberson and Davidoff’s conclusions and the results of Experiments 1 and 2 support this view. As outlined above, the working memory theory and subsequent research on the capacity of verbal short term memory (see Baddeley, Lewis & Vallary, 1984; Neath, 2000; Salame & Baddeley, 1982) suggest that a single colour word could be retained across the ISI whilst carrying out an additional verbal task such as reading out words or judging whether two words rhyme. Therefore, it could be argued that a working memory model would in fact predict that, if the discrimination task is done purely on the basis of verbal labels, neither visual nor verbal interference should eliminate CP.

The colour-to-labels matching task presented in Experiment 3 directly tested this. Participants were explicitly instructed to verbally encode a target colour and retain just the colour name. The results showed that verbal interference did not selectively affect accuracy in blocked or unblocked conditions. These results strongly suggest that the results of verbal interference in Roberson and Davidoff were not on name retention per se. If the effects were on name retention they should also have been found when the task was to explicitly retain the name of the target.
The results of the visual search task in Experiment 4 support a more perceptual account of CP. Decisions about the presence or absence of targets were made relatively quickly, before verbal labels of all stimuli would have been available. (Boynton and Olson (1990) report mean response times for colour naming using basic colour terms such as blue and green to be approximately 1.5 seconds). However, it is possible that the categorical separation reinforced pop-out and that the target was named. Hence, target name codes may have facilitated search, but search was not done using just verbal labels. Thus, the findings suggest that the language account of colour CP is not generalisable to other types of task as the importance and use of verbal and perceptual codes may vary significantly. Overall, the results suggest that CP can result from perceptual and/or verbal coding and that the codes may reinforce each other.

When using dual-task designs to attempt to locate a phenomenon (such as CP) within the information-processing system it is important that the interference tasks used are equally hard. Looking at the accuracy scores and the reaction times to the interference tasks in Experiments 1 to 3 it becomes apparent that accuracy is higher and reaction times shorter for the verbal task than the visual task in all three experiments. This could be taken to suggest that the verbal interference task was easier than the visual interference task. This may have influenced the relative strength of the CP effects in the different interference conditions. However, it should be noted that the interference tasks used here were the same as those used by Pilling et al. (2003, see also Pilling, 2001) in a colour discrimination task with blocked interference conditions. With blocked interference, CP was eliminated in the verbal interference condition. Pilling (2001) reports that taking the performance on the visual and verbal interference tasks into account did not change the overall pattern of results and therefore concluded that the tasks were of equal difficulty. The somewhat better overall performance on the verbal compared to the visual interference task in unblocked conditions is therefore unlikely to be responsible for the finding that neither task removed CP. Furthermore, as mentioned earlier, the majority of errors on the interference tasks were not due to incorrect responses, but to a failure to respond within the given time. Hence, it is possible that the shape task took on average slightly longer to complete, but that this was also unfairly represented in the accuracy measure of the interference tasks.
There is a possible further problem with the interference tasks used in the current experiments (also Pilling et al., 2003). Spatial working memory may be separate from visual working memory (Hecker & Mapperson, 1997; Tresch, Sinaamon & Seamon, 1993). Vuontela, Rämä, Raninen, Aronen and Carlson (1999) have argued that short-term memory for location is dissociated from short-term memory for colour. This highlights the importance of carefully selecting the interference tasks. The visual interference task in the current experiments, judging whether one shape fit into a second shape, may be a spatial task. Therefore, it is possible that the ‘visual’ interference task in Experiments 1 and 2 did not selectively interfere with CP because the task was not purely visual. Thus the possibility that CP was perceptual cannot be ruled out. However, the comparison of blocked and unblocked experiments makes a truly perceptual account of CP in these tasks unlikely. Firstly, the finding that verbal interference does have an effect on CP in blocked conditions suggests that verbal codes are used in these tasks. Secondly, in the visual interference conditions in Roberson and Davidoff’s experiments pure visual interference was used (a multicoloured dot-pattern), however this also had no selective effect on CP.

Alternatives to the direct language account
Overall the evidence suggests that the effect of verbal interference found by Roberson and Davidoff (2000) and Pilling et al. (2003) in blocked delayed discrimination tasks was not on working memory per se but was rather an effect of strategy. Different encoding strategies may have been adopted in the verbal and the visual interference conditions; a verbal strategy was more effective for cross-category comparisons and hence CP was only found in the no interference and the visual interference condition. In the verbal interference condition the target was encoded perceptually, hence the target name was not generated, and therefore there was no advantage for cross-category discrimination. In the current unblocked experiments, a blanket verbal strategy may have been used on all trials and CP survived in all conditions. This account is still in line with colour CP as a language effect; however it suggests that the working memory framework may not be appropriate to test the effect, since a single colour name can be retained whilst carrying out an additional verbal task. Furthermore, Winaver et al. (2003) showed that verbal interference eliminated CP in
2AFC tasks with no memory component. This supports the idea that the effect is located at encoding rather than during retention.

If target name generation is necessary for CP, the effect could be due to a simple matching of labels at the response stage. A second possibility is that target name generation activates some kind of 'category code' which is in turn used to match the stimuli at the test stage of the discrimination task. This would be in line with Bornstein and Korda’s (1984) view that the target colour activates a physical and a categorical code in parallel. Their work followed on from Posner and Keele’s (1967) work on physical and category codes for letters (see § 2.1.2). Bornstein and Korda measured reaction times to same-different judgements of blue and green colour stimuli. The task was to respond same if the stimuli were from the same category. Consistent with Posner and Keele, ‘same’ responses were faster for A-A (physically identical) than A-a (same category) pairs. Furthermore, when the task was to respond ‘same’ only if the stimuli were physically identical, ‘different’ responses were faster to A-B (cross-category) compared to A-a (within category) pairs. This shows that people were better able to discriminate pairs of stimuli that straddled a category boundary than stimuli taken from the same colour category. Bornstein and Korda suggested that physical and categorical stimulus information was processed in parallel and that categorical information may be available early on in processing. It is possible that this categorical information is represented as a category code. The category code will here be defined as a representation of a colour which contains some, but not the exact, visual information about the colour. It is thought to be possibly represented as a mental image of the concept of a colour, or an approximation of e.g. a prototypical blue. This category code would facilitate cross-category but not within category discrimination by providing an additional level of analysis. It is possible that target name generation in colour discrimination tasks activates or reinforces the category code.

The idea of target name generation activating a category code could also be related to work done by Rosch (1975) and Neuman and D’Agostina (1981). Rosch studied the effects of priming on colour discrimination and found that priming with the colour category name had the effect of facilitating the discrimination of good but not poor members of that category. Furthermore, this priming was only effective when
presented prior to the test pair which led to the conclusion that the information contained in the category prototype facilitated the encoding rather than the later decision process. Hence Rosch argued that the mental representation must be in the form of a concrete visual code. Neuman and D’Agostino were interested in the structure and specificity of mental colour codes and their work followed on from Rosch’s previous findings. They used both the category name and the prototype colour as primes and found them to have significantly different effects. When primed with the prototype chip, responses to good members were facilitated whereas responses to poor members were inhibited. Priming by the name facilitated the response on all levels. From these findings Neuman and D’Agostino concluded that the information contained in mental representations generated by the category name is less specific than that contained in the physical code. However, the important point is that priming with a colour name is thought to result in a mental representation of the colour that is visual in nature.

Thus, if target name generation activates or reinforces a category code, it is plausible that colour CP is more complex than just the matching of colour labels\(^6\). If the category code generated on presentation of the target produces a representation\(^7\) that is less specific than the physical code but contains enough categorical information to facilitate discrimination of cross-category pairs, the resulting account of CP is both verbal, in that CP requires target name generation, and perceptual as target name generation activates a visual categorical code. Therefore, target name generation may be necessary for CP, the effect however may not be a direct language effect based solely on the comparison of labels.

To conclude, the current experiments lend support to a language account of CP in colour discrimination tasks. However, it is suggested that Roberson and Davidoff’s direct language account needs modifying since the effect of verbal interference on CP (in Pilling et al., 2003; Roberson & Davidoff, 2000) is unlikely to have been directly on name retention. The results of Experiments 1 and 2 suggest that if the target name

\(^6\) For which, incidentally, not only target but also test name generation would be necessary. This issue will be addressed in Chapter 3.

\(^7\) This point it related to work on the shift towards prototype in memory tasks (Fluttenlocher, Hedges & Duncan, 1991; Fluttenlocher, Hedges & Vevea, 2000). The direction and effect of this representational shift is addressed in Chapter 4.
is generated, CP survives in both visual and verbal interference conditions. The results of the colour-to-name matching task in Experiment 3 support this view. Furthermore, the results of Experiment 4 support a perceptual account of CP in more perceptually loaded tasks such as visual search. Thus, a language account of CP is likely to be task specific.

For colour discrimination (such as 2AFC and same-different) tasks, an account of colour CP in which target name generation is necessary is proposed. Whether CP is then a simple matching-to-labels process or if target name generation activates a category code (Bornstein & Korda, 1984) – possibly in a similar way that priming with a category name generated a mental representation (Neuman & D'Agostino, 1981; Rosch, 1975) – will be investigated in the next chapter.
Chapter Three

The effect of Stroop interference on the categorical perception of colour

3.1. General Introduction

The previous chapter addressed questions pertaining to the nature and the locus of colour categorical perception effects in discrimination and memory tasks. By studying the effects of interference on CP it was thought possible to locate the effect in either visual or verbal short-term memory. For blocked conditions verbal interference during storage (ISI) had been found to eliminate CP (Pilling et al., 2003; Roberson & Davidoff, 2000). However, the results of the delayed colour discrimination tasks discussed in Chapter 2 showed that CP can survive verbal interference if conditions are unblocked making the type of interference unpredictable. Thus, it was argued that the verbal interference tasks used in the blocked experiments did not affect verbal memory per se but rather the choice of strategy. Retaining the name of the colour may have been effortful or believed to be ineffective when having to complete a secondary verbal task. If this was the case the target may not have been named in the first place and the effect of verbal interference would hence have been an indirect effect at target presentation rather than during storage.

If this account is correct, naming the target is necessary for CP and therefore verbal interference during target presentation should reduce or even prevent CP. Baddeley, Lewis and Vallar (1984) reported that articulatory suppression (having to repeatedly utter an irrelevant sound, for example the word ‘the’) made the encoding of visually presented material difficult and argued that this prevented the material from being registered in the phonological store. In Roberson and Davidoff (2000, Experiment 4) participants had to count backwards from a random number presented either shortly before (a) or shortly after (b) target presentation. Hence, the interference was either at encoding (a) or during storage (b) or both. The results showed that interference during storage caused a dramatic drop in cross-category discrimination accuracy and that most interference was evident when counting was required both during encoding and
storage. Furthermore, although cross-category discrimination was not as strongly affected by interference at encoding as it was by interference during storage, the effect was strong enough to remove CP in the encoding condition also. Hence, it is possible that having to count backwards during encoding prevented participants from generating the correct target name.

However, since counting does not require rehearsal it could be argued that it is not a purely phonological loop activity and therefore other types of verbal interference at encoding need to be considered. One way of testing whether target name codes are necessary for CP is to use a type of interference known to strongly disrupt colour naming: Stroop interference. Using Stroop-like stimuli ensures concurrent processing of the target (the colour stimuli) and the verbal interference (the word). The current chapter presents a series of experiments using Stroop-like interference in 2AFC colour memory tasks.

3.1.1. Stroop interference and colour memory

In the classic Stroop colour-word task subjects have to name the colour of the ink in which an incompatible colour name is printed. Stroop (1935) and many researchers since (see MacLeod, 1991 for review) have shown this incongruity of stimulus dimensions to lead to a significant decrease in the accuracy and speed with which the colour can be named. Many variations of the original task have been found to be equally susceptible to Stroop interference. However, the vast majority of studies have focused on the relevant stimulus dimension, the word and not the irrelevant dimension, the colour. It is likely that a colour discrimination task with Stroop interference will not involve all the same mechanisms as the original Stroop task. Nevertheless, if colour naming is involved in both tasks, both should be susceptible to Stroop interference. Key aspects of the general theoretical framework that need to be considered when using Stroop-like interference in colour memory or discrimination experiments include the response modality; the use of visual and verbal codes in

8 But see Flowers & Blair (1976) and Flowers & Dutch (1976) for studies focusing on hue variations in Stroop stimuli and Bradlyn & Rollins (1980) on incidental memory of the colour in Stroop tasks.
colour processing; congruency effects; predictability of interference and use of strategies.

3.1.2. Response Modality

The original Stroop task required a verbal response – the colour of the stimulus had to be named. It was thought possible that the compatibility of verbal response and irrelevant dimension (word) accounted for the interference found in Stroop tasks (MacLeod, 1991). In studies testing non-verbal responses to Stroop stimuli (Keele, 1972; Logan, Zbrodoff & Williamson, 1984; Redding & Gerjets, 1977) the task was typically to indicate the colour of the ink by pressing a button corresponding to the appropriate colour name. Thus, although no overt naming was required, the task still required the use of the colour name and findings generally indicate that Stroop interference is also found when response is manual, albeit somewhat weaker. Therefore, Stroop interference does not seem to be response modality specific.

Researchers have also been interested in the effect of Stroop interference on different tasks such as colour sorting and matching. Chmiel (1984) asked subjects to sort incongruently labelled colours into colour-labelled bins and found evidence for Stroop-like interference. Flowers (1975) showed subjects sequentially presented colour words which had to be matched to colour patches. Matching of incongruent stimuli was significantly slower. Both results show that the presence of incongruent colour labels has an effect not just on the naming of colours but also on tasks which do not require the colours to be named explicitly. However, Durgin (2000) recently found Stroop effects to disappear when the task was to point to a matching patch of colour. Interestingly, for this type of task the author reported a strong reverse Stroop effect; when the task was to point to the patch of colour congruent with the colour word of the Stroop stimulus, responses were significantly slower.

Corballis and Luthe (1971) were interested in Stroop-like stimuli in relation to memory and recall; the study investigated how accurately the words and colour attributes of Stroop stimuli could be recalled. Three incongruent Stroop stimuli were presented on each trial and the task was to immediately recall the colours and the
colour words. The experiment was also done with coloured number words. Instructions were either to give a channel-by-channel recall (grouping words and colours together) or a temporal recall (the two attributes of each item had to be reported together). The results showed that when words had to be reported first, channel-by-channel recall was better than temporal recall. However, when colours had to be reported first, the effect was reversed. The results show that both attributes of incongruent Stroop stimuli can be encoded, but how well they can be recalled depends on the type of task. Importantly, the authors found no significant difference between the recall of colour-word/colour items (Stroop stimuli) and number-word/colour items. However, since participants were required to recall both word and colour as opposed to ignoring one dimension (as in the original Stroop task), it is perhaps not surprising that no Stroop interference was reported. In the context of the current set of experiments it is interesting that subjects were able to recall the colour of the stimulus.

3.1.3. Visual and verbal codes

In Chapter 2 the question of whether colours are encoded visually and/or verbally was reviewed in the context of colour memory tasks. The evidence suggests that given sufficient time, colour stimuli are encoded both visually and verbally (Allen, 1990; Garro, 1986; Schooler & Engstler-Schooler, 1990; Schooler, Fiore & Brandimonte, 1997). It is also possible that colours, especially focal colours, are easy-to-name visual stimuli (Brandimonte, Schooler & Gabbino, 1997) and hence may be named automatically. Whether verbal labels are useful for recall is debatable (Schooler and Engstler-Schooler for example have argued that verbal labels overshadow the veridical visual code), however verbal labels certainly facilitate cross-category discriminations – and thus possibly CP. The question of whether verbal and/or visual codes are used for processing, discriminating and retaining colours is equally relevant to Stroop and colour discrimination tasks. The original Stroop task required only the use of the verbal code, the name of the ink colour. However, is the verbal code the only way the colour is encoded or is there also a (intact?) visual code?
Flowers and Blair (1976) conducted a series of studies investigating the use of verbal and visual codes in colour classification tasks with Stroop interference. The focus was on the hues of the to-be-remembered inks and the effect of varying the similarity of these hues. In a speeded classification task they found evidence for Stroop interference when non-adjacent hues had to be grouped together but not when adjacent hues were grouped together. The authors suggested that the categorisation of non-adjacent hues requires access to verbal and not just perceptual codes. Grouping adjacent hues, on the other hand, can be accomplished on the basis of rapid visual perceptual decisions alone. These are not disrupted by verbal response competition - hence no Stroop interference. Therefore, verbal and visual codes of the relevant stimulus dimension (the colour) may be present and either form of encoding may be used to make perceptual decisions, ideally the most effective code is used.

The results reported by Flowers and colleagues (Flowers & Blair, 1976; Flowers & Dutch, 1976) highlight some interesting points concerning the use of Stroop interference in colour memory and discrimination tasks. The colour stimuli used in the current experiments are adjacent hues, they are perceptually quite similar and hence the memory task is relatively hard. The 2AFC task requires similar colours to be retained and discriminated, not grouped together. It is possible that this task is 'perceptually' more difficult and using a verbal code may well be more efficient. However, this is only going to be true for cross-category discriminations, within category discriminations should not be aided by verbal labels.

Furthermore, Flowers and Dutch (1976) argued that it is only rapid visual perceptual decisions and tasks with small memory loads that are not disrupted by Stroop interference. This could be related to Luck & Vogel's (1997) findings that verbal interference did not affect colour short-term memory in a same-different task in which colour arrays were presented for only 100ms. Taken together this suggests that verbal labels are only used when there is sufficient time to encode all the stimuli verbally. The current experiments require a memory representation to be retained for the duration of the ISI, for it then to be compared to two test stimuli. By manipulating the duration of stimulus presentation and the length of the ISI the relevance of Flower and Dutch's account to delayed colour discrimination tasks can be tested. Tasks with long ISIs and long test colour presentations may not have small memory loads nor require
rapid visual decisions. Again, the relative efficiency of visual codes may be smaller than that of verbal codes – especially for cross-category discriminations. However, when presentation times and ISIs are shortened, much quicker decisions should occur which should, according to Flower and Dutch, be less susceptible to Stroop interference.

3.1.4. Congruency effects

The presence of incongruent colour words leads to substantial interference in colour naming, sorting and matching tasks but what effect do congruent colour words have on colour naming? The original Stroop task did not include a congruent condition but many variations of the task since have. Congruency can lead to facilitation, with the speed of naming the colour increasing compared to a neutral condition (e.g. Glaser & Glaser, 1982; Duncan-Johnson & Koppel, 1980). However, there is usually less facilitation than interference (MacLeod, 1991). Furthermore, it has been pointed out that both facilitation and interference may be to some extent task dependent in that the ratio of congruent to incongruent trials affects the strength of interference/facilitation (MacLeod, 1991). This point is closely related to the question of task strategies and predictability of trial type discussed below.

3.1.5. Strategy

In a task in which colour and word never match, there is no benefit to be gained from attending to the irrelevant dimension (the word). As mentioned above, the original Stroop task did not include congruent trials, but many studies since have included both congruent and incongruent trials. The more congruent trials there are, however, the more important the irrelevant dimension becomes as there is a greater probability that attending to the word will pay off. The presence of congruent trials tends to invoke a tactic of splitting attention over the two dimensions (MacLeod, 1991) and hence leads to an increase in interference on the incongruent trials.
Does this mean that knowing what kind of interference to expect or which condition is most likely allows subjects to adopt specific strategies? In the previous chapter it was argued that delayed colour discrimination tasks with verbal and visual interference are open to strategic choices; participants may adopt a verbal or a visual encoding strategy depending on type of interference. Elliot, Cowan and Valle-Inclan (1998) present evidence against such condition specific strategies in Stroop tasks. In a colour-word versus non-colour-word interference task they found no difference between unblocked and blocked conditions and argued that the predictability of colour-word conflict is not critical to the task. Logan, Zbrodoff and Williamson (1984) however, found interference to vary as the relative frequency of congruent and incongruent stimuli was manipulated. The authors argued that when incongruent stimuli are more frequent, subjects attend to both dimensions to give a response. They suggested that interference in Stroop tasks may be strategy dependent and an effect of the task design itself.

This idea finds further support in a recent study by Dishon–Berkovits and Algom (2000). The authors argued that the design of most Stroop studies is biased towards finding an interference effect. The inclusion of congruent trials results in a positive correlation between the Stroop dimensions since, if there are equal numbers of congruent and incongruent trials, the probability of a word given its matching colour is greater than any other colour. Therefore, the correlation between the irrelevant dimension and the target dimension means that the word is predictive of the colour and hence provides a cue.

3.1.6. Stroop effects and colour CP

The aim of the following series of experiments was to further test the direct naming account of CP, and more specifically to test the idea that target name generation is necessary for CP. If CP requires naming, and if Stroop weakens name generation even if no explicit name is required (Chmiel, 1984; Flowers, 1975) then introducing Stroop interference at the various stages in a 2AFC task should provide some insight into the processes involved.
The experiments that follow vary in how the Stroop interference is implemented. By introducing congruent and incongruent Stroop interference at encoding, Experiment 5 tests the account put forward in Chapter 2 that target name generation is necessary for CP. Also, by varying the degree of separation of stimulus dimensions the question of how easily the word dimension can be ignored is addressed. Experiment 6 tests the matching-to-labels account of CP by introducing Stroop interference at the response stage of a discrimination task. Hence, if CP is solely based on matching colour labels, the two test stimuli at the response stage in a 2AFC task must also be named correctly. By shortening the inter-stimulus interval and stimulus presentation times, Experiments 7 and 8 test the hypothesis that tasks with lower memory loads allow quick perceptual decisions and are not susceptible to Stroop interference (Flowers & Dutch, 1976); this is again tested at encoding (target presentation, Experiment 7) as well as at response (test presentation, Experiment 8).
3.2. Experiment Five

Stroop interference at target presentation

3.2.1. Introduction

The advantage of cross-category discrimination was found to survive verbal interference during retention in Experiments 1 and 2. The current experiment addresses the question of whether CP can survive a different type of verbal interference, Stroop interference. Using Stroop interference in colour discrimination tasks should allow a test of whether visual or verbal codes are generated at encoding and hence test whether the locus of CP is at encoding rather than during storage. If congruent and incongruent Stroop interference were found to have no selective effect on CP, a perceptual account of CP would be supported. However, if CP were the result of comparison of verbal labels, facilitation and interference from congruent and incongruent stimuli respectively should be evident.

The two dimensions of Stroop stimuli, the colour and the colour word, can be more or less integrated. The stimuli can be presented as colour-in-word stimuli (as in the original Stroop task) or the dimensions can be separated. Studies have manipulated the spatial separation of colour and word, for example by presenting the word above a coloured bar or square. The more separated the dimensions are, the less Stroop interference there is (see MacLeod, 1991). In a card-sorting task Flowers and Stoup (1977) found that Stroop effects disappeared with practice for non-integrated stimuli but not for integrated stimuli. Furthermore, La Heij, van der Heijden and Pollij (2001) manipulated the temporal separation of Stroop stimuli, by removing either the colour or the word before response. Stroop interference was reduced when the colour was removed from the display before the word. The current experiment was run with superimposed Stroop words (5a) and colour-in-word stimuli (5b) to test whether the degree of separation influences how easily the word can be ignored. The bias towards Stroop effects (as identified by Dishon-Berkovits & Algom, 2000) was avoided by using a set of colours which were either blue or green and correspondingly only two
possible labels (blue or green). Therefore the chance of a colour being congruently labelled is equal to that of it being incongruently labelled.

The current experiment measured within and cross-category discrimination accuracy in a 2AFC task. Target presentation time was shortened to 150 ms (compared to 1000 ms in Experiment 1) to ensure that the visual code and the verbal code did indeed conflict in the incongruent condition. With longer presentation times, subjects may have been able to generate an alternative verbal label for the colour. In a same-different colour memory task Luck & Vogel (1997) presented colour arrays for 100ms and found evidence that the task was done in visual and not verbal short-term memory. Hence, this suggests that 100ms are sufficient to encode an array of colours visually, but not to generate labels. Therefore, in the current experiment 150 ms should be sufficient to generate a visual code, but it should be difficult to name the target correctly in the incongruent condition.

As in previous experiments with no interference, a CP effect should be found in the neutral Stroop condition. If the better discriminability of cross-category pairs over within category pairs is a language and not a perceptual effect, Stroop target stimuli should either reduce or enhance CP relative to the neutral condition: CP should be reduced in the incongruent condition since word-colour incongruency weakens name generation. However, as congruent Stroop pairs (for example Blue-blue) can facilitate a name response, CP should be enhanced on congruent trials.

In standard Stroop tasks, reaction time (naming latencies) is the main dependent measure and a response is required immediately on presentation of Stroop stimuli. In the current experiment, however, target naming was not an explicit task requirement and the time of Stroop interference was dissociated from time of response. Hence, although target name generation in the congruent condition should be significantly faster than in the incongruent condition, this should not be reflected in the reaction time measure to the colour discrimination task.
3.2.2. Method

Subjects
Twenty-eight subjects (6 male, 22 female, age range 18-34) participated in the superimposed condition (5a), a further twenty-eight subjects (5 male, 23 female) completed the colour-in-word condition (5b). All were students at the University of Surrey; some received course credits for participation. None of the subjects had participated in the previous experiments.

Stimuli
5a: Superimposed interference
The stimulus set was made up of the sixteen Munsell stimuli described in Experiment 1. There were four cross-category pairs (A-B) and eight within category pairs (A-a). A colour word (‘BLUE’ or ‘GREEN’) or ‘XXXX’ in black uppercase lettering (font size 16) was superimposed in the centre of the target colour. The colour stimuli were 55mm² and subtended a visual angle of approximately 6.3°. The colour word or XXXX measured 15mm x 5mm (1.7° x 0.6°). The congruency and incongruency of the colour-word stimuli was based on the naming data reported in §2.2.2. Hence all stimuli of Munsell hue 1.44B and 8.76BG were classified as congruent when combined with the label BLUE and as incongruent when combined with the label GREEN. All stimuli of 6.06BG and 3.46BG hue were congruent when labelled GREEN and incongruent when labelled BLUE. Viewing distance was approximately 500mm.

5b: Colour-in-word interference
The colour stimuli were the same as those used in 5a. Stroop interference however was presented as ‘colour-in-word’ stimuli, XXXX, BLUE or GREEN were filled with the to-be-remembered colour and presented in large, uppercase Impact font (font size 120). The words (or XXXX) measured 90mm x 40mm (approximately 10.3° x 4.5°).
Procedure

5a: Superimposed interference

Discrimination was measured using a delayed 2AFC task. Stimulus pairs were the same as in Experiment 1 apart from the addition of the Stroop word; pairs were chosen at random on each trial but all twelve pairs were shown with equal frequency in all three conditions. Each stimulus of the pair was shown as the target and distractor stimulus with equal frequency. A trial consisted of a fixation cross for 500ms, followed 400ms later by the target plus Stroop word for 150ms, followed by a 5000ms unfilled ISI after which the two test stimuli were presented for 1000ms or until a response was made. In the neutral condition, ‘XXXX’ was superimposed on the target, in the congruent condition ‘BLUE’ (on a blue target stimulus) or ‘GREEN’ (on a green target stimulus) and in the incongruent condition ‘BLUE’ (on a green target stimulus) and ‘GREEN’ (on a blue target stimulus). There were equal numbers of congruent, incongruent and neutral trials in each block. Schematic representations of neutral, congruent and incongruent trials are presented in Figure 3.1. The task took approximately 40 minutes to complete.

Fig. 3.1. Schematic representations of neutral (a), congruent (b) and incongruent (c) cross-category trials in 5a. The Stroop interference was superimposed on the centre of the target colour.
At test presentation the target stimulus was presented as the left and right stimulus with equal frequency. The inter-trial interval was 600ms. Subjects were instructed to ignore the word and attend to the colour. Instructions also emphasised speed and accuracy. There were a total of 216 trials, presented in six blocks of thirty-six trials. Stroop interference was not blocked; the type of condition was randomised across all trials so that each block consisted of neutral trials, trials with congruent colour words and trials with incongruent colour words. Three blocks of ten practice trials were completed before starting the experimental trials. Auditory feedback was given for incorrect responses.

5b: Colour-in-word interference
The procedure was the same as in 5a, apart from the Stroop stimuli being presented as colour-in-word stimuli as shown in Figure 3.2.

Fig. 3.2. Examples of a congruent trial. Stimuli were presented as colour-in-word stimuli to ensure concurrent processing of the two dimensions (colour and colour word). The example given is a cross-category trial.
3.2.3. Results

5a: Superimposed interference

Reaction times

Median reaction times for within and cross-category discriminations were calculated for each participant. Table 3.1. presents mean reaction times for all conditions across participants.

Table 3.1. Mean reaction time (standard error) ms in neutral, congruent, and incongruent Stroop conditions for cross-category (A-B) and within category (A-a) discriminations.

<table>
<thead>
<tr>
<th>Stroop Interference</th>
<th>Neutral</th>
<th>Congruent</th>
<th>Incongruent</th>
<th>( \bar{x} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-B</td>
<td>871.1 (33.5)</td>
<td>849.1 (28.2)</td>
<td>874.5 (32.2)</td>
<td>864.9 (29.9)</td>
</tr>
<tr>
<td>A-a</td>
<td>921.5 (33.9)</td>
<td>924.9 (33.5)</td>
<td>904 (25.7)</td>
<td>916.8 (29.9)</td>
</tr>
<tr>
<td>( \bar{x} )</td>
<td>8963.3 (32.2)</td>
<td>887.0 (29.9)</td>
<td>889.3 (27.9)</td>
<td></td>
</tr>
</tbody>
</table>

As Table 3.1 shows, cross-category discriminations were faster than within category discrimination in all interference conditions. However, neither cross-category nor within category discriminations varied substantially across interference type. A two-way ANOVA for pair-type (cross, within) and Stroop interference (neutral, congruent, incongruent) confirmed this. There was a significant effect of pair-type \( (F[1,27] = 19.19, \text{MSE} = 5895.17, p<0.001) \), but the main effect of Stroop interference was not significant \( (F[2,54] = 0.34, \text{MSE} = 3851.03, p>0.05) \). The pair-type-interference interaction was also non-significant \( (F [2,54] = 2.14, \text{MSE} = 3505.84, p>0.05) \).

Post-hoc testing by protected t-test showed that reaction times were significantly faster for cross-category pairs compared to within category pairs in both the neutral
(t[26] = 2.45, p<0.05) and the congruent condition (t[26] = 3.61, p<0.01). The difference was not significant in the incongruent condition (t[26] = 1.36, p>0.05).

**Accuracy**

The mean accuracy scores for within and cross-category pairs are presented in Figure 3.3. As the graph shows, Stroop interference did not appear to significantly affect within category discrimination: mean accuracy was approximately 65% in the neutral, the congruent and the incongruent condition. Cross-category discriminations, however, seemed to be strongly affected by changes in type of interference; CP effects were dependent on type of interference. In the neutral interference condition a CP effect was found, cross-category discrimination was better than within category discrimination. Compared to neutral condition CP was facilitated in the cross-congruent Stroop condition but reduced in the incongruent Stroop condition.

![Figure 3.3](image)

**Figure 3.3.** Mean discrimination accuracy (error bars show +/- 1 standard error) for within (A-a) and cross-category (A-B) pairs.
Results were analysed by a two-way ANOVA, on pair-type (cross, within) and interference condition (neutral, congruent and incongruent). Both factors were repeated measures and both were highly significant (pair-type: $F[1, 27] = 19.72, \text{MSE} = 72.36, p<0.001$; interference: $F[2, 54] = 7.81, \text{MSE} = 46.50, p<0.01$). The pair-type x interference interaction was also significant ($F[2, 54] = 14.5, \text{MSE} = 54.43, p<0.001$).

Post-hoc testing by protected t-tests showed that there was a significant advantage for cross-category pairs over within category pairs in the neutral condition ($t[26] = 2.23, p<0.05$) and a highly significant advantage in the congruent condition ($t[26] = 6.02, p<0.001$). However, there was no advantage of cross-category pairs over within category pairs in the incongruent condition ($t[26] = 0.55, p>0.05$).

For cross-category pairs only, the results show that accuracy was significantly higher in the congruent condition compared to both neutral ($t[25] = 4.41, p<0.001$) and incongruent ($t[25] = 6.77, p<0.001$) interference. Accuracy was also significantly better in the neutral compared to the incongruent condition ($t[25] = 2.37, p<0.05$). For within category pairs only, there was no significant difference in accuracy between the neutral and the congruent condition ($t[25] = 0.33, p>0.05$), the neutral and the incongruent condition ($t[25] = 1.1, p>0.05$) or the congruent and the incongruent condition ($t[25] = 1.42, p>0.05$).

A re-analysis of the effects of interference and pair-type on accuracy was performed taking reaction times into account. A two-way ANCOVA with reaction time as a covariate was performed. However, this did not significantly alter the pattern of results.
Experiment 5b (colour-in-word interference):

Reaction times
Mean reaction times across participants for each interference condition are presented in Table 3.2.

Table 3.2. Mean reaction time (standard error) ms in neutral, congruent, and incongruent Stroop conditions for cross-category (A-B) and within category (A-a) discriminations.

<table>
<thead>
<tr>
<th>Stroop Interference</th>
<th>Neutral</th>
<th>Congruent</th>
<th>Incongruent</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-B</td>
<td>843.3 (28.2)</td>
<td>849.9 (24.8)</td>
<td>834.5 (22.7)</td>
<td>842.6 (23.9)</td>
</tr>
<tr>
<td>A-a</td>
<td>877.8 (24.6)</td>
<td>884.5 (26.3)</td>
<td>874.7 (26.7)</td>
<td>879.0 (25.0)</td>
</tr>
<tr>
<td>Mean</td>
<td>860.6 (25.4)</td>
<td>867.2 (24.8)</td>
<td>854.6 (23.6)</td>
<td></td>
</tr>
</tbody>
</table>

Cross-category discriminations were again faster than within category discrimination in all interference conditions. However, this effect was smaller than in Experiment 5a. Furthermore, not only did both cross-category and within category discriminations not vary substantially across interference type, congruent cross and within category discriminations were actually slower than neutral discriminations. Also surprising is that the fastest discriminations were made in the incongruent cross-category conditions. Mean reaction times were analysed by a two-way ANOVA. The results showed that reaction times were significantly faster for cross-category pairs than within category pairs ($F[1,27] = 12.7, \text{MSE} = 4387.62, p<0.01$). The main effect of Stroop interference, however, was clearly not significant ($F[2,54] = 0.84, \text{MSE} = 2619.69, p>0.05$). The pair-type-interference interaction was also non-significant ($F[2,54] = 0.7, \text{MSE} = 2080.46, p>0.05$). Post-hoc analysis by protected t-tests revealed that the cross-category discriminations were significantly faster than within
category discriminations in all three interference conditions (minimum $t[26] = 1.95$, $p<0.05$)

Accuracy

The mean accuracy scores for within and cross-category pairs in each interference condition are presented in Figure 3.4. As the graph shows, the pattern of results was very similar to that found in the superimposed condition (5a), there was however less facilitation with colour-in-word stimuli. Results were analysed by a two-way ANOVA. There was a significant effect for pair type ($F[1, 27] = 16.63$, MSE = 67.22, $p<0.001$), interference, however was not significant ($F[2, 54] = 1.2$, MSE = 73.98, $p>0.05$). The pair-type x interference interaction was also again highly significant ($F[2, 54] = 10.64$, MSE = 50.58, $p<0.001$).

![Figure 3.4. Mean discrimination accuracy (error bars show +/- 1 standard error) for cross-category (A-B) and within (A-a) pairs.](image)

Post-hoc testing by protected t-tests showed that there was a significant advantage for cross-category pairs over within category pairs in the neutral condition ($t[26] = 2.65$, $p<0.05$) and, a highly significant advantage in the congruent condition ($t[26] = 5.62$, $p<0.001$).
Again, there was no advantage of cross-category pairs over within category pairs in the incongruent condition ($t[26] = 0.61, p>0.05$).

Looking at just the cross-category pairs, accuracy was significantly better in the congruent condition compared to both neutral and incongruent interference (minimum $t[25] = 2.2, p<0.05$). There was no significant difference in accuracy between the neutral and the incongruent condition ($t[25] = 1.22, p>0.05$). For within category pairs only, none of the comparisons was significant (maximum $t[25] = 1.94, p>0.05$).

Re-analysis of the data by means of a two-way ANCOVA with reaction time as a covariate again did not change the pattern of results significantly.

Superimposed vs. colour-in-word interference

Further analysis was performed to test whether the type of Stroop presentation had selective effects on accuracy and reaction times scores. The data of Experiments 5a and 5b were combined; type of presentation was treated as a factor. A three-way ANOVA with pair-type (cross, within), interference (neutral, congruent, incongruent) and type of presentation (superimposed, integrated) was performed. There were significant effects of pair-type ($F[1,27] = 36.6, \text{MSE} = 69.37, p<0.001$) and interference ($F[2,54] = 5.4, \text{MSE} = 70.41, p<0.01$), the interaction between pair-type and interference was also significant ($F[2,54] = 22.3, \text{MSE} = 58.55, p<0.001$). The effect of type of Stroop presentation, however, was not significant ($F[2,54] = 0.9, \text{MSE} = 186.77, p>0.05$). The analysis of reaction time showed that only pair-type had a significant effect ($F[1,27] = 42.2, \text{MSE} = 3884.47, p<0.001$). This confirmed that the type of Stroop presentation did not selectively affect accuracy or reaction times.
Summary

The analysis of reaction times found that cross-category discriminations were faster than within category discrimination, independent of type of Stroop interference. As predicted discrimination was significantly better for cross-category pairs compared to within category pairs in the neutral conditions in Experiments 5a and 5b. This effect was significantly increased in the congruent conditions; there was a highly significant advantage of cross-category pairs over within category pairs. However, this advantage was lost in the incongruent conditions where no difference in accuracy between cross and within-category pairs was found.

3.2.4. Discussion

CP was evident in the neutral and the congruent Stroop conditions: recognition of stimuli from cross-category pairs was superior to within category recognition. This advantage was also reflected in the reaction time measure: discrimination of pairs that straddled the category boundary were faster than within-category discrimination in all conditions. Furthermore, CP for accuracy was greater in the congruent than in the neutral condition. Thus, providing the correct label at target presentation had the effect of making cross-category discriminations more accurate. In contrast, in the incongruent condition CP was eliminated. When the target was presented with an incongruent label, cross-category accuracy fell to the level of within-category accuracy. These results support the suggestion made in Chapter 2 that target name generation is necessary for CP. As predicted, reaction times were not affected by different types of Stroop interference. It is likely that this was due to Stroop interference not being presented at time of response.

The finding that CP did not survive incongruent Stroop interference suggests that the short presentation time of 150ms was too brief to allow generation of an alternative, the correct, colour label to use in the discrimination task. Furthermore, it suggests that the cross-category advantage found in the neutral and the congruent conditions is not based on visual codes, since incongruent Stroop interference should have no effect on these. However, the results also show that visual codes must be used in discrimination
tasks, since within category discriminations were above chance level in all conditions and these were largely unaffected by Stroop interference.

Although presenting the correct target name in the congruent condition led to high cross-category discrimination accuracy and hence substantial CP effects, the results do not fully support a simple language account of CP. If target name generation leads to CP, it is not clear why the CP effect in the neutral condition, where labelling the target should have been easy, was relatively small. This may mean the targets in the neutral condition were not labelled correctly all the time. As the naming data in §2.2.2 shows, naming agreement for the boundary stimuli was not perfect. Hence, it is possible that the boundary stimuli were labelled ‘incorrectly’ on some neutral trials which would then have made the labels unhelpful for cross-category discriminations.

On the other hand, as the instructions were to attend to the colour of the stimulus and ignore the word (to use a visual strategy on the task) subjects may not have tried to name the colour and may have defaulted to a perceptual strategy, producing only moderate (but possibly perceptual) CP in the neutral condition. Thus, the possibility that CP in the neutral and CP in the congruent conditions may be due to different encoding strategies should be considered. Whatever the reason for the small CP effect in the neutral condition, the crucial findings are that having the correct target name strongly facilitated CP and preventing target name generation eliminated CP.

Experiments 5a and 5b produced very similar findings, whether Stroop interference was presented in integrated or non-integrated stimuli did not significantly change the pattern or results. Contrary to the findings reviewed by MacLeod (1991) suggesting that congruent dimensions produce less facilitation than incongruent dimensions produce interference, the congruency effect in the current experiments (especially Experiment 5a) was very strong. However, as mentioned above, the effect in the neutral condition was not large. Thus, the effect of interference did not need to be very strong to eliminate the advantage. It is possible that the current experiment offered more scope for facilitation than other Stroop tasks. There was also less facilitation in Experiment 5b where stimuli were presented as integrated colour-in-word stimuli. It may have been harder to read the word (and hence easier to ignore) because the word was much bigger in Experiment 5b.
Overall, Experiments 5a and 5b suggest that target name generation is necessary or at least highly advantageous for CP in delayed colour discrimination tasks. Whether test as well as target labelling is necessary for CP will be examined in the next experiment.
3.3 Experiment Six

Stroop interference at test presentation

3.3.1. Introduction

The results of Experiment 5 support the argument that CP requires naming of the target stimulus and that verbal interference in Roberson and Davidoff (2000) and Pilling et al. (2003) may not have directly affected CP, but rather indirectly prevented the target from being labelled. Therefore, it seems likely that no CP effect was found in the incongruent conditions in the previous experiment because target name generation was weakened. The aim of the current experiment was to investigate the effect of verbal interference at the response stage of a delayed discrimination task. As outlined in the previous chapter, there are two possible routes which could lead to CP after target name generation. Firstly, CP could arise due to direct matching of target with test labels: CP could arise because the target (e.g. blue) as well as the test stimuli (e.g. one blue, the other green) are labelled and the task is done by matching these labels. Secondly, target name generation could activate or reinforce the category code (see §2.1.2.) which facilitates cross-category, but not within category discrimination. If this leads to CP, test name generation would not appear to be necessary.

If it is true, as argued by Roberson and Davidoff (2000), that CP arises because of the comparison of verbal codes, the test stimuli must also be named correctly. This would suggest that not only target, but also test stimulus name generation is necessary for CP. In the current experiment there was no interference during target presentation, nor during ISI. Thus name generation and retention should be easy. However, there was interference at the response stage, enabling a test of the importance of name generation at this stage.

Stroop interference was introduced at the response stage by labelling the target and the distractor either congruently or incongruently. If CP is due to a simple matching of labels, congruent Stroop interference at the response stage should facilitate CP, incongruent interference should reduce CP. However, as mentioned above there is
another possibility. Name generation at target presentation stage may activate a
category code, and CP may arise because this category code reinforces the physical
code comparison. If this is the case, CP should not be affected by verbal Stroop
interference at the response stage since test name generation would not be necessary
for CP. The activated category code may reinforce the visual representation (possibly
similar to the mental representations referred to by Rosch (1975) and Neuman
&D'Agostina (1981)). If such visual representations are used in discrimination tasks,
the congruency and incongruency of Stroop interference at test presentation should
not selectively affect discrimination.

Integrated and non-integrated Stroop stimuli had very similar effects on within and
cross-category colour discrimination in Experiment 5. Since presenting the colour
with Stroop interference superimposed, rather than as an integrated stimulus, is closer
to the experimental design of the 2AFC and same-different tasks presented in Chapter
2, the current experiment used the non-integrated interference. The effect of Stroop
interference on within and cross-category colour discrimination was measured in a
2AFC task; Stroop interference was presented concurrently with the two test stimuli.
The two test stimuli were randomly labelled either blue or green, resulting in five
possible Stroop conditions. As in Experiment 5, there was neutral, congruent and
incongruent interference. Furthermore, there was a target-congruent/distractor-
incongruent and a target-incongruent/distractor-congruent condition. This was
necessary as the Stroop words would have been predictive of pairtype if only neutral,
congruent and incongruent conditions had been used. If both stimuli were either
congruently or both incongruently labelled, the labels would differ on cross-category
trials (one would always be blue, the other green) and match on within-category trials
(either both blue or both green). This was avoided by including mixed conditions.

In line with previous studies with no interference, a CP effect should be found in the
neutral interference condition. If CP is due to direct matching-to-labels, congruent
Stroop interference at response should facilitate CP relative to the neutral condition.
In contrast, incongruent Stroop interference should weaken name generation and
hence reduce CP. If, however, test name generation is not necessary for CP, the
advantage of cross-category over within category discriminations should survive
across all conditions. This would support an account of CP in which target name
generation activates a category code which facilitates cross-category, but not within category discrimination.

In Experiment 5 it was found that cross-category discriminations were faster than within category discriminations. Stroop interference, however, did not selectively affect reaction times. Presenting Stroop stimuli at test presentation in a 2AFC task is closer to the standard Stroop task in that an immediate response to the Stroop stimulus is required. Hence, if test name generation is necessary for CP, Stroop conditions should have selective effects on reaction time: congruency facilitates naming and should hence quicken response. Incongruency, on the other hand, weakens naming and should therefore slow down the response.

3.3.2. Method

Subjects
Twenty-eight subjects (11 male, 17 female, age range 18-42) participated in this experiment. All were students or staff at the University of Surrey; some received course credits for their participation. None had participated in the previous experiments.

Stimuli
The stimulus set was the same set of eight within category (A-a) and four cross-category pairs (A-B) used in Experiments 5. Stroop interference was presented at test presentation: the target and the distractor were presented concurrently with either a colour name (BLUE or GREEN) or XXXX in black uppercase lettering (font size 16) superimposed in the centre. This meant that there were now five interference conditions: neutral, congruent, incongruent, target-congruent/distractor-incongruent and target-incongruent/distractor-congruent. Congruency and incongruency of colour stimuli and colour words were again based on the naming data reported in Experiment 1. In the neutral condition XXXX was superimposed on both test stimuli. In the congruent conditions BLUE or GREEN (depending on whether the stimuli were from the blue or green category) were superimposed on the test stimuli; and in the incongruent conditions BLUE was presented with the green and GREEN with the...
blue stimulus. Accordingly, the target-congruent/distractor-incongruent and target-congruent/distractor-incongruent were mixed: one of the stimuli was congruently labelled whereas the other was incongruently labelled.

Procedure
As there were now five conditions, it was necessary to increase the number of trials to 240 in order to have equal number of trials in each interference condition whilst presenting each stimulus pair with equal frequency. The trials were split into six blocks of forty trials. Interference was again unblocked. Stimulus presentation times and ISI were the same as in the previous experiment. Hence, the target stimulus was presented for 150 ms, followed by a 5000 ms ISI after which the two test stimuli (together with the colour words or XXXX) were presented for 1000 ms or until a response was made. Presentations were balanced so that for each pair both stimuli were presented as the target equally often. The combination of position of target (left or right) and Stroop caption (XXXX and XXXX, BLUE and BLUE, GREEN and GREEN, BLUE and GREEN or GREEN and BLUE) determined which of the five conditions any given trial was in. There were 48 trials in each of the five interference conditions. Examples of trials are schematically presented in Figure 3.5.

Subjects had to choose which of the two test stimuli was the same as the target. Instructions were to ignore the colour words and attend to the colour of the stimuli only. As in the previous experiments, subjects responded by pressing the corresponding mouse-button (left, right). Instructions emphasised accuracy and speed. Three blocks of ten practice trials were completed before starting the experimental trials. Auditory feedback to incorrect response was given on the practice trials. Overall, the experiment took approximately 50 minutes to complete.
Fig. 3.5. Schematic representations of neutral (a), congruent (b), incongruent (c), and mixed (d,e) cross-category trials presented in Experiment 6.

3.3.3. Results

Reaction times
The mean reaction times across participants for each interference condition are presented in Table 3.3.
Table 3.3. Mean reaction time (standard error) ms in neutral, congruent, incongruent and mixed Stroop conditions for cross-category (A-B) and within category (A-a) discriminations.

<table>
<thead>
<tr>
<th>Stroop interference</th>
<th>Neutral</th>
<th>Congruent</th>
<th>Incon.</th>
<th>Target congruent</th>
<th>Distractor Congruent</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-B</td>
<td>938.7</td>
<td>977.2</td>
<td>989.3</td>
<td>949.0</td>
<td>969.0</td>
<td>964.7</td>
</tr>
<tr>
<td></td>
<td>(46.6)</td>
<td>(45.4)</td>
<td>(49.6)</td>
<td>(40.3)</td>
<td>(42.1)</td>
<td>(40.7)</td>
</tr>
<tr>
<td>A-a</td>
<td>985.5</td>
<td>1041.9</td>
<td>1031.1</td>
<td>1006.9</td>
<td>1031.1</td>
<td>1019.3</td>
</tr>
<tr>
<td></td>
<td>(42.0)</td>
<td>(57.0)</td>
<td>(49.7)</td>
<td>(55.4)</td>
<td>(49.7)</td>
<td>(49.6)</td>
</tr>
<tr>
<td>X</td>
<td>962.13</td>
<td>1009.5</td>
<td>1010.2</td>
<td>978.0</td>
<td>1000.0</td>
<td>1000.0</td>
</tr>
<tr>
<td></td>
<td>(41.7)</td>
<td>(49.0)</td>
<td>(46.4)</td>
<td>(46.7)</td>
<td>(44.2)</td>
<td></td>
</tr>
</tbody>
</table>

As shown above, cross-category discriminations were faster on average than within category discrimination in all interference conditions. Furthermore, Stroop interference seemed to affect both within and cross-category discriminations. Discriminations in the neutral conditions were fastest; hence the presentation of congruent and incongruent Stroop interference slowed down reaction times. It is interesting to note that reaction times were faster for both within and cross-category discriminations when the target was labelled congruent and the distractor incongruently compared to both target and distractor being congruently labelled. Mean reaction times were analysed by a two-way analysis of variance; the results showed that reaction times were significantly faster for cross-category pairs than within category pairs ($F[1, 27] = 6.7, \text{MSE} = 25.64.87, p<0.05$). The main effect of Stroop interference, was also significant ($F[4, 108] = 2.6, \text{MSE} = 9795.82, p<0.05$). The pair-type-interference interaction, however, was not significant ($F[4, 108] = 0.17, \text{MSE} = 7960.68, p>0.05$).
For within category pairs, discrimination in the neutral interference condition was significantly faster than discrimination in the congruent, the incongruent and the target-incongruent/distractor congruent conditions (minimum \( t[25] = 1.72, p<0.05 \)). For cross-category pairs, the discrimination in the neutral condition was only significantly faster than discrimination in the incongruent condition \( (t[25] = 1.91, p<0.05) \).

**Accuracy**

Mean accuracy for within and cross-category pairs is presented in Figure 3.6. As the graph shows, Stroop interference at the response stage did not produce clear results. Cross-category discrimination appeared to be slightly better than within category discrimination in the neutral condition\(^9\), however this effect was lost in all other conditions. In fact, within-category judgements were slightly better than cross-category judgements in all conditions. Results were analysed by a two-way repeated measures ANOVA, on pair-type (cross, within) and interference condition (neutral, congruent, incongruent, target-congruent/distractor-incongruent and target-incongruent/distractor-congruent). Neither pair-type \( (F[1,27] = 0.25, \text{MSE} = 232.35, p>0.05) \) nor interference \( (F[4,108] = 1.89, \text{MSE} = 97.56, p>0.05) \) or their interaction were significant \( (F[4,108] = 1.95, \text{MSE} = 66.08, p>0.05) \).

The effects of pair-type and interference on accuracy were re-analysed using a two-way ANCOVA with reaction time as a covariate. However, this did not change the overall pattern of results.

\(^9\) A paired-sample t-test showed that the difference in the neutral interference condition between cross-category and within category discrimination was significant \( (t[27] = 2.14, p<0.05) \).
Figure 3.6. Mean discrimination accuracy (error bars show +/- 1 standard error) for cross-category (A-B) and within category (A-a) pairs.

**Summary**

Cross-category discrimination was significantly faster than within category discrimination. Furthermore, reaction times were fastest in the neutral conditions compared to all Stroop conditions. Thus, congruent Stroop interference did not facilitate reaction times relative to the neutral condition. Whereas both main effects were significant when reaction time was the dependent measure, neither factor significantly affected accuracy. Discrimination was significantly better for cross-category compared to within category pairs in the neutral interference condition. However, this advantage was lost in all four Stroop conditions. It should be noted that this was due not only to cross-category discriminations being less accurate but also within category discriminations being more accurate in the Stroop interference conditions compared to the neutral condition.
3.3.4. Discussion

Stroop interference at target presentation and Stroop interference at test presentation did not produce easily comparable results. Whereas congruent Stroop interference at target presentation strongly facilitated CP in Experiment 5, congruent interference at response had no such clear effects. A CP effect was found in the neutral condition: discrimination of cross-category pairs was significantly better and faster than discrimination of within category pairs. For accuracy, however, this advantage was lost in all other conditions. Hence, the presence of colour words – whether congruent or incongruent – at the response stage eliminated the CP effect. For reaction times, CP was found in all conditions. Furthermore, responses to incongruently labelled targets were slower than to neutral targets but there was no difference between congruent and incongruent targets.

There are at least two possible explanations for these findings. Firstly, it is possible that the stimuli presented at the test stage in delayed discrimination tasks (such as Experiment 5) are not encoded and processed verbally. Labelling the test stimuli in the current experiment may have led to more confusion than facilitation because the discriminatory decision is usually done without the use of name codes. This would mean that the matching-to-labels account is incomplete. Schooler and Engstler-Schooler (1990) have shown that verbal processing does not always improve memory performance and argued that this is due to verbal overshadowing of the veridical visual image. The second possibility is that usually a verbal strategy is used (by comparing the target and test stimuli labels) but that the presence of many incongruent stimuli led to greater uncertainty at the response stage and hence the verbal strategy was abandoned. As was suggested earlier in relation to the blocked interference experiments (Roberson & Davidoff, 2000; Pilling et al., 2003), the target may not have been labelled in the first place. This would explain why the congruent and incongruent interference conditions had no differential effects on CP. Furthermore, in the mixed conditions the interference and facilitation may have cancelled each other out. Although the results are not clear-cut, the lack of evidence of facilitation in the congruent conditions does suggest that a direct matching of labels account of CP is unlikely.
There were, however, significant differences between Experiment 5 and Experiment 6. In both experiments the target was presented for 150ms and the two test stimuli were on screen for 1000ms or until a response was made. This meant that Stroop exposure duration differed across experiments. LaHeij et al. (2001) found less Stroop interference with short presentations times (120ms), compared to a continuous condition in which the stimuli were presented until response. The longer Stroop presentation (1000 ms) in Experiment 6 may have caused more interference than the short Stroop presentation (150ms) in Experiment 5 and so eliminated CP in all Stroop conditions in the current experiment. Also, the fact that there were two Stroop stimuli at the response stage and only one at target presentation may have meant that, overall, there was more interference and less facilitation in the current experiment.

A further difference between the two experiments was that the probability of congruent trials differed. The probability of a 100% congruent trial was .5 in Experiment 5 (target labelled correctly) and .25 in Experiment 6 (target and distractor labelled correctly). This meant that there were more incongruent trials in Experiment 5 and as MacLeod (1991) has pointed out, the ratio of congruent to incongruent trials can affect the strength of facilitation and interference respectively.

A third point is that the name codes given to target and test stimuli may have differed. In relation to the ‘perceptual magnet effect’ in auditory perception, Lotto, Kluender and Holt (1998) have argued that stimuli presented and labelled in isolation may be categorised differently than when they are presented in pairs. Hence, it is possible that the stimuli used in the current experiments were named differently when viewed in pairs (at the response stage) compared to in isolation (at target presentation). This could mean that in effect the congruent and incongruent conditions may not always have been appropriate. In Experiment 5 the label was given to the participants at target presentation. It is likely that, when presented again at the test stage, the target colour was labelled accordingly. However, in experiments with interference at the response stage, the naming uncertainty is added. The participant may label the target differently, especially if it is a boundary stimulus (see § 2.2.2.). If this is the case, the congruent labels given to the test stimuli would not be helpful. There would be no facilitation because the labels did not match the label given to the stimulus by the participant at target presentation.
Taken together the results of Experiments 5 and 6 support a language account of CP. If within and cross-category discriminations were based on perceptual codes, and CP was due to warping of perceptual space, Stroop interference should not affect discrimination accuracy or reaction times. Overall the results suggest that CP is based on target name generation, but not necessarily test name generation. It is possible, however, that the current experiments strongly favoured or invited a verbal strategy and hence that the effects of Stroop interference on CP are not generalisable to colour discrimination tasks per se (as was suggested in Experiment 4). It is also possible that Stroop interference had a general performance lowering effect because resources were used in trying to ignore the names. However, this does not explain the selective effects of congruent and incongruent interference on accuracy in Experiment 5.

Flowers and Dutch (1976) argued that the relative efficiency of verbal and visual codes in colour discrimination and classification tasks depends on the memory requirement of the task. The next experiment tests the effect of Stroop interference on CP when the memory load is significantly reduced and hence the efficiency of visual codes increased.
3.4. Experiment Seven

Stroop interference at target presentation in quick discrimination tasks

3.4.1. Introduction

Flowers and colleagues (Flowers & Blair, 1976; Flowers & Dutch, 1976) reported a series of experiments examining the effect of Stroop interference on colour classifications, colour scanning and colour matching. In Flowers and Dutch (1976) the task was essentially a visual search task; an array of 125 Stroop stimuli (printed in five columns of twenty-five colours) had to be scanned for target colours as quickly as possible. The stimulus lists were either presented as colour patches printed as XXXX or the colours formed incongruent colour names. The colours and colour names used were red, orange, yellow, green, blue, purple and green. Targets were either a combination of adjacent (red, orange and yellow or green, blue and purple) or non-adjacent colours (red, yellow and blue or orange, green and purple). The task was a disjunctive search for three different targets. The relationship among the three targets was varied in terms of whether the colours came from adjacent regions of colour space (red, orange and yellow; green, blue and purple) or from non-adjacent regions (red, yellow and blue; orange, green and purple). The task was to search for all three colours amongst the stimulus lists. The results showed that search was slowed by incongruency only when non-adjacent hues were targets. Flowers and Dutch argued that the search for three non-adjacent hues could not be done perceptually and hence the task required access to name codes. On the other hand the search for adjacent hues could be accomplished using only perceptual codes. They also found that search for single targets and pairs of targets (adjacent and non-adjacent) were not affected by Stroop interference. They argued that the memory load is a critical factor in determining whether the use of verbal or visual codes is more efficient.

If only tasks that require the use of verbal codes are susceptible to Stroop interference, the design of Experiments 5a and 5b (target retention over 5 seconds ISI) may have favoured the use of verbal codes and thus the finding that Stroop interference affects colour CP may be task specific. Experiments 7 and 8 looked at the effects of
decreasing the memory load by reducing the length of the inter-stimulus interval. The experiments tested the possibility that the task could now be done on a more perceptual level, by comparing or matching visual codes. If so it should not be susceptible to Stroop interference.

In Experiment 7, discrimination accuracy was again measured in a 2AFC task with neutral, congruent and incongruent Stroop interference at target presentation. The presentation times and length of inter-stimulus interval were reduced (relative to Experiment 5) in order to test Flowers and Blair's (1976) suggestion that quick discrimination tasks are not susceptible to Stroop interference. The ISI used was 1000 ms, based on Posner and Keele’s (1967) finding that in categorisation tasks visual codes were available for up to 1500 ms.

On the other hand, if, irrespective of stimulus duration and length of ISI, target name generation is necessary for CP, similar effects of congruency and incongruency on accuracy to those found in Experiment 5 should also be found. As Stroop interference was concurrent with the target and temporally separated from the response, Stroop conditions should not selectively affect reaction times, as in Experiment 5.

3.4.2. Method

Subjects
Twenty-one subjects (7 male, 14 female; age range 20-38) participated in this experiment. All were students at the University of Surrey. None of the subjects had participated in the previous experiments.

Stimuli and apparatus
The stimuli and apparatus were the same as those used in Experiment 5.
Procedure
The procedure was very similar to that of Experiment 5, except that test stimuli presentation times and the length of the ISI were reduced. The target plus Stroop interference was presented for 150ms, followed by a 1000ms ISI, after which the two test stimuli were presented for 150ms only. An example of a trial is presented in Figure 3.7. All other aspects of the procedure were the same as in Experiment 5. The experiment took approximately 20 minutes to complete.

Figure 3.7. Example of a congruent cross-category trial, both target and test stimuli were presented for 150ms and the ISI was reduced to 1000ms.

3.4.3 Results

Reaction times
Table 3.4. presents mean reaction times for all conditions across participants.
Table 3.4. Mean reaction time (standard error) ms in neutral, congruent, and incongruent Stroop conditions for cross-category (A-B) and within-category (A-a) discriminations.

<table>
<thead>
<tr>
<th>Stroop interference</th>
<th>Neutral</th>
<th>Congruent</th>
<th>Incongruent</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-B</td>
<td>764.9 (29.2)</td>
<td>756.2 (23.8)</td>
<td>764.7 (26.3)</td>
<td>761.9 (25.7)</td>
</tr>
<tr>
<td>A-a</td>
<td>760.9 (22.9)</td>
<td>767.7 (24.1)</td>
<td>780.1 (27.2)</td>
<td>769.7 (23.7)</td>
</tr>
<tr>
<td>X</td>
<td>762.9 (25.60)</td>
<td>761.9 (23.3)</td>
<td>772.7 (25.4)</td>
<td></td>
</tr>
</tbody>
</table>

As shown in Table 3.4, reaction times did not differ much across conditions. A two-way repeated measures ANOVA (pair-type by interference) confirmed that there were no significant effects (pair-type: $F[1,20] = 0.97$, MSE = 1983.32, $p>0.05$; interference; $F[2,40] = 1.22$, MSE = 1211.97, $p>0.05$); interaction ($F[2,40] = 0.62$, MSE = 1874.94, $p>0.05$).

**Accuracy**

Mean accuracy scores for within and cross-category pairs are presented in Figure 3.8. for each interference condition. Accuracy was generally higher than in the previous experiments. It appears that cross-category discrimination was better than within category discrimination in the no interference condition, however the CP effect was relatively small. In the congruent interference condition, the difference in accuracy between cross and within category discrimination was strongly facilitated, whereas in the incongruent condition the CP effect was lost.
Figure 3.8. Mean discrimination accuracy (error bars show +/- 1 standard error) for cross-category (A-B) and within (A-a) pairs.

Results were analysed by a two-way repeated measures ANOVA, on pair-type (cross, within) and Stroop interference (neutral, congruent, incongruent). The main effect of pair-type was highly significant ($F[1,20] = 18.18$, MSE = 64.86, $p<0.001$), as was the pair-type by interference interaction ($F[2,40] = 12.12$, MSE = 61.97, $p<0.001$), the main effect of interference, however, was not significant ($F[2,40] = 2.32$, MSE = 41.98, $p>0.05$).

Post-hoc testing by protected t-tests showed that there was a significant advantage for cross-category pairs over within category pairs in the neutral interference condition ($t[19] = 1.8$, $p<0.05$) and a highly significant advantage in the congruent condition ($t[19] = 6.71$, $p<0.01$). In the incongruent condition, however, the advantage of cross-category discrimination over within category discrimination was lost ($t[19] = 0.55$, $p>0.05$).

For cross-category pairs only, the results showed that accuracy was significantly higher in the congruent condition than in the neutral ($t[18] = 2.083$, $p<0.05$) and the incongruent conditions ($t[18] = 5.06$, $p<0.01$). Accuracy was also significantly higher in the neutral condition than in the incongruent condition ($t[18] = 3.01$, $p<0.01$). For
within category pairs only, accuracy in the congruent condition was significantly lower than accuracy in both the neutral and the incongruent condition (minimum \( t[18] = 3.3, p<0.01 \)).

A re-analysis of the effects of pair-type and interference on accuracy was performed taking reaction time into account. However, the two-way ANCOVA using median RT as a covariate found the same pattern of results.

**Summary**

Reaction times showed no CP effects; responses were no faster for cross-category discriminations; also congruency and incongruency did not have selective effects on reaction times. A significant CP effect for accuracy was found in the neutral condition, this effect was strongly facilitated in the congruent condition, but no CP was found in the incongruent condition.

**3.4.4 Discussion**

Discrimination was significantly more accurate for cross-category pairs compared to within-category pairs in the neutral interference condition. This effect was also found in the congruent condition, but CP was eliminated in the incongruent condition. No CP was found for reaction times in any interference condition. Reaction times were also unaffected by type of Stroop interference. The pattern of results for accuracy was very similar to those of Experiment 5, suggesting that the facilitation and elimination of CP by congruent and incongruent Stroop interference respectively, are not specific to long ISIs. If only tasks that invite or require verbal coding are susceptible to Stroop interference, then the results of the current experiment suggest that verbal labels are used even in the quick version of the task. However, the CP effect in the neutral condition was again relatively small. As mentioned in the discussion of Experiment 5, it is possible that different encoding strategy were used in the different conditions. Thus, discriminations in the neutral condition may have been based on perceptual codes.
It is also important to note that overall both cross-category and within category accuracy was higher than for a 5 second ISI (cross-category discrimination was on average 5.96%, within category discrimination 4.90% more accurate). Assuming that within category discriminations are primarily based on the comparison of perceptual codes (since comparison of verbal codes would result in a score at chance level), this suggests that perceptual codes are used more than in the 5000 ms ISI. They cannot, however, be responsible for the cross-category advantage in the neutral and the congruent condition since perceptual codes should not be susceptible to Stroop interference. Therefore, if this task was done on the basis of perceptual comparisons alone, incongruent Stroop interference should not have had a selective effect on CP. Reaction times were shorter in the current experiment (cross-category discrimination was on average 103.0ms faster; within-category search 147.0ms faster), however the analysis of reaction times showed there was no selective effect of Stroop interference in either experiment. This supports the suggestion made earlier that since no response was required immediately after Stroop presentation, the mechanisms underlying the effects found here may be different to those of standard Stroop tasks.

The finding that presenting the correct colour label at target presentation strongly facilitated CP, whereas preventing the correct name to be generated reduced CP was also found for short presentation times and ISIs, lends further support to a language account of colour CP in which target name generation plays a particularly important role. The results do not support Flowers and Dutch’s suggestion that reducing the memory load in Stroop tasks leads to processing on a perceptual level.
3.5. Experiment Eight

Stroop interference at test presentation in quick discrimination tasks

3.5.1. Introduction

The previous experiment tested the argument put forward by Flowers and colleagues (Flowers & Blair, 1976; Flowers & Dutch, 1976) that tasks with low memory loads and requiring quick perceptual decisions are not susceptible to Stroop interference. However, the results of Experiment 7 showed that congruent and incongruent Stroop interference affects colour CP even in tasks with shortened presentation times and short ISI. Both target and test stimuli were only presented for 150ms, the inter-stimulus interval was 1000ms, hence memory load was small and the discrimination judgements had to be made quickly. Congruency facilitated CP, incongruency reduced CP which suggests that target name generation was important even in quick discrimination tasks. LaHeij, Heiden & Plooij (2001) argued that the magnitude of interference is a function of the duration of the Stroop stimulus. Hence, if the length of Stroop presentation rather than other task variables (such as length of inter-stimulus delays) determines the strength of the effect, the similar patterns found in Experiments 5 and 7 could be simply due to Stroop presentation times being the same in both experiments. Furthermore, the long Stroop interference (1000ms) at test presentation in Experiment 6 may have meant that interference was increased. Hence, if exposure duration determines the effects of Stroop interference, reducing the Stroop interference at response to the same as that at target presentation (see Experiment 5) should produce comparable results.

Within and cross-category discrimination was measured using a 2AFC task with neutral, congruent, incongruent, target-congruent/distractor-incongruent and target-incongruent/distractor-congruent Stroop interference. As in the previous experiments, discrimination of cross-category pairs should be more accurate than discrimination of within category pairs in the neutral condition. If CP is due to a direct matching-to-labels process, congruency and incongruency at test stage should increase CP or reduce CP respectively: congruency should help name generation, thus facilitating
cross-category discrimination. Incongruency, on the other hand, should weaken name generation and so reduce the advantage of cross-category discrimination. Furthermore, if test name generation is necessary for CP, Stroop conditions should have selective effects on reaction time. If however, test name generation is not necessary for CP, Stroop interference at the response stage should affect neither within nor cross-category discrimination accuracy or reaction times and hence CP should be unaffected.

3.5.2. Method

Subjects
Twenty-eight subjects (13 male, 15 female, age range 19-46, mean age 25.7 years) participated in this experiment. All were students at the University of Surrey and were either paid or received course credits for their participation.

Stimuli and apparatus
The stimuli and apparatus were the same as those used in Experiment 6.

Procedure
The procedure was the same as that of Experiment 6 except for the length of the ISI and the presentation time of the test stimuli. The target was presented for 150ms, followed by a 1000ms ISI, after which both test stimuli plus Stroop interference were presented for 150ms. An example is presented in Figure 3.9. All other aspects of the procedure were identical to that of Experiment 6. The experiment took about 25 minutes to complete.
Fig. 3.9. Example of an incongruent cross-category trial used in Experiment 8.

### 3.5.3. Results

*Reaction times*

Mean reaction times across participants for each interference condition are presented in Table 3.5.

Table 3.5. Mean reaction time (standard error) ms in neutral, congruent, incongruent, and mixed Stroop conditions for cross-category (A-B) and within-category (A-a) discriminations.

<table>
<thead>
<tr>
<th>Stroop interference</th>
<th>Neutral</th>
<th>Congruent</th>
<th>Incon.</th>
<th>Target congruent</th>
<th>Distractor Congruent</th>
<th>( \bar{x} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-B</td>
<td>738.8</td>
<td>743.1</td>
<td>732.5</td>
<td>735.7</td>
<td>743.3</td>
<td>738.7</td>
</tr>
<tr>
<td></td>
<td>(23.6)</td>
<td>(25.8)</td>
<td>(25.9)</td>
<td>(22.9)</td>
<td>(29.2)</td>
<td>(24.0)</td>
</tr>
<tr>
<td>A-a</td>
<td>741.9</td>
<td>751.7</td>
<td>751.4</td>
<td>741.4</td>
<td>751.4</td>
<td>747.6</td>
</tr>
<tr>
<td></td>
<td>(23.8)</td>
<td>(25.5)</td>
<td>(23.5)</td>
<td>(19.8)</td>
<td>(23.5)</td>
<td>(22.5)</td>
</tr>
<tr>
<td>( \bar{x} )</td>
<td>740.4</td>
<td>747.4</td>
<td>741.9</td>
<td>738.5</td>
<td>747.3</td>
<td>(23.1)</td>
</tr>
<tr>
<td></td>
<td>(23.1)</td>
<td>(24.6)</td>
<td>(24.2)</td>
<td>(20.8)</td>
<td>(25.6)</td>
<td></td>
</tr>
</tbody>
</table>
As it appears in Table 3.7, reaction times did not differ much across Stroop conditions and cross-category discriminations were again faster than within category discriminations in all conditions. Analysis by a two-way repeated measures ANOVA (pairtype by interference) confirmed that neither main effect was significant (pair-type: $F[1,27] = 1.52$, MSE = 3623.42, $p>0.05$; interference: $F[4,108] = 0.38$, MSE = 2424.95, $p>0.05$).

**Accuracy**
Mean accuracy scores for within and cross-category pairs are presented in Figure 3.10. It appears that reducing the ISI and the test presentation duration made overall discrimination more accurate than Experiment 6, especially cross-category discriminations. Furthermore, cross-category discriminations appear more accurate than within category discriminations in all conditions. However, cross-category discrimination accuracy in the target-incongruent/distractor-congruent condition seems lower than in the other conditions.

---

10 Average cross-category accuracy was 71.0% in Experiment 5 and 75.6% in the current Experiment.
Fig. 3.10. Mean discrimination accuracy (error bars show ±1 standard error) for cross-category (A-B) and within category (A-a) pairs.

Results were analysed by a two-way repeated measures ANOVA, pair-type (cross, within) and interference condition (neutral, congruent, incongruent, target-congruent/distractor-incongruent and target-incongruent/distractor-congruent). There was a highly significant effect of pair-type ($F[1,27]=16.75$, $MSE = 289.10$, $p<0.001$), but neither interference ($F[4,108]=1.47$, $MSE = 61.54$, $p>0.05$) or the interaction ($F[4,108]=0.77$, $MSE = 70.35$, $p>0.05$) was significant.

Post-hoc testing by protected t-tests showed that there was a significant advantage for cross-category pairs over within category pairs (hence a CP effect) in the neutral ($t[26] = 1.98$, $p<0.05$), the congruent ($t[26] = 1.88$, $p<0.05$), the incongruent ($t[26] = 2.19$, $p<0.01$), and the target-congruent/distractor-incongruent condition ($t[26] = 2.07$, $p<0.05$).

For cross-category pairs only, the results show that accuracy was significantly lower in the target-incongruent/distractor-congruent condition compared to both the incongruent ($t[23] = 2.47$, $p<0.05$) and the target-congruent/distractor-incongruent
\(r[23] = 2.66, p<0.05\) conditions. All other differences in accuracy were not significantly different. For within category pairs, there were no significant differences between any of the Stroop interference conditions (maximum \(r[23] = 0.64, p>0.05\)).

The results were re-analysed by means of a two-way ANCOVA using reaction time as a covariate; this did not change the overall pattern of results.

**Summary**

Cross-category discriminations were not significantly faster than within category discriminations. Furthermore, reaction times did not vary significantly across Stroop conditions. Cross-category discrimination was, however, significantly better than within-category discrimination in the neutral Stroop condition. CP was also found in the congruent, the incongruent and the target-congruent/distractor-incongruent conditions. Overall, Stroop interference at test presentation did not have a significant effect on CP.

### 3.5.4. Discussion

The current experiment and Experiment 6 differed only in the length of the inter-stimulus interval and presentation times of the test stimuli. The results of the two experiments, however, differed markedly. In Experiment 6 CP was lost in all Stroop conditions. In the current experiment, cross-category discrimination was better than within category discrimination in all conditions. This CP effect was significant in all but the target-incongruent/distractor-congruent condition. Crucially, CP was found independent of type of Stroop interference. The prediction that Stroop interference at test presentation would affect reaction times was not supported. It seems likely that the short test presentation time (150ms compared to 1000ms in Experiment 6) meant that Stroop interference was easier to ignore making reaction times more similar across conditions. The finding that CP as indicated by accuracy survived in all conditions supports this idea. The results further support LaHeij et al.’s theory that long Stroop presentations cause more interference than do short presentations.
The results of the current experiment partly support Flowers and Dutch's (1976) suggestion that tasks with lower memory loads and quick processing demands are not be susceptible to Stroop interference. As there was no evidence for facilitation or interference across Stroop conditions, the results suggest that test name generation is not necessary for CP. However, the fact that Stroop interference had an effect on CP in Experiment 7 suggests that different levels of processing are used at different stages of the task. Presumably, target name generation was also necessary for CP in the current experiment. These results support the idea that CP may be a language effect in that target naming is necessary, but it may not be a simple matching-to-label process.
3.6. General Discussion

In the previous chapter it was argued that Roberson and Davidoff’s (2000) account of colour CP was incomplete as CP was found to survive verbal interference when conditions were unblocked. This suggested that tasks were open to different strategies, the efficiency of visual and verbal codes may be task dependent. Chapter 3 addressed two questions: firstly, is CP due to naming the target and the test stimuli and a subsequent comparison of labels. Secondly, does the use or efficiency of verbal codes on tasks vary depending on design (such as length of ISI) and hence on demands of the task. Stroop interference was used to test the role of colour naming in CP. This type of interference was used as it is tasks which require the use of or access to verbal labels that are susceptible to Stroop interference (Flowers & Dutch, 1976; MacLeod, 1991).

To address the first question, Stroop interference was either presented concurrently with the target stimulus (Experiment 5) or the test stimuli (Experiment 6). If target name generation is necessary for CP, preventing the target from being named correctly should reduce CP. Effects of verbal interference at encoding have previously been reported to reduce colour CP (Roberson & Davidoff, 2000, Experiment 4). The results of Experiment 5 supported this prediction. CP was eliminated in the incongruent interference condition but strongly facilitated in the congruent interference condition. Thus, having the correct target name code resulted in CP.

If CP is due to a comparison of target and test stimuli labels, the test stimuli must also be named correctly and CP should be susceptible to Stroop interference at test presentation. The results of Experiment 6 were not as straightforward. All types of Stroop interference, whether congruent or incongruent, reduced CP. Various possible explanations for these findings were discussed. It is possible that this was due to more interference being caused by longer Stroop presentations (in line with LaHeij et al, 2001) and two Stroop stimuli being present simultaneously. Alternatively, it is possible that the congruent/incongruent conditions did not always match the name given to the target on target presentation. Thus, the results of Experiment 6 did not answer the question of whether test name generation is essential for CP.
The second question, whether the use and efficiency of name codes depends on aspects of the task design was tested in Experiment 7 and 8. According to Flowers and colleagues (Flowers & Blair, 1976; Flowers & Dutch, 1976) tasks with low memory loads that can be done on the basis of quick perceptual decisions are not susceptible to Stroop interference. Stimuli presentation times and the length of the ISI were reduced in Experiments 7 and 8 (relative to Experiment 5 and 6) in order to test this.

The pattern of results for accuracy found in Experiment 7 paralleled those of Experiment 5. Hence, verbal codes seemed to be used even when the memory load was substantially reduced. However, CP was not affected by Stroop interference in Experiment 8 (interference at response with short ISI), congruent interference did not facilitate CP, incongruency did not reduce CP. This suggests that no access to the verbal codes of the test stimuli was required. Taken together the results presented in the current chapter support the idea that target name generation is necessary for CP. The findings also suggest that it is not a simple matching-to-labels process. Furthermore, the finding that Stroop has selective effects at target but not at test presentation was not an artefact of the long inter-stimulus interval and presentation times used in Experiments 5 and 6.

There are some potential problems that come with introducing Stroop interference to such colour discrimination tasks. Firstly, the fact that the colours used in the current experiments were boundary and not focal colours makes a comparison to other Stroop studies difficult. In standard Stroop tasks there is no uncertainty as to whether the stimulus dimensions are congruent or incongruent since the colours are usually prototypical colours. However, for boundary stimuli naming consistency is lower and certain boundary stimuli may be categorised as blue by some and as green by others. Hence, it is likely that there was some degree of uncertainty about the congruency and incongruency of Stroop stimuli which may have affected performance on the task.

It is also likely that presenting Stroop interference at target and at test presentation affected different processes or mechanisms and hence had differential effects on discrimination accuracy and reaction times. Stroop stimuli at test presentation could be said to be closer to the standard Stroop task since the response had to be given
immediately after Stroop presentation. Stroop interference, however, only affected reaction times in Experiment 6. When all stimulus presentation times and length of ISI were matched (Experiment 7 and 8), reaction times were not affected by Stroop interference or pair-type. Furthermore, in a comparison of Experiment 7 and 8, Stroop interference at test presentation affected CP less than Stroop interference at target presentation. It is possible that Stroop interference at test presentation was easier to ignore since a response was required immediately and the stimuli did not have to be held in memory.

Introducing Stroop interference at target presentation may have effectively cued the verbal code and so encouraged or even forced participants to use a verbal strategy on the discrimination task in the congruent and incongruent conditions. It could therefore be argued that the strong facilitating effect of congruent Stroop interference on CP does not demonstrate that naming the target is necessary for CP, but rather that CP arises because the labels were provided in the first place. The fact that in both Experiments 5 and 7 the CP effect in the neutral condition was relatively small supports this idea. As mentioned earlier it is possible that participants did not have enough time to generate a verbal label, or that they chose to adopt a perceptual strategy. The evidence of CP in the neutral condition could be taken to suggest that CP was sometimes based on the comparison of perceptual codes. Overall, the interpretation of these findings is not straightforward, the results do, however, show that when the target stimulus cannot be named correctly, there is no evidence for CP.

The results of the Stroop experiments are inconsistent with a perceptual warping account of CP (Goldstone, Lippa & Shiffrin, 2001; Harnad, Hanson & Lubin, 1991). Within category judgements which must be primarily based on the comparison of visual codes, were unaffected by Stroop interference. Hence, warped perceptual codes should also be available in incongruent cross-category conditions to produce CP, this however was not the case. The findings support a verbal account of CP in colour discrimination tasks and are in line with the idea that verbal interference prevents or weakens the verbal coding of the colour and/or discourages attempts to use a verbal strategy. Being given or being able to generate the correct verbal colour code in a CP colour memory task strongly facilitates CP and the results are consistent with the idea that CP requires the naming of the target stimulus. The results further suggest that
explicit naming of the two test stimuli at the response stage is not necessary for CP. Taken together these results also suggest the matching-to-labels account of CP is incomplete.

If CP is not the result of a simple matching-to-labels process but target name generation is necessary, what leads to CP? As outlined earlier it is possible that target name generation activates a category code which facilitates cross-category discrimination. The work on priming by Rosch (1975) and Neuman & D’Agostino (1981) may be relevant here. It suggests that priming with a colour word in colour discrimination tasks leads to a mental representation that is visual in nature but is not as specific as the actual physical code. The Stroop word at target presentation could be thought of as acting in a similar fashion as the word prime. If this were the case, cross category discrimination would be facilitated as the label may reinforce the category code.
Chapter Four

Is categorical perception due to a shift towards the prototype?

4.1. General Introduction

CP was found in the 2AFC, same-different and Stroop tasks presented in Chapters 2 and 3: in conditions with no interference discrimination accuracy was higher for cross-category compared to within category pairs. However, certain types of interference seemingly remove this CP effect. Roberson and Davidoff (2000) and Pilling et al. (2003) have found verbal interference tasks in blocked conditions to reduce the advantage of cross-category pairs and incongruent Stroop interference also eliminated CP (Chapter 3). These results do not support a perceptual account of CP; if the advantage of cross-category over within category pairs stemmed from warped perceptual space, these perceptual codes should also have been available in verbal interference conditions. However, the results of Chapter 3 also make a direct language account of colour categorical perception unlikely since the naming of test stimuli at the response stage in discrimination tasks does not seem necessary for CP. Overall the findings presented in Chapters 2 and 3 suggest target name generation to be necessary for CP. The previous chapter raised the possibility that naming the target may activate a category code, which would facilitate cross-category but not within-category discrimination. However, the question remains of how these category effects arise during processing in delayed discrimination tasks.

All tasks in the above mentioned experiments had a memory component, hence the possibility that CP is a memory phenomenon rather than a perception or a language effect needs to be considered. The shift towards prototype (Huttenlocher, Hedges & Vevea, 2000), reported in the memory literature, and the perceptual magnet effect, reported in the speech perception literature, may be closely related to categorical perception. All three phenomena involve non-uniform judgements of uniform stimulus dimensions, thus the discriminability of stimuli varies as a function of their positions within a category. The aim of the current chapter is to test whether the shift towards prototype, the perceptual magnet effect and CP are related phenomena.
Furthermore, the shift towards prototype account will be used to test whether CP is a memory effect rather than a perceptual or a language effect.

4.1.1. Perceptual magnet effect

The perceptual magnet effect is a phenomenon reported in the speech perception literature and is characterised by a warping of perceptual space around phonemic category centres (Guenther & Gjaja, 1996). The discriminability of speech sounds has been found to be lower around category prototypes and higher towards category boundaries. For example Kuhl (1991) and Lotto, Kluender and Holt (1998) have found discriminability of vowels as measured by $d'$ to be worse in the centre of phonemic categories, than towards boundaries. According to Kuhl (1991) the prototype is defined as the best example of a vowel and other vowels of the same category are drawn towards this prototype increasing the similarity of stimuli around the prototypical example. Since the perceptual space seems to shrink around the prototype, the effect has been termed the perceptual magnet effect.

4.1.2. Shift towards prototype

Huttenlocher, Hedges and Vevea (2000) argued that the non-uniformities found in judgements of stimuli are due to a representational shift, or bias, towards the category prototype. According to this account, the goal of stimulus judgements is to achieve high accuracy and in order to achieve this goal the inexact memory representation is placed in the context of prior information. Huttenlocher et al. proposed a Bayesian model of this procedure. The premise of the model is that, as uncertainty increases, it is beneficial to assume the stimulus was the prototype, where the prototype is the average of all the stimuli seen. Thus, categorical information is used to adjust inexactlly represented stimuli and this gives rise to a bias that improves accuracy overall by reducing the variability of estimates.

Uncertainty about a remembered stimulus, either due to sensory limits, or memory decay, results in a bias towards the best example of the category. Huttenlocher and
colleagues have found shift towards prototype effects in estimates of spatial location (Crawford, Regier & Huttenlocher, 2000; Huttenlocher, Hedges & Duncan, 1991) and simple shapes as well as line length (Huttenlocher, Hedges & Vevea, 2000). This bias towards the prototype implies that the memory representations of stimuli are shifted towards the prototype. In the above mentioned studies the stimulus had to be reproduced and hence the shift was measured as the distance between the original stimulus and the reconstructed stimulus. However, if the shift is a characteristic of perceptual memory, it should also be evident in delayed discrimination tasks.

In colour recognition tasks, Belli (1988) found that an object’s typical colouration influences recognition and that this resulted in memory representations that blend the actual and the typical colour. Also, recognition may be simultaneously influenced by typical knowledge, event information and post-event information. However, these studies required the recall or recognition of the colour of real objects. An important question then is whether similar processes also determine the recall, recognition and reconstruction of simple stimuli such as the colours used in hue discrimination tasks. According to Huttenlocher et al. (2000), comparable processes develop during memory tasks for simple stimuli or spatial locations where prior information is the average of all stimuli seen during the experiment. The distributions and frequencies determine where the prototype is located and therefore the direction of the bias in memory.

The shift towards prototype and categorical perception have been studied with different methods in different stimulus domains. However, they may have common underlying mechanisms and both may be related to the perceptual magnet effect reported in the speech literature. To test whether a shift towards prototype occurs in colour memory tasks and whether CP found in these tasks could be due to a memory bias towards the prototype, the results of the previous experiments were re-analysed.
4.1.3. CP, perceptual magnet effect and shift towards prototype

Both CP and the shift towards prototype offer accounts of the apparent warping of similarity space. The result of warping is that discrimination accuracy peaks at the category boundaries and discrimination minima are found around the prototypes (Harnad, 1987). This is also consistent with a perceptual magnet effect. Using the blue-green region of colour space as an example, this is illustrated in Figure 4.1. The space is a continuum represented by stimuli at equal intervals ranging from blue to green (denoted as B4-G4). Although CP and the shift towards prototype seem to be describing related phenomena, the proposed loci of these effects differ: in perceptual accounts of CP the locus is at encoding. The shift towards prototype account locates the categorical effects in reconstructive processes in short term memory.

![Figure 4.1: Warping of perceptual colour space](image)

Fig. 4.1. Warping of perceptual colour space. The discrimination maximum is at the blue-green boundary, the minima around the prototypes. P_b represents the location of the blue category prototype, P_g the location of the green category prototype. Memory errors in the direction of the arrows.

Özgen and Davies (2002) tested whether CP could be induced by training participants to split a pre-existing colour category into two new categories. Participants were either trained to split the green category into two new categories or to perform the respective separation in the blue category. The blue stimuli ranged from Munsell 3.75B to 1.25PB. The mid-point (around 7.5B) was the best example of blue; the blue
The green stimuli ranged from 3.75G to 1.25BG; the prototype was around 7.5G. The pre-training data allowed a test of whether there were non-uniformities in the discriminability of colour stimuli as a function of category position (as pictured in Figure 4.1.). These data were reported by Davies, Özgen, Pilling and Wiggett (2003). The discrimination accuracy for pairs around the prototype and near the boundary are presented schematically in Figure 4.2.

![Diagram showing accuracy around prototypes](image)

**Fig. 4.2.** Accuracy ($A'$) around the prototypes (dashed line) compared to either side of the prototypes (solid line). The data is collapsed across blue and green (reproduced from Davies, Özgen, Pilling & Wiggett, 2003).

As Figure 4.2 shows, Davies et al. found discrimination around the prototypes of the blue and green categories to be worse (0.55) than discrimination either side of the prototype (0.63/0.64). Thus, discrimination around the centre of the colour categories was considerably less accurate than discrimination of peripheral stimuli. This is consistent with a perceptual magnet effect and led Davies et al. to suggest that perceptual magnet effects are found not only in speech but also in colour perception. The results show that the discriminability of colour stimuli varies as a function of category position.

The next question addressed by Davies et al. (2003) was whether these non-uniformities in the discriminability of colour stimuli were driven by a shift towards prototype. If a shift towards prototype occurs in colour discrimination, order effects of stimulus presentation should be found. In successive same-different judgements, as
the memory representation becomes less accurate across the ISI, the estimate of these stimuli should shift towards the prototype. This idea is illustrated in Figure 4.3.

Figure 4.3. The effect of a shift towards prototype. $P_b$ denotes the location of the blue category prototype; target indicated by box and (a) and (b) show different stimulus orders, (a') and (b') show the respective separations in similarity space. (a') shows a decreased separation for peripheral pairs, (b') shows an increased separation.

The effect of a shift towards prototype on distances in representational space depends on order of stimuli presentation. The effect of presenting the stimuli furthest from the prototype as the target is shown in (a). The representation of B4 or B1 will shift towards the category prototype (a'). This has the effect of decreasing the separation between B4-B3 and B1-B2. However, as shown in (b) if the stimulus closer to the prototype is presented first, the resultant separation between (B4-B3) increases (b'). Thus, the order of presentation affects the task differently and should lead to order effects.

In order to test whether this shift actually occurs in colour discrimination tasks, Davies et al. analysed same-different accuracy as a function of stimulus order. On different trials one of the stimuli in a pair was presented first (as the target), this can be either the stimulus closer to the category centre or the peripheral stimulus. Davies et al. re-analysed the pre-training data (from Özgen & Davies, 2002, presented in
Figure 4.2) by order of presentation for peripheral pairs. The result of the analysis of order effects is presented in Figure 4.4.

![Diagram](image)

Fig. 4.4. Same-different accuracy (A') for peripheral pairs. Box indicates target, the target is linked to test stimulus by horizontal line. Scores are collapsed across both pairs within a colour category (e.g. B4 → B3 and B1 → B2). The difference between order 1 and order 2 was statistically significant (p<0.05) (reproduced from Davies et al., 2003).

Davies et al. report a clear order effect for within-category pairs. Discrimination accuracy was significantly higher when the stimulus closer to the category prototype was the target and hence presented first. These findings are consistent with the idea that the distance between the two stimuli in representational space either increases or decreases depending on the order of presentation. A shift towards prototype would predict that central targets (B3, B2, G2, G3) should lead to greater accuracy (increased separation) than non-central targets (B4, B1, G4, G1; decreased separation). The results support the predictions from a shift towards prototype account. Therefore, the results support the possibility that a shift towards prototype occurs in colour memory tasks.
4.1.4. Categorical Perception and Shift towards prototype

An interesting question that arises from Davies et al.'s analysis is whether a shift towards prototype can account for CP. If so, CP may be a memory rather than a perception or a language effect. Figure 4.5. demonstrates how a shift towards prototype predicts CP and the order effects predicted for within category discriminations.

Fig. 4.5 Schematic representation of stimuli used in the categorical perception experiments in Chapter 2 and 3. The stimuli are collapsed across the four lightness levels. Stimuli representations shift towards the prototype. Shift towards prototype arguments predict CP.

When testing CP in Chapters 2 and 3, the stimulus set straddled a category boundary. As shown above this meant that the category prototypes (Pg and Pb) were not part of the stimulus set. However, a shift towards prototype predicts not only order effects for within category discriminations but also CP. For cross-category discriminations, the target (either B1 or G1) should shift towards the respective prototype and hence the separation between B1-G1 should lead to increased discrimination accuracy. For within category pairs, the effect depends on order of presentation. Hence, B1 followed by B2 would decrease the separation whereas B2 followed by B1 would increase the
separation. Average within category separations should be less than cross-category separations, thus producing categorical perception effects.

4.2. Re-analysis of 2AFC and same-different colour discrimination tasks

In order to test the shift towards prototype conjecture, the results of Experiment 1 (2AFC), Experiment 2 (same–different) and Experiments 5-8 (the Stroop tasks) were re-analysed. In the original analysis, order was treated as a balancing factor, and not analysed. Here, the tasks are re-analysed with order of presentation as a factor. The experiments testing CP included verbal and visual interference conditions. It was thought possible that different types of interference may selectively affect order effects. Therefore, data are presented separately for each interference condition as in Chapters 2 and 3 and so the graphs do not correspond directly to the figures used by Davies et al. (2003).

4.2.1. Re-analysis of Experiment One: 2AFC

The primary task in Experiment 1 was a 2AFC discrimination task which was combined with no, verbal and visual interference. If a shift towards prototype occurs, within category discrimination accuracy should depend on order of presentation.

(for Method see § 2.2.3.)

Results
The results were re-analysed by looking at within category discrimination accuracy for boundary and non-boundary targets separately. As in the original analysis in Chapter 2, results were collapsed across blue and green stimulus pairs. The results are presented in Figure 4.6. Cross-category discrimination accuracy is also presented.
As the graph shows, no order effects were found in Experiment 1. Since CP effects were found in the experiment, the results suggest that CP is not driven by a shift towards prototype in a 2AFC discrimination task.

4.2.2. Re-analysis of Experiment Two: same-different

The primary task in Experiment 2 was a same-different discrimination task which was combined with the same verbal and visual interference tasks as the 2AFC discrimination task in Experiment 1.

(for Method see § 2.3.2.)
Results

The re-analysis of Experiment 2 for order effects is presented in Figure 4.7. The results were collapsed across blue and green stimulus pairs.

As the graph shows, strong order effects were found for within category discriminations. Discrimination accuracy was higher when the target was the boundary stimulus compared to when the non-boundary stimulus was presented as the target. Results were analysed by a two-way repeated measures ANOVA, on order of presentation (boundary, non-boundary) and interference (no, visual, verbal). Both order of presentation ($F[1,25] = 48.62$, $MSE = 0.0018$, $p<0.001$) and interference ($F[2,50] = 4.35$, $MSE = 0.001$, $p<0.05$) were significant main effects. Post-hoc analysis by protected t-tests showed the order effect to be highly significant in all interference conditions (minimum $t[22] = 10.15$, $p<0.001$).

The re-analysis of Experiment 2 showed clear order effects, however the order effect was in the opposite direction from that predicted. This is demonstrated in Figure 4.8.
Fig. 4.8. Schematic representation of order effects found in Experiment 2. $A'$ for within category pairs as a function of stimulus order. Data is collapsed across interference conditions. Cross-category accuracy was 0.73. Box indicates target.

Discrimination for within category pairs was more accurate when the boundary stimulus (B1 or G1) was the target, compared to when the non-boundary stimulus was the target (B2 or G2). The shift was in the opposite direction to the predicted shift. A shift in this direction, i.e. towards the category boundary, is not consistent with CP since the separation across the boundary (B1-G1) would decrease. Therefore, the categorical perception effect in Experiment 2 cannot be due to a shift towards prototype.

To summarise the results of the re-analysis of Experiments 1 and 2, no within category order effects were found when the task was run as a 2AFC discrimination task, but strong order effects were found for the same-different task. The effects, however, were in the opposite direction from that predicted: within category discrimination accuracy was higher when the boundary colour was the target compared to when the non-boundary colour was the target.
Discussion

In the re-analysis of Experiments 1 and 2, it was assumed that the prototypes were the pre-existing prototypes, the best example of blue or green (as shown in Figure 4.5). The prototypical blue or green were assumed to be further from the boundary than any of the test stimuli. It is important to note that Huttenlocher et al.'s theory of a shift towards prototype was intended to apply to temporary categories rather than pre-existing categories. In the shift towards prototype studies the prototype was defined as the average of the stimulus set experienced during the experiment.

If we apply this definition of a prototype to the re-analysis of the current experiments, the prototype would fall on the category boundary and the predicted shifts would be in the opposite direction to those shown in Figure 4.5. If the prototype was the average of all stimuli used in the experiment, the shift would have the effect of reducing cross-category separations and, for within category judgements, boundary-first separations would be greater than non-boundary first separations. The two possible prototypes and the respective directions of the shift are presented are presented in Figure 4.9.

Fig. 4.9. Diagram showing two possible directions of the shift towards prototype. The direction of the shift is dependent on whether the prototype is defined as the average of all stimuli (P_t) or the pre-existing prototypes (P_b, P_g).

A shift towards the temporary prototype (P_t) would predict discrimination of boundary targets to be more accurate than discrimination of non-boundary targets. A shift towards the pre-existing prototypes, however, would predict the opposite: the shift would increase stimulus separation on non-boundary target trials.
The predictions that follow from a shift towards the temporary prototype are met for within category discriminations in the re-analysis of Experiment 2; however cross-category discriminations should be less accurate than within category discriminations. Given that significant CP effects were found in all conditions this was clearly not the case. Thus, the fact that no order effects were found in the 2AFC discrimination task and that the effects found in the same-different task cannot account for the CP effects, suggests that categorical perception is dissociated from a shift towards prototype.

Although the order effect found in Experiment 2 is unlikely to be due to a shift towards prototype, it is nevertheless an interesting result since it shows CP - the advantage of cross-category over within category pairs - to be entirely due to the low discrimination accuracy for within category discriminations for non-boundary targets. Note also that there is evidence to suggest that this is a robust effect. Significant order effects for within category discrimination in same-different tasks were also found in the re-analysis of Özgen & Davies’ (2002) as well as Pilling’s (2001) results (reported by Davies, et al., 2003).

The comparison of the re-analysis of Experiments 1 and 2 showed that task design affects within category order effects; order effects were found for the same-different but not the 2AFC task. The order effects found in the same-different task, however, do not predict CP.
4.3. The effects of Stroop interference on within category order effects

The question of whether the shift in colour discrimination is towards the pre-existing prototype (prototypical blue or green in the current experiments) or a temporary prototype (the average of all stimuli in the set) can also be addressed by re-analysing Experiments 5-8. Colour CP effects have been found in range of colour discrimination tasks (Bornstein & Korda, 1984; Boynton, Fargo, Olsen & Smallman, 1989; Laws, Davies & Andrews, 1995; Uchikawa & Shinoda, 1996); note however that a shift towards a temporary prototype – in CP experiments the category boundary – would predict cross-category discrimination to be worse than within category discrimination. Hence, it is unlikely that the category boundary was the prototype in the above mentioned experiments, or, if it was, memory representations did not shift in this direction. However, it is of course possible that different processes are responsible for cross-category and within-category discrimination. It is also possible that both the temporary and the pre-existing prototype affect discrimination. For instance, naming the stimuli may invoke the pre-existing prototype and using a visual strategy may mean that the visual memory representation shifts more towards the temporary prototype. If having the target name in a colour memory tasks acts in similar fashion to priming (Neuman & D’Agostino, 1981; Rosch, 1975), as suggested in the previous chapter, it is possible that Stroop interference affects the direction of the shift of the mental representation and hence the order effects in within-category judgements.

If we assume that the Stroop word acts as a kind of prime, the mental representation should shift towards the (pre-existing) prototype. This would, however, predict that the type of Stroop interference (congruent, incongruent) should selectively affect within category order effects. With incongruent Stroop interference the representation should shift towards the other category prototype (denoted by the Stroop caption) and hence order effects should be reversed. This idea is illustrated schematically in Figure 4.10.
Fig. 4.1. Possible effects of incongruency on within category discrimination. The example given is for a blue target (either B1 or B2) and assumes the target is incongruently labelled GREEN. Incongruent Stroop could have the effect of shifting the representation towards the other category prototype.

Figure 4.10 demonstrates the possibility that incongruent Stroop interference in effect primes the other category prototype. The example in (a) shows a blue within category pair, assuming it is incongruently labelled, the memory representation may shift towards the green category prototype. Hence, if as shown in the top half of (a) B1 is the target, the shift towards the green memory prototype would mean the distance between B1 and B2 would increase, as shown in (b). If, however, B2 is the target, the shift towards P6 would decrease the distance between B1 and B2 (c). For incongruently labelled stimuli this would predict that boundary target is more accurate than non-boundary target.

The effects of congruent and incongruent Stroop interference on within category order effects were tested by re-analysing Experiments 5-8.
4.3.1. Re-analysis of Experiment Five: Stroop interference at target presentation

Experiment 5a:
The aim of Experiment 5a was to test the direct language account of CP (Roberson & Davidoff, 2000) by introducing Stroop interference to a colour discrimination task. The results showed that congruent Stroop interference strongly facilitated CP, whereas incongruent Stroop interference reduced CP.

(for Method see § 3.2.2.)

Results
The results of Experiment 5a were re-analysed taking within category stimulus order into account and are presented in Figure 4.11. Results are collapsed across blue and green stimulus pairs.

![Figure 4.11](image)

Fig. 4.11. Mean accuracy (error bars show +/- 1 standard error) for within category discrimination (analysed by order of stimulus presentation) and cross-category discrimination.

As the graph shows, order of presentation for within category pairs affected discrimination accuracy. However, the results show that these order effects seem to be
dependent on type of Stroop interference. No order effect was found in the neutral condition, in the congruent condition presenting the non-boundary stimuli as the target appears to be more accurate than presenting the boundary stimulus as the target. Furthermore, the effect seems to be reversed in the incongruent condition. The results were analysed by a two-way repeated measure ANOVA (order of presentation, interference). Neither main effect was significant (order: $F[1,27] = 0.33, \text{MSE} = 607.10, p>0.05$; interference: $F[2,54] = 0.14, \text{MSE} = 47.66, p>0.05$), the interaction, however, was significant ($F[2,54] = 4.43, \text{MSE} = 110.22, p<0.05$).

The difference between boundary and non-boundary target discrimination did not reach significance in any Stroop condition (maximum $t(26) = 1.29, p>0.05$). When the boundary stimulus was the target, discrimination in the congruent condition was significantly worse than in the neutral ($t[25] = 1.86, p<0.05$) and the incongruent condition ($t[25] = 3.39, p<0.01$). On non-boundary target trials, accuracy on congruent trials was significantly better than in the neutral ($t[25] = 2.10, p<0.05$) and the incongruent conditions ($t[25] = 2.90, p<0.01$).

Although the effect of order of presentation did not reach significance, the results of the re-analysis of Experiment 5a show order effects in the predicted direction, i.e. towards the pre-existing prototype in the congruent condition. Furthermore, non-boundary target discrimination was significantly better in the congruent compared to the incongruent and the neutral condition, which suggests that congruent Stroop interference reinforced the shift towards the blue or green prototypes relative to the other Stroop conditions. The reversal of the order effect in the incongruent Stroop condition was not significant.

**Experiment 5b**

Experiment 5b used integrated Stroop stimuli in a colour discrimination task, all other aspects were the same as in Experiment 5a.

(for Method see § 3.2.2)
Results

The results were re-analysed for within category order effects and are presented in Figure 4.12.

![Figure 4.12](image)

Figure 4.12. Mean Accuracy scores (error bars show +/- 1 standard error) for within category (analysed by order of presentation) and cross-category discrimination.

As shown above, clear order effects were found for within category discriminations in Experiment 5b and, as in Experiment 5a, the order effects were dependent on Stroop interference. However, now order effects seem to be apparent in the congruent and in the neutral condition, but not in the incongruent interference condition. Within-category scores were analysed by a two-way repeated measures ANOVA, on order of presentation and interference. The effect of order approached significance ($F[1,27] = 3.95$, MSE = 637.52, $p=0.057$) and the interaction was highly significant ($F[2,54] = 7.78$, MSE = 87.51, $p<0.01$). The effect of interference was not significant ($F[2,54] = 1.44$, MSE = 43.35, $p>0.05$).

Post-hoc testing (protected t-tests) showed that the difference between boundary and non-boundary target accuracy was significant in the congruent interference condition ($t[27]= 2.01$, $p<0.05$), but not in the neutral condition ($t[27] =1.4$, $p>0.05$) or the incongruent condition ($t[27] =0$, $p>0.05$). For just the boundary target condition,
discrimination in the incongruent interference condition was significantly better than in the neutral ($t[25] = 3.89, p<0.001$) and the congruent condition ($t[25] = 4.74, p<0.001$). For non-boundary discriminations, accuracy in the congruent condition was significantly better than in the incongruent condition ($t[25] = 2.96, p<0.01$).

The re-analysis of Experiment 5b showed that within category discrimination accuracy was dependent on order of presentation in some interference conditions. Discrimination accuracy was higher when the stimulus further from the boundary was presented first than when the boundary stimulus was the target. This was the case in the neutral and the congruent conditions, but the effect was only significant with congruent Stroop interference. The order effect was lost in the incongruent condition. As outlined above, it is possible that the incongruent Stroop interference has the effect of shifting the memory representation towards the other category prototype. The results of Experiment 5b are consistent with this idea. In the boundary target condition discrimination in the incongruent condition was better than in both the neutral and the congruent condition. This suggests that relative to the neutral and the congruent condition, the shift in the incongruent condition was more towards the other category prototype (the colour name). Also, discrimination accuracy was lower in the incongruent condition compared to the congruent condition for non-boundary targets which supports this idea. Overall, the order effect found in the congruent condition and the loss of this order effect in the incongruent condition are consistent with the idea that verbal labels shift the representation towards the corresponding (pre-existing) prototypes.
4.3.2. Re-analysis of Experiment Six: Stroop interference at test presentation

To test the matching-to-labels account of CP, Stroop interference was presented at the response stage of a colour discrimination task in Experiment 6. Target congruent Stroop shifts the memory presentation towards the pre-existing prototype. Since the Stroop interference in Experiment 6 was presented after a possible shift of the memory representation, the different Stroop conditions should not selectively affect within category order effects. Nevertheless, there could still be a shift towards prototype if the targets were labelled.

(for Method see § 3.3.2.)

Results

The re-analysis of Experiment 6 for within category order effects is presented in Figure 4.13. Results were collapsed across blue and green stimulus pairs.

![Figure 4.13](image)

Fig. 4.13. Mean accuracy (error bars show +/- 1 standard error) for within category discriminations (analysed by order of presentation) and cross-category presentation.
There appears to be a within category order effect in all Stroop conditions, when the non-boundary colour was presented as the target discrimination was better than when the boundary colour was presented first. Interestingly, non-boundary target within-category discrimination seems to be better than cross-category discrimination in all but the neutral interference condition.

Within category discrimination accuracy was analysed by a two-way repeated measures ANOVA (order of presentation, interference). The order effect was not significant ($F[4,108] = 0.83$, $\text{MSE} = 1273.25$, $p>0.05$), however the main effect of interference reached significance ($F[1,27] = 2.47$, $\text{MSE} = 82.06$, $p=0.049$). The interaction was not significant ($F[4,108] = 0.16$, $p>0.05$).

**Summary of Experiments 5 and 6**

Order effects were found in delayed discrimination tasks with Stroop interference in the congruent conditions in Experiments 5a and 5b. The order effects were consistent with a shift towards the pre-existing prototype. The effects were lost in the incongruent Stroop conditions which suggests that incongruent Stroop interference may have had the effect of shifting the memory representation further towards the other category prototype. When Stroop interference was presented at response stage (Experiment 6), no within category order effects were found.
4.4. Order effects in quick discrimination tasks with Stroop interference

In Experiments 7 and 8 the target and test presentation times and the inter-stimulus interval were shortened relative to the previous experiments. Both target and test stimuli (and Stroop interference) were presented for 150ms, the inter-stimulus interval was 100ms. When testing for CP this did not change the pattern of results when Stroop interference was presented at target presentation (Experiment 7), it did however when Stroop interference was presented at response (Experiment 8). Huttenlocher, Hedges and Vevea (2000) have found the bias towards the prototype to be smaller if the memory load is decreased. In Experiments 7 and 8 the ISI, the time the target representation had to be retained in memory, was reduced from 5 seconds to 1 second. Therefore, if the order effects found in the previous experiments were due to similar mechanisms or processes as those reported by Huttenlocher et al., any order effects found in Experiments 7 and 8 should, due to the decrease in memory load, be smaller than those found in Experiments 5 and 6.

4.4.1. Re-analysis of Experiment Seven

In Experiment 7 Stroop interference was presented in a quick colour discrimination task. The analysis showed similar effects to those found in Experiment 5, hence congruent Stroop interference facilitated CP and incongruent Stroop interference reduced CP. In Experiment 5 there were order effects, but these were dependent on type of Stroop interference. Congruent interference increased the shift towards the pre-existing prototype; incongruent interference eliminated or even reversed (though not significantly) this order effect. If, as Huttenlocher et al. argued, the shift decreases as the memory load is reduced, the shorter inter-stimulus interval in the current experiment should produce less bias and therefore smaller order effects.

(for Method see § 3.4.2.)
Results

The results of the re-analysis for within category order effects in Experiment 7 are presented in Figure 4.14.

Fig. 4.14. Mean accuracy (error bars show +/- 1 standard error) for within category discrimination (analysed by order of presentation) and cross-category discrimination.

As the graph shows, there seems to be a strong order effect in all interference conditions. Within category discrimination was more accurate when the non-boundary stimulus was presented as the target compared to when the boundary stimulus was the target. The results were analysed by a two-way repeated measures ANOVA (order or presentation, interference). Accuracy was greater when the non-boundary stimulus was the target than when the boundary stimulus was the target ($F[1,20] = 104.14$, $MSE = 238.98$, $p<0.001$). The effect of interference was not significant ($F[2,40] = 1.45$, $MSE = 59.36$, $p>0.05$), nor was the interaction ($F[2,40] = 1.36$, $MSE = 61.23$, $p>0.05$). Post-hoc (protected t) tests showed that non-boundary target discrimination was significantly better than boundary target discrimination in all Stroop conditions (minimum $t[19] = 5.40$, $p<0.001$).
The results suggest that length of the inter-stimulus interval (ISI) – the time the target has to be kept in memory – has a strong effect on order effects in within category discriminations: the order effect increases as the ISI decreases. This is not consistent with Huttenlocher et al.’s findings that the shift towards prototype is less the smaller the memory load. Further analysis was done to test whether the order effect found for a short ISI was significantly bigger than that found with a longer ISI in Experiment 5. A three-way ANOVA on order of presentation (boundary, non-boundary) and interference (neutral, congruent, incongruent) as repeated measures and length of ISI (1000ms, 5000ms) as a between subjects factor was run. There was a highly significant effect of order of presentation ($F[1,47] = 36.67$, MSE = 450.45, $p<0.001$) and of length of ISI ($F[1,47] =18.49$, MSE = 172.05, $p<0.001$). Significant interactions were found between order and ISI ($F[1,47] 26.86$, $p<0.001$) and between order and interference ($F[2, 94] = 5.06$, MSE = 89.37, $p<0.05$).

Post-hoc testing (protected t-test) showed that non-boundary discrimination was significantly more accurate than boundary discrimination in the 1000ms ISI ($t(19) = 4.29$, $p<0.01$), but not the 5000ms ISI condition ($t(26) = 0.38$, $p>0.05$). Furthermore, non-boundary discrimination was significantly better in the 1000ms compared to the 5000ms ISI condition ($t(47) = 5.18$, $p<0.01$); discrimination accuracy in the boundary target conditions did not differ significantly ($t(47) = 1.67$, $p>0.05$). The order effect was significantly bigger in the short ISI condition.

The results also show that CP (better cross-category than average within category discrimination) was entirely due to boundary first within category discrimination accuracy being extremely low. Since CP effects were dependent on type of Stroop interference, but order effects were found in all conditions, the re-analysis suggests that CP effects are dissociated from the order effects in this experiment.

### 4.4.2. Re-analysis of Experiment Eight

In Experiment 8 Stroop interference was again presented concurrently with the test stimuli at the response stage. Presentation times and ISI were matched to Experiment 7, hence test and target presentation were 150ms and ISI was reduced to 1000ms. The
results of Experiment 8 showed that CP survived in all Stroop interference conditions. The re-analysis of Experiment 6 showed within boundary target discrimination to be better than boundary target discrimination, but this difference was not significant. The results of Experiment 7 showed greater order effects for shorter ISIs. Therefore, in Experiment 8 non-boundary discrimination should be better than boundary discrimination, independent of type of Stroop interference.

(for Method see § 3.5.2.)

Results
Results were analysed for within category order effects and are presented in Figure 4.15.

![Figure 4.15](image)

Fig. 4.15. Mean accuracy (error bars show +/- 1 standard error) for within category discrimination (analysed by order of presentation) and cross-category discrimination.

Figure 4.14 seems to show that the strong order effects found in Experiment 7 were also present in within category discriminations in Experiment 8. Discrimination of non-boundary first pairs was better than discrimination of boundary first pairs. The
order effect was so strong that non-boundary discrimination was better than cross category discrimination. Results were analysed by a two-way repeated measures ANOVA (order of presentation, interference). There was a highly significant main effect of order \((F[1,27] = 74.9, \text{MSE} = 746.34, p<0.001)\). The effect of interference \((F[2,54] = 0.68, \text{MSE} = 43.62, p>0.05)\) and the interaction \((F[2,54] = 0.91, \text{MSE} = 64.93, p>0.05)\) were not significant. Post-hoc protected t-tests showed that non-boundary discrimination accuracy was significantly better in all Stroop interference conditions \(t[27] = 3.52, p<0.01\).

Further analysis was done to test whether the order effects found in Experiment 8 (short ISI) were significantly greater than those found in Experiment 6 (long ISI). A three-way ANOVA with order of presentation and interference as repeated measures and length of ISI as a between subjects factor showed that only the main effect of order was significant \((F[1,54] = 35.81, \text{MSE} = 1009.79, p<0.001)\). The effect of length of ISI was not significant \((F[1,54] = 0.08, \text{MSE} = 408.23, p>0.05)\). The interaction between order and length of ISI was again highly significant \((F(1,54) = 20.59, p<0.001)\).

Post-hoc testing by protected t-tests showed that non-boundary discrimination was significantly better than boundary discrimination in the 1000ms ISI \(t(26) = 3.33, p<0.01\), but not the 5000ms ISI condition \(t(26) = 0.46, p>0.05\). Furthermore, non-boundary discrimination was significantly better in the 1000ms compared to the 5000ms ISI condition \(t(26) = 2.17, p<0.05\). This was reversed in the boundary target condition where discrimination accuracy in the 5000ms ISI was better than in the 1000ms ISI condition \(t(26) = 2.34, p<0.05\). Thus, as with the comparison of Experiments 5 and 7 (Stroop at target presentation), the order effects here were stronger for shorter ISIs. This was due both to non-boundary discrimination accuracy being higher and boundary discrimination accuracy being lower in the 1000ms compared to 5000ms condition.
Summary of Experiments 7 and 8

Strong order effects were found in all Stroop conditions whether Stroop interference was presented concurrently with the target stimulus (Experiment 7) or with the test stimuli (Experiment 8). The order effects were consistent with a shift towards the pre-existing blue and green prototypes. The order effects found with short ISIs were stronger than those found with long ISIs in Experiments 5 and 6.

4.5. General Discussion

The re-analysis of Experiments 1 and 2 and 5-8 suggest that the direction and strength of order effects depend on the type of tasks. The findings are summarised in Table 4.1.

Table 4.1. Summary of the re-analysis of Experiments 1 and 2 and 5-8 for within category order effects in Chapter 4. (* the numbers in brackets correspond to the target stimuli exposure duration, the length of ISI and the test stimuli exposure duration respectively)

<table>
<thead>
<tr>
<th>Exp.</th>
<th>Type of interference</th>
<th>Order effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No, visual and verbal interference during ISI (1000ms, 5000ms, 1000ms)*</td>
<td>- No order effects</td>
</tr>
<tr>
<td>1AFC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| 2    | No, visual and verbal interference during ISI (1000ms, 5000ms, 1000ms) | - Order effects found in all conditions  
- Opposite direction than predicted  
- Consistent with a shift towards the temporary prototype |
Table 4.1. cont.

<table>
<thead>
<tr>
<th></th>
<th>Stroop interference at target presentation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5a</td>
<td>2AFC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(150ms, 5000ms, 1000ms)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Order effects found only for</td>
<td></td>
</tr>
<tr>
<td></td>
<td>congruent Stroop interference</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Consistent with shift towards the</td>
<td></td>
</tr>
<tr>
<td></td>
<td>pre-existing blue and green prototypes</td>
<td></td>
</tr>
<tr>
<td>5b</td>
<td>2AFC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>As 5a except integrated interference</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(150ms, 5000ms, 1000ms)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Same results as 5a</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2AFC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stroop interference at test presentation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(150ms, 5000ms, 1000ms)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• No order effects</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2AFC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stroop interference at target presentation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(150ms, 1000ms, 150ms)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Highly significant order effects in</td>
<td></td>
</tr>
<tr>
<td></td>
<td>all Stroop conditions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Consistent with shift towards pre-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>existing blue and green prototypes</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2AFC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stroop interference at test presentation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(150ms, 1000ms, 150ms)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Same results as 7</td>
<td></td>
</tr>
</tbody>
</table>

Re-analysis of the 2AFC and same-different discrimination tasks

The re-analysis of Experiments 1 and 2, the 2AFC and same-different discrimination tasks with verbal and visual interference, showed that there were order effects but only in the same-different task (Experiment 2). For the 2AFC task there was no difference in discrimination accuracy whether the boundary or the non-boundary stimulus was the target. In the same-different task strong order effects were found in all conditions. However, the effect was in the opposite direction to that predicted. The shift was not towards the pre-existing green and blue prototypes, but towards the green/blue boundary. Huttenlocher and colleagues (Crawford, Regier & Huttenlocher, 2000; Huttenlocher, Hedges & Duncan, 1991; Huttenlocher, Hedges & Vevea, 2000) define the prototypes as the average of all stimuli used in the experiment. In their experiments the categories were induced for the purpose of the experiments and
therefore the prototypes were temporary prototypes. However, pre-existing categories such as colours have ‘natural’ prototypes (Rosch, 1978). The prototypes are the focal colours, the best examples of the colour. For the current re-analysis of the same-different and the 2AFC discrimination task it was assumed that the prototypes would be the pre-existing blue and green prototypes. However, it is possible that this assumption extended Huttenlocher et al.’s model inappropriately.

The results of the re-analysis of Experiment 2 could be interpreted as suggesting the shift was towards the temporary prototype, the average of all stimuli used in the experiment. If this was the case, the order effects found were consistent with the predictions of a shift towards the temporary prototype; boundary stimuli first resulted in a greater separation and hence greater accuracy. However, it is important to note that a shift in this direction would reduce cross-category separations and hence CP is not accounted for. Thus, the categorical perception effect found in Experiment 2 is dissociated from the possible shift responsible for the order effects found. The comparison of Experiments 1 and 2 also suggests that there are differences in the sensitivities of same-different and 2AFC tasks that did not show up in the analysis of CP effects in Chapter 2. When looking specifically at within category judgements there are clear differences between the pattern of results. It is possible that in a 2AFC task the representational shift is to a certain degree cancelled out by the presence of both the target and the distractor at the response stage and that this has the effect of ‘correcting’ the inexact memory representation.

*Re-analysis of Stroop experiments 5 and 6*

The re-analysis of Stroop experiments 5 and 6 showed a very different pattern of results. All experiments were 2AFC tasks; hence from the re-analysis of Experiment 1 no order effects were expected. However, Stroop interference unlike the verbal and visual interference tasks used in Experiments 1 and 2 had selective effects on the order effects found in Experiments 5a and 5b. The order effects found here were in the opposite direction to those found in Experiment 2, the shift was towards the pre-existing prototypes. Interestingly, the shift was significant only in the congruent interference conditions, but not in the neutral or the incongruent conditions. Overall, Stroop interference clearly affected within category order effects. It is possible that the presence of Stroop interference, hence the words BLUE or GREEN presented
concurrently with the target colour induced a representational shift towards the corresponding prototype. Consistent with this is the finding that the order effects were lost in the incongruent condition. Hence, for example when a blue within category pair was presented with the colour word GREEN it is possible that the representation shifted towards the prototype of the colour word and not the colour itself. In Experiment 6 Stroop interference was presented concurrently with the test stimuli. The re-analysis of within category discrimination accuracy in this experiment showed no significant order effects.

Re-analysis of Stroop experiments 7 and 8
The re-analysis of Experiments 7 and 8 showed that presentation times and length of ISI also influenced order effects. Strong order effects were found in both experiments, the direction was consistent with a shift towards pre-existing category prototypes. The order effects in these quick discrimination experiments did not vary across Stroop conditions, thus in the incongruent conditions the shift was also towards the prototype of the stimulus colour. This differs from Experiment 5 where incongruency reduced the order effect. The order effects were stronger for short ISIs (Experiments 7 and 8) than long ISIs (Experiments 5 and 6). This is not consistent with Huttenlocher et al. (2000) who argued that the uncertainty, and hence the shift, should be increased when the memory load increases.

There is another point that needs to be mentioned. According to Huttenlocher et al., as stimulus uncertainty increases, so does the shift towards prototype. Uncertainty increases as the memory trace fades (increasing ISI) but also with initial uncertainty. Short target presentation will increase uncertainty. As target presentation was 1000ms in the 2AFC and the same-different task, but reduced to 150 ms in the Stroop tasks, it is possible that the stimulus uncertainty, and hence the strong order effect, was due to short exposure rather than to length of ISI. Huttenlocher et al. also proposed that visual interference ought to reduce the exactness of the memory trace. However, the results of the re-analysis of Experiments 1 and 2 found no evidence for type of interference affecting the strength of the order effect, which again suggests that the CP effects found in these experiments were not due to a shift towards prototype. Experiments testing the shift towards prototype and CP will need to systematically manipulate these variables.
Shift towards prototype and CP?

The re-analysis presented in this chapter does not readily offer an explanation for CP found in the experiments discussed. If a shift towards prototype causes CP, it can only be a shift towards the pre-existing prototypes and not the temporary prototype. A shift towards the temporary prototype would predict cross-category discrimination to be worse than average within category discrimination. Hence, in Experiment 2 where order effects were found consistent with a shift towards a temporary prototype, this shift was clearly dissociated from and cannot explain CP.

The order effects found in the Stroop experiments were consistent with CP, however there is also evidence that the two phenomena are dissociated. Stroop interference at target presentation affected CP even in quick discrimination tasks; the order effects found in Experiment 7, however, did not show selective effects of Stroop interference. Thus, CP, especially in Experiments 7 and 8, was due to the low discrimination accuracy for boundary stimuli presented first. This is problematic for theories of CP in general since it shows that within category discrimination is not inherently worse than cross-category discrimination, but is entirely dependent on order of presentation. These findings also undermine a warping account of CP since warped perceptual codes should be unaffected by order of presentation of within category pairs.

It is possible that the order effects found for within category discrimination in Experiment 5-8 could be due to a possible effect of the imbalance of non-boundary and boundary stimuli rather than a representational shift. In the experiments used here, the within and cross-category pairs were taken from a matrix of 16 colours. There were four cross-category and eight within category pairs. The boundary stimuli were part of the within as well as the cross-category pairs. Therefore, when the non-boundary stimulus was the target, there was only one possible test stimulus at response. Thus, non-boundary target implied for certain what the test stimuli will be. This implies that discrimination should be easier when the non-boundary stimulus is the target. The order effects found in the Stroop tasks are consistent with this possibility. However, the results of the same-different task (Experiment 2) as well as the results presented by Davies, Pilling, Wiggett and Özgen (2003) show opposite order effects, despite the same imbalance of non-boundary and boundary stimuli.
These results are consistent with Huttenlocher et al.'s account that there should be less uncertainty about the more frequent stimuli. Hence, discrimination of boundary targets should be easier because the stimuli are more familiar.

In Chapter 3 the possibility was raised that Stroop interference in discrimination tasks may have similar effects to the priming effects reported by Rosch (1975) and Neuman & D'Agostino (1981). The results suggested that target, but not test name generation was necessary for CP, which suggests that not only was target name generation necessary it also seemed to be sufficient. It was proposed that target name generation may activate a category code that is then available to use as part of the choice process. It was further argued that this category code may be similar to the mental representation activated by a colour word prime. If we assume that this 'priming' effect shifts the memory representation towards the prototype, the resultant effects on discrimination accuracy could be similar to the processes described in other stimulus domains by Huttenlocher and colleagues. However, the effects in Stroop tasks are likely to be due to labelling, the effects described by Huttenlocher et al. are memory effects. Therefore the shifts observed in the current experiments and those reported by Huttenlocher et al. may not be due to the same mechanisms.

The re-analysis of Experiments 1 and 2 and 5-8 for order effects in within category discrimination suggests that CP cannot be due to a shift towards prototype for two reasons. Firstly order effects and hence a shift towards prototype were not found in all experiments that showed CP effects. Secondly, if Huttenlocher et al.'s model is applied as intended, the prototype being the temporary prototype, a shift towards prototype would predict cross-category discriminations to be worse than average within category discrimination. If a shift towards (temporary) prototype happens on target presentation, no CP effects should be found. The fact that significant CP effects were found in Experiment 2 suggests that the better discriminability of boundary targets is not due to a shift. Thus, overall the results suggest that the categorical perception effects found in colour discrimination tasks are not due to a shift towards prototype. A shift towards prototype seems to occur when the pre-existing prototype and the temporary prototype coincide, hence for just within category stimulus sets (see Figure 4.4.). However, in tasks that compare within and cross-category discriminations and therefore stimulus sets that straddle a category boundary, both
order effects and categorical effects were found under certain conditions. However, the results of the re-analysis presented in this chapter suggest that these effects were not due to processes analogous to the shift towards prototype in Huttenlocher, Hedges and Vevea (2000) and Crawford, Regier and Huttenlocher (2000).
Chapter Five

Cross-cultural examinations of language and colour

5.1. Introduction

The linguistic relativity hypothesis proposes that the language we speak influences the way we think and perceive. Thus, differences in languages across cultures are thought to be paralleled by cognitive non-linguistic differences. Much of the experimental work done to test this hypothesis has focused on colour cognition as colour vocabularies differ widely across languages (Berlin and Kay, 1969). English has eleven basic colour terms, many African languages have four to five terms (Davies and Corbett, 1997) and the Dani of New Guinea have been reported to have only two basic terms (Rosch, 1972). The differences in the number of colour terms means that it is possible to test whether language influences perception by comparing the performance of speakers of languages with different sets of colour terms on colour perception tasks such as similarity judgements or discrimination accuracy.

Language could affect performance on a colour perception task in two ways. Firstly, language could affect performance by changing or warping perceptual colour space (Özgen & Davies, 2002). If this were the case, having, for example, separate blue and green terms would mean that the differences across the category boundary were enhanced whereas within category differences were reduced. This would lead to better discrimination accuracy for speakers of the language with separate terms; the effect of language would be on colour perception per se. Secondly, language could have online effects on performance due to verbal strategies being used. If two languages were compared, one of which linguistically differentiated green and blue, and one which did not, the use of verbal labels on a discrimination or memory task should disadvantage the latter language. If stimuli have different labels (e.g. blue and green) these verbal codes will be more useful for remembering and discriminating the colours than if both have the same name. This is consistent with the findings discussed in Chapter 2 suggesting that target name generation is necessary for CP.
For languages with a different set of colour terms, category distributions and positions of category boundaries also differ. If a task requires a verbal strategy, or if a verbal strategy is useful on the task, speakers of different languages will have different name codes for colour stimuli. Furthermore, the comparison of colours may be within category for one language, but cross a category boundary in another language. The colour name codes for colours are going to differ across languages, but the crucial question is whether the physical colour codes differ, too. If so, it would suggest that colour language had warped perceptual space.

The current chapter tests this possibility by comparing languages with different colour nomenclatures: English and Kwanyama and Balantu, both languages of the Owambo, Namibia. Data for Experiment 9 were collected on Kwanyama speakers, data for Experiment 10 on Balantu speakers. The languages differ from English by having fewer basic colour terms, thus making Balantu and Kwanyama colour categories broader than English categories. If the language we speak and the categorisations our language performs influences our perception of the world (Whorf, 1956), differences between the language groups should be found. The differences in colour language mean that colours taken from separate English categories will often be within category for Owambo speakers and so allow a cross-cultural test of CP (Harnad, 1987).

The tasks used in the current chapter were chosen because they differ in the locus of the cognitive load and in the potential usefulness of a naming strategy. They were a visual search and colour triads task. The results of Chapter 2 and 3 suggested that the locus of CP depends on the type of task. In tasks such as 2AFC delayed discrimination tasks where verbal strategies are likely to have been useful for cross-category comparisons, language effects on CP were found. However, Experiment 4 also found evidence for CP in a visual search task for which verbal strategies are unlikely to have been useful.

The visual search task (Experiment 9) is likely to be a relatively pure perceptual task since it requires a quick response to a relatively large array of colours. The task was to detect target colours amongst distractors as accurately and as quickly as possible. It is unlikely that naming each colour to determine whether it was a target or a distractor
would facilitate search time or accuracy. The short response times to the search task in Experiment 4 support this point; it is unlikely an attempt was made to name all stimuli in the search array.

The speed with which targets can be detected in an array of distractors depends on the similarity between the targets and the distractors. The more similar the colours are, the harder search is. If the distance between target and distractors is over 20 CIE $L^*u^*v^*$ ($\Delta E$) units, targets tend to 'pop-out' making search easy (Carter, 1982). However, search has also been found to depend on the categorical identity and relations of stimuli. Pilling and Davies (2003) reported that search was faster when the target and distractors were from separate categories compared to within category search. Furthermore, as discussed in Experiment 4, it is possible that categorical separations facilitate or reinforce pop-out.

If language has warped perceptual colour space and cross-category differences have become stretched, targets should be easier to detect when the target and the distractors are in separate linguistic categories. Warping could also result in within category equivalence rather than cross-category distinctiveness (as reported by Goldstone, Lipppa & Shiffrin, 2001) which could affect search for example by making it easier to classify all same-category distractors as distractors. If differences between language groups were found in line with these predictions it would suggest that colour language influences colour perception and that the representations of colour space differ across languages.

Experiments 10 used a triads task. Three colours were presented and the task was to choose the most different colour of the three. The task was less perceptually loaded than the visual search task. If two colours had the same name but the third differed (for example blue, blue, green) and the task was done by naming, the colour with the different name is most likely to be chosen. Thus, in a comparison of two languages for which the naming of the tiles differ, triad choices should differ in line with colour naming. However, differences in triad choices across languages should also be found if language had affected colour perception per se. A perceptual warping account of CP proposes that cross-category differences are exaggerated; the differences between colours with different names should be increased. Thus, the tile given a different
name would be chosen as the most different because it was the most (perceptually) isolated colour out of the three. If perceptual warping across category boundaries had occurred, differences between language groups would be found whether a verbal strategy was used on the task or not. Furthermore, the verbal and perceptual codes might add or reinforce each other.

The comparison of the visual search and the triads task allows a test of whether the differences in colour cognition are direct language effects or whether perceptual colour space itself is warped. If no difference between the English and the Owambo groups were found on either task, the idea of cognitive universals driving colour perception would be supported. However, if language effects were found, linguistic relativity would be supported.

Depending on whether effects were found on the triads task where the use of verbal strategies in relatively useful and likely or the visual search task which was more perceptually loaded, inferences could be made about the location of the effects of language. If differences were found only on the triads task it would suggest that the language effect was a direct effect of naming. If however, differences between English and Owambo speakers were also found on the visual search task, it would suggest that colour language may influence colour perception per se. This would suggest that by having a different set of basic colour terms, the perceptual representation of colour for English and Owambo had come to differ significantly.
5.2. Experiment Nine

Visual search task

5.2.1. Introduction

If having a certain set of colour terms warps perceptual colour space accordingly and so enhances language specific categorical differences, the effect of language on colour cognition would be indirect. Hence, if language influences colour perception *per se* and the effects are not merely online effects due to stimulus naming, differences between languages should be found when verbal strategies are unlikely to be advantageous. This could include tasks which require quick perceptual decisions and tasks in which many colours are present simultaneously such as visual search tasks. Response latencies in visual search tasks are typically short (about 500ms, Bauer, Jolicoer & Cowan, 1996) which suggests the decision whether a target is present or not is made before verbal labels are available.

In Experiment 4, CP was found in a computerised visual search task; search was quicker and more accurate if the target and the distractor were in different categories compared to same category target and distractors. Crucially, cross-category and within category distances were held equal. The results support a perceptual account of CP in visual search tasks and are consistent with Pilling (2001) suggesting that other factors alongside the similarity of target and distractor such as categorical identity facilitate search.

This finding has implications for cross-cultural investigations of colour cognition: if colour space is warped according to the language spoken, differences between languages should also be found on tasks that do not require or allow the use of verbal strategies. Thus, if the target and distractor are in separate categories, the target should appear perceptually more isolated and hence search should be facilitated for the speakers of the language making the categorical distinction.
Naming

To test whether language influences colour perception per se a set of colours for which English and Kwanyama naming differed significantly was chosen. Kwanyama is an Owambo language spoken in northern Namibia. The Kwanyama language does not have an orange term and accordingly the yellow and red colour categories are broader and the position of the category boundary differs. Four stimuli were used in the search task in Experiment 9, the CIE L*u*v* coordinates and most common naming responses for English and Kwanyama speakers are presented in Table 5.1.

Table 5.1. Colours used in the visual search task, dominant term given by English and Kwanyama speakers and CIE L*u*v* co-ordinates for each colour.

<table>
<thead>
<tr>
<th>Code</th>
<th>English</th>
<th>Kwanyama</th>
<th>L*</th>
<th>u*</th>
<th>v*</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>Red</td>
<td>Shitilyana</td>
<td>49.05</td>
<td>102.02</td>
<td>25.5</td>
</tr>
<tr>
<td>O1</td>
<td>Orange</td>
<td>Shitilyana</td>
<td>63.14</td>
<td>87.27</td>
<td>46.81</td>
</tr>
<tr>
<td>O2</td>
<td>Orange</td>
<td>Oshunga shei</td>
<td>60.86</td>
<td>46.39</td>
<td>68.26</td>
</tr>
<tr>
<td>Y</td>
<td>Yellow</td>
<td>Oshunga shei</td>
<td>80.92</td>
<td>50.52</td>
<td>80.14</td>
</tr>
</tbody>
</table>

The Kwanyama language uses only two colour terms to cover the four colours, whereas English uses three colour terms. Thus, naming responses show that Kwanyama colour language does not have a basic colour term for orange, the orange tiles were either labelled shunga shei (yellow) or shitilyana (red). For English speakers, the stimulus set crossed two category boundaries: yellow-orange and orange-red. The aim of the present study was to test whether these linguistic differences were paralleled by non-linguistic behavioural differences on a perceptual task for which the use of a verbal strategy was unlikely.

9 The naming data was taken from Davies' (unpublished manuscript) study of the colour languages of northern Namibia.
The current experiment was a visual search task with four search conditions. There were two one-target conditions (IT-pure and IT-mixed) and two two-target conditions (2T-pure and 2T-mixed). The colour stimuli used were: Yellow (Y), Orange 1 (01), Orange 2 (02), and Red (R). In Condition IT-pure there was just one target (01) and red and yellow distractors. The other search conditions varied in the number and categorical nature of distractors (condition IT-mixed) and in the number of targets (conditions 2T-pure and 2T-mixed). The search conditions are denoted according to the English categorical separations within the search lists. In the pure list conditions distractors were always from different English categories than targets. In the mixed conditions (for the English) targets were distinguished from only some distractors on the basis of category.

The main dependent variable was the time taken to find all targets in the visual search array (search time). If colour perception itself does not differ between English and Kwanyama, there should be no difference in the performance on the search task between the two groups. If language has changed or warped perception this should be evident in the performance of English and Kwanyama speakers on the task. However, it is also important to note that simple main effects may be masked by uncontrolled factors such as differences in schooling or familiarity with the type of task. This makes predictions about absolute differences in search performance across languages difficult. Therefore, the more important comparison is to look at the relative influences of the search variables. Second order effects (interactions) are more likely to survive extraneous variables.

The search list in condition IT-pure comprised an orange target (01) and yellow (Y) and red (R) distractors. English speakers should be advantaged compared to Kwanyama speakers as the target was categorically separate from either distractor. For Kwanyama speakers however, the target was in the same category as one of the distractors. Therefore, if language has affected perception, targets should be more distinguishable from distractors for English speakers. The stimuli used in condition IT-mixed were identical to those used in IT-pure except for the addition of a further distractor (O2). Since some of the distractors were now in the same (English) category as the target, any advantage in IT-pure for English speakers over Kwanyama speakers...
should be reduced. For the Kwanjama, any loss in performance on IT-mixed compared to IT-pure should be less than for English speakers.

Similar predictions were made for the two target conditions. Thus, in condition 2T-pure where there were two orange targets (O1 and O2) and yellow (Y) and red (R) distractors, English speakers should again be advantaged since the targets were categorically separate from either distractor. For Kwanjama speakers the target and the distractor categories overlapped. Furthermore, if search was done categorically (for English speakers all orange stimuli were targets), performance on IT-pure and 2T-pure should not differ substantially.

Search in 2T-mixed should be significantly slowed for English speakers as the targets were now in two separate categories (one orange and one red target). However, for Balantu speakers this change in target position should make search easier since both targets were now categorically separate from the distractors (both targets were shitilyana; the distractors shunga shei). As in the one-target condition, the cost of changing the search array from 2T-pure to 2T-mixed should be greater for English than Kwanjama speakers.

Thus, if categorical factors influence search as suggested by Pilling (2001) and the results of Experimient 4, the fact that English and Kwanjama make different categorical distinctions should be reflected in differing search patterns across the four conditions.
5.2.2. Method

Subjects
Nineteen English speakers (twelve male, seven female; mean age 30) and eighteen monolingual Kwanyama speakers took part in this experiment (ten male, eight female; mean age 35). Kwanyama is a dialect of Owambo, a language spoken in Northern Namibia.

Stimuli
The stimulus set consisted of four visual search cards plus a practice card. Each card was A4 (210 x 297 mm) and the colour stimuli were printed on a neutral grey background (see Figure 5.1.).

![Fig 5.1. Schematic representation of the lists used in the visual search task. The example shows condition 1T-pure. O1 was the target and the distractors were either red or yellow. The target was in a separate category from both distractors for English speakers but not for Kwanyama speakers.](image)

Each card contained 168 items of which 149 were distractors and 19 targets. The target was always presented at the top of the list and remained in view throughout the task. The search array was a 14x12 grid of 15mm² colour patches, adjacent patches were spaced 1mm apart. On the practice card the target was red, all distractors were green. The location of stimuli and perceptual distances between colour stimuli in CIE
L*u*v* in the four search conditions are presented in Figure 5.2. In both one-target conditions the target was orange (O1), in Condition 1T-pure the distractors were red (R) and yellow (Y); in Condition 1T-mixed an orange distractor (1T-mixed) was added. In condition 2T-pure both targets were orange (O1 and O2), the distractors were yellow (Y) and red (R). In condition 2T-mixed the targets were taken from separate English categories (one target was orange (O1), the other red (R)), but both targets were in the same Kwanyama category (shitilyana). The distractors were orange (O2) and yellow (Y).

Figure 5.2. Location and distances of colours in CIE L*u*v* space for all four search conditions. Target(s) indicated by box.
Procedure
The task was conducted either by the experimenter (English data) or by a native Kwanyama speaker (Kwanyama data). The English data were collected in the laboratory (using a MacBeth daylight lamp), the Kwanyama data were collected in natural day light avoiding direct sun and deep shade. The practice card was used to explain the task. Instructions were that the colour at the top of the grid was the target, that the grid of colours had to be searched for colours identical to this target and that all identical colours were to be marked with a pen. Participants had to start the task immediately on presentation of the card, to mark all target colours and to put the pen on the last colour when they had finished the task. Instructions were to complete the task as quickly as possible without compromising accuracy. Search time was recorded with a stopwatch by the experimenter. A clear acetate sheet was attached on top of each of the search cards to allow the stimulus cards to be re-used. All participants completed all four conditions of the task. Order of search tasks was randomised across participants.

5.2.3. Results

Search time
Median search times were calculated for each participant for each search condition. The means and estimated standard errors for the one-target conditions are presented in Figure 5.3.
Fig. 5.3. Search time in seconds (error bars shown +/- 1 standard error) for one-target conditions (1T-pure and 1T-mixed).

As shown above, both English and Kwanyama search times appear to be slower in the mixed than in the pure search condition. Furthermore, Kwanyama search times seem to be less affected than English search by adding a further distractor in the 1T-mixed condition.

The means and estimated standard errors for the two-target conditions are presented in Figure 5.4.
Fig. 5.4. Search time in seconds (error bars show +/- 1 standard error) for the two target conditions (2T-pure and 2T-mixed).

As Figure 5.4 shows, English and Kwanyama search times in the two target conditions appear to be slower in the mixed than the pure conditions. As for the one-target condition, Kwanyama search appears to have been slowed less by the changes in search list.

Analysis was performed by running a three-way analysis of variance across all conditions. The ANOVA was run with language (English, Kwanyama) as a between subjects factor and list-type (pure, mixed) and number of targets (1,2) as within subject factors. The results showed a significant main effect of language ($F[1,35] = 8.02$, $MSE = 834.73$, $p<0.01$), a significant effect of list-type ($F[1,35] = 58.66$, $MSE = 441.59$, $p<0.001$) and number of targets ($F[1,35] = 38.02$, $MSE = 315.72$, $p<0.001$). The list-type x language interaction was significant ($F[1,35] = 6.24$, $p<0.05$), as was the number of targets x language interaction ($F[1,35] = 12.42$, $p<0.01$) and the list-type x number of targets interaction ($F[1,35] = 12.03$, $MSE = 258.20$, $p<0.01$). The three way language x list-type x number of targets was also significant ($F[1,35] = 4.96$, $MSE = 258.20$, $p<0.05$).
Post-hoc testing (protected t-tests) showed that for the pure list conditions, there was no significant difference between English and Kwanyama search times ($t(35) = 0.51$, $p > 0.05$). However, there was a significant difference for the mixed-list conditions ($t(35) = 2.33$, $p < 0.05$). Furthermore, search times were faster in the pure compared to the mixed search conditions for both English ($t(17) = 5.14$, $p < 0.01$) and Kwanyama ($t(16) = 2.54$, $p < 0.05$) speakers.

Looking at just the number of targets, it was found that English and Kwanyama search times were not significantly different in the one-target conditions ($t(35) = 0.33$, $p > 0.05$), but there was a significant difference in the two-target conditions ($t(35) = 2.50$, $p < 0.05$). For English speakers only, search was significantly faster in the one-target conditions compared to the two target conditions ($t(17) = 4.91$, $p < 0.01$). This was not the case for Kwanyama speakers ($t(35) = 1.30$, $p > 0.05$).

**Accuracy**

A measure of accuracy was computed for each participant for each condition using A-prime (see Swets, 1996). Both the English sample and the Kwanyama sample made very few errors in the one-target conditions, both language groups made slightly more errors in the two-target conditions.

**One-target conditions (1T-pure and 1T-mixed)**

The mean $A'$ score across participants in condition 1T-pure was 0.997 for English speakers and 0.989 for Kwanyama speakers. In condition 1T-mixed, the average English participant score was 0.975, the average Kwanyama score 0.948.

**Two-target conditions (2T-pure and 2T-mixed)**

The mean $A'$ score across participants in search condition 2T-pure was 0.962 for English speakers and 0.961 for Kwanyama speakers. In condition 2T-mixed the average English participant score was 0.892, the average Kwanyama score 0.898.

A three-way ANOVA was run across all conditions with language as a between subjects factor and list-type and number of distractors as within subjects factors. Both within subjects factors were significant main effects (list-type: $F(1,35) = 21.94$, MSE
number of targets: $F[1, 35] = 27.11$, MSE = 0.0088, $p < 0.01$). The effect of language however, was not significant, nor were any of the interactions (maximum $F[1, 35] = 1.76, p > 0.05$).

Summary
The analysis showed all three factors to influence performance on the search tasks. As predicted, English search was slowed significantly more than Kwanyama search in the mixed compared to the pure conditions. Furthermore, English search times were significantly faster in the one-target compared to the two-target conditions; for Kwanyama speakers search times did not differ significantly across one-target and two-target conditions. The analysis of accuracy on the search task showed that the manipulations across search lists did not affect English and Balantu speakers differentially.

5.2.4. Discussion
Based on the fact that targets and distractors were in separate English colour categories in the pure list conditions, search should have been relatively easy for English speakers. However, the results showed that there was no significant difference between English and Kwanyama search times in the one- or the two-target pure condition. As mentioned in the introduction, a test of absolute differences between English and Balantu search is problematic since uncontrolled variables may also have affected performance. The second prediction, that the cost of adding further targets and distractors to the search list should be greater for English speakers, was supported. Hence, although overall English and Kwanyama search did not differ significantly on the pure search lists, search times did differ significantly on the mixed lists. This suggests that changing the visual search task from a cross-category search (for English) to a search where targets and distractors were no longer categorically separate, affected search times in line with the language spoken. It is likely that the cost of these changes was smaller for Kwanyama speakers because the pure conditions were only in line with English, but not Kwanyama colour categories.
The finding that overall search was slowed less for Kwanyama speakers compared to English speakers is consistent with warping resulting in within category equivalence. Adding a further distractor (O2) to the search list in 1T-mixed made it harder for English speakers since the additional distractor was in the same category as the target. If warping had occurred, the target and the distractor would have appeared to look more similar. For Kwanyama speakers however, the added distractor was in the same category as one of the original distractors. Therefore within category equivalence would predict that it is relatively easy to mark them both as distractors.

The results in search condition 2T-mixed are not straightforward. For 2T-mixed, both targets were in the same Kwanyama colour category, the two distractors were in a separate category. Thus, search should have been relatively easy. However, compared to the 2T-pure condition Kwanyama search was significantly slowed. For English speakers, the targets were in separate categories (one target was red, the other orange) and target and distractor categories overlapped. As predicted, search was significantly slowed for English speakers relative to 2T-pure and relative to Kwanyama search.

The findings are consistent with categorisation resulting in within category equivalence but they do not support the idea that perceptual space across category boundaries is warped. This supports Goldstone et al. (2001) who argued that cross-category advantages may be due to strategic biases (verbal strategies being used) rather than to warping of perceptual space. Although it was assumed that the visual search task was unlikely to invite a verbal strategy (see Experiment 4), the finding that English speakers were not advantaged suggests that the tasks may have been done verbally. Search times, especially for English speakers, were relatively slow (over 90 seconds in condition 2T-mixed) which would allow for a serial, verbal search (half a second per item). Thus, the results are consistent with the task being done, at least in part, by using verbal labels to distinguish between target and distractor. However, there is another possible explanation for the much longer search times in this type of search task. The current experiment required multiple manual responses which are relatively slow compared to the computerised target present/non-presents response required in Experiment 4. Therefore, the slow response times may be partly due to the type of response required.
5.3. Experiment Ten

Colour triads task

5.3.1. Introduction

Experiment 10 tested whether colour language influences the perceived similarity of colours in a triads task. In a triads design three colours are presented simultaneously and the task is to choose the most different of the three colours. Using this type of task, cross-cultural differences have been reported by Kay and Kempton (1984), Davies et al. (1998) and Pilling (2001). Comparing languages with different sets of colour terms allows a test of whether colours in separate linguistic categories appear to be more dissimilar than colours within a category.

Kay and Kempton (1984) reported a comparison of English and Tamauhara speakers. Tamauhara is an indigenous language of Mexico. It has a single term covering the blue and green region of colour space. The authors compared similarity judgements using triads which for English speakers included two stimuli from one category (for example blue) and one stimuli from a different category (for example green). For Tamauhara speakers all three stimuli were within the same category. The results showed clear effects of language: English speakers chose in line with English colour naming by selecting the tile with the different name as the most different. Tamauhara speakers, however, did not choose in line with English categories. They chose the (English) within category and the cross-category colour with almost equal frequency.

These findings suggest that the perceived distance between stimuli across the blue-green category boundary was stretched for English speakers and hence the perceived distances between blue and green stimuli were larger for English than for Tamauhara speakers. However, Kay and Kempton further report that the increase in perceived distance across the blue-green boundary for English speakers disappeared when a naming strategy was prevented.
Davies et al. (1998) compared English and Setswana speakers on a triads task and also found that similarity judgements were made in line with linguistic predictions. Thus, the more nominally isolated a colour was, the more likely it was to be chosen as the most different. Davies et al. also included triads for which the nominal predictions did not differ across languages. English and Setswana triad choices were more similar on these triads.

These category effects are unlikely to be independent of the perceptual distances between the colour stimuli in a triad. The more perceptually isolated one of the stimuli is (for example in CIE L°u°v° space), the more likely speakers of different (colour) languages should be to choose the same stimulus. Furthermore, as Shepp (1991) pointed out, the perception of the individual dimensions of colour (hue, lightness and saturation) is in part determined by the type of triad classification required (whether the stimuli are all from the same category or whether the judgement is across categories). In the triads experiment reported by Shepp, subjects were first instructed on the three dimensions of colour (hue, lightness and saturation) and told that two stimuli within a triad always shared a common hue. The task was to report these two stimuli. For English speakers, Shepp found that subjects failed to identify hue when stimuli in a triad were chosen from the same category, but they were successful when the stimuli used crossed a category boundary. If this is true, the perception of dimensions should be related to linguistic categorisations and hence should differ across languages.

The current experiment compares English and Balantu speakers (a language spoken in northern Namibia) on a triads task. If there is a relationship between colour language and colour perception, colour naming should have an effect on triad choice. Within a triad, a tile that is linguistically separate from the other two tiles should be more likely to be chosen as the most different as categorical relations should be primary (Shepp, 1991). When naming predictions between languages differ significantly, the frequency of choices should also differ significantly since choices should be made in line with the own language’s colour categories. As outlined above, language could affect triad choices either directly if a naming strategy was used on the task. Alternatively, language could also affect triad choice through warping of perceptual
space. When colour categorisation across languages does not differ, English and Balantu triad choices should be more similar.

The triads used here also addresses the influence of dimensional similarity on triad choice by varying the stimuli on hue and lightness. This allows a test of what similarity judgements are based on when there are no categorical separations between stimuli. Burns and Shepp (1988), following on from Garner (1974), argued that, for within category triads, similarity judgements are based solely on the overall similarity of colours: the two stimuli that are closest in perceptual colour space are grouped together. However, using a triads task, Boyles (2001) found that Ndonga (a language spoken in northern Namibia) children tended to group colours by lightness more than English children. The variations in hue and lightness allow a test of whether this salience of lightness is also found in adult speakers of northern Namibian languages.

The aim of the experiment and the subsequent analysis was to determine the influence of categorical and dimensional factors on triad choice and how these may differ across languages. Colours used in the current task were chosen so that some triads were cross-category for English speakers, but within category for Balantu speakers. As in Davies et al. (1998), the set also included triads that were within category for both language groups. English and Balantu triad choice was expected to be more similar when naming predictions between the two languages did not conflict (in the within category triads). Furthermore, in order to be able to test the influence of hue and lightness similarities, the triads were designed so that two stimuli in each triad shared a common hue and two stimuli shared a common lightness. This manipulation meant that triad classifications could either be done by grouping colours by hue or by lightness. For the triads that were cross-category for English speakers, the stimuli sharing hue were in the same category, the stimuli sharing lightness were in different categories.

The results will be tested correlationally; English and Balantu within category triad choices should be correlated more strongly than cross-category triad choices since the naming predictions of the two languages do not conflict. The results will be further analysed by looking at the possible predictors of triad choice: categorical isolation, hue isolation and lightness isolation. A tile in a given triad is categorically isolated if
it is in a different category than the other two stimuli. Furthermore, a colour within a
triad can be separated from the other colours either because it is a hue isolate (the
other two stimuli share a common hue) or a lightness isolate (the other stimuli share a
common lightness). If there is an effect of language on colour perception, English
triad choice should be better predicted by the categorical separations than Balantu
triad choice. Furthermore, if for Balantu speakers lightness similarities are weighted
more than hue similarities, Balantu triad choices should be predicted by the lightness
separations more than the hue separation.

The following predictions were made: firstly, naming data was expected to confirm
that colour nomenclature differs across the two languages. Secondly, in line with
findings of previous research (Davies, et al., 1998; Kay & Kempton, 1984; Pilling,
2001), a relationship should be found between naming and triad choice. Thirdly,
Balantu and English triad choices should differ most on those triads for which English
and Balantu naming predictions differ. Fourthly, if the lightness dimension is more
salient to Balantu speakers, Balantu triad choice should be better predicted by
lightness separations than English triad choice.

5.3.2. Method

Subjects

Twenty-one English (five male and sixteen female; age range 18-46, mean age 24.9)
and twenty (ten male, ten female; age range 17-58, mean age 29.6) Balantu speakers
took part in this experiment. The English participants were all students at the
University of Surrey and data was collected at the University; the Balantu data was
collected in the Omusati region of northern Namibia by a native Balantu speaker.

Stimuli

The stimuli were Munsell colours; the Munsell codes for all stimuli used in the triads
and naming task are presented in Table 5.2.
Table 5.2. Triad sets used in Experiment 10. The table shows triad type (within, between) for each set for Balantu and English speakers, Munsell codes (hue, value, chroma) for each tile and location of English category boundaries for cross-category sets.

<table>
<thead>
<tr>
<th>Triad set</th>
<th>Balantu</th>
<th>Munsell code</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red/Pink 1</td>
<td>Within</td>
<td>2.5 R 4/14, 2.5 R 5/14</td>
<td>Pink</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8.75 R 4/14, 8.75 R 5/14</td>
<td>Red</td>
</tr>
<tr>
<td>Red/Pink 2</td>
<td>Within</td>
<td>2.5 R 4/14, 2.5 R 5/14</td>
<td>Pink</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.5 R 4/14, 7.5 R 5/14</td>
<td>Red</td>
</tr>
<tr>
<td>Blue/Green</td>
<td>Within</td>
<td>5 BG 4/8, 5BG 5/8</td>
<td>Blue</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10 BG 4/8, 10 BG 5/8</td>
<td>Green</td>
</tr>
<tr>
<td>Green</td>
<td>Within</td>
<td>2.5 G 4/10, 2.5 G 5/10</td>
<td>Within</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.5 G 5/10, 7.5 G 4/10, 7.5 G 5/10</td>
<td></td>
</tr>
<tr>
<td>Pink</td>
<td>Within</td>
<td>2.5 R 5/10, 2.5 R 6/10</td>
<td>Within</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.5 R 6/10, 7.5 RP 5/10, 7.5 RP 6/10</td>
<td></td>
</tr>
<tr>
<td>Red</td>
<td>Within</td>
<td>5 R 4/14, 5 R 5/14</td>
<td>Within</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.5 R 4/14, 7.5 R 5/14</td>
<td></td>
</tr>
</tbody>
</table>

For each set, triads consisted of all possible sets of three tiles selected from the set of four. As shown in Table 5.2, triads were within and across category for English, but all were within category for the Balantu speakers. The triad sets are denoted by the English basic term for the colours within a set. There were two Red/Pink, a Blue/Green, a Green, a Pink and a Red set. The stimuli within a triad varied on hue and value at constant chroma. The variation of hue and value meant that two stimuli in each triad shared a common hue and two stimuli shared a common value.
For cross-category triads two of the stimuli belonged to the same English colour category whereas the third stimulus was categorically separate. An example is presented in Figure 5.5, which shows a schematic representation of the design of the Red/Pink 2 triad set. In the triad made up of tiles A, B and C, tiles A and B are pink and tile C is red.

Figure 5.5. Schematic representation of the stimuli used for the Red/Pink 2 triads. An example of a triad would be A-B-C.

| Table 5.3. Perceptual hue and value CIE distances (ΔE) for each triad set. |
|---------------------------------|-----------------|-----------------|
| Triad Type                     | Hue distance (dh) | Value distance (dv) |
| Red/Pink 1                     | 27              | 10.5           |
| Red/Pink 2                     | 23.1            | 10.5           |
| Green                          | 14.8            | 10.5           |
| Pink                           | 12.7            | 10.4           |
| Blue/Green                     | 12.7            | 10.4           |
| Red                            | 11.9            | 10.3           |
The average $\Delta E$ for the hue and value components ($dh$, $dv$ in Figure 5.5) are presented in Table 5.3. The value step was held constant across all triad sets, but the hue step varied. The stimulus differing on hue was perceptually the most isolated on all triads.

**Naming task**

The stimuli used to elicit naming responses consisted of twenty individual tiles, consisting of a Munsell colour (approximately 2x2 cm) mounted centrally on a square of white card (approximately 5x5 cm). The tiles were numbered from one to twenty on the back of the card.

**Triad task**

The triads were constructed by mounting the three Munsell colour tiles (approximately 2x2 cm) on round, grey card disks (approximately 10 cm diameter). The centres of three colours were equally spaced on the disk, approximately 2.5 cm apart, arranged in a triangular fashion (see Figure 5.6.). Triads were numbered on the back of the disks.

![Figure 5.6. Schematic representation of the triads used.](image_url)
Procedure

The tasks were administered either by the experimenter (English data) or by a native Balantu speaker (Namibian data). The English data were collected in the laboratory using a Macbeth daylight lamp, and the Namibian data were collected in natural daylight (direct sun and deep shade were avoided). All subjects were tested for colour vision problems using the City University colour vision test (Fletcher, 1980). Balantu instructions were verified by a process of back translation.

All participants completed both the naming and the triad task. Order of tasks was randomised across participants.

Naming task
Each colour was presented individually. The instructions were to name the colour using the simplest possible term which described the colour adequately. As much time as needed could be taken on each colour. Each response was recorded by the experimenter before presenting the next colour. The stimuli were shuffled after each trial ensuring the order of presentation was randomised across subjects.

Triads task
The triads task consisted of twenty-four triads (plus a practice disk made up of two green and one red tile) which were presented individually. Using the practice triad the task was explained. Instructions were to choose the colour that was the most different out of the three stimuli. The response was recorded before presenting the next triad. There was no time constraint for the response. To ensure the order of presentation of triads was randomised across all subjects, the disks were shuffled after each testing. Discs were used to try to randomise the orientation they were presented in.

On average the tasks took approximately 15 to 20 minutes to complete.
5.3.3. Results

*Naming task*

The most frequent and second most frequent naming response for each tile is presented in Table 5.4 for English speakers and in Table 5.5 for Balantu speakers.

Table 5.4. English Naming Data.

<table>
<thead>
<tr>
<th>Triad set</th>
<th>Munsell Code</th>
<th>Colour name response</th>
<th>Most frequent (%)</th>
<th>2(^{nd}) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red/ Pink 1</td>
<td>2.5 R 4/14</td>
<td>Pink</td>
<td>66.6</td>
<td>Red</td>
</tr>
<tr>
<td></td>
<td>2.5 R 5/14</td>
<td>Pink</td>
<td>85.9</td>
<td>Red</td>
</tr>
<tr>
<td></td>
<td>8.75 R 4/14</td>
<td>Red</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>8.75 R 5/14</td>
<td>Red</td>
<td>66.6</td>
<td>Orange</td>
</tr>
<tr>
<td>Red/ Pink 2</td>
<td>2.5 R 4/14</td>
<td>Pink</td>
<td>66.6</td>
<td>Red</td>
</tr>
<tr>
<td></td>
<td>2.5 R 5/14</td>
<td>Pink</td>
<td>85.7</td>
<td>Red</td>
</tr>
<tr>
<td></td>
<td>7.5 R 4/14</td>
<td>Red</td>
<td>95.2</td>
<td>Pink</td>
</tr>
<tr>
<td></td>
<td>7.5 R 5/14</td>
<td>Red</td>
<td>95.2</td>
<td>Pink</td>
</tr>
<tr>
<td>Blue/ Green</td>
<td>5 BG 4/8</td>
<td>Green</td>
<td>80.9</td>
<td>Blue</td>
</tr>
<tr>
<td></td>
<td>5BG 5/8</td>
<td>Green</td>
<td>76.2</td>
<td>Blue</td>
</tr>
<tr>
<td></td>
<td>10 BG 4/8</td>
<td>Blue</td>
<td>71.4</td>
<td>Green</td>
</tr>
<tr>
<td></td>
<td>10 BG 5/8</td>
<td>Blue</td>
<td>71.4</td>
<td>Green</td>
</tr>
<tr>
<td>Green</td>
<td>2.5 G 4/10</td>
<td>Green</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>2.5 G 5/10</td>
<td>Green</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>7.5 G 4/10</td>
<td>Green</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>7.5 G 5/10</td>
<td>Green</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td>Pink</td>
<td>2.5 R 5/10</td>
<td>Pink</td>
<td>85.7</td>
<td>Purple</td>
</tr>
<tr>
<td></td>
<td>2.5 R 6/10</td>
<td>Pink</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>7.5 RP 5/10</td>
<td>Pink</td>
<td>76.2</td>
<td>Purple</td>
</tr>
<tr>
<td></td>
<td>7.5 RP 6/10</td>
<td>Pink</td>
<td>95.2</td>
<td>Purple</td>
</tr>
<tr>
<td>Red</td>
<td>5 R 4/14</td>
<td>Red</td>
<td>95.2</td>
<td>Pink</td>
</tr>
<tr>
<td></td>
<td>5 R 5/14</td>
<td>Red</td>
<td>95.2</td>
<td>Pink</td>
</tr>
<tr>
<td></td>
<td>7.5 R 4/14</td>
<td>Red</td>
<td>95.2</td>
<td>Pink</td>
</tr>
<tr>
<td></td>
<td>7.5 R 5/14</td>
<td>Red</td>
<td>95.2</td>
<td>Pink</td>
</tr>
</tbody>
</table>
Table 5.5. Balantu Naming Data (IDK = “I don’t know”)

<table>
<thead>
<tr>
<th>Triad Set</th>
<th>Munsell Code</th>
<th>Colour name response&lt;sup&gt;10&lt;/sup&gt;</th>
<th>Most frequent (%)</th>
<th>2&lt;sup&gt;nd&lt;/sup&gt; (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red/Pink 1</td>
<td>2.5 R 4/14</td>
<td>Oshitiligane</td>
<td>95.0</td>
<td>Oshipingi</td>
</tr>
<tr>
<td></td>
<td>2.5 R 5/14</td>
<td>Oshitiligane</td>
<td>95.0</td>
<td>Oshipingi</td>
</tr>
<tr>
<td></td>
<td>8.75 R 4/14</td>
<td>Oshitiligane</td>
<td>95.0</td>
<td>Violet</td>
</tr>
<tr>
<td></td>
<td>8.75 R 5/14</td>
<td>Oshitiligane</td>
<td>95.0</td>
<td>Orange</td>
</tr>
<tr>
<td>Red/Pink 2</td>
<td>2.5 R 4/14</td>
<td>Oshitiligane</td>
<td>95.0</td>
<td>Oshipingi</td>
</tr>
<tr>
<td></td>
<td>2.5 R 5/14</td>
<td>Oshitiligane</td>
<td>95.0</td>
<td>Oshipingi</td>
</tr>
<tr>
<td></td>
<td>7.5 R 4/14</td>
<td>Oshitiligane</td>
<td>90.0</td>
<td>Violet</td>
</tr>
<tr>
<td></td>
<td>7.5 R 5/14</td>
<td>Oshitiligane</td>
<td>85.0</td>
<td>Orange</td>
</tr>
<tr>
<td>Blue/Green</td>
<td>5BG 4/8</td>
<td>Oshimbulau</td>
<td>70.0</td>
<td>Oshizizi</td>
</tr>
<tr>
<td></td>
<td>5BG 5/8</td>
<td>Oshimbulau</td>
<td>65.0</td>
<td>Oshizizi</td>
</tr>
<tr>
<td></td>
<td>10 BG 4/8</td>
<td>Oshimbulau</td>
<td>95.0</td>
<td>Oshizizi</td>
</tr>
<tr>
<td></td>
<td>10 BG 8</td>
<td>Oshimbulau</td>
<td>95.0</td>
<td>Oshizizi</td>
</tr>
<tr>
<td>Green</td>
<td>2.5 G 4/10</td>
<td>Oshizizi</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>2.5 G 5/10</td>
<td>Oshizizi</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>7.5 G 4/10</td>
<td>Oshizizi</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>7.5 G 5/10</td>
<td>Oshizizi</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td>Pink</td>
<td>2.5 R 5/10</td>
<td>Oshitiligane</td>
<td>85.0</td>
<td>IDK</td>
</tr>
<tr>
<td></td>
<td>2.5 R 6/10</td>
<td>Oshitiligane</td>
<td>65.0</td>
<td>IDK</td>
</tr>
<tr>
<td></td>
<td>7.5 RP 5/10</td>
<td>Oshitiligane</td>
<td>85.0</td>
<td>Oshipingi</td>
</tr>
<tr>
<td></td>
<td>7.5 RP 6/10</td>
<td>Oshitiligane</td>
<td>70.0</td>
<td>IDK</td>
</tr>
<tr>
<td>Red</td>
<td>5 R 4/14</td>
<td>Oshitiligane</td>
<td>90.0</td>
<td>Violet</td>
</tr>
<tr>
<td></td>
<td>5 R 5/14</td>
<td>Oshitiligane</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>7.5 R 4/14</td>
<td>Oshitiligane</td>
<td>90.0</td>
<td>Violet</td>
</tr>
<tr>
<td></td>
<td>7.5 R 5/14</td>
<td>Oshitiligane</td>
<td>85.0</td>
<td>Orange</td>
</tr>
</tbody>
</table>

The naming data support the prediction that the Balantu language does not have terms directly corresponding to the English pink and red terms, rather, the term oshitiligane was predominantly used for both colours. Furthermore, the naming suggests the distribution of category boundaries in the blue/green region of colour space differs. As can be seen in Table 5.5. Balantu speakers did have a term corresponding to green (oshizizi is the name given to all tiles in the Green set) and a term corresponding to blue (oshimbulau). The Blue/Green set was originally intended to be a cross-category triad set for both language groups.

<sup>10</sup> Some Balantu respondents used borrowed English terms for some colours.
However, Balantu naming showed all four tiles were predominantly labelled oshimbula making this set of triads also within category. In the remaining three sets of triads, all four tiles were labelled green (or oshizizi), pink (oshitiligane) or red (oshitiligane) respectively, thus confirming the within category status in both languages.

**Triad task**
Frequency of choices for each tile on all triads were calculated and are presented in Table 5.6. The tile most frequently chosen as the most different is presented in bold. The results show that on the majority of triads the most frequent choices made by English speakers differed from those made by Balantu speakers. For cross-category triads, the most frequently chosen tile was the same for English and Balantu on only four out of the twelve triads. For within category triads, English and Balantu speakers agreed as to which tile was the most different on seven out of the twelve triads.
Table 5.6. Frequency of choice for each tile (with position of English category boundary) for English and Balantu.

<table>
<thead>
<tr>
<th>Triad</th>
<th>Tile</th>
<th>English</th>
<th>Balantu</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.5 R 4/14</td>
<td>5.9 %</td>
<td>30.8 %</td>
</tr>
<tr>
<td>1</td>
<td>2.5 R 5/14</td>
<td>0 %</td>
<td>0 %</td>
</tr>
<tr>
<td></td>
<td>8.75 R 5/14</td>
<td>94.1 %</td>
<td>69.2 %</td>
</tr>
<tr>
<td>2</td>
<td>2.5 R 4/14</td>
<td>0 %</td>
<td>30.8 %</td>
</tr>
<tr>
<td></td>
<td>2.5 R 5/14</td>
<td>8.8 %</td>
<td>33.3 %</td>
</tr>
<tr>
<td></td>
<td>8.75 R 4/14</td>
<td>91.2 %</td>
<td>35.9 %</td>
</tr>
<tr>
<td></td>
<td>2.5 R 5/14</td>
<td>88.2 %</td>
<td>35.9 %</td>
</tr>
<tr>
<td></td>
<td>8.75 R 4/14</td>
<td>5.9 %</td>
<td>64.1 %</td>
</tr>
<tr>
<td></td>
<td>8.75 R 5/14</td>
<td>5.9 %</td>
<td>0 %</td>
</tr>
<tr>
<td>3</td>
<td>2.5 R 4/14</td>
<td>91.2 %</td>
<td>43.6 %</td>
</tr>
<tr>
<td></td>
<td>8.75 R 4/14</td>
<td>0 %</td>
<td>23.1 %</td>
</tr>
<tr>
<td></td>
<td>8.75 R 5/14</td>
<td>8.8 %</td>
<td>33.3 %</td>
</tr>
<tr>
<td></td>
<td>2.5 R 5/14</td>
<td>19 %</td>
<td>95 %</td>
</tr>
<tr>
<td></td>
<td>7.5 R 4/14</td>
<td>81 %</td>
<td>5 %</td>
</tr>
<tr>
<td></td>
<td>2.5 R 5/14</td>
<td>85.7 %</td>
<td>15 %</td>
</tr>
<tr>
<td>4</td>
<td>7.5 R 4/14</td>
<td>14.3 %</td>
<td>80 %</td>
</tr>
<tr>
<td></td>
<td>7.5 R 5/14</td>
<td>0 %</td>
<td>5 %</td>
</tr>
<tr>
<td></td>
<td>2.5 R 4/14</td>
<td>71.4 %</td>
<td>10 %</td>
</tr>
<tr>
<td>5</td>
<td>7.5 R 4/14</td>
<td>9.5 %</td>
<td>0 %</td>
</tr>
<tr>
<td></td>
<td>7.5 R 5/14</td>
<td>19 %</td>
<td>90 %</td>
</tr>
<tr>
<td>6</td>
<td>2.5 R 4/14</td>
<td>0 %</td>
<td>95 %</td>
</tr>
<tr>
<td></td>
<td>2.5 R 5/14</td>
<td>0 %</td>
<td>0 %</td>
</tr>
<tr>
<td></td>
<td>7.5 R 5/14</td>
<td>100 %</td>
<td>10 %</td>
</tr>
<tr>
<td>7</td>
<td>5 BG 4/8</td>
<td>17.6 %</td>
<td>82.1 %</td>
</tr>
<tr>
<td></td>
<td>5 BG 5/8</td>
<td>0 %</td>
<td>0 %</td>
</tr>
<tr>
<td>8</td>
<td>10 BG 4/8</td>
<td>82.4 %</td>
<td>17.9 %</td>
</tr>
<tr>
<td></td>
<td>10 BG 5/8</td>
<td>0 %</td>
<td>25.6 %</td>
</tr>
<tr>
<td></td>
<td>5 BG 4/8</td>
<td>0 %</td>
<td>59 %</td>
</tr>
<tr>
<td>9</td>
<td>5 BG 5/8</td>
<td>17.6 %</td>
<td>59 %</td>
</tr>
<tr>
<td></td>
<td>10 BG 4/8</td>
<td>82.4 %</td>
<td>15.4 %</td>
</tr>
<tr>
<td></td>
<td>10 BG 5/8</td>
<td>0 %</td>
<td>0 %</td>
</tr>
<tr>
<td>10</td>
<td>5 BG 5/8</td>
<td>82.4 %</td>
<td>64.1 %</td>
</tr>
<tr>
<td></td>
<td>10 BG 4/8</td>
<td>17.6 %</td>
<td>35.9 %</td>
</tr>
<tr>
<td></td>
<td>10 BG 5/8</td>
<td>0 %</td>
<td>0 %</td>
</tr>
<tr>
<td>11</td>
<td>5 BG 4/8</td>
<td>79.4 %</td>
<td>56.4 %</td>
</tr>
<tr>
<td></td>
<td>10 BG 4/8</td>
<td>0 %</td>
<td>0 %</td>
</tr>
<tr>
<td></td>
<td>10 BG 5/8</td>
<td>20.6 %</td>
<td>43.6 %</td>
</tr>
</tbody>
</table>

189
<table>
<thead>
<tr>
<th>Triad</th>
<th>English</th>
<th>Balantu</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>2.5G 5/10</td>
<td>61.9%</td>
</tr>
<tr>
<td></td>
<td>7.5 G 4/10</td>
<td>28.6%</td>
</tr>
<tr>
<td></td>
<td>7.5 G 5/10</td>
<td>9.5%</td>
</tr>
<tr>
<td>14</td>
<td>2.5G 4/10</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>7.5 G 5/10</td>
<td>76.2%</td>
</tr>
<tr>
<td></td>
<td>7.5 G 4/10</td>
<td>23.8%</td>
</tr>
<tr>
<td>15</td>
<td>2.5G 4/10</td>
<td>57.1%</td>
</tr>
<tr>
<td></td>
<td>7.5 G 4/10</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>7.5 G 5/10</td>
<td>42.9%</td>
</tr>
<tr>
<td>16</td>
<td>2.5G 5/10</td>
<td>47.6%</td>
</tr>
<tr>
<td></td>
<td>7.5 G 5/10</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>2.5G 4/10</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>7.5 G 4/10</td>
<td>76.2%</td>
</tr>
<tr>
<td></td>
<td>7.5 G 5/10</td>
<td>42.9%</td>
</tr>
<tr>
<td></td>
<td>2.5G 4/10</td>
<td>57.1%</td>
</tr>
<tr>
<td></td>
<td>7.5 G 4/10</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>7.5 G 5/10</td>
<td>23.8%</td>
</tr>
<tr>
<td>17</td>
<td>2.5 R 5/10</td>
<td>42.9%</td>
</tr>
<tr>
<td></td>
<td>2.5 R 6/10</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>7.5 RP 6/10</td>
<td>57.1%</td>
</tr>
<tr>
<td>18</td>
<td>2.5 R 5/10</td>
<td>14.3%</td>
</tr>
<tr>
<td></td>
<td>2.5 R 6/10</td>
<td>42.9%</td>
</tr>
<tr>
<td></td>
<td>7.5 RP 5/10</td>
<td>42.9%</td>
</tr>
<tr>
<td>19</td>
<td>2.5 R 5/10</td>
<td>76.2%</td>
</tr>
<tr>
<td></td>
<td>7.5 RP 5/10</td>
<td>9.5%</td>
</tr>
<tr>
<td></td>
<td>7.5 RP 6/10</td>
<td>14.3%</td>
</tr>
<tr>
<td>20</td>
<td>2.5 R 6/10</td>
<td>42.9%</td>
</tr>
<tr>
<td></td>
<td>7.5 RP 5/10</td>
<td>57.1%</td>
</tr>
<tr>
<td></td>
<td>7.5 RP 6/10</td>
<td>0%</td>
</tr>
<tr>
<td>21</td>
<td>5 R 4/14</td>
<td>79.4%</td>
</tr>
<tr>
<td></td>
<td>5 R 5/14</td>
<td>2.9%</td>
</tr>
<tr>
<td></td>
<td>7.5 R 5/14</td>
<td>17.6%</td>
</tr>
<tr>
<td>22</td>
<td>5 R 4/14</td>
<td>2.9%</td>
</tr>
<tr>
<td></td>
<td>5 R 5/14</td>
<td>73.5%</td>
</tr>
<tr>
<td></td>
<td>7.5 R 4/14</td>
<td>23.5%</td>
</tr>
<tr>
<td>23</td>
<td>5 R 5/14</td>
<td>44.1%</td>
</tr>
<tr>
<td></td>
<td>7.5 R 4/14</td>
<td>52.9%</td>
</tr>
<tr>
<td></td>
<td>7.5 R 5/14</td>
<td>2.9%</td>
</tr>
<tr>
<td>24</td>
<td>5 R 4/14</td>
<td>38.2%</td>
</tr>
<tr>
<td></td>
<td>7.5 R 4/14</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>7.5 R 5/14</td>
<td>61.8%</td>
</tr>
</tbody>
</table>
Correlations between English and Balantu triad choice

To compare English and Balantu triad choice, the frequency of choice of each tile across participants was calculated. Correlational analysis of tile choices on each triad confirmed that English and Balantu triad choices were more similar on the within category compared to the cross-category triads. English and Balantu within category triad choice was strongly correlated ($r = 0.51, p < 0.01$), but the correlation between English and Balantu cross-category triad choice was much lower ($r = 0.052, p > 0.05$). The correlations were significantly different (Fisher’s $r'z = 2.51, p < 0.05$).

It could be argued that the stronger correlation of English and Balantu triad choices on within category triads suggest that within category choice was more obvious than cross-category choice. However, the fact that the hue distance was on average greater on cross-category triads compared to within category triads (see Table 5.3.) suggests that choices on the cross-category triads should have been more obvious.

Cross-category triads

A triad was made up of three colour stimuli, hence on each triad three possible choices could be made. On cross-category triads, the first possibility was to make a ‘category’ choice, in line with the English categorisation. This was also a ‘hue’ choice since the two stimuli that belonged to the same linguistic category also shared a common hue. The second possibility was to group the colours by lightness and choose the tile differing on value as the most different. Thirdly, a ‘haphazard’ choice could be made.

On all cross-category triads, one of the three tiles was in a separate English linguistic category. English naming predicted this tile to be chosen. Thus, for these triads a category score was calculated, each time the triad choice was in line with the English nominal prediction, a score of one was awarded. A maximum of four could be obtained on each set of triads. Mean category scores across participants are presented in Table 5.7. The proportion of hue (category) and value choices for English and
Balantu speakers is presented in Figure 5.7. (missing values indicate haphazard choices).11

Table 5.7. Mean Category Scores (and SD) for each cross-category triad set type for English and Balantu.

<table>
<thead>
<tr>
<th>Triads</th>
<th>English</th>
<th>Balantu</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red/Pink 1</td>
<td>3.65 (.65)</td>
<td>1.85 (1.47)</td>
</tr>
<tr>
<td>Red/Pink 2</td>
<td>3.38 (.81)</td>
<td>0.35 (.99)</td>
</tr>
<tr>
<td>Blue/Green</td>
<td>3.26 (.99)</td>
<td>1.54 (.72)</td>
</tr>
</tbody>
</table>

Figure 5.7. Mean category (hue) and value scores for each cross-category triad set type for English and Balantu speakers (value scores shown in dotted line).

11 The possibility that triad choice may differ in line with individual cross-category and within category classifications was also considered. Frequency of choice was calculated for each tile for English and Balantu speakers depending on whether the participant had given all three tiles within a triad the same name (within category) or not (cross-category). However, this did not change the distribution of tile choices significantly (see Appendix 4).
As shown above, Balantu speakers made less category choices than English speakers in all three triad sets, which suggests that Balantu speakers followed the English naming prediction less than English speakers. The difference between English and Balantu triad choices was most pronounced on the Red/Pink 2 triads. This is not surprising since perceptual distances between the red and the pink tiles were much greater in the Red/Pink 1 than in the Red/Pink 2 set and the Blue/Green set. When Balantu speakers did not make category choices, the choices were predominantly lightness choices.

Within category triads
When all three tiles in a triad were taken from same linguistic category none of the tiles was nominally isolated and hence language was unlikely to affect within category judgements. Since the stimuli in a triad were all categorically the same, choices between languages should be more similar. As mentioned above, the stimuli within a triad varied on hue and value, hence two colours shared a common hue and two a common value. For within category triads the possible choices were: the colour differing on hue could be chosen as the most different; the stimuli differing on value could be chosen or a haphazard choice could be made.

How similar were English and Balantu choices when nominal predictions did not differ? And what were choices based on when the task could not be done categorically? For the within category triads, hue and value scores were calculated. Each time the tile differing on hue was chosen as the most different a score of one was added to the hue score; each time a value choice was made, a score of one was added to the value score. Both hue and value scores for within category triads are presented in Figure 5.8 and Table 5.8. Hue and value scores should add up to a total of four per triad set; missing values indicate haphazard choices.
Figure 5.8. Mean Value and hue scores for each within category triad set for English and Balantu speakers.

Table 5.8. Mean Value and Hue scores (and SD) for each triad set type for English and Balantu

<table>
<thead>
<tr>
<th>Triads</th>
<th>English hue</th>
<th>English value</th>
<th>Balantu hue</th>
<th>Balantu value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Green</td>
<td>1.95 (.92)</td>
<td>1.95 (.92)</td>
<td>0.4 (.82)</td>
<td>3.4 (.88)</td>
</tr>
<tr>
<td>Pink</td>
<td>2.19 (.75)</td>
<td>1.57 (.81)</td>
<td>0.1 (.31)</td>
<td>3.9 (.31)</td>
</tr>
<tr>
<td>Red</td>
<td>1.24 (1.05)</td>
<td>2.68 (1.12)</td>
<td>0.36 (0.54)</td>
<td>3.33 (.98)</td>
</tr>
</tbody>
</table>

English and Balantu triad choices differed considerably on the within category triads also. As can be seen in Table 5.9 (and Figure 5.8) Balantu speakers frequently made value choices. This indicates that similarity judgements were made on the basis of shared value, by choosing the tile differing on value as the most different. English speakers made considerably less value choices than Balantu speakers, however it is also worth noting that English speakers made more value choices in the within category condition than in the cross-category condition. This is likely to be due to the
fact that the hue separations in the within category triads were on average smaller than in the cross-category triads (see Table 5.3).

It should also be noted that for cross-category triads, category choices were hue choices: the stimulus that was categorically separate from the other two stimuli was separated by a hue step in all triads. It could be argued that the finding that English triad choice was in line with English colour categories was not a category effect but an effect of hue separation. However, the hue separation in the Blue/Green set was no bigger than the average within category hue separation, yet English speakers made more hue (category) choices in the Blue/Green set than in any within category triad set.

To test the differences between Balantu and English tile choices on within and cross-category triads, the category scores for the cross-category triads were added up and compared to the sum of within category triads hue scores. Results were analysed by a two-way ANOVA. The two factors were language (between subjects: English, Balantu) and triad type (across versus within). There was a significant effect of language ($F[1,39] = 314.12, \text{MSE} = 1.95, p<0.001$) and a significant effect of triad type ($F[1, 39] = 119.91, \text{MSE} = 2.57, p<0.001$). The language x triad type was also significant ($F[1, 39] = 14.8, p<0.01$). As Figure 5.9 shows, cross category choices were more different than within category choices across languages. However, language also strongly affected within category choices.
Post-hoc testing by protected t-tests showed that the difference between English and Balantu triad choices was highly significant for cross-category ($t[39] = 14.87, p<0.001$) as well as within category triads ($t[39] = 10.16, p<0.001$). Furthermore, both language groups made significantly more hue choices on cross-category compared to within category triads (English: $t[19] = 9.91, p<.001$; Balantu: $t[18] = 5.62, p<0.001$).

English and Balantu triad choices clearly differed on both cross-category and within category triads. On cross-category triads, English speakers grouped colours in line with the English colour categorisation. Balantu triad choices did not follow English naming predictions. On within category triads, for which the naming predictions did not differ across languages, English and Balantu triad choices were also significantly different. This suggests that other factors besides categorical separations influenced triad choice. Furthermore, as the most perceptually isolated colour (in CIEL*u*v*) in each triad was the hue isolate, the finding that Balantu speakers made lightness choices suggests that they were not attending to overall perceptual isolation.
Predictors of triad choice

The analysis of cross-category and within category triad choices showed clear differences between the English and the Balantu performance on the task. The findings suggest that Balantu speakers attended more to shared lightness than the overall similarity of colour stimuli. Further analysis was done by isolating the individual factors tested in this experiment and assessing how well they predicted English and Balantu triad choices. The three predictors tested were: firstly whether the tile was categorically separate, secondly whether the tile was separated by hue and thirdly whether it was separated by lightness. Each tile within each triad was given a category, a hue and a lightness score (1 to indicate the tile was the isolate within the triad, 0 to indicate it was not isolated). Stepwise regressions were run on English and Balantu triad choice with category, hue and lightness isolations as the independent factors.

For English triad choices, the regression went through three steps. R was significantly different from zero at the end of the analysis (R = 0.87, $F[3,68] = 67.29$, $p<0.001$). After step one was hue isolation was added to categorical isolation, this improved $R^2$ by 0.096. After step two, lightness isolation was added, improving $R^2$ by 0.13. For categorical isolation $\beta = 0.45$ ($t = 5.65$, $p<0.001$), for hue isolation $\beta = 0.60$ ($t = 6.99$, $p<0.001$), for lightness isolation $\beta = 0.40$ ($t = 5.89$, $p<0.001$).

For Balantu triad choices, the regression went through two steps. R was significantly different from zero at the end of the analysis (R = 0.82, $F[2,69] = 68.65$, $p<0.001$). After step one was categorical isolation was added to lightness isolation, this improved $R^2$ by 0.031. For lightness isolation $\beta = 0.84$ ($t = 11.71$, $p<0.001$), for category $\beta = 0.18$ ($t = 2.53$, $p<0.05$).

The regression analysis confirmed that English and Balantu triad choices were best predicted by different factors. The strongest predictor of English triad choice was categorical isolation, followed by hue isolation. The regression showed that Balantu triad choice was most strongly predicted by lightness isolation. Hue isolation was not a significant predictor of Balantu triad choice. As mentioned above, the perceptually most isolated colour on all triads was the hue isolate. Hence, if choices were made in
line with CIE $L^*u^*v^*$ colour space, hue should be a significant predictor of triad choice. The finding that Balantu speakers did not respond to the overall measure of perceptual isolation, but only the lightness isolation suggests that the CIE $L^*u^*v^*$ colour space is not appropriate for mapping their perceived similarity of colours.

**Summary**

The analysis of triad choices showed significant differences between English and Balantu triad choice. Triad choices were most different on cross-category triads for which naming predictions differed. However, the difference between English and Balantu triad choices was also significant on within category triads. Furthermore, type of triad influenced choice. More lightness choices were made on the within category triads by both English and Balantu speakers. Overall Balantu triad choice was best predicted by lightness isolation, English by categorical and hue isolation. This suggests that the salience of hue and value dimensions may differ across languages.

**5.3.4. Discussion**

Consistent with previous studies using triads to elicit colour similarity judgements (Davies et al, 1998; Kay & Kempton, 1984; Pilling, 2001), a significant relationship between colour naming and colour perception was found. The first prediction was supported: colour naming differed significantly across the languages tested. Balantu speakers did not have separate red and pink terms and the distribution of the blue/green category boundary differed significantly. The second prediction was also supported, a relationship was found between colour naming and triad choice. English triad choice was in line with English colour categorisation, Balantu triad choice was not in line with English categories.

The third prediction was only partly supported. Although triad choice differed significantly for cross-category triads, English and Balantu triad choices were also significantly different when naming predictions did not conflict. The analysis of choices on triads that were within category for both English and Balantu speakers revealed that Balantu speakers were significantly more likely to select the colour
differing in value as the most different. It is important to note that such a value choice was not in line with overall perceptual distances (in CIE L*u*v*); a choice based on perceptual isolation predicted a hue choice. English triad choice was significantly predicted by hue isolation, Balantu triad choice however was not. When looking at the dimensional components of perceptual isolation, the results showed that the strongest predictor of Balantu triad choice was the lightness isolation. This suggests that lightness similarities are more salient to Balantu speakers rather than overall similarity of colours (in L*u*v* colour space). One possible explanation for this result can be found in an early study by Luria (1976). Luria (1976) reported a series of colour perception experiments conducted in Uzbekistan comparing non-literate traditional farming people with others from the same village who had undergone some literacy training. Luria found that the non-literate group was much less likely to group colours by hue, even when instructed to do so. This suggests that factors besides cross-cultural differences in colour nomenclature may be affecting performance on the tasks presented here.

Triad type had a significant effect on both English and Balantu triad choice. Whereas Balantu speakers made more value choices than English speakers in both triad conditions it is also worth noting that value choices were less frequent in the cross category condition for both languages. Thus, when no category boundary was present within a triad Balantu and English speakers made more value choices. The fact that Balantu speakers made more hue choices in the cross-category condition may suggest that red-pink boundary is covertly present although the language does not accord it a separate colour term.

It should be noted that there were some problems with the design of the triads task. The experiment attempted to test the influence of categorical and dimensional factors on triad choice. Thus, two questions were addressed: the question of whether triad choices differed in line with colour language and the question of whether there were across-language differences in the grouping of colours by hue or lightness. The fact that both variables were manipulated within a triad meant that it was difficult to assess the effect of one variable independent of the other. The effect of category found for English speakers on cross-category triads is probably confounded by the fact that the
hue separations were on average bigger than in within category triads. Furthermore, it is difficult to assess whether Balantu speakers made less category choices because their language does not make the categorical distinctions or because they attend more to the lightness dimension.

The finding that English and Balantu triad choices differed not only on cross-category but also on within category triads is nonetheless an interesting result. It suggests that language is affecting colour perception not only directly by the use of direct naming strategies. It suggests that the perceptual representation of colours may vary across languages. The notion of the perceptual integrality of colour dimensions (Burns & Shepp, 1988) may need testing cross-culturally.
5.4. General Discussion

The cross-cultural comparisons presented in Chapter 5 found a significant relationship between language and colour cognition. Significant differences in the performance of English speakers and Ovambo speakers were found on the naming, the visual search and the triads tasks. The naming tasks confirmed that the languages used in this comparison differed on the number of basic colour terms and the position of category boundaries.

The visual search task used in Experiment 9 was thought to be the perceptually purer out of the two tasks presented in this chapter. The task was to search a large array of colours for targets. The number of different distractors and targets were varied across conditions and based on English and Kwanyama naming this affected the categorical relations of target and distractors differently for the language groups. In line with previous findings (Pilling & Davies, 2003) performance on the search task was expected to depend in part on the categorical distinctions made in the languages tested.

The results of the visual search task showed that Kwanyama and English search times were not significantly different in the pure search conditions. This is not consistent with the results of the search task in Experiment 4 which showed cross-category search to be faster than within category search for English speakers. This was thought to have implications for cross-cultural comparisons on search tasks: discriminations should be easier for languages that make the categorical distinction compared to languages that do not. It is possible that the computerised search task with only one target in each search array and the multi-target pen and paper task used for the cross-cultural comparison differed in their sensitivities or indeed that the tasks were simply too different to allow direct, meaningful comparisons. Not only did the experiments require quite different responses (detection of single target vs. location of all targets in an array), Experiments 4 and 9 also manipulated different variables. Set size was manipulated only in the computerised search task, whereas number of targets varied only in Experiment 9.
The prediction that changing the (English) categorical separations within a search list should slow English search significantly more than Kwanyama search was supported. The findings suggest that performance on a visual search task is influenced by the categorical distinctions given by the language spoken. Furthermore, the results of the search task are consistent with language having warped within category representations making within category stimuli more similar. For Kwanyama speakers who have fewer linguistic categories, this would have meant that more of the colours looked more alike, making search easier by being able to eliminate similar distractors faster.

It is possible that the task was done at least in part using a verbal strategy. Search times were relatively slow, which suggests that verbal labels may have been used. However, a verbal strategy would have predicted an advantage for English speakers on some conditions (1T-pure and 2T-pure). However, it is possible that the different language groups used different strategies, that English speakers used a verbal strategy more than Balantu speakers. This would be consistent with the greater costs for English speakers of adding additional targets or distractors since a verbal strategy should have slowed down search.

In support of linguistic relativity, English and Balantu triad choices in Experiment 10 differed significantly. On triads that were cross-category for English speakers and within category for Balantu speakers, English speakers were more likely to choose in line with English nominal isolation than Balantu speakers. However, the results of Experiment 10 also suggest that the differences in colour cognition between the groups were more complex than a direct naming or language effect would suggest. Triad choices also differed significantly on those triads that were within category for both language groups, hence for which the nominal isolation scores did not differ. This suggests that the relationship between language and colour cognition is unlikely to be a direct effect of using verbal strategies on such tasks as triad tasks. Furthermore, it was found that Balantu speakers were more likely than English speakers to base their similarity judgements on value. This could be related to Luria's (1976) finding that non-literate people are less likely to group colours by hue than literate people. The results also raise the question of whether the integrality of stimulus dimensions reported by Burns and Shepp (1988) is a cognitive universal.
In terms of testing CP by comparing languages that make different categorical distinctions, the results of Chapter 5 were mixed. Experiment 10 did show that triad choices were most different when nominal predictions between languages conflicted. This is consistent with either a perceptual or a verbal account of CP. However, the fact that there were significant differences between English and Balantu on within category triads also, suggests that the difference on cross-category triads were not necessarily CP effects. The results of the comparison of English and Kwanyama speakers on visual search were also not consistent with CP. Overall, the comparison of English and Owambo speakers suggests that there are small differences in colour perception across languages. However, the effects are unlikely to be direct effects of naming.

Taken together, the results of the cross-cultural comparison in Chapter 5 and their implications for CP are not clear-cut. This may in part be due to the designs of the task used. In the search task in Experiment 9 the stimuli crossed two category boundaries for English speakers and one category boundary for Kwanyama speakers. The design allowed for a test of CP within languages, however, it complicated the comparison between languages. A simpler test of CP (and the account proposed in Chapters 2 and 3) would have been to include a search condition that was cross-category for English speakers (for example orange targets amongst red distractors), but within category for Kwanyama (target and distractors shitilyana). This test would also allow a direct comparison with the search task used in Experiment 4.

The triads task in Experiment 10 attempted to test the influence of various factors on triad choice. Thus, not only did the triads differ in terms of the categorical identity of stimuli (across versus within), an attempt was also made to test the effects of dimensional relations by varying stimuli on hue and lightness. This is likely to have complicated the interpretation of possible category effects. The issues of categorical effects and the influence of shared dimensions on category choices need to be tested either separately or by including additional conditions. If both factors were tested in the same experiment, the stimulus set would need to be made up of a matrix of stimuli in which the size of the hue separations equalled that of the lightness separations.
Ideally the set would need to include cross-category triads defined by a lightness separation and triads that were also cross-category for Balantu speakers.

Overall, the results of Chapter 5 present some evidence that colour language affects the performance on colour perception tasks. Significant effects of language were found for both similarity judgements in the triads task and search times in the visual search task.
15) (now p.99, para 1) the two hypothetical effects or routes of CP are outlined in
more detail to make the argument clearer that even though target name generation
may be necessary for CP, it is not necessarily a direct matching to labels process.

16) Typo corrected

17) (now p.128) Some more explanation was added to distinguish between test stimuli
(referred to in the sentence starting “however”) and the target stimuli (referred to
in the sentence starting “overall”).

18) (now p.135): has been changed to B1-G1

20) (now p.198) Discussion of triads experiments tasks now cites Luria’s finding that
non-literates prioritise grouping by lightness over grouping by hue as a possible
explanation for the differences found here between English and B

21) (now p.201) differences between the computerised search task used in Experiment 4
and the multiple-target search of Experiment 9 are emphasised. The tasks may
have been too different to allow direct comparison.

22) Some additions have been made to the general discussion. Firstly, each summary
starts with the answer to the questions stated in the overview of the discussion
(p.205) (is CP as misnomer – p.209; is CP a memory effect – p.213; cross-cultural
CP – 217). Hence, an attempt has been made to answer these questions. A
paragraph evaluating the thesis work in relation to CP in general has been added
to the discussion (p. 218)
Chapter Six

Discussion

6.1. Overview

This thesis addressed questions relating to the nature and the origin of categorical colour perception. Several different experimental approaches were used in an attempt to examine the relationship between language and colour cognition, both within and across languages. There were two main questions: (i) is categorical perception a misnomer? Hence, is CP a direct language, or memory, effect rather than truly perceptual? (ii) Is CP only found for speakers of languages that make certain categorical distinctions? The principal findings relating to these questions are discussed below.

6.2. Is categorical perception a misnomer?

Is CP a direct language effect?

Typically, evidence of better cross-category than within category colour discrimination has been attributed to underlying perceptual mechanisms, i.e. warping of perceptual space (Bornstein & Korda, 1984; Boynton, Fargo, Olson & Smallman, 1989; Uchikawa & Shinoda, 1996). Roberson and Davidoff (2000), however, argued that CP is not a perceptual, but rather a language effect based on the comparison of verbal labels. In order to test this conjecture, they tested CP within a dual-task framework and found verbal interference to reduce CP. However, as argued in Chapter 2, it seems likely that a single name code could be retained over an inter-stimulus interval whilst carrying out an additional verbal task. The experiments presented in Chapter 2 directly tested Roberson and Davidoff’s (2000) language account of CP.

Roberson and Davidoff argued that the effect of verbal interference was on name retention: the target was labelled and an attempt to retain the label across the inter-
stimulus interval was disrupted by verbal, but not visual interference. The results presented here show there to be a problem with this account as CP survived verbal interference when interference conditions were unblocked (Experiment 1 and 2). Furthermore, verbal interference did not affect performance in a colour-to-name matching task (Experiment 3) where the instructions were explicitly to label the target and retain the name code. Taken together these findings suggest that the effect found by Roberson and Davidoff was not on name retention per se, but possibly on choice of strategy. Tasks such as delayed discrimination tasks may be open to strategic biases (Goldstone, Lippa & Shiffrin, 2001); different strategies, i.e. attempting to retain either perceptual or verbal codes, may be used depending on aspects of the task design. In blocked interference conditions, it is possible that subjects thought a verbal strategy would be ineffective when the task was presented with verbal interference, and hence switched to a visual strategy. This is consistent with Schooler and Engstler-Schooler (1990) who showed that participants can be instructed to use either verbal or visual codes on a colour discrimination and recognition task. With unblocked interference (Experiments 1 and 2), the type of interference is not predictable and hence a blanket strategy has to be used.

A delayed discrimination task may invite or encourage a verbal strategy. However, this does not mean that perceptual CP could not be found in other types of tasks. The use of verbal codes is unlikely to be useful in a visual search task as it is a purer perceptual task. Experiment 4 found evidence for CP in a search task; cross-category search was faster and more accurate than within category search. This suggests that explanations of CP have to consider that the mechanisms leading to CP may be different in different types of tasks.

The findings reported in Chapter 2 question Roberson and Davidoff’s conclusions as verbal interference was found to not affect retention of a single name code. However, the account of CP being the result of a comparison of name codes could still be correct. In the verbal interference condition in blocked interference tasks, the target may not have been labelled in the first place which may have prevented CP.
The experiments presented in Chapter 3 tested the question of whether target name generation is necessary for CP. The type of interference used was Stroop interference (Stroop, 1935; MacLeod, 1991). If the target in a delayed discrimination task has to be named, presenting incongruent Stroop interference concurrently with the target should reduce or even eliminate CP. The results of Experiment 5 supported this. Furthermore, when the target colour was presented with congruent Stroop interference, CP was strongly facilitated. Flowers and colleagues (Flowers & Blair, 1976; Flower & Dutch, 1976) argued that it is tasks which require access to verbal codes that are susceptible to Stroop interference. If this is the case, the findings suggest that CP requires access to colour name codes.

If Roberson and Davidoff are right in proposing that CP arises because name codes are compared and this comparison facilitates cross-category (e.g. blue-green), but not within category (e.g. blue-blue) discrimination, not only the target but also the test stimuli would have to be named. Experiment 6 tested this by introducing Stroop interference at the response stage. The results were not clear-cut, but CP was lost in all Stroop interference conditions. Although the results were inconclusive in answering the question whether test name generation was necessary for CP, the results did suggest that congruent and incongruent Stroop interference at response stage did not selectively affect CP.

The visual search task in Experiment 4 found evidence for CP in a quick perceptual task. It is possible that the 5 second inter-stimulus interval used in Experiments 5 and 6 meant that the use of verbal codes was necessary or primary, as the perceptual codes had decayed. In a series of visual perception experiments using letters, Posner and colleagues (Posner, Boies, Eichelman & Taylor, 1969; Posner & Keele, 1967) found that visual codes were available for up to 1500ms. Reducing the inter-stimulus delay to 1000ms showed the same pattern of results as the long ISI when Stroop interference was presented concurrently with the target. When Stroop interference was at test presentation, CP survived in the congruent and incongruent conditions. This suggests that test name generation is not necessary for CP and hence that CP is not merely a comparison of verbal labels. Taken together, these findings suggest that Roberson and Davidoff’s language account of CP is incomplete.
The results of the Stroop experiments, especially the finding that Stroop interference at test presentation does not selectively affect colour discrimination, suggest that even though targets are encoded verbally (or else Stroop interference at target presentation should have no effect), a categorical code is also generated along with the name code. This is consistent with Bornstein and Korda (1984) who argued that a category code for colours is available early on in visual processing. As Stroop interference at test did not affect CP, the categorical code is likely to contain more information than the verbal code. It suggests it contains some perceptual information also.

In Chapter 3 the possibility was raised that Stroop interference at target presentation may be similar to the priming effects reported by Rosch (1975) and Neuman and D’Agostina (1981). These studies showed that priming colour discrimination judgements with the colour word facilitated discrimination. Rosch argued that in response to a category name, subjects generate a representation of the category prototype and that the representation is in the form of a concrete visual code. Thus, it is possible that presenting the target colour with Stroop labels similarly activates a mental code which contains information about the colour. This idea is consistent with Stroop interference at test not selectively affecting CP, as it suggests that no access to the verbal codes was necessary.

The effects of Stroop interference on CP are not consistent with a perceptual warping account. Warped perceptual codes should have been available in the incongruent conditions also. However, it is important to note that within category discrimination in all experiments was above chance level. This shows that perceptual codes must also play a part in the discrimination task; however, they are unlikely to be responsible for the cross-category advantage.

As target name generation was found to be important for CP in colour discrimination tasks, the account of the effect is still likely to be primarily verbal rather than perceptual. Furthermore, although the effects of verbal interference found in blocked tasks (Pilling et al., 2003; Roberson & Davidoff, 2000) are unlikely to have been on name retention, CP in delayed discrimination tasks could still be a memory effect of some sort. CP may be an artefact of the memory component used in these tasks.
Summary
The results presented in Chapters 2 and 3 of the current thesis suggest that categorical perception is a misnomer, but only in certain type of tasks. The results suggest that CP can be based on perceptual or verbal codes and that the locus of CP depends on the demands of the task. The visual search task (Experiment 4) showed a clear category effect; however it is unlikely to be a direct effect of naming since visual search tasks are relatively pure perceptual tasks. It is unlikely that the task was done using only verbal codes and hence the finding supports a perceptual account of CP in visual search tasks. Thus, a language account of CP is likely to apply only to certain tasks. Secondly, the results of Chapter 2 and 3 suggest that Roberson and Davidoff's account of CP in delayed discrimination tasks is incomplete. Experiment 3 presented evidence against their account of verbal interference on name retention. The Stroop tasks showed that CP is not due to a direct matching of labels. Future studies could try to determine which type of tasks and which aspects of task designs allow or even encourage a verbal strategy. Tasks which are thought to be highly perceptually loaded, the visual search task for example, could be combined with Stroop interference to test whether access to verbal codes is necessary for CP. If tasks are found not to be susceptible to Stroop interference, but still show evidence of CP, stronger claims about the perceptual basis of CP could be made. Furthermore, psychophysical techniques such as measuring discrimination thresholds across (linguistic) colour boundaries could help determine whether colour CP has an actual perceptual basis.

Is CP a memory effect?
Memory for colours of objects can be influenced by typical information, prior knowledge or post-event information (Belli, 1988). This suggests that our memory representations are not perfect matches of the original stimulus. It is possible that CP arises because of, and during, memory retention in the inter-stimulus interval. A phenomenon reported in the memory literature is the shift towards prototype (Crawford & Huttenlocher, 2000; Crawford, Huttenlocher & Engebretson, 2000; Huttenlocher & Hedges, 1992; Huttenlocher, Vedges & Vevea, 2000). Huttenlocher and colleagues have found that
memory representations of visual stimuli are biased towards the category prototype. CP and the shift toward prototype have clear parallels. Both phenomena describe non-uniform discrimination functions of uniform stimulus dimensions. The question addressed in Chapter 4 was whether the two phenomena are based on the same mechanisms and, specifically, whether CP could be a memory effect due to a shift towards prototype.

A shift towards the blue and green prototypes in the experiments presented in Chapters 2 and 3 would predict CP. If, during memory retention, the stimulus representation shifted towards the prototype, the distance between target and test stimulus would increase, thus facilitating cross-category discrimination. However, such a shift also has implications for within category discriminations. Stimuli in 2AFC and same-different tasks are sets of stimulus pairs and the stimuli can be presented in different orders. Depending on the order of presentation, the distance between the two stimuli should either decrease or increase. Thus, if a shift towards prototype occurs, within category discrimination accuracy should depend on order of presentation. This was tested by re-analysing the 2AFC, the same-different and the Stroop tasks.

No within category order effect was found in the 2AFC task, but significant order effects were found in the same-different task. However, the order effect was in the opposite direction from that predicted. Accuracy was better when the boundary colour was the target. This is not consistent with a shift towards the blue and green prototypes. A re-analysis of the Stroop tasks showed evidence for order effects in all four tasks. Interestingly, they were dependent on Stroop interference in Experiments 5 and 7 (when Stroop interference was presented at target presentation), but not in Experiments 6 and 8 (Stroop interference at test).

These results support the idea that the target is labelled and that the shift occurs towards the prototype of the corresponding category. When interference was presented with the target, the effects were dependent on type of Stroop interference. Strong order effects were found in the congruent conditions but the effect was reversed in the incongruent conditions. What this suggests is that the Stroop word ‘primes’ the shift and, more
importantly, the direction of the shift. The results are consistent with the shift in the incongruent conditions being towards the other prototype (e.g. if a blue target was labelled *green*, the order effect is consistent with the shift being more towards the green category). When Stroop interference was presented at test, the shift was towards the pre-existing blue and green prototypes in all Stroop conditions which suggest that the target was labelled easily and the interference at response had no effect.

However, it is also important to note that in the short ISI conditions (Experiments 7 and 8) strong order effects were found in all Stroop conditions. In these experiments especially, CP was entirely due to very poor boundary-first discrimination accuracy. This poses problems for any accounts of CP since it suggests that cross-category discrimination is not inherently better than within category discrimination.

In Chapter 4 the possibility was raised that the better non-boundary target discrimination accuracy may have been due to the presentation imbalance of boundary and non-boundary stimuli. The boundary stimuli were part of the within, as well as the cross-category pairs. Therefore, when the boundary colour was the target, it was not predictive of type of the test stimuli. When a non-boundary stimulus was presented, there was no such uncertainty. The results of the re-analysis of Stroop experiments support this point. However, there is another possible explanation. If the shift responsible for the order effect is initiated by target name generation, the higher cross-subject naming agreement for non-boundary compared to boundary colours (see Table 2.1.) should strengthen the order effect. On target presentation, non-boundary colours are likely to be named more easily than boundary colours. Thus, on non-boundary trials, the shift may have been initiated quicker and unequivocally. Although this idea is not necessarily consistent with Huttenlocher et al.’s model since stimulus labelling is not a factor in their model, it does suggest that discrimination accuracy of non-boundary stimuli may have been higher due to less uncertainty about category membership.

The fact that no order effect was evident in the 2AFC task (Experiment 1) and an order effect in the opposite direction was found in the same-different task (Experiment 2) makes the results and interpretation of the findings less straightforward. The order effect
found in the same-different task is consistent with a shift in the other direction, i.e. towards the blue-green category boundary. Huttenlocher et al.'s (2000) model is based on temporary categories invoked for the purpose of the experiment. In these temporary categories, the prototype is defined as the average of all stimuli, i.e. the central value. If this definition is applied to the stimulus set used in the same-different task, the results are consistent with a shift towards the (temporary) prototype. However, a shift in this direction is not consistent with CP since it would reduce cross-category distances.

Taken together, the findings of Chapter 4 suggest that different tasks may invoke different processes and hence lead to different shifts. As previously discussed, the colour words used with Stroop interference may reinforce the shift towards the pre-existing prototypes. Thus, using verbal codes may cause a shift to the pre-existing prototype; using visual or perceptual codes may cause a shift to the temporary prototypes. Note however, that this is not consistent with the suggestion made that a blanket verbal strategy was used on the same-different task in Experiment 2.

The relationship between CP and a shift towards prototype account of category effects needs testing by systematically manipulating a number of variables. Huttenlocher et al. (2000) argued that stimulus uncertainty reduced the shift. The uncertainty can either be due to initial uncertainty or the trace may weaken due to the memory demands of the task. However, in the experiments presented here these two factors were mixed. The same-different and 2AFC tasks had long target presentations, which should have reduced uncertainty, and long ISIs, which should have increased stimulus uncertainty. These uncertainty factors may add up and hence, in the above example, cancel each other out. A further factor reducing uncertainty in the 2AFC task may be the presence of both target and test stimuli on test presentation. This may explain why order effects were found in the same-different task, but not in the 2AFC task. However, according to Huttenlocher et al.'s model visual interference should also increase uncertainty, but no selective effects of type of interference were found in either Experiment 1 or 2.

In the Stroop tasks (Experiments 5 and 6) there were short stimulus presentations and long ISIs. This should have increased uncertainty, causing strong order effects compared
to Experiments 7 and 8 where uncertainty should have been reduced by short ISI. However, order effects in Experiments 7 and 8 were stronger than in 5 and 6 suggesting that the mechanisms are different to those proposed by Huttenlocher et al. (2000).

Summary

The results of the analysis presented in Chapter 4 suggest that CP in delayed colour discrimination tasks are not memory effects. The results of the re-analysis of earlier experiments suggested that CP is not due to a shift towards prototype for two reasons. Firstly, CP was found in all experiments, order effects were not. Secondly, the model of a shift towards prototype as proposed by Huttenlocher et al., a shift towards the temporary prototypes, does not predict CP. However, it is possible that different processes and mechanisms are involved in within category and cross-category effects. The data presented by Davies, Özgen, Pilling, and Wiggett (2003) showed a shift towards prototype for colours when just within category discriminations were tested. Hence, Huttenlocher et al.’s model may apply to colour discriminations only when temporary and pre-existing prototypes correspond.
6.3. Cross-cultural investigations of CP

The influence of language on colour discrimination and specifically on colour categorical perception was tested cross-culturally in Chapter 5. This was motivated by recent findings suggesting that colour language can affect colour perception (Roberson, Davies & Davidoff, 2000). The relevance of a cross-cultural comparison of colour cognition to the questions addressed in this thesis is that it has implications for CP. If colour perception tasks were done primarily by coding colours verbally, there should be clear differences in the performance of speakers of different languages. This would be consistent with a language account of CP, but it would only tell us something relatively trivial about the relationship between language and colour cognition. However, if CP is a perceptual effect due to warping of similarity space across category boundaries, speakers of different languages should not only have different verbal, but also different perceptual codes. Hence, the representation of colours should differ. This second point is based on a perceptual learning account of CP (Goldstone, 1994). As Özgen and Davies (2002) have shown, CP is found across acquired colour boundaries (for example light green - dark green). Thus, speakers of different languages should show language specific CP: CP should be evident only across those boundaries encoded in a given language. If, however, colour perception is independent of language and perceptual colour categories are universal, the performance of speakers of different languages on colour tasks should not differ.

Two types of tasks were used in the cross-cultural comparison of English and Owambo speakers, a visual search and a triads task. The visual search task is perceptually the purer task of the two, and hence a verbal strategy is less likely. Differences in search times between speakers of different languages are more likely to be due to perceptual effects than differences found on triads tasks. Categorical separation of target and distractor was found to facilitate search in Experiment 4. If search is faster if the target is in a separate category from the distractors compared to when all stimuli are taken from the same category, this should only be true for speakers of languages that make the categorical distinction. Pilling (2001) found evidence for language affecting search times. The predictions for Experiment 9 were that English speakers should have been advantaged in
some conditions as targets were categorically separated from distractors for English but not for Kwanyama. Furthermore, adding further distractors that were from the same English category should have been more detrimental to English than to Kwanyama speakers. The results did not support the first prediction. However, it was noted that it is difficult to assess absolute differences in cross-cultural comparisons as performance may also be influenced by uncontrolled, extraneous variables. The second prediction was supported; English search was slowed more than Kwanyama search by adding further distractors. Thus, consistent with Hunt and Agnoli’s (1991) account of linguistic relativity, the effects found can be assessed in terms of differential costs for speakers of different languages.

The findings do not support the idea that cross-category distances were stretched. However, the results do, to a certain extent, support the idea of within category equivalence. Adding further distractors slowed English search times more than Kwanyama search times. The new distractors were from a different English colour category than the existing distractors. For the Kwanyama, however, all distractors were from the same category. This could suggest that within category equivalence meant it was relatively easy for Kwanyama speakers to identify both stimuli as distractors.

It is important to note that the visual search task used in Experiment 9 was very different to the computerised task used in Experiment 4. The visual search had to be adapted to be used as a paper and pencil task in the field. The task was to search an array of colours for a number of targets; however, in the computerised visual search task, there was only one target and the set size was varied. Although both tasks were thought to have relatively high perceptual loads, it seems likely that the pencil and paper version was more open to verbal strategies. Search times were relatively slow which suggests that the task may have been done, at least in part, by using verbal labels. However, use of verbal codes should also have advantaged English speakers in some conditions. It is possible that English speakers used a verbal strategy and Kwanyama speakers (because there were less categorical separations between the stimuli and thus naming was less useful), used a predominantly perceptual strategy. Also, as mentioned earlier it is possible that the longer search times were partly the result of multiple manual responses required on the task.
As categorical separations did not facilitate search, no evidence of CP was found in Experiment 9. The effects of pop-out (Bauer, Jolicoer & Cowan, 1996) and categorical separations (Pilling, 2001) may differ when applied to one-target search compared to an array containing several targets. The search tasks required different search strategies. The search for multiple targets (Experiment 9) had to be done serially; each target had to be marked. When search is for a single target (as in Experiment 4) search is only done serially when the target and the distractors are similar to each other (Treisman & Gelade, 1980) and when they are linearly non-separable (D'Zmura, 1991). Thus, the tasks may have been too dissimilar to allow for a direct comparison.

The second task used in the cross-cultural comparison was a triads task (Experiment 10). Triads tasks require subjects to make similarity judgements between pairs of colour stimuli and chose the colour that is most different. When all three colours are taken from the same linguistic category (e.g. red), the task cannot simply be done by labelling and choosing the colour with a different name, but must be done perceptually. Hence, triads that were within category for both languages enabled a test of whether, independent of the influence of different colour categorisations, the perceived similarity of colours was universal. Furthermore, the use of triads that were cross-category for English speakers, but within category for Balantu speakers allowed a test of whether distances across boundaries become stretched and/or within category stimuli more alike in line with the linguistic distinctions made.

Similarity judgements elicited using triads have been shown to vary depending on the language spoken (Davies et al. 1998; Kay & Kempton, 1984; Pilling, 2001). Consistent with this, English and Balantu triad choices were found to be most different on triads for which the nominal predictions conflicted. English speakers chose in line with English colour categorisation, Balantu speakers did not. However, triad choices differed significantly not only on cross-category but also on within category triads.

The triads were also designed to test what similarity judgements were based on when choices could not be made on the basis of the categorical identity of stimuli. Sahli...
(1976) argued that brightness discrimination may have been an evolutionary precursor of hue discrimination and there is some evidence to suggest that the colour categories of some languages are lightness rather than hue based (MacLaury, 1992; 1997). Furthermore, Boyles (2001) found Ndonga children to be more likely to group colours by lightness than English children. The stimuli in the triads used here varied on hue and lightness, hence similarity judgements could either be made by grouping the two colours sharing hue or the colours sharing lightness. The results showed that Balantu speakers made significantly more lightness choices than English speakers. This suggested that not only did English and Balantu colour language differ, but also their dimensional representations of colour. Furthermore, it was found that English speakers also made less hue choices when there was no category boundary within a triad. For English colour categories, category boundaries are hue boundaries (MacLaury, 1992). Hue differences may become more salient when they signify a category boundary. The work of Luria (1976) on colour naming and colour grouping is also relevant here. Luria found that a group of non-literate traditional farming people were much less likely to group colours by hue than people from the same village who had undergone some literacy training. Thus, it is possible that differences in levels of literacy affected the results of the triads task. Another point which was not considered in the studies presented here is that of perceptual diets affecting colour perception. Conditions in countries like Namibia are on average much brighter than in European countries and it is possible that over time this significantly affects the perceptual system (see Bornstein, 1973). Future research could compare speakers of different languages, who nonetheless have been exposed to similar perceptual diets; comparing for example Owambo speakers with South African English speakers.

Summary
The results of Chapter 5 do not offer a straightforward answer to the question of whether CP is only found for speakers of languages that make certain categorical distinctions. The cross-cultural comparison of English and Owambo speakers did suggest that colour cognition may vary cross-culturally. However, as with CP, these differences are unlikely to be direct effects of colour naming. The results of the triads task suggested that the lightness dimension may be more salient to Balantu than English speakers. This suggests
that the representation of colour differs along with differences in colour language. The results of the visual search task were inconclusive. It was pointed out that the pencil and paper version may be open to verbal strategies more than the computerised search task. This highlights the importance of finding tasks suitable for conducting in the field without compromising the reliability of the task. Future cross-cultural experiments should attempt a more straightforward test of CP than those presented in Experiments 9 and 10.

Both the triads and the visual search experiments could be run varying just the categorical identity of stimuli across languages. Thus, triads and search arrays could be designed to be cross-category for English speakers, but within category for speakers of another language. Ideally, there would also be a condition where this was reversed. The effects of varying both hue and lightness within a triad (Experiment 10) should be looked at separately. The finding that Balantu speakers made more lightness choices on the triads suggests that Burns and Shepp’s (1988) account of the integrality of stimulus dimensions for colour needs testing cross-culturally.

6.4. Colour CP?

According to Harnad (1987) categorical perception is a fundamental characteristic of cognition. CP is defined by a warping of similarity space: differences between instances within categories appear smaller and cross-category differences appear stretched. The earliest examples of CP were in speech perception (Liberman, Harris, Hoffman & Griffith, 1957), but interest was soon extended to other areas of perception, including colour perception. The findings presented in the current thesis suggest that there are some potential problems with the approach used in colour CP experiments. The tasks used to study CP in colour perception have typically been designed to include a memory component (Bornstein & Korda, 1984; Boynton, Fargo, Olson & Smallman, 1989; Uchikawa & Shinoda, 1996) and therefore it is possible that the effects found are not purely perceptual (see Roberson & Davidoff, 2000). The results of the experiments presented here suggest that CP is due to stimulus labelling in tasks such as delayed colour discrimination. Evidence for more perceptual CP was found in a visual search task. Hence, careful account of the task design needs to be taken before conclusions about the
perceptual nature of category effects in colour tasks can be made. Furthermore, as highlighted by the re-analysis of experiments in Chapter 4, cross-category discrimination is not inherently superior to within category discrimination. This finding poses problems for perceptual and language accounts of CP.

6.5. Conclusion

This thesis addressed the question of whether categorical perception is a language or a memory effect rather than a perceptual phenomenon. The question was investigated in the laboratory in a series of dual-task and interference experiments and cross-culturally in a comparison of English and Owambo speakers. The results of the first two empirical chapters support a language account of CP in which target name generation is necessary for the advantage of cross-category over within category discrimination. However, the findings do not support a direct language account of CP in which the advantage is due to a comparison of labels as proposed by Roberson and Davidoff (2000). The process leading to CP is likely to be more complex than that. An alternative explanation of CP - a shift towards prototype - was considered in Chapter 4; however the findings suggest that CP is not entirely due to the same mechanisms a shift towards prototype (Huttenlocher, Hedges & Vevea, 2000). CP in colour discrimination tasks is likely to be due to the use of verbal and categorical codes. The category code is defined here as a mental representation of the colour which contains some perceptual information. A model is proposed in which target name generation reinforces this category code which is then used as part of the decision process at the response stage. The category code facilitates cross-category discrimination, but not within category discrimination. The cross-cultural comparison of English and Owambo speakers, suggests that, along with colour naming, there may also be some variation in how colours are internally represented across languages.

The work presented in this thesis has contributed to the understanding of where and how categorical colour perception arises within the information processing system. The findings suggest that name codes are important for CP, however the results also
emphasise the importance of task design as different task may invoke different processing mechanisms. The results of the cross-cultural comparison suggest that the effects of language on colour cognition are not direct effects of naming and language may lead to differences in the representation of colour. Taken together, the results support the idea that colour cognition is not independent of colour language.
References


Cambridge: Cambridge University Press.


Bulletin and Review, 7,* 3, 403-423.


*Cognition, 85,* 113-143.

Özgen, E, & Davies, I.R.L. (2002). Acquisition of categorical color perception: a
perceptual learning approach to the linguistic relativity hypothesis. *Journal of
Experimental Psychology: General, 131,* 477-493.

Carterette & M.P. Friedman (Eds.) *Handbook of Perception, VIII Perceptual Coding,*

Oxford University Press.

human subjects: the role of task difficulty. In: Ramscar, M., Hahn, U.,
Cambouropoulos, W. & Pain, H. (Eds.) *Proceedings of SimCat 1997: Interdisciplinary
Workshop on Similarity and Categorization.* Department of Artificial Intelligence,
Edinburgh University: 189-195.


Appendix 1

Perceptually uniform colour spaces

In the CIE (Commission Internationale de l'Eclairage) colour specification system, each colour is defined by a set of three stimuli (X, Y, and Z) which are in turn defined by a linear combination of their red, green and blue components. The problem with this specification system is that a small change in any of these variables does not necessarily produce the same change in the perceived colour at any point or in any direction in colour space. Equally noticeable colour differences vary in size depending on the region of colour space. In the blue region very small changes can be detected, in the green region, however, changes in saturation are harder to detect (Malacara, 2002). In order to be able to carefully control colour stimuli and the distances between colours, a perceptually uniform colour space is needed. This is done by transforming or modifying CIE (X,Y,Z) space. The objective of such transformations is to have a colour space in which the minimum perceived colour changes are equal in any direction so that one step at any given point is equal to one step at any other point in the system. Several colour systems have been designed to obtain an approximately uniform colour space. A brief outline of the CIE L*u*v* and the Munsell system is given below. Both systems are widely used in colour research and both are based on psychophysical judgements of colour differences. For detailed descriptions and conversion formulas see Hunt (1987) and Malacara (2002).

The Munsell System

The Munsell system was devised by Albert Munsell in 1905 and was designed to give standardised, printable colours.

The Munsell system represents colour as an irregular cylinder. The axis represents the perceived lightness (value) of colour and is segmented into 10 steps ranging from zero to nine. Thus, at the lowest level the axis represents black, at the highest white.
Chroma is defined by the saturation of a colour and is defined by the horizontal plane. Chroma increases (from 0 to 8) in perpendicular direction to the axis towards the edge of the cylinder. Hue is represented by the horizontal angle in the cylinder. The Munsell circle is divided into ten angular sectors represented by a hue letter (R, RY, YG, G, BG, B, PB, P or RP) which are each again split into ten sub sectors.

The 3D colour solid is made up of the representation of all colours as a series of radial planes with constant hue. Hence, all colours on a given radial plane have the same hue, but chroma increases towards the edge of the solid and value increases vertically. Note that not all hue planes have the same shape. This is because not all colours have the same range of saturation levels. For example, the highest level of saturation for blue and purple colours is found at low values, for yellow, however, it is at high values. These differences mean that colour space is represented as an irregular solid.

Each Munsell colour is represented by a standard code consisting of first the hue letter, followed by a number denoting value (lightness) and another number representing chroma (saturation). Each hue letter has 10 units, however the Munsell book of colours shows colour samples at 2.5 units intervals. Hue. Value ranges from 0 to 9, how many saturation levels there are depends on the given hue. An example of a Munsell notation is 5BG 5/10.

The perceptual uniformity of the Munsell system means that adjacent pairs are perceptually equidistant from each other, hence the differences between, for example, 5BG 5/10 and 7.5BG 5/10 is the same as the differences between 5R5/14 and 7.5R5/14. The system was developed by getting observers to compare differences between pairs of colours against a grey scale. The advantage of such a system is that it describes colours in tangible dimensions. However, there are also some problems with the Munsell system. Firstly, there is no simple way of equating colour differences across the different axis. One horizontal hue step does not equal one vertical value step. Furthermore, as colour space was standardised with respect to adjacent pairs only, it is likely to be less precise in matching larger colour differences.
The CIE (1976) L*u*v* system

Any colour uniquely defined by its hue, saturation and luminance, can be specified by the tristimulus values X, Y, and Z (or by the chromaticity coordinates x and y, plus the tristimulus value Y). However, as outlined above, it has been shown that for standard observers, the Y x y chromaticity space is perceptually uneven. The L*u*v* system is a transformation of the CIE Y x y chromaticity space to a CIE system which is perceptually uniform. Any colour difference, whether from reflective surfaces or lights, can be described converting measurements into L*u*v*.

However, the dimensions of u* and v* do not map easily onto the perceived colour dimensions and lines along a constant dimension do not always equal equivalent hues. The Bezold-Brücke effect (Malacara, 2002) shows that for impure colours, changes in luminance also involve a shift in perceived hue. Hence, in a u*v* diagram, colours with certain hue and chroma are not uniquely represented with the same values of u*v* for all possible values of lightness.

The data for both the L*u*v* and the Munsell system were collected on ‘standard observers’. However, as the standard observers tend to be American or European, these systems do not take possible cross-cultural variations in the representation of colour into account.
Appendix 2

List of word pairs used in the verbal interference tasks in Experiments 1-3.

<table>
<thead>
<tr>
<th>Rhyming</th>
<th>Non-Rhyming</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aisle</td>
<td>Pile</td>
</tr>
<tr>
<td>Answer</td>
<td>Cancer</td>
</tr>
<tr>
<td>Break</td>
<td>Rake</td>
</tr>
<tr>
<td>Brooch</td>
<td>Coach</td>
</tr>
<tr>
<td>Browse</td>
<td>Cows</td>
</tr>
<tr>
<td>Bull</td>
<td>Wool</td>
</tr>
<tr>
<td>Castle</td>
<td>Hassle</td>
</tr>
<tr>
<td>Chord</td>
<td>Broad</td>
</tr>
<tr>
<td>Clear</td>
<td>Beer</td>
</tr>
<tr>
<td>Craft</td>
<td>Draught</td>
</tr>
<tr>
<td>Dough</td>
<td>Low</td>
</tr>
<tr>
<td>Farm</td>
<td>Calm</td>
</tr>
<tr>
<td>Flout</td>
<td>Drought</td>
</tr>
<tr>
<td>Flower</td>
<td>Sour</td>
</tr>
<tr>
<td>Home</td>
<td>Comb</td>
</tr>
<tr>
<td>Horse</td>
<td>Course</td>
</tr>
<tr>
<td>Isle</td>
<td>Mile</td>
</tr>
<tr>
<td>Limb</td>
<td>Brim</td>
</tr>
<tr>
<td>Lord</td>
<td>Reward</td>
</tr>
<tr>
<td>Lute</td>
<td>Hoot</td>
</tr>
<tr>
<td>Mute</td>
<td>Fruit</td>
</tr>
<tr>
<td>Nigh</td>
<td>Why</td>
</tr>
<tr>
<td>Page</td>
<td>Gauge</td>
</tr>
<tr>
<td>Plough</td>
<td>Now</td>
</tr>
<tr>
<td>Proof</td>
<td>Sleuth</td>
</tr>
<tr>
<td>Quay</td>
<td>See</td>
</tr>
<tr>
<td>Quest</td>
<td>Zest</td>
</tr>
<tr>
<td>Say</td>
<td>Neigh</td>
</tr>
<tr>
<td>Shoot</td>
<td>Chute</td>
</tr>
<tr>
<td>Tall</td>
<td>Crawl</td>
</tr>
<tr>
<td>Theme</td>
<td>Seem</td>
</tr>
<tr>
<td>Torque</td>
<td>Fork</td>
</tr>
<tr>
<td>Tough</td>
<td>Stuff</td>
</tr>
<tr>
<td>Trial</td>
<td>Pile</td>
</tr>
<tr>
<td>Wood</td>
<td>Could</td>
</tr>
<tr>
<td>Yacht</td>
<td>Plot</td>
</tr>
</tbody>
</table>
Munsell codes, CIE Yxy and L*u*v* co-ordinates of colour stimuli used in Experiments 1-8.

<table>
<thead>
<tr>
<th>Hue</th>
<th>Lightness</th>
<th>Chroma</th>
<th>Y</th>
<th>x</th>
<th>y</th>
<th>L*</th>
<th>u*</th>
<th>v*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.44 B</td>
<td>5.40</td>
<td>7.40</td>
<td>23.5</td>
<td>.21</td>
<td>.28</td>
<td>55.6</td>
<td>-44.3</td>
<td>-25.0</td>
</tr>
<tr>
<td>8.76 BG</td>
<td>5.40</td>
<td>7.40</td>
<td>23.5</td>
<td>.21</td>
<td>.30</td>
<td>55.6</td>
<td>-45.7</td>
<td>-16.2</td>
</tr>
<tr>
<td>6.06 BG</td>
<td>5.40</td>
<td>7.40</td>
<td>23.5</td>
<td>.22</td>
<td>.32</td>
<td>55.6</td>
<td>-46.2</td>
<td>-7.3</td>
</tr>
<tr>
<td>3.46 BG</td>
<td>5.40</td>
<td>7.40</td>
<td>23.5</td>
<td>.23</td>
<td>.34</td>
<td>55.6</td>
<td>-45.7</td>
<td>1.7</td>
</tr>
<tr>
<td>1.44 B</td>
<td>5.80</td>
<td>7.40</td>
<td>27.8</td>
<td>.21</td>
<td>.29</td>
<td>59.7</td>
<td>-45.5</td>
<td>-24.8</td>
</tr>
<tr>
<td>8.76 BG</td>
<td>5.80</td>
<td>7.40</td>
<td>27.8</td>
<td>.22</td>
<td>.30</td>
<td>59.7</td>
<td>-46.9</td>
<td>-15.7</td>
</tr>
<tr>
<td>6.06 BG</td>
<td>5.80</td>
<td>7.40</td>
<td>27.8</td>
<td>.22</td>
<td>.32</td>
<td>59.7</td>
<td>-47.3</td>
<td>-6.6</td>
</tr>
<tr>
<td>3.46 BG</td>
<td>5.80</td>
<td>7.40</td>
<td>27.8</td>
<td>.23</td>
<td>.34</td>
<td>59.7</td>
<td>-46.9</td>
<td>2.4</td>
</tr>
<tr>
<td>1.44 B</td>
<td>6.20</td>
<td>7.40</td>
<td>32.4</td>
<td>.22</td>
<td>.29</td>
<td>63.7</td>
<td>-46.4</td>
<td>-24.0</td>
</tr>
<tr>
<td>8.76 BG</td>
<td>6.20</td>
<td>7.40</td>
<td>32.4</td>
<td>.22</td>
<td>.31</td>
<td>63.7</td>
<td>-47.7</td>
<td>15.1</td>
</tr>
<tr>
<td>6.06 BG</td>
<td>6.20</td>
<td>7.40</td>
<td>32.4</td>
<td>.23</td>
<td>.32</td>
<td>63.7</td>
<td>-49.2</td>
<td>-6.0</td>
</tr>
<tr>
<td>3.46 BG</td>
<td>6.20</td>
<td>7.40</td>
<td>32.4</td>
<td>.24</td>
<td>.34</td>
<td>63.7</td>
<td>-47.7</td>
<td>3.0</td>
</tr>
<tr>
<td>1.44 B</td>
<td>6.60</td>
<td>7.40</td>
<td>37.5</td>
<td>.22</td>
<td>.29</td>
<td>67.6</td>
<td>-47.0</td>
<td>-23.3</td>
</tr>
<tr>
<td>8.76 BG</td>
<td>6.60</td>
<td>7.40</td>
<td>37.5</td>
<td>.23</td>
<td>.31</td>
<td>67.6</td>
<td>-48.3</td>
<td>-14.3</td>
</tr>
<tr>
<td>6.06 BG</td>
<td>6.60</td>
<td>7.40</td>
<td>37.5</td>
<td>.23</td>
<td>.33</td>
<td>67.6</td>
<td>-48.7</td>
<td>-5.2</td>
</tr>
<tr>
<td>3.46 BG</td>
<td>6.60</td>
<td>7.40</td>
<td>37.5</td>
<td>.24</td>
<td>.34</td>
<td>67.6</td>
<td>-48.2</td>
<td>3.7</td>
</tr>
<tr>
<td>Focal Blue</td>
<td>8.3</td>
<td>.14</td>
<td>34.6</td>
<td>-15.9</td>
<td>-132.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Focal Green</td>
<td>65.2</td>
<td>.29</td>
<td>8.46</td>
<td>-86.8</td>
<td>109.0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 4

Triad choice by individual cross-category or within category classification. Tables show the frequencies of choice of each tile for subjects for whom all three tiles were within category compared to those for whom they were cross-category triads (based on individual naming responses). Most frequent choices are presented in bold. The green triad set is not included as all participants labelled all stimuli green (or oshizizzi).

**Red/Pink 1**

<table>
<thead>
<tr>
<th></th>
<th>English</th>
<th></th>
<th>Balantu</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.5R 4/14</td>
<td>2.5R 5/14</td>
<td>8.75R 5/14</td>
<td>2.5R 4/14</td>
</tr>
<tr>
<td>cross</td>
<td>1</td>
<td>17</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>within</td>
<td>3</td>
<td></td>
<td></td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>English</th>
<th></th>
<th></th>
<th>Balantu</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.5R 4/14</td>
<td>2.5R 5/14</td>
<td>8.75R 4/14</td>
<td>2.5R 4/14</td>
<td>2.5R 5/14</td>
<td>8.75R 4/14</td>
</tr>
<tr>
<td>cross</td>
<td>2</td>
<td>15</td>
<td></td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>within</td>
<td>4</td>
<td></td>
<td></td>
<td>4</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>English</th>
<th></th>
<th></th>
<th>Balantu</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.5R 5/14</td>
<td>8.75R 4/14</td>
<td>8.75R 5/14</td>
<td>2.5R 5/14</td>
<td>8.75R 4/14</td>
<td>8.75R 5/14</td>
</tr>
<tr>
<td>cross</td>
<td>16</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>within</td>
<td>3</td>
<td></td>
<td></td>
<td>6</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>English</th>
<th></th>
<th></th>
<th>Balantu</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.5R 4/14</td>
<td>8.75R 4/14</td>
<td>8.75R 5/14</td>
<td>2.5R 4/14</td>
<td>8.75R 4/14</td>
<td>8.75R 5/14</td>
</tr>
<tr>
<td>cross</td>
<td>15</td>
<td>2</td>
<td></td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>within</td>
<td>4</td>
<td></td>
<td></td>
<td>7</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Triad</td>
<td>English</td>
<td></td>
<td></td>
<td></td>
<td>Balantu</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---------</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>2.5R</td>
<td>2.5R</td>
<td>7.5R</td>
<td>2.5R</td>
<td>2.5R</td>
<td>7.5R</td>
</tr>
<tr>
<td>cross</td>
<td>4/14</td>
<td>5/14</td>
<td>4/14</td>
<td>cross</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>within</td>
<td>1</td>
<td></td>
<td></td>
<td>within</td>
<td>18</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2.5R</td>
<td>7.5R</td>
<td>7.5R</td>
<td>2.5R</td>
<td>7.5R</td>
<td>7.5R</td>
</tr>
<tr>
<td>cross</td>
<td>17</td>
<td>3</td>
<td></td>
<td>Cross</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>within</td>
<td>1</td>
<td></td>
<td></td>
<td>within</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>2.5R</td>
<td>7.5R</td>
<td>7.5R</td>
<td>2.5R</td>
<td>7.5R</td>
<td>7.5R</td>
</tr>
<tr>
<td>cross</td>
<td>10</td>
<td>3</td>
<td>2</td>
<td>Cross</td>
<td>2</td>
<td>17</td>
</tr>
<tr>
<td>within</td>
<td>5</td>
<td>1</td>
<td></td>
<td>within</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.5R</td>
<td>2.5R</td>
<td>7.5R</td>
<td>2.5R</td>
<td>2.5R</td>
<td>7.5R</td>
</tr>
<tr>
<td>cross</td>
<td>20</td>
<td></td>
<td></td>
<td>Cross</td>
<td>19</td>
<td>1</td>
</tr>
<tr>
<td>within</td>
<td>1</td>
<td></td>
<td></td>
<td>within</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
### Blue/Green

<table>
<thead>
<tr>
<th>Triad 9</th>
<th>English</th>
<th>5BG</th>
<th>5BG</th>
<th>10BG</th>
<th>Balantu</th>
<th>5BG</th>
<th>5BG</th>
<th>10BG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cross</td>
<td>3</td>
<td>14</td>
<td></td>
<td>cross</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>within</td>
<td>4</td>
<td></td>
<td></td>
<td>within</td>
<td>13</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Triad 10</th>
<th>English</th>
<th>5BG</th>
<th>5BG</th>
<th>10BG</th>
<th>Balantu</th>
<th>5BG</th>
<th>5BG</th>
<th>10BG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cross</td>
<td>4</td>
<td>13</td>
<td></td>
<td>cross</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>within</td>
<td>4</td>
<td></td>
<td></td>
<td>within</td>
<td>2</td>
<td>9</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Triad 11</th>
<th>English</th>
<th>5BG</th>
<th>10BG</th>
<th>10BG</th>
<th>Balantu</th>
<th>5BG</th>
<th>10BG</th>
<th>10BG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cross</td>
<td>14</td>
<td>4</td>
<td></td>
<td>cross</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>within</td>
<td>3</td>
<td></td>
<td></td>
<td>within</td>
<td>10</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Triad 12</th>
<th>English</th>
<th>5BG</th>
<th>10BG</th>
<th>10BG</th>
<th>Balantu</th>
<th>5BG</th>
<th>10BG</th>
<th>10BG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cross</td>
<td>14</td>
<td>3</td>
<td></td>
<td>cross</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>within</td>
<td>3</td>
<td>1</td>
<td></td>
<td>within</td>
<td>9</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>English</td>
<td>Balantu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cross</td>
<td>2.5R 5/10</td>
<td>2.5R 5/10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>within</td>
<td>9 6/10</td>
<td>10 6/10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cross</td>
<td>7.5RP 5/10</td>
<td>7.5RP 5/10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>within</td>
<td>13 6/10</td>
<td>13 6/10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>English</th>
<th>Balantu</th>
</tr>
</thead>
<tbody>
<tr>
<td>cross</td>
<td>2.5R 5/10</td>
<td>2.5R 5/10</td>
</tr>
<tr>
<td>within</td>
<td>3 6/10</td>
<td>6 6/10</td>
</tr>
<tr>
<td>cross</td>
<td>7.5RP 5/10</td>
<td>7.5RP 5/10</td>
</tr>
<tr>
<td>within</td>
<td>12 6/10</td>
<td>12 6/10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>English</th>
<th>Balantu</th>
</tr>
</thead>
<tbody>
<tr>
<td>cross</td>
<td>2.5R 6/10</td>
<td>2.5R 6/10</td>
</tr>
<tr>
<td>within</td>
<td>4 6/10</td>
<td>2 6/10</td>
</tr>
<tr>
<td>cross</td>
<td>7.5RP 6/10</td>
<td>7.5RP 6/10</td>
</tr>
<tr>
<td>within</td>
<td>14 6/10</td>
<td>14 6/10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>English</th>
<th>Balantu</th>
</tr>
</thead>
<tbody>
<tr>
<td>cross</td>
<td>2.5R 6/10</td>
<td>2.5R 6/10</td>
</tr>
<tr>
<td>within</td>
<td>7 6/10</td>
<td>8 6/10</td>
</tr>
<tr>
<td>cross</td>
<td>7.5RP 6/10</td>
<td>7.5RP 6/10</td>
</tr>
<tr>
<td>within</td>
<td>14 6/10</td>
<td>14 6/10</td>
</tr>
<tr>
<td>Triad 21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>--------------</td>
<td>--------------</td>
</tr>
<tr>
<td></td>
<td>English</td>
<td>5R 4/14</td>
</tr>
<tr>
<td></td>
<td>cross</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>within</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>