Language and Culture in Perception:

A three-pronged investigation of phylogenetic, ontogenetic and cross-cultural evidence.

Thesis presented by
Julie Natanela Goldstein

For the degree of Doctor of Philosophy

Goldsmiths, University of London

I declare that the work presented in my thesis is my own.

Signed:

Dated: 18th Dec 2007

Acknowledgements

To Jules Davidoff for supporting my professional development and providing me with great research opportunities.

To Joel Fagot and his team at the animal lab for guidance in comparative work and hospitality during my visits at the CNRS in Marseille, France.

To Laura Shapiro for introducing me to field work in Namibia.

To Kemuu Jakurama for guiding me in the field and helping me negotiate my stay with the Himba in various villages.

To the Himba for their warm welcome and cheerful participation in my studies.

To my parents and brother for their care and support.

To Patti Owens for throwing me a lifeline when I needed it.

To Miguel Andrés-Clavera for many reasons too close to my heart to put into words.

Abstract

Brown and Lenneberg (1954) and Rosch Heider (1972) were among the first to conduct psychological investigations to test the Whorfian view that language affects thought. They both asked about colour categories. The debate has continued with some research supporting a relativist (Whorfian) account (Davidoff, Davies & Roberson, 1999; Borodistsky, 2001), and some supporting a universalist account (e.g., Kay & Regier, 2003; Spelke & Kinzler, 2007). The present thesis adds to the debate by taking three different approaches i.e., cross-cultural, ontogenetic and phylogenetic frames in which to carry out investigations of categorization of various perceptual continua.

Categorical Perception's hallmark is the effect of mental warping of space such as has been found for phonemes (Pisoni & Tash, 1974) and colour (Bornstein & Monroe, 1980; Bornstein & Korda, 1984). With respect to colours, those that cross a category boundary seem more distant than two otherwise equally spaced colours from the same category. Warping is tested using cognitive methods such as two-alternative-forced-choice and matching-to-sample. Evidence is considered for the continua under investigation i.e. colour and animal patterns.

Experiments 1 and 2 find evidence of categorical perception for human-primates and not for monkeys. Experiment 3 finds that Himba and English human adults categorize differently, particularly for colours crossing a category boundary, but also show broad similarity in solving the same matching-to-sample task as used with the monkeys (experiment 1) who showed clear differences with humans. Experiment 4 and 5 tested Himba and English toddlers and found categorical perception of colour mainly for toddlers that knew their colour terms despite prior findings (Franklin et al., 2005) indicative of universal colour categories. In experiment 6, Himba and English categorical perception of animal patterns was tested for the first time, and result indicate a cross-category advantage for participants who knew the animal pattern terms.

Therefore, a weak Whorfian view of linguistic relativity's role in obtaining categorical perception effects is presented. Although there is some evidence of an inherent human way of grouping drawn from results of experiment 1 and 3, results in all experiments (1,2,3,4,5,and 6) show that linguistic labels and categorical perception effects go hand-in-hand; categorization effects are not found when linguistic terms are not acquired at test and have not had a chance to affect cognition. This was true for all populations under observation in this set of studies, providing further support for effects of language and culture in perception.

Table of contents

1. **General Introduction** 15
  - 1.1 Grouping behaviour 16
  - 1.2 Linguistic relativity: The philosophy behind it 18
    - 1.2.1 Anthropological pursuit 18
    - 1.2.2 Whorf's Hypothesis 19
    - 1.2.3 The Paradox of the Heap 21
  - 1.3 Cognitive methods in colour research 23
    - 1.3.1 Codability 23
    - 1.3.2 Communicability 25
    - 1.3.3 Categorical Perception 26
  - 1.4 Whorfian evidence 27
    - 1.4.1 Supporting evidence 28
    - 1.4.2 Linguistic relativity in the spatial domain 29
    - 1.4.3 Linguistic relativity in the temporal domain 31
    - 1.4.4 Cognitive modelling and robot-behaviour 33
    - 1.4.5 Learned colour categories 35
  - 1.5 Evidence of an innate view 36
    - 1.5.1 The eleven basic colours 36
    - 1.5.2 Prototype and cognitive dominance over language 37
    - 1.5.3 Neurophysiological determinism 39
    - 1.5.4 Biologically determined colour categories 41
  - 1.6 Current findings related to universal and linguistic influences on colour categorization. 43
    - 1.6.1 Relative and linguistically influenced colour categories 43
      - 1.6.1.1 Relativity's developmental approach 45
    - 1.6.2 Universal and innate colour categories 47
      - 1.6.2.1 Universality's developmental approach 49

2. **Current study** 51
- 2.1 Motivation for current study 51
- 2.2 The current set of experiments 52
  - 2.2.1 Methodological considerations 52
  - 2.2.2 The participant populations of this study 55

3. **Experiment 1 and 2** - A cross-species comparison of colour categorization. 63
- 3.1 General Introduction of experiments 1 and 2 63
- 3.2 Experiment 1- Colour categorization in matching-to-sample 66
  - 3.2.1 Method 66
    - 3.2.1.1 Participants 66
    - 3.2.1.2 Apparatus 66
    - 3.2.1.3 Stimuli 67
    - 3.2.1.4 Procedure 69
  - 3.2.2 Results 71
  - 3.2.3 Discussion 78
- 3.3 Experiment 2- Colour discrimination and categorization in 2 alternative-forced-choices. 79
- 3.3.1 Experiment 2a- Discrimination data 80
  - 3.3.1.1 Method 80
    - i. Participants 80
    - ii. Apparatus and Stimuli 80
    - iii. Procedure 80
  - 3.3.1.2 Results 81
  - 3.3.1.3 Discussion 83
- 3.3.2 Experiment 2b- Colour categorization in 2 alternative-forced-choices. 83
  - 3.3.2.1 Method 86
    - i. Participants 86
    - ii. Stimuli 86
    - iii. Procedure 86

3.3.2.2 Results 87
3.3.2.3 Discussion 89
3.4 General discussion of experiments 2 and 3 90

4. **Experiment 3**- A cross-cultural comparison of colour categorization in maching-to-sample. 92
4.1 Introduction 92
4.2 Method 93
4.2.1 Participants 93
4.2.2 Apparatus and Stimuli 94
4.2.3 Procedure 95
4.3 Results 96
4.6.1 Naming 96
4.6.2 Test 100
4.4 Discussion 105

5. **Experiment 4 and 5**- A developmental study of effects of colour term acquisition and a cross-cultural comparison of colour categorization in toddlers. 107
5.1 General introduction of experiments 5 and 6 107
5.2 Experiment 5- A developmental study of effects of colour term acquisition. 108
5.2.1 Method 108
5.2.1.1 Participants 108
5.2.1.2 Stimuli 108
5.2.1.3 Procedure 112
5.2.2 Results 114
5.2.3 Discussion 128

5.3 Experiment 5- A cross-cultural comparison of colour categorization in toddlers. 129

5.3.1 Method 129

5.3.1.1 Participants 129

5.3.1.2 Stimuli, design and procedure 130

5.3.2 Results 130

5.3.3 Discussion 131

5.4 General discussion of experiments 5 and 6 132

6. **Experiment 6**- A cross-cultural comparison of animal pattern categorization in two-alternative-choices 136

6.1 Introduction 136

6.2 Production of the animal continua 138

6.3 Two-alternative-forced-choice test 143

6.3.1 Method 143

6.3.1.1 Participants 143

6.3.1.2 Stimuli 143

6.3.1.3 Procedure 145

6.3.2 Results 146

6.4 Discussion 152

7. **General discussion**- From grouping behaviour in perception to conceptual grouping in categorization 156

7.1 Cross-species grouping differences 156

7.2 Some cross-cultural linguistic influences in colour grouping 158

7.3 Conceptual grouping influenced by language in developmental work on colour. 159

7.4 Cross-cultural findings of conceptual grouping of animal patterns influenced by language 161

7.5 Implications for universality 162

7.6 Implications for relativity 165

7.6.1 A three-pronged investigation's implications for relativity versus conclusions of previous two-pronged investigations. 167

7.7 Future research 169

8. **Conclusion** 171

Index of Figures and Tables

Figure 1.1 Berlin & Kay's (1969) colour stages of evolution. 36

Figure 1.2 Distribution of Himba naming and choices of best exemplars for the 160 chip saturated array (for 31 observers) compared to those English speakers for the same array. 58

Figure 3.1 Munsell stimuli plotted in CIE L*u*v* space. 69

Figure 3.2 Responses choices (% of "green" responses) and standard errors for human and baboon participants. 71

Figure 3.3 Observed latency data and standard errors for human and baboon participants. 73

Figure 3.4 Discrimination thresholds obtained in humans (dash lines) and baboons (continuous lines) for 8 equally spaced Munsell colours in the green-blue range. 82

Figure 3.5 Percentage of correct responses for the two species, and for cross- and within-category pairs. 88

Figure 4.1 Equidistant (Saturation 5, Brightness 6) Munsell colours 5Y, 7.5Y, 10Y, 2.5GY, 5GY, 7.5GY, 10GY, 2.5G, 5G, 7.5G, 10G, 2.5BG represented in CIE L*u*v* colour space. 95

Figure 4.2 Mean green responses (+/- 1 SE) for Himba and English participants for Similarity Matching trials. 102

Figure 4.3 Mean dumbu responses (+/- 1 SE) for Himba and English participants on similarity matching trials. 104

Figure 5.1 Munsell codes, categorical status, and Munsell distances of the stimuli of the category pairs used in Experiment 5 as in Franklin et al. (2005). 111

Figure 5.2 Munsell colours of figure 1 represented in CIE Lu*v* colour space. Same values used in Franklin et al. (2005). 111

Figure 5.3 Mean accuracy (±1 SE) for within- and between-category pairs for the blue-purple and blue-green sets. 115

Figure 5.4 Mean accuracy (±1 SE) of Name Boundary for the Blue-purple set on Within- and Between-category pairs. 119

Figure 5.5 Mean accuracy (±1 SE) of Name Boundary for the Blue-green set on Within- and Between-category pairs. 120

Figure 5.6 Mean numbers correct on naming and comprehension tasks for each age band. Bars represent ±1 SE. 122

Figure 5.7A and 5.7B Mean accuracy (±1 SE) of Linguistic Terminology for Within- and Between-category pairs for blue-purple and blue-green sets 127

Figure 6.1 A-E Endpoint pictures of continua used for 2AFC test. 139

Figure 6.2 A-E Naming agreement from 30 Himba participants on five continua for animal patterns. Each continuum consisted of 21 pictures (19 equally spaced morphed images from the two endpoints). 142

Figure 6.3 A-E Subsets of the 5 continua containing the pictures used in the 2AFC test. 144

Figure 6.4 Mean accuracy scores (max = 20) and standard errors for English and Himba participants on continuum 1 for Between-Category and Within-Category trials in the 2AFC test. Comparisons giving p < .001 are indicated. 147

Figure 6.5 Mean accuracy scores (max = 20) and standard errors for English and Himba participants on continuum 2 for Between-Category and Within-Category trials in the 2AFC test. Comparisons giving p < .1 are indicated. 148

Figure 6.6 Mean accuracy scores (max = 20) and standard errors for English and Himba participants on continuum 3 for Between-Category and Within-Category trials in the 2AFC test. 149

Figure 6.7 Mean accuracy scores (max = 20) and standard errors for English and Himba participants on continuum 4 for Between-Category and Within-Category trials in the 2AFC test. Comparisons giving p < .05 are indicated. 150

Figure 6.8 Mean accuracy scores (max = 20) and standard errors for English and Himba participants on continuum 5 for Between-Category and Within-Category trials in the 2AFC test. 151

Table 1.1	Animal patterns used in Goldstein & Davidoff (in press) that also correspond to description found in Eckl (2000).	59
Table 3.1	Best-fitting models and parameter values for individual human and baboon participants.	77
Table 3.2	Munsell-metric-distance for experiment 1,2, and 3 between the target and foils for within- and between-category pairs of Sandell, Gross & Bornstein's (1979) study.	85
Table 4.1A	Mean % Colour Term Used for English in the naming task of green-blue colour range.	97
Table 4.1B	Mean % Colour Term Used for Himba in the naming task of green-blue colour range.	98
Table 4.2A	Mean % Colour Term Used for Himba in the naming task of dumbu-buru colour range.	99
Table 4.2B	Mean % Colour Term Used for English in the naming task of dumbu-buru colour range.	99
Table 4.3	Mean green responses (+/- 1 SE) for Himba and English participants for Similarity Matching trials.	102
Table 4.4	Mean dumbu responses (+/- 1 SE) for Himba and English participants on similarity matching trials.	104
Table 5.1	Designations of the 22 Color Aid sweaters in the CIE Lu*v* metric under Illuminant C as measured by Roberson et al (2005)	109
Table 5.2	Mean accuracy (±1 SE) for within- and between-category pairs for the blue-purple and blue-green sets.	114
Table 5.3	Frequencies (percentages) of the colour terms offered for the stimuli of each set	117
Table 5.4	Mean accuracy (±1 SE) of Name Boundary for the Blue-purple set on Within- and Between-category pairs.	119
Table 5.5	Mean accuracy (±1 SE) of Name Boundary for the Blue-green set on Within- and Between-category pairs.	119

Table 5.6 Percentages correctly named focals in the naming task for each age band 121

Table 5.7 Percentages correctly identified focals in the comprehension task for each age band using Franklin et al's criteria. 122

Table 5.8 Number of children who were allocated to Franklin et al.'s (2005) Linguistic Boundary, Reverse Linguistic Boundary and No Boundary groups according to naming and knew either blue and purple or blue and green according to prior naming and comprehension tests. 125

Table 5.9 Number of children that qualified for full, partial or no colour term knowledge. 126

Table 5.10A Mean accuracy (±1 SE) of Linguistic Terminology for Within- and Between-category pairs for the blue-purple set 126

Table 5.10B Mean accuracy (±1 SE) of Linguistic Terminology for Within- and Between-category pairs for the blue-green set 126

1. General Introduction

How groups are formed is an essential question in this thesis. Some categories, for example numbers, can be defined a priori and without any reference to experience. Most categories are not like numbers and so 'What defines a category' is a central question in their examination. Some researchers have argued that perceptual similarity is what drives membership to a category (e.g. Rosch & Mervis, 1975; Tversky, 1977; Murphy & Medin, 1985). However, similarity may just be a by-product rather than a cause of categorization; similar things may share the same properties but similarity may not explain *why* they are in the same category (Medin & Ross, 1996).

This thesis suggests that in some instances there is more to categorization than perceptual similarity. One of the first to suggest categories were relative was Whorf (1956), punctuating absolute relativity to language. In accordance, non-perceptual mechanisms such as labelling (Davidoff, 2001) have been found to be commonly used in identification and discrimination of category members (Pisoni, 1975). Discrimination differences have been more closely studied in the experiments that showed evidence of a sharp peak in discriminability of stimuli at category boundaries (e.g. Bornstein & Korda, 1984).

Nevertheless, the different takes on the universality and origins of colour categories have provided some debate and are outlined with particular reference to colour. In some instances colour categories are found to be universal (e.g. Rosch, 1975) and innate (e.g. Sandell, Gross & Bornstein, 1979) and the origins of colour categorization are explained biologically/neurophysiologically (Kay & McDaniel, 1978; Bornstein, 1987). In other instances nurtured colour categories are found, and the origins are argued to be of a non-biological or non-perceptual nature (Sapir, 1921; Whorf, 1956). This thesis re-evaluates the debate up to date to further understand how the study of grouping behaviours in illusion, colour and animal patterns fit and are relevant to explaining categorization's nature.

The experimental chapters are divided into 5 sections and cover phylogenetic, ontogenetic, and cross-cultural approaches to grouping behaviour in the context of visual illusion, colour and animal patterns. The first, third and fifth sections address these various contexts by looking at the cross-cultural comparison of various grouping behaviours for two different cultures that have resulted from these experiments. The second section covers comparative work with monkeys in two experiments looking at the emergence of colour categories. The fourth section addresses colour categories in ontogenetic work with the study of toddlers both within and across two different cultures. A series of experiments are reported that examined grouping behaviour of both perceptual and conceptual natures and in particular whether categories are nurtured by language. The concluding chapter examines the implications of these experiments for the current debate between universality and relativity and possible future directions for research.

1.1 Grouping behaviour

Perceptual grouping seems vital to the way we organize different parts of the visual world. Gestalt psychologists (e.g. Wertheimer, 1912) seem to have implemented this idea into their investigation of visual perception and what becomes figural (a group) in terms of the ground possibilities which are as good as the environment allows (Law of Pragnanz). Seeing a figure as more than the sum of its parts seems to be a very strong perceptual property, which has the effect of making things in the environment similar. Similarity has been argued to permeate all grouping behaviour and is therefore perhaps not the best theoretical construct to differentiate between different grouping processes (Goldstone, 1994; Hampton, 2004). Nevertheless similarity has been used in different ways to explain grouping. One of the consequences is that researchers tackle the difference between perceptual and conceptual similarity as grouping has been identified to be sensitive to both perceptual similarity (Coren & Miller, 1974; Choplin & Medin, 1999) and conceptual similarity (Coren & Enns, 1993). Other times reports note the effect of conceptual knowledge on perceptual similarity or family resemblance as facilitating grouping (Medin, Wattenmaker & Hampson, 1987) and a further interest has been expressed (e.g., Murphy & Medin, 1985) about concepts that influence grouping behaviour in 'relations, operations and transformations that perceptual features participate in' (p. 277).

Although the distinction between what drives perceptual or conceptual grouping behaviour in humans is not yet clear, research on compound visual forms in the field of comparative psychology may give some clues about perceiving and processing visual objects across species (Fagot & Deruelle, 1998; Fagot & Tomonaga, 1999). For example, global precedence effects have been identified in humans (Navon, 1977; Kimchi, 1988), and a processing discontinuity between monkeys and humans has also been identified, showing monkeys are local processors (e.g., Spinozzi, De Lillo & Salvi, 2006). Cross-species grouping differences have further revealed our human nature and have also been emphasized by other comparative psychologists (Tomasello, 1999; Tomasello & Call, 1997) who have focused on cultural learning based on understanding intentions only found in humans and not inherent to monkeys. Although the latter is debated (Horner, Bonnie & de Waal, 2005; Horner & Whiten, 2005; Whiten et al., 1996) in for e.g. instances where emulation behaviour is found in apes (versus imitation found in children) indicative of apes having accumulated experiences in the wild, conceptual grouping can be quite abstract when it involves for e.g. language that cannot be so readily found in non-human primates (Matsuno, Kawai & Matsuzawa, 2004).

Grouping differences among people of different cultures are reported pertaining to more abstract thought (Roberson et al., 2005a; Kitayama et al., 2003), indicating that there are constraints on grouping that are culturally specific and require more than just perceptual mechanism to arrive to these differences. The nature of cultural constraints on grouping are also of particular interest in this thesis and are discussed possibly as a combination of minimal perceptual similarity and linguistic influence (Davidoff, Davies & Roberson, 1999; Roberson et al, 2004, 2005a). The thesis explains further why it is believed that conceptual grouping is influenced by language (Steels & Belpaeme, 2005; Braisby & Dockrell, 1999; Harnad, 1987).

Goldstone (1998) gives the metaphor of a bridge for perception, that implodes once language is learned i.e. perception is in pieces. The perceptual bridge needs to slowly be re-built with the idea that now perception is influenced by language. In whichever way the bridge is re-built, grouping seems to include a discrete flexibility that allows one to still group perceptually or conceptually in different situations after the bridge has been built at least in the human adult case. This thesis has attended to findings of different grouping behaviour and attempts an explanation of the context in which conceptual or perceptual grouping behaviour can occur.

1.2 Linguistic relativity: the philosophy behind it

1.2.1 Anthropological pursuit

Ideologies of how language and thought interrelate were discussed extensively in a past tradition that relates to the debate that is advanced in this thesis. Historically as far back as the 6th Century AD, the Indian mystic Bhartrihari said that language anchored thought and all mediated processes of experience such as perception. This opened a debate on the subject to last for decades in the Indian linguistic tradition. The Bhartrihari's following quote is an example of a relational anchor for the link between language and cognition: "Language infuses all cognition. Without this close relationship between awareness and language nothing would become known; it makes identification of things possible. Language is the basis of all branches of knowledge, or all crafts and arts. The consciousness of all migrating beings does not go beyond language...." (in Vakyapadiya Kanda, 1983, p115). Later on, intellectuals and theoreticians in Europe shared similar thoughts. Philosophically Humboldt (1820) was the initiator of linguistic relativity which holds the axiom that language controls thought. Humboldt gave a strong version of linguistic determinism, in his 'Comparative Study of Languages', that man lives in the world about him principally and exclusively as language presents it to him.

Further, the German philosopher Cassirer made a strong claim for the symbolic world that language provides human beings with, and claimed that myth and language were once inseparable, and created worlds through which we were able to understand experience in symbolic form. Myth and language grew apart because of functional evolution but their means of expression are still equal in the way they represent the world so that no detachment can be considered between an object and the description of it: "...the distinctions which here are taken for granted, the analysis of reality in terms of things and processes, permanent and transitory aspects, objects and actions, do not precede language as a substratum of given fact, but language itself is what initiates such articulations, and develops them in its own sphere" (Cassirer, 1946, p. 12).

Linguistic relativity, bound to its philosophical claims, took an investigatory form in the field of anthropology in the work of Boas. Founder of Anthropology in America, Boas believed that there was an overarching biological unity of psychological processes that all humans shared, but they also shared differences that took their form in culture. These cultural areas influenced language which in turn influenced thought. Boas made three important claims on the basis of his work with Native Americans; languages firstly classify experience into categories; secondly different languages classify experiences differently; and thirdly, linguistic phenomena are unconscious in structure, so much so we are unaware of the process of classification. This said, the Boasian tradition still follows the idea that culture and thought influence language, but also suggests some cases where language influences culture and thought. Hence, the relativity of language remains quite controversial. It is this controversy that is presented in this thesis. This debate is also known as the Sapir-Whorf Hypothesis.

1.2.2 Whorf's Hypothesis

Sapir, a student of Boas, elaborated on his arguments, and felt that language channelled thought in a more conclusive and absolute way, acting as a symbolic complete system of reference for thought and our perceptions of the world; thought arises from a 'reading into' linguistic classifications. He believed that language made thoughts possible and thoughts in turn (and only after) refined language to a lesser extent; "the instrument makes possible the product, the product refines the instrument" (Sapir, 1921, p 17, in Lucy 1992). It follows that if language is the instrument for thinking and making conceptual interpretations of reality (also social function of shared formal classification), there exists a relativity of concepts and a relativity of the form of thought depending on which language is being studied. Consequentially, relativity operates at the level of conceptual thought and NOT at the level of basic psychological processes or functions. Hence, Sapir moved away from his mentor. By respecting Boas' work, this cautious separation between overarching loci of psychological processes and the relativity of conceptual thought determined by language proved to be a distinction of great importance that attracted Whorf in his studies.

Whorf, who came into contact with Sapir in his lectures on American Indian linguistics, elaborated on Sapir's careful distinction even further. Whorf first found his inspiration in religious texts, among which he first carefully studied the Hebrew language and looked into root-ideas that he thought underlies the whole language as suggested by Fabre d'Olivet (1921, in Whorf, 1956). He believed that these underlying ideas that help convey meaning through language were a structure that lay at the heart of understanding the meaning of our own behaviour. He thus adopted two ideas that were the first stepping stones into his study of semantic influences. The first idea was *Binary Grouping*, how one word can carry a meaning for many exemplars; "A group of semantic roots having in common a certain sequence of two consonants, containing all the roots with this sequence in one language, and having these roots with but few exceptions associated to a "few certain kinds of meaning" (Whorf, 1956, p13). The second idea was *Oligiosynthetic Languages* which assumes that there are languages of which their vocabularies are built on a very small amount of elements, which lead to phonetic symbolism or more simply the suggested connection between ideas and their related linguistic elements. Whorf wanted to find evidence for these ideas and realized he needed to start objectifying semantic influences for research purposes.

He started working with Sapir, at which point there was talk about carrying out experiments. Working in the realm of linguistics, a wish for extensions to more systematic areas of study were primal and adapted to the study of the Hopi Indians. Lots of these more structured examples are given in his paper on 'The relation of habitual thought and behaviour to language' (Whorf, 1939; in Carroll, 1956) that supported the idea that language was at the heart of perception. Whorf noted that "The strange grammar of Hopi might betoken a different mode of perceiving and conceiving things on the part of the native speaker" (Whorf, 1956, p.17). Whorf further focused on distinctive linguistic analogies in the Hopi and English language and found that the cognitive orientation differed completely, depending on the individual interpretation of experience which in turn shapes specific cultural patterns of behaviour. These linguistic analogies have been analyzed according to their structure going from nonlinguistic observables, to mental interpretation, to linguistic meaning, to linguistic forms in both languages (in Lucy, 1992). Whorf's findings quickly found favour amongst scientists.

An editor at the Technological Review encouraged Whorf to publish an article of his work: "Inasmuch as the analysis of reality is matter of language, and the relativity of such analyses can only be appreciated through studies that show the immense range of possible diversity in linguistic expression it will be seen that there is a connection here with the attempts of science to understand the universe and man- I think it would be very interesting to see the ideas implicit in this statement from your october letter expanded in an article aimed at the Review's group. Is this project of interest?" Thus, Whorf came to make his statement public, into what we know today as the Linguistic Relativity Hypothesis in both its strong and weaker forms; Language influences thought which influences behaviour and the less deterministic version; language is related to thought, some influence of language on behaviour does exist.

In other words, Whorf made his position on linguistic relativity open to further research by claiming firstly that structural differences between language systems will in general be paralleled by non-linguistic cognitive differences in the native speakers of two languages, and secondly that the structure of anyone's native language strongly influences or fully determines the world-view he will acquire as he learns the language.

1.2.3 The Paradox of the Heap

Whorf's hypothesis has been extended to categorization (e.g., Brown & Lenneberg, 1954). The link between Whorfian thought and categorization has been debated on the grounds of similarity and typicality versus deeper relationships or rules that could affirm category membership (Hampton, 1998). Whorf's ideas may parallel the type of categorization that requires more than similarity based on surface features, and requires more in depth conceptual knowledge. Thus, Whorf also initially suggested that perceptual and conceptual thought could be dissociated with the latter qualifying as evidence of language's influence on thought. Those two types of grouping behaviours may indeed be dissociated (Rips, 1989). However, a complete lack of similarity is perhaps not the best way to explain categorization membership in particular because similarity has also been argued to be involved in the emergence of perceptual categories (Goldstone, 1994; Hampton, 2005). Therefore, a weaker Whorfian view, endorsing a milder form of perceptual similarity combined with a linguistic influence seems plausible under constraints suggested by the Paradox of the Heap that point out that non-perceptual mechanisms are at the heart of obtaining a perceptual category (Wright, 1975; Dummett, 1975). Through the paradox one understands that if perceptual similarity were the only mechanism one could use to obtain a perceptual category one would never achieve it.

The paradox asks the question, 'what constitutes a heap?' In the instance of this thesis, a heap might be likened to a category because currently the same is asked of categories i.e., what constitutes a category? Further, the fact that concepts can have unclear and variable boundaries justifies the similarity between a heap and a category because both a heap and a category are vague predicates and in the paradox they are also considered as perceptually vague (Braisby & Dockrell, 1999; Davidoff et al., 2000). The answer to the question of the paradox when using a colour category is that one colour does not constitute a category, and if that colour is not a category, adding another colour will not make it into a category, therefore there are no categories of colour. The answer is paradoxical because we know that colour categories exist and that it is not only because two things are similar that they necessarily are of the same category (e.g., Gelman & Markman, 1986).

Another way of looking at the paradox is to look at how one gets to discriminate between two colours (Davidoff, 2001). According to perceptual similarity, if I take a red, and find its closest coloured neighbour in the continuum, the neighbouring colour will be so similar that the two colours will not be distinguished. One could cover the whole continuum in this way and all colours would then be considered as red. One would thus arrive at the paradoxical knowledge that this is impossibly so, because blue colours for instance would also be considered as red. So perceptual similarity may be useful for some judgments but not for the whole continuum of colour as a hole, and certainly cannot be held responsible for the emergence of categories in its entirety.

Clearly the paradox is concerned with how vague one allows the concept of a heap or a category to remain and how similar exemplars cannot lead to anything but vague categories. If non-perceptually or theoretically one were to decide one needs ten grains of sand to form a heap, then that is fixed. One could then reproduce/learn the rule by similarity (Hampton, 2005). One could also say the same about colour categories; by acquiring a colour category one sets constraints to the colours that belong to its name and creates a boundary between this set of colours and other colours that allows one to further refer to differences as well as similarities between exemplars. The latter is in accordance with a weaker form of Whorf's (1956) suggestion of linguistic influence as a possible non-perceptual mechanism, to defy perceptual similarities' vagueness and the Paradox of the Heap, and that nevertheless does not exclude the usefulness of perceptual similarity once boundaries have been fixed. Thus, one can say that there are non-trivial constraints on colour categories that are beyond grouping by similarity in order for perceptual categories to emerge.

1.3 Cognitive Methods in Colour Research

1.3.1 Codability

The research community struggled to find ways of investigating the Whorfian hypothesis beyond analysis of linguistic analogy. Lenneberg was the first researcher to reformulate the relativity hypothesis so as to make it adaptable to psychological concerns and methods, developing a procedure for investigating, comparing and characterizing language categories more objectively in terms of their denotational referents.

Taking an intra-cultural view, Brown and Lenneberg (1954) made a big contribution to the advancement of the relativity hypothesis in "A study on Language and Cognition" and Lenneberg is much praised for his contribution to the monograph in a memorial tribute to him written by Brown (1976). Lenneberg thought Whorf's approach was rigorous enough for psychology's experimental purposes. However his view was that Whorf did not explain 'how' language influenced thought. Lenneberg sought to improve scientific knowledge about this connection. To support his view, he suggested the study of cognitive processes as this meant that there was a chance for more controlled assessment of the hypothesis (laboratory experiment). Moreover, Lenneberg placed emphasis on finding a research technique that would assess linguistic and non-linguistic data so that a direct relation could be demonstrated. This linking mechanism was named 'codability', a composite measure of agreement in naming, length of name, and response latency in naming. Codability also expressed the inter-subject agreement on naming tasks.

For their first set of experiments, based on the opinion of 5 participants, they created a colour selection of 24 colours (using glossy chips from the Munsell Company) including 8 of the most frequently used colour terms. In their test on colour cognition and language they made the link between names and recognition. On the naming test participants saw the whole array, and were then given colours separately to name one at a time. Degree of consensus was measured; the higher the consensus the higher the codability was to be for that colour. In the recognition task they were given a colour to remember (1-4) and shown the array in which they had to point to the correct colour. There was a positive correlation between codability and recognition accuracy indicating a relationship between naming and memory, in other words between language and cognition.

In Lenneberg's (1957) thesis on the approach to language learning, he tested 27 participants on their colour naming and found a frequency of occurrence expressed in varying symmetrical curves. A trend seemed to appear; the larger curves represented high frequency and matched category centres, overlapping with smaller ones at the intersection between two colours, indicating low name frequency. The fact that curves were overlapping seemed to also imply that colour names as concepts "are best characterised as areas of waxing and waning typicality on a continuum" (Lenneberg, 1957, p.2). This implied a movement away from fixed semantics to a 'fuzzy set' of colours in which category membership was not an all-or-nothing matter. This first expression of colours in approximation to category boundaries, was also iterated in Lenneberg and Roberts's (1956) "the language of experience" – a study on differentiating between foci and boundaries comparing the Zuni language (no distinction between orange and yellow) with English. This was a first attempt to transfer a controlled experiment onto cross-cultural work that unfortunately brought forward some inconclusive results.

Following in Lenneberg's footsteps, there have been further developments of Lenneberg's view of linguistic relativity- some of which are the availibility of controlled stimuli (Munsell colours) and evidence from intra-cultural study to draw on other populations' perception of colour in inter-cultural study, i.e. English lexical items as metacategories for comparative work (Berlin and Kay, 1969), and findings of behaviour at boundaries between colours has certainly been taken up in research to show the influence of language on cognition (Davidoff et al., 1999). According to Lucy (1992), Lenneberg produced a scientific approach tending towards universality of thought/cognitive processes and cutting out differences of linguistic content, which is also the position that Lenneberg adopted in his work in later years (e.g. Lenneberg, 1967), and the position favoured by researchers in the next couple of decades.

1.3.2 Communicability

Lantz and Streffle (1964) revisited Lenneberg and Brown's work on language and cognition and found Lenneberg's measure for assessing colour names -codability- was not related to all colour ranges (such as Burnham and Clark's range, 1955, in actual fact a negative correlation) and added that although codability was associated to within-participant variation it was not a good measure to make between-participant conclusions about linguistic relativity. They used "communicability" in which participants related colour names to each other and identified colours on that basis. This measure was positively related to both colour ranges and had implications for its relation to recognition in the way it captured both communicability between people in a community but also communicability within one person over time. Lantz and colleagues found that communication accuracy was positively correlated with colour recognition, and replicated this result across cultures (Spanish and Mayan participants). However they did not similarly find that communication errors correlated with a particular set of colours. This did not fit the positive correlation between recognition memory and communication because if recognition memory was higher for colours that were more viable in communication, others should be less well remembered. Moreover, much debate arose around how much communication accuracy explains intrasubject processes related to non-linguistic behaviour, and was finally abandoned in psychological circles to give way to codability as known in prototype theory further explored by Berlin and Kay (1969) and Rosch Heider (1972). These theories more thoroughly explained cognitive processes but showed a relative indifference to the influence of language structure. However, the importance of a measure of communication is revived in robot experiments mimicking human verbal interchange by Steels & Belpaeme (2005), in Ozgen & Davies' (2002) who found language to be a form of learning, and in a recent paper by Agrillo & Roberson (2007) indicating that cognitive advantages are naturally obtained when names are used consistently and with greatest consensus. However, before we get onto the supporting evidence, further influence of peak communication and highly coded boundary colours need to be addressed because they lead to the important findings on category demarcations.

1.3.3 Categorical Perception

Harnad (1987) and Bornstein (1987) eradicated the idea of the paradox of perceptual similarity to create categories (Wright, 1975; Dummett, 1975), by finding some fixed characteristics by which a category can be recognized. The characteristic foundation of Categorical Perception is a perceptual warping effect so that some colours are seen as being closer and some as further apart in space, even though all these colours are in reality equidistant in a continuous array of colour (as quantified in for e.g. the Munsell continuum of colour).

Harnad (1987) coined Categorical Perception, and relates his model of Categorical Perception to the Whorfian hypothesis as follows: 'The perceptual and conceptual discriminations we make are governed by the categories we name and by our representations of the invariant features underlying the categorization…To the extent that discrete categorical differences could be demonstrated to be influenced by learned names and descriptions this version of the Whorf Hypothesis would be supported' (p.546). He further found evidence for high-level perceptual processes and language involved in the grounding of categories by neural network models because they fit his top-down view of Categorical Perception (e.g. Congelosi & Harnad, 2001). Although pre-dating Harnad's coining of the term and work mainly done in the realm of audition, Bornstein and Korda (1984, 1985) findings provided support for his model's extension to colour cognition.

Bornstein and colleagues provided evidence for Categorical Perception in same/different matching and discrimination tasks of between and within hue pairs measured by reaction times. The paradigm Bornstein and colleagues have used (Bornstein & Korda, 1984; Bornstein & Monroe, 1980) related to previous work on audition (Posner, 1969) where latencies in a paired matching task have reflected which level of perceptual processing is necessary and in turn the type of information required for the comparison; mainly there is evidence of an immediate visual coding, however with a delay, more abstract, conceptual coding is used, and the advantage of visual coding is no longer present. These findings are also referred to as the distance rule of classification. Similarly, Bornstein's results indicate that matching by similarity brings up some minor differences on pairs of within-category judgements, namely physical similarity judgments required more time on pairs from same category name, but participants reliably showed faster latencies on between category pairs overall. In other words, the spatial warping of Categorical Perception showed improved discrimination responses on between-category pairs versus within-category pairs (Bornstein & Korda 1984), and those colours at the centre of a category were classified together faster than colours that were nearer the boundary (Bornstein & Monroe, 1980). These findings are in accordance with Harnad's (1987) view of Categorical Perception, and the Paradox of the Heap that says that visual coding is not sufficient for categorization; if it was so one would address all stimuli pairs equally, but we now know that there are fixed characteristic for Categorical Perception. These experiments provided a grounding method to demonstrate the effectiveness of Categorical Perception and that there was something more to categorization than perceptual similarity could provide for.

1.4 Whorfian evidence

Evidence has been collected in favour of language's influence on cognition. Five sections of such evidence are described in detail. The first section looks at Kay and colleagues supporting evidence in a cross-cultural study and a colour detection study in the context of colour categories. The second section looks at language's influence on spatial cognition as studied cross-culturally by Levinson and colleagues. The third section looks at thought processes surrounding time and the role language plays as studied by Boroditsky and colleagues. The fourth section covers work on colour categories as simulated in robots by Steels and Belpaeme (2005) and finally the fifth section talks of particular type of learning needed for colour categories to emerge as studied by Ozgen & Davies (2002). All sections have in common that they provide results of conceptual grouping with particular attention to language's role in this kind of grouping behaviour.

1.4.1 Supporting evidence for colour categories

In support of the Whorfian hypothesis, and one of the first studies of its kind, Kay and Kempton (1984) have shown that the Tarahumara speaking Native South Americans and English perceived colours differently using a Categorical Perception paradigm. The English, having colour names for blue and green, judged two out of three colours to be more similar because they shared the same name irrespective of the fact they were physically, in Munsell notation, further apart. The Tarahumara, not under the same linguistic influence, made their judgments on a perceptual level, judging two chips to be more alike because they were physically more alike. Kay and Kempton suggested that although there was a linguistic influence when using colour names, this influence could be eliminated. When participants were refrained from using a naming strategy in the same task, they made their judgments on the basis of perceptual distance. Kay and Kempton's results therefore showed that English and Tarahumara speakers were primarily bound to language to group conceptually and grouping perceptually by similarity when deprived from accessing colour terms.

Gilbert, Regier, Kay and Ivry (2006) have equally been successful in showing linguistic influence in a cognitive colour detection task, but this time with an additional biological effect of lateralization. The hypothesis is that perceptual discrimination may be affected by language through the use of unspoken lexical codes in the left hemisphere (LH) of the brain responsible for language processing but not the right hemisphere (RH). A lateralized colour discrimination task with between- and within-category trials was administered to 11 participants who judged whether the target colour was in the right or left part of a circle. When targets appeared in the right visual field (RVF or LH) participants were faster on between-category trials and slower on within-category trials, as compared to when targets were presented to the left visual field (LVF or RH). With verbal interference during this task, the effects were reversed, subjects became slower on between-category trials and quicker on within-category trials for targets presented to RVF, and on the whole, faster on between category trials when targets were shown in the LVF. Non-verbal interference did not reverse the Categorical Perception effect. To make sure that the laterality effects were solely caused by the LH and not some inter-hemispheric communication, a callosotomy patient JW was tested. Laterality effects were replicated showing, with more conviction, that lexical categories influence perception at a cortical level, in the RVF. Thus, these findings confirm that linguistic relativity and categorical perception can be found to have an impact on language-related areas of the brain confirming categorical perception is relative to language.

1.4.2 Linguistic relativity in the spatial domain

The spatial directions most commonly known to us such as right/left, front/behind, up/down used in an egocentric coordinate system of reference (Miller and Johnson-Laird, 1996), are not preferably used by all. Some other languages such as Guugu Yimithirr (Queensland, Australia) and Tzeltal speakers (Mayan, Mexico near Guatemalan border) use an allocentric coordinate system with fixed bearings such as cardinal orientation points. The difference in use of cognitive systems of reference is studied cross-culturally whereby relative frames of reference (FoR) and absolute FoR are allocated in speech and produce very different ways of thinking about space. The differences point to a non-universal account of spatial cognition, with linguistic origins at the basis of these differences (e.g. Levinson 1996c, Levinson et al, 2002).

Levinson (1996a) argues that the FoRs are essential to locate oneself and other objects around us, but are also used as a means to locate buildings and areas out of sight (Brown and Levinson, 1993). Those FoRs vary in language and also vary in cognition. For example, there seems to be an adaptation of 'spoken' FoR to a horizontal plane, an example would be the adaptation of the cardinal directions to a table top, if a notebook would be lying to the right of a cup on the table it would be lying to the east (for us). For Guugu Yimithirr speakers, the notebook would be relative to fixed cardinal coordinates ('real' North-South-East-West like having an internal compass), so that having fixed bearings alters the perception of the location of the notebook. Evidence of this is found in a rotation tasks e.g. Animals-in-a-row (developed by Levinson and Schmitt, 1993; or Maze Task developed by Pederson and Schmitt, 1993) in which a participant sees animals in a row and has to select and remember those animals once he has turned 180 degrees to face another table. The participant is scored for remembering the correct sequence but also for the reconstructions of the 'spatial time-line'. Cross-cultural findings indicate that Dutch speakers use a relative FoR whereas Guugu Yimithirr speakers use an absolute/intrinsic FoR to reconstruct the row of animals (Levinson, 1997). Similarly, the Tzeltal Maya have also been found to have a discrete absolute FoR made evident in their use of cardinal directions (geocentric i.e. uphill/downhill) (Levinson, 1993). Levinson's conclusion is that non-linguistic FoR found in cognitive rotation tasks that are devoid of the use of language, align nicely FoR of people's own language and show that a mental model is created based on FoR's cultural expression (Levinson, 2003; Majid et al, 2004). As further support for this work, some developmental studies (e.g. Brown and Levinson, 2000; Bowerman, 2000; Choi et al., 1999) have shown that language specific spatial categories extend to children. Bowerman & Choi (2001) suggest that the earlier a child acquires meaning through language, the more defined the role of language in structuring children's early spatial semantic categories, and cognition. In accordance, children accustomed to different languages (English, Korean, Dutch, Tzoltil Mayan) do show evidence of different spatial categories in line with the linguistic use of these categories. In studying their development, Brown and Levinson (2000) showed that children as young as 2 seemed to be tuning in with thoughts aligned to spoken spatial concepts found in Tzeltal Mayan adults, such as uphill and downhill orientations. The children seem to adopt and produce an absolute FoR in speech. Furthermore, Brown and Levinson (2000) report that there is early although varied acquisition of conceptually complex FoRs but that Relative FoRs show the slowest developmental understanding (around 8 years of age; Brown and Levinson, 2000; replicated from De Leon for Tzoltil speakers, 1994) indicating a non-universal approach in FoR acquisition as well. Choi et al. (1999) further contributed to the investigation of the origins of spatial categories. They found that children at an age of 16-20 months (Choi & Bowerman, 1991), showed evidence of spatial categories in a preferential looking paradigm when target words were present. Language use was of particular importance because English and Korean children were found to be sensitive to different spatial morpheme use (English categories 'in' and 'on', Korean categories for same actions 'nehta', 'kkita', 'ssuta', 'kelta', 'nohta', all delineated by verbs and usually related to 'tight or lose fit'). Thus, infants' spatial categories were relative to cultural experience and shows that there is a process of capturing language to orient oneself that happens according to the culture one belongs to.

1.4.3 Linguistic relativity in the temporal domain

Boroditsky asks: Do we rely on sensory and perceptual experiences or do we have conceptual understandings related to deeper semantic structures that enable us to understand we are perceiving 'red'? Boroditsky suggests that what seems biological in nature demands a process of deeper metaphoric structuring. Boroditsky means that mental models are created under the influence of language. Metaphoric structuring (e.g. Boroditsky, 2000; Boroditsky, Ramscar & Frank, 2005) is under investigation in the study of space-time mapping. For example, participants were shown to be affected by ego-moving metaphors when making time judgements (Boroditsky et al, 2005). Participants who took part in an office chair rodeo, showed that whether they were 'riding' on a chair or roping the chair to the finishing line did not make a difference to participants' ego-moving thoughts about time, thus showing it is the concept for moving through time that allows a person to make accurate time estimates. Boroditsky further looked at influences of language on thought in 'thinking for speaking' (by which language is secondary to a brain process affected by the language spoken, in accordance with Slobin, 1996) type experiments that did not require explicit language resources to predict the effect of language on cognition. She further argues that time is more free to vary across languages and cultures because they are not as heavily constrained by perception as colour or other perceptual categories may be.

In an experiment using ego-moving or object-moving priming information (such as 'the flower is in front of me' and 'the hat-box is in front of the Kleenex', Boroditsky, 2000), results indicated the use of spatial metaphors to think about time when answering temporal questions ('Next Wednesday's meeting has been moved forward two days. Which day is the meeting now it has been moved?'). Moreover, doing so is not necessary (can also use temporal information to think about time), but this cross-domain effect cannot be reversed i.e., temporal information cannot be used as a metaphor to think about space. If thinking about time is a high-level process that requires conceptual thought, is it also Whorfian? The answer is yes and Boroditsky has adopted the Whorfian view in cross-cultural experiments.

In one of her first investigations of this kind, Boroditsky (2001) showed that conception of time varies cross-culturally. Mandarin speakers rely on vertical spatial metaphors for estimates about time whereas English speakers commonly use horizontal spatial metaphors. When participants were given a vertical spatial prime and probed with a temporal question, Mandarin speakers were faster to respond than English speakers were. When probed with a horizontal spatial prime, English speakers were faster to respond than Mandarin speakers. This was shown to be an effect of exposure to one's native language, whether Mandarin or English. To remove any cultural factors that might be impinging on this task, English speakers received Mandarin linguistic training in the vertical concept of time, and showed same results as Mandarin speakers at test, concluding that differences in talking do lead to differences in thinking.

Other linguistic effects on thought are demonstrated in a paper by Boroditsky, Schmidt, & Phillips (2002) in which German and Spanish speakers had different effects of grammatical gender on thought. German and Spanish speakers have inverse conceptions of grammatical gender use e.g., the sun is feminine in German and masculine in Spanish. Boroditsky and colleagues showed memory effects induced by preferred grammar use, so that consistent grammar pairs would be remembered better than inconsistent ones. Grammatical gender was also found to have an effect on object descriptions. Different adjectives were used by both language speakers for a set of words, and in learning a new distinction that represented male/female categories, English speakers used adjectives in the same way Spanish and Germans were depending on which language they were exposed to. It can be concluded that differences in grammar are enough to influence how people think about objects and it can therefore be concluded that grammar affects thought. Concept acquisition is therefore seen as direct evidence of deeper language structure.

Developmental work (Imai & Gentner, 1997; Imai & Mazuka, 1997) has further supported the idea that language influences object descriptions (Boroditsky, Scmidt, & Phillips (2002). Cross-cultural developmental findings of the influence of object and substance extensions in Japanese and English participants found that English have a shape bias when perceiving complex objects, and that Japanese have a material/substance bias. This leads participants to make similarity judgments in a triad task based on the corresponding word-extension patterns, with English focusing on common shape and Japanese focusing on common substance. Further these cognitive effects appear quite late as 4-year-olds do not extend their bias to similarity judgments indicating considerable linguistic experience is needed in accordance with other studies on shape classification patterns (e.g., Smith & Sera, 1992; Lucy & Gaskins, 2001).

Thus, in accordance with developmental work, Boroditsky and colleagues are suggesting language has a deeper influence on thought by administering indirect tests of the use of space to think about time; Boroditsky and colleagues have made language evident in cognition without on-line language use. The latter is quite important because it suggests that in some cases perceptual categories are accessed via mental models that operate according to the language they originate from.

1.4.4 Cognitive modelling and robot-behaviour

In agreement with Jameson (2005), Steels and colleagues base their experiments on a model that highlights that colour categories are constructed via colour perception under cultural pressures that rely on 'cultural consensus and linguistic charity' as a way of ending up with somewhat similar categories in a particular culture while maintaining some individual differences that are not relevant to the communication of these categories. This approach points at a complex interaction of linguistic and non-linguistic mechanisms of categorical perception. Steels and colleagues (Belpaeme, 2001; Steels & Belpaeme, 2005) give such evidence for the role of language in cultural evolution. Steels considers "language to be a distributed, complex dynamic system; in which self-organization in the representation of the individual and in the dynamics of the language community is responsible for stable states" (Steels, 1997, 1998). This aspect of language responsible for producing stable states, has been verified in the study of how meanings can be associated to words under a computational model as an alternative to 'neurophysiological prototype theory' (Lammens, 1994; Steels, 1997; Lamberts, 1997). In robot studies, robots perceived, categorised perceptions, and finally lexicalised the resulting categories (Steels & Belpaeme, 2005). Author's predictions were made with computer models for robots' category outcomes that addressed the debate of the origin of colour categories. The computational experimental account consisted of robots solving discrimination and guessing games to make these predictions. First, the discrimination game allowed for robots to acquire a colour category set on the basis of discriminating a colour from its context successfully. The robot does this by matching colour to internal representations. If there is an internal category, all colours will be placed in the same one and discrimination will not be possible. If the robot managed to allocate a 'topic' colour to a category to which all other colours do not belong the discrimination game is successful. Second, the guessing game is played between 2 agents, hearer and speaker. The speaker knows what the topic is and plays a discrimination game. Once the relevant category is found the speaker relays the category name or word to the hearer. The hearer plays an inverted game where he looks for word associated to particular category, which can be mapped onto the colours in turn. It then points to the colour that matches best with the category. The speaker gives 'correct/incorrect' feedback. The game is successful when two robots manage to communicate category exemplars to each other. The game fails if speaker fails the discrimination game, if speaker has no word associated with category and has to create new word, if hearer does not know word, and if the hearer fails to pick out the topic in which cases strength of colour-word decreases. Thus, the robots are responsible for self-organising vocabularies and create perceptually grounded meaning if they are successful. Discriminative and communicative successes are analysed.

Belpaeme's (2001) and Steels and Belpaeme's (2005) results suggest that when there is no communication (as in genetic evolution and individual learning without linguistic interaction or cultural learning), categorization of the colour space is not at its optimal condition (i.e. the eleven basic colour categories are not achieved). A better coherence of categories is achieved through shared environments where robots are subjected to the same colours, although coherence between categories stays low because there is no communication. More accurately, it is the communication between agents that results in a shared lexicon and also increases agents' coherence of categories to the full set of basic colours. The conclusion shows that the role of communication is to increase differences to find optimal use of a continuum such as colour and confirms a weak Whorfian hypothesis because the work shows that perceptual similarity can also amount to some colour groupings but not in the most optimal way as language can. Moreover the simulation that included successful communication in robots resulted in the closest set of colours to those found in humans, indicating humans conceptualise further through language.

1.4.5 Learned colour categories

Communication being crucial to categorization, Ozgen and Davies (2002) suggest that language affects Categorical Perception and is evidence for a form of category learning involved in colour recognition. The authors suggest this route may be interesting to discriminate between categories that might be acquired at birth and others that are learned as a member of a language community so that language is not a pre-requisite for colour categories, only a probable route to take when colour terms are being acquired.

In the set of experiments, participants show improvement on hue and lightness judgments with training on a successive same-different judgment task. When participants were then exposed to novel categories and not trained, they were more accurate at within/new category judgments than between/new category judgments. Next, category learning by training reversed this effect; participants were found to be more accurate on between/new category judgments than within/new category judgments. Some have argued this is evidence for adaptation (Goldstone, 1994, 1998) i.e. being able to unlearn pre-existing colour categories while learning novel ones. Nevertheless, linguistic effects are prominent during training to induce Categorical Perception effects. Moreover, participants were also quite flexible in their learning, as category effects seem to emerge after just one day indicating a possible linguistic learning mechanism. In their discussion, Ozgen and Davies (2002) point out that names can be used to learn colour categories, although it is arguable and not very probable that names were used directly at test seeing that training and test phase were not matched for colour. This means that this was an indirect test of the effects of language on colour categories, and the results strongly suggest that without language very little can be learned and cognition would not be affected at test. The authors do not suggest learning is exclusive to Categorical Perception. Their results rather suggest that Categorical Perception effects are facilitated by naming, and these results are again in accordance with a weaker form of the Whorfian hypothesis.

1.5 Evidence of an innate view

1.5.1 The eleven basic colours

Berlin and Kay's (1969) research characterises a period of colour research in which colour space was shown to have an innate evolution in which 11 basic terms could be acquired. Berlin and Kay (1969) founded their results upon investigation of colour terminology systems of 20 languages using the stimulus array of Lenneberg and Roberts (1956) consisting of 320 colour chips reconstructing the Munsell colour array. Three tasks were administered to mostly bilingual subjects in California in the US: a basic-colour-term inventory, naming and focus task (point to the best example of). Data analysed, found significant clustering in particular areas of the colour array and conclusion of the research were: "The referents for the basic colour terms of all languages appear to be drawn from a set of eleven universal perceptual categories, and these categories become encoded in the history of a given language in a partially fixed order" (Berlin & Kay, 1969, p.4). Moreover, Berlin and Kay suggested the evolutionary order of colour term acquisition was the same across cultures so that a particular language would possess a number of terms predicted by its stage of acquisition. Different languages are therefore suggested to have different levels of terminology-development (for the criteria see figure 1.1); Stage 1 would suggest a language with 2 colour terms; stage 2 with three colour terms; stage 3 with five colour terms; stage 4 with 6/7 colour terms, stage 5 consisting of the acquisition of the eleven colour terms as we know them to be in the English language and several other languages.

Stage 1 Stage 2 Stage 3 Stage 4 Stage 5

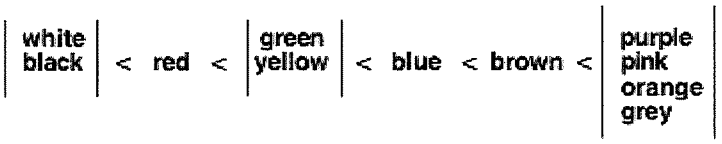


Figure 1.1 Berlin & Kay's (1969) colour stages of evolution.

Although various criticism on the research i.e. non-authentic sampling (bilinguals with English as their second language) and 20 non-industrial languages which were however tested in industrial California (e.g. Hickerson, 1971, Conklin, 1973; in Saunders & Van Brakel, 2002) did not lessen the impact of Berlin and Kay's findings that implied that something other than language determines colours that are perceived, probably responsible for an innate order of colour terminology. Thus, there is something about the world that humans perceive to be constant, even though they live in different environments and speak different languages.

1.5.2 Prototype and cognitive dominance over language

Rosch (1971) introduced the 'prototype', probed by innate colour categorisation advanced by Berlin and Kay (1969), adopting the idea of rigorous cognitive study set by Lenneberg (1967). It is with her input of systematic methods of experimentation (Heider, 1972a) giving the colour domain as an example of ethnographic study in her work with the Dani, that linguistic codes were shown to be 'artefacts' for colours and other domains of natural categories (Rosch, 1973). The latter meant that linguistic terms were just codes to communicate that were a result of deeper underlying representations of best examples of different categories (Rosch, 1975a), and these best examples were suggested to be biologically determined anchors that directed our perception and language use. Therefore, continuous natural categories, such as colour and shape, were thought to be universally agreed upon and to shape human cognitive representation.

Cross-cultural research (Rosch Heider, 1972b) of groups of American and Dani people of New Guinea (who only had two colour names) included tasks of finding the best examples, naming, recognition, memory, and learning. Rosch found that focal and non focal codes for colour (best examples of for instance 'red' and other exemplars that are less good examples of 'red') do not vary across languages, and memory co-varies with changes in this 'universal' code; focals were given shorter names, named more rapidly, and remembered more accurately than non-focals in both cultures. Hence memory for colour seems to elicit same cognitive responding across cultures irrespective of colour terms acquired. In addition, Rosch (Rosch Heider & Olivier, 1972) looked more closely at the relationship between naming and memory for the Dani and the American English speakers as a way of uncovering whether the structure of colour terminology, obviously varying for both cultures (the Dani dividing colour on the basis of brightness and the American on the basis of hue), influenced internal cognitive colour space. Rosch addressed the comparison by multi-dimensional scaling of the 'confusions' made in colour naming and memory tasks. The schematic drawing and statistical analysis confirmed that there was a significant difference in naming structures for both cultures. However, the naming structures did not seem to carry over in parallel fashion to the memory structures within cultures. The important finding was that memory structures were more similar than naming structures across cultures. The only difference found was that overall performance was lower for the Dani than the Americans.

Next Rosch tried to find a theory that could explain universal categories in more detail. Other models of categorization e.g., concept-formation, i.e. correct learning of combination of discrete attributes (e.g. Katz & Postal, 1964; Bourne, 1968; in Rosch, 1973), and abstraction-processing, i.e. arbitrary abstraction of central tendency on basis of culture or givens (e.g. Posner, 1969), were refuted as explanations of the development of categories and Rosch (1973) started advocating for 'natural categories' that were structured in a non-arbitrary semantic way by perceptually salient 'natural prototypes'. A prototype is natural in all languages and characterised by its central position in a basic colour category. All other members of that category are said to cluster around this prototype. Investigation of the Dani (Rosch, 1973) showed that colours and geometric forms that were prototypical were more easily and rapidly learned, and more often chosen as typical exemplars of particular categories in naming and triad-pointing tasks. Further looking at the nature of the mental codes for colour categories, using priming in a matching paradigm, Rosch (1975b) showed that the prime facilitated responses to prototypical examples of that category and inhibited responses to colours that were poor examples. It was concluded that colour names generate a cognitive representation of colour to help encode physically coloured stimuli, and that this representation reflects the prototype structure of colour categories. Rosch argued that there was no reason to believe that this prototype structure is any different across cultures because of different colour taxonomies, as mental codes are translated to cognitive representations, which are prototypical in nature.

To summarise, Rosch is a strong advocate of the universality hypothesis of categorisation. She made this assumption on the basis of her findings that reflect cross-cultural similarities in perception of colour space according to cognitive organisation of natural categories that have inbuilt focals exemplars cross-referencing Berlin and Kay (although Rosch never managed to show any evidence in favour of the evolutionary stages). Subsequently, a complementary line of investigation addressing Berlin and Kay's work was taking place, preoccupied with the nature of prototypes, Kay and McDaniel suggested physiologically determined colour categories.

1.5.3 Neurophysiological determinism

Kay and McDaniel extended the findings of semantic and developmental colour universals to clarify the origins of colour categories. A bridge is created between phenomenal and biological understanding of colour category encoding. Fuzzy set theory is suggested with the basic assumptions that colour categories are continuous in colour space and a degree of membership can be allocated to every colour. The membership functions that represent the semantic structures of these categories can be derived directly from the neural response functions that make up the physiological code for colour.

Prior evidence for the physiological code of colour (Jameson & Hurvich, 1968) suggests that a tripartite response of the retina's 3 cone system is transformed into a set of opponent neural responses from the retina to higher points in the visual pathway. Accordingly, opponent response cells (phenomenally described by Hering and physiologically studied by DeValois and others) were identified in the Lateral Geniculate Nucleus (DeValois, Abramov, & Jacobs, 1966; DeValois & Jacobs, 1968). According to Kay and McDaniel, this physiological code of colour in the Lateral Geniculate Nucleus are triggered by functional neural responses (FNRs), and can best be treated as fuzzy sets that give a precise representation of colour. Membership has a degree, with a maxima and a minima of 0 for colour category membership to colour space i.e. green will find its maxima at the peak response (510 nm), and its minima of 0 in all other areas of colour space that are outside the region of 475nm to 575nm.

Fuzzy sets predict universal foci and universal boundaries (McDaniel, 1974; in Kay and McDaniel, 1978). Kay and McDaniel suggest there are 3 colour groups that can be perceived as fuzzy sets in their model. Primary colours are the four FNRs derived from opponent processes of neural mechanisms for which there is semantic evidence and hence underlies colour perception in its most basic form. Composite categories are those that are created by fuzzy union and arise at the early stages of the basic colour term system, before primary basic colours do. These categories are characterized by their membership in functional response categories from which they are formed, and the absence of focal choices in those colour regions. There are four composite categories, 'warm' (red or yellow), 'light-warm' (white or red or yellow), cool/Grue, and dark-cool (black or dark colours). Derived categories are at fuzzy intersections and are found at later stages of the basic colour term system. These colours are brown, purple, pink, orange, and grey, orange for example being a derived category from red and yellow, but also a basic category in its own right. Although there is some discussion around orange and its status (Boynton, 1997), orange seems to be derived from primary colours and is part of the fuzzy logical operation of successive differentiation of pre-existing basic colour categories.

The evolution based on fuzzy partitioning of colour space is revised from Berlin & Kay (1969), and suggests derived categories are the last colours to evolve and therefore do not carry the differentiation characteristic. Composites arise first in evolution, followed by primary colours, which are a refinement of composite colours, followed by colours derived form the primaries. Kay and McDaniel suggested these stages were not finite; as an example Kay and McDaniel described colours such as chartreuse and maroon to perhaps become derived basic categories in the future such that the '11 basic colour categories' are not a necessary condition.

These are the premises for a neurophysiological basis of colour categories as suggested by Kay and McDaniel, supported by some researchers (Kay, Berlin and Merrifield, 1991; Kay, Berlin, Maffi and Merrifield, 1997; Kay and Maffi, 1999) and refuted by others (MacLaury, 1997; Saunders and Van Brakel, 1997; Lucy, 1997) who have questioned the simplicity of the premises of their work. Nevertheless, from Kay and McDaniel's analysis it is suggested that there are pan-human colour categories modelled by fuzzy set theory, that are further endorsed as support for biological determinism in later research.

1.5.4 Biologically determined colour categories

Bornstein has argued that if children at a pre-linguistic stage and monkeys show evidence of colour categories using psychological methods that are equivalent to those used with adults or in cross-cultural contexts, then it could be concluded that colour categorization is innate and probably has a neurophysiological locus. Bornstein and colleagues have put forward strong evidence in favour of this argument and it is therefore appropriate to look at their work in more detail.

In a study on infants using a preferential looking paradigm, Bornstein (1975) subjected 4-month-olds to stimuli presentations, either colours presented in pairs (all possible combination of centre and boundary colour pairs), and infants were monitored and taped for looking time to a particular colour. Adults were asked to rate the same colours for pleasantness on a scale from 0 to 20. It was found that adults' pleasantness ratings and infants' fixation times seemed to correspond. More importantly, infants preferred looking at category centres ('prototype', Rosch, 1973) than category boundary colours. This finding was in concordance with adults shown to identify category centres faster than category boundaries that are between hues which are more psychologically complex. This infant study is a preliminary study that gives a first indication of fixation times as evidence for colour vision and a hint of a similar preferential mechanism in both infants and adults. More sophisticated studies were carried out using the habituation paradigm in which infants' recognition memory for hue was evident.

Bornstein, Kessen and Weiskopf (1976) first of all demonstrated that infants, who were habituated to a particular colour, chose to look at a colour with a different colour name when given the choice to look at this colour or a colour equally distant but from the same category. Infants did this reliably for all parts of the continua provided by the four primary colours i.e. blue-green, green-yellow, and yellow-red. In a similar vein, Bornstein (1976) showed more evidence for memory of colour being organised around hue categories. Using a single-stimulus habituation paradigm, infants saw a triplet of colours serially, including the habituated-to colour, one colour from the same category, and a colour that crossed the category boundary. What was so remarkable about this study was that infants looked longer at the cross-boundary colour despite interference in the form of delays. This shows more deeply rooted categorisation than one merely acquired during habituation.

To address the question of human uniqueness of colour categories, a study of colour perception was carried out with macaque monkeys (Sandell, Gross and Bornstein, 1979). These monkeys were also made to habituate to a colour, and then responded to different colours. Responses were analysed. The experimenters used a generalization/extinction paradigm in which the monkey was reinforced (with juice) every time he responded to a training colour, and kept from reward at test at which time the monkey saw new colours. Monkeys responded to colours that fell under the same category name, and failed to respond to colours that were from a different category. Hence, macaques recognised a category boundary that matched human colour naming boundaries and therefore colour categories were not unique to humans.

In a further attempt to clarify the neurophysiological locus of Categorical Perception, Bornstein and Korda (1985) used identification and adaptation of hue for further investigation. The authors showed that Categorical Perception varied according to adaptation of a particular hue; adapting to for e.g. blue, in a 2AFC test shifted inflection points and peaked reaction times were found to have shifted from green to blue wavelengths. Bornstein concludes that the locus of Categorical Perception is neurological and relies on mechanisms that do not include language, even though his tests did not assess linguistic mechanisms.

Although there are several criticisms to Bornstein's work that will be discussed later in this thesis, Bornstein has researched colour perception using a two-pronged approach. This approach parallels with the current direction in this thesis, in that the same paradigm is used to draw comparative conclusions on the origins of Categorical Perception.

1.6 Current findings related to universal and linguistic influences on categorization

1.6.1 Relative and linguistically influenced colour categories

The ideas suggesting Categorical Perception's relativity to language have been expressed by Davidoff (2001). Davidoff (2001) starts by claiming that observation alone cannot account for colour categorization as there is no evidence for neuronal selectivity to any of the four basic colours, or any other secondary colours (e.g. Webster, M.A, Miyahara, E., Malkoc, G., and Raker, V.E., 2000). Moreover, Categorical Perception is argued to be paradoxical if based on a perceptual mechanism (Wright, C., 1975; Dummett, M., 1975; Paradox of the Heap) and is particularly applied to colour to explain colour categories are impossible unless a non-perceptual mechanism is involved in cognition. Linguistic relativity is therefore suggested as the basis for categories, because evidence of thresholds at boundaries between colours is found that indicates how one would parse the colours into discrete categories based on linguistic differences. Davidoff (1999; 2001) further elaborates his intentions by cross-cultural investigations.

Davidoff, Davies & Roberson (1999), in 'Color categories in a stone-age tribe', showed that the Berinmo, living in Papua New Guinea and possessing 5 colour terms, have a 5-term language-specific cross-category advantage that resembles the cross-category advantage the English have with respect to their 11 colour terms. CP effects were found in their 2AFC judgments relative to their language, and double dissociation was reported between the Berinmo and the English in a learning paradigm where both languages showed preferential linguistic advantages. Berinmo participants learned paired associations more easily for the nol/wor distinction than the yellow/green boundary and found blue / green and green (shade 1) / green (shade 2) distinction equally difficult to learn. The English participants found association to blue/green distinction easier than learning green (shade 1) / green (shade 2) and found yellow/green learning easier than nol/wor. Results indicate that familiarity with different colour terms may lead to different cognitive results across cultures.

Roberson, Davies & Davidoff (2000) elaborated their findings in the form of an investigation of Rosch's findings that were argued to be in need of replication. Rosch (1972; Rosch Heider & Olivier, 1972) argued that the Dani possessing just two colour terms, showed a closer memory pattern to the English memory pattern for colour, than to their own colour naming pattern. Therefore it was accepted that cognitive patterns did not vary across culture like naming did and rather reflected an innate disposition to a universal colour parsing. However, among other flaws, the colour array used by Rosch was concluded to be biased towards focals (Lucy & Schweder, 1979) and the statistics to back the multi-dimensional scaling conclusions were not as clear-cut with some questions about the multi-dimensional scaling looking at the relationship between naming and memory (Roberson et al., 2000). Thus, this study was revised.

In the first 3 experiments meant to serve as replication of Rosch's work, none of the findings were replicated; a relationship between colour memory and colour name vocabulary was confirmed over and above a relationship between colour memory across cultures, no superior short-term recall of focal colours and no facilitation of associative learning of these focal colours was found. The results took on a cultural significance rather than a universal one. Signs of incipient colour term evolution were also found; Berinmo disagreed on where best examples for basic colour categories should be and the secondary terms used by the Berinmo were not in concordance with the evolutionary path suggested by Berlin and Kay (1969). Although, disagreement with the evolution of basic colour terms was also found by Rosch (1972), the data were not interpreted as culturally significant variations. Further, in the latter three experiments of the study, Roberson et al tried to extend Rosch's findings to shed new light on the matter of the universality of colour categories. New paradigms were used based on Categorical Perception literature (Harnad, 1987; Bornstein, 1987). The first test required participants to make similarity judgments in a method of triad presentation of colours. Triads either crossed the category boundary or exemplars in the triad were called the same name. In the second test, participants had to learn English and Berinmo categories by association to pictures. In the third test participants were subjected to a 2AFC recognition memory task. The results indicated that two colours were judged as more similar when called the same name, learning demonstrated facility for existing linguistic distinction, and a colour was more easily remembered when presented in a cross-category boundary pair. These results all together point towards a linguistic relativity of the kind that does not make universal focals salient, neither in similarity judgments, memory recognition or learning. Do these results carry-over to other non-Western cultures?

Same results were obtained in a cross-cultural study using the same experimental procedure comparing Himba to English and Berinmo speakers (Roberson, Davidoff, Davies and Shapiro, 2005). As with the Berinmo, the Himba, a semi-nomadic tribe from Namibia in Southern Africa, equally have 5 colour terms and although at first glance they showed a similar parsing of the colour space, they had small boundary position differences that lead to salient cross-cultural cognitive differences in colour perception. These results were not exclusively found for colour. A replication of Rosch's finding of cross-cultural 'prototypical geometric shapes' with English and Dani speakers, demonstrated that the Himba did not show any evidence of prototypical geometrical shapes like circle, square, triangle. Roberson, Davidoff, and Shapiro (2002) argued that the absence of straight lines in nature was reflected in the Himba's linguistic absence of these concepts that explained the equal difficulty in learning paired associations of shapes whether typical or a-typical. Thus, findings show generalizing across different cultures and for different perceptual continua.

1.6.1.1 Relativity's developmental approach

Results in favour of linguistic relativity were also supported by developmental work with the Himba. A longitudinal study tracking colour term acquisition was explored to this avail. In Roberson, Davidoff, Davies, and Shapiro (2004), colour category acquisition was studied over 3 years among Himba and English children tracked from the age of 3-4 years. Naming, comprehension, memory and a list task were administered in both populations over the period of study. The aim of the study was two-fold; the first aim was to compare the universalist and relativist argument for the development of colour categories; the second aim was to examine the trajectory of colour term acquisition in both cultures. The first finding was that in the recognition memory task, children that did not master colour names, made errors based on perceptual distance rather than errors reflecting pre-determined colour categories. Children that knew colour terms made more language specific errors (of the within-category type). This finding suggests that language plays a role in re-organisation of colour space reflected in memory errors. The second finding was that there was no predictable order in which the colour terms were acquired. Colour term acquisition was acquired slowly and with great individual variation in both cultures. This suggests that colour terms are not acquired in the ordered fashion according to innate basic colour categories (Berlin & Kay, 1969; Kay & McDaniel, 1978) as shown by for example early naming confusions made by English (white called pink, orange called yellow, brown called black) and Himba children (zoozu called dumbu or vinde, serandu called dumbu or honi). The third finding that bears on 'basic order' as well, is that children who knew no colour names did not show an advantage for focal colours in memory, hereby counter-replicating Rosch's (1973) findings with the Dani adults. Focality is concluded to be language dependent; when children learn colour names, effects of focality arise in memory tasks in function of learned terms and colour boundaries. Thus there is a drastic colour space re-structuring during colour term acquisition that is linguistically triggered.

The second aim of the study, to outline the trajectory of colour term acquisition, indicated that children still refine their colour categories after first colours are learned. It seems that focality, as seen in the advantage to remember focal colours better than non-focal ones, improves over time as colour term acquisition improves. Thus no evidence of biological influence is distinguished. For both populations colour memory performance is determined by the number of terms acquired. Children that know few will remember the few they know, and older children that know many will remember the many colours they have learned. Memory error patterns also vary over time; older children make less of them and make more within-category memory errors regardless of perceptual distance whereas perceptual categorization contributed to a lot of errors made by young children early on in the study. In sum, this work confirms a linguistically formed internal space of colour representations (Davidoff, 1997). Evidence from cross-cultural adult and developmental work, correspond and support linguistic relativity of colour perception.

1.6.2 Universal and innate colour categories

A revision of Berlin and Kay's (1969) study , the 'World Color Survey' (WCS; Kay, Berlin & Merrifield, 1991), was made in the hope of more rigorous data analysis on the universality of colour categories. Shortcomings of Berlin & Kay's (1969) study were re-addressed, e.g. sampling and testing in country of origin by linguist-missionaries rather than testing in California, samples consisted of at least 25 participants per language rather than smaller samples of sometimes just two people. The data encoding and analysis were also a lot more rigorous; naming arrays and term maps were produced to construe a WCS data archive that would reflect more acceptable universal tendencies. The purpose of the new database is to investigate '...two broad questions, one concerning universals, and the other addressing variation of colour naming' (Cook, Kay and Regier, 2005, p 9).

Using WCS data, Kay and Regier (2003) gave a first objective demonstration of universal tendencies in colour naming. They set two questions; Is WCS clustering of colour terms greater than chance and do WCS colour terms fall near colour terms of industrialized societies? They demonstrated WCS clustering principles on the basis of statistical analyses on WCS; calculating one centroid per term and distances between centroids within each language, summing all distances across languages and terms, and adapting a measure of dispersion, low dispersion values showing strong clustering which was their result after 1000 WCS data rotation calculations that maintained a within language structure and randomized cross-language structure. Kay and Regier (2003) confirmed that WCS colour terms fell near terms of industrialized societies. They additionally used a Monte-Carlo rotation to measure the separation between WCS data and Berlin and Kay's data and showed that as compared to the Monte-Carlo expectation the separation was much less than predicted by chance, validating Berlin and Kay's data (1969). Moreover, Kay and Regier (2003) showed that the WCS centroid-clustering were significantly found at the location of the 11 basic colours in English, vs. the remaining 319 colour positions in the Munsell array. These data signify universals are present in the data.

The nature of these universals were further studied by Kay and colleagues (Kay, cook & Regier, 2005; Kay & Regier, 2006b) by investigating whether the universals in naming data stem from universals in the foci or in the category extensions. First, Kay and colleagues show interest in disproving linguistic relativity supported by the study of the

Berinmo in Papua New Guinea (Davidoff, Davies & Roberson, 1999) and find that Berinmo colour centroids are in agreement with English foci so that colour universals are found to be best examples foils of categories across written and unwritten languages. Their findings are based on contour plots (by pooling best-example data and calculating number of hits on each chip) and compared best examples choices from WCS data with best examples of English colour terms. Contour plots matched 6 most basic English colour terms. The second part of the paper focuses on whether the foci from WCS data were derived from category boundaries or foci, by comparing whether best examples of categories cluster more tightly across all languages than do the centres of those categories' extensions. Again using centroid calculation and measures of dispersion the finding were that best examples were clustered more tightly together across languages. The affirmation of foci's role for the universality of colour perception is still not absolute due to cross-cultural investigations refuting universals. Therefore, Kay and colleagues further question the basis of these universally-favoured regions with case studies.

In a further case study of the Berinmo, Kay and Regier (2006a) demonstrated the same colour universals as the ones found in WCS thus offering a further counter-argument to previous studies carried out (Davidoff, Davies, & Roberson, 1999; Roberson, Davies & Davidoff, 2000) that emphasised linguistic relativity with respect to colour categories and their boundary locations. These differences were demonstrated non-linguistically through memory experiments. Using statistical rotation techniques, the original Berinmo naming data seem to align (accumulating boundary chips with same scoring in boundary map) more closely to WCS boundaries than did any of the rotated versions, showing that there are universal constraints on colour term boundaries. The second step was to demonstrate that not only were the Berinmo boundaries typical of WCS boundaries but also in particular alignment with eight other 5-term languages. This was successfully shown, e.g. Berinmo vs. Bauzi (Irian Jaya) had a 73% boundary match. From a look at the maps, the boundary maps for all these languages are incredibly similar to the Berinmo's and boundary matches that lay between 69% and 80%. Kay and Regier drew from findings that the Berinmo were not in any way exempted from having universal constraints on their colour categories and are moreover similar in evolutionary stage to 8 other WCS 5-term languages. As discussed in the previous section, these findings are in sharp contrast to linguistic influence on categorical perception of colour found for two 5-term languages (Roberson et al., 2005). Nevertheless, Kay and colleagues argue for colour universals using the WCS, the largest cross-cultural colour survey ever carried out, and convincingly show that universal pressures on colour categories exist.

1.6.2.1 Universality's developmental approach

As with Bornstein, Franklin's approach to the nativist account is to show that infants as young as 3 to 4 months of age show effects of CP in categorical responding. Seeing that they do not have verbal labels, they are a perfect population to test under the hypothesis that if they were to respond categorically to colour, colour categorization could be concluded to be innate. According to Franklin and colleagues (e.g. Franklin & Davies, 2004) this account leaves two options to the origin of categorization open to inspection and the following rationale is reflected in their studies. The first option takes the incompatibility of cross-cultural findings into account, following which it could be implied that learning verbal labels re-shapes colour categories; the second option implies that we are predisposed with this ability from birth despite language- in this case infant categories should be universal (for which there is no evidence yet).

After highlighting concerns on poor methodology used by Gerhardstein, Renner, and Rovee-Collier (1999) to show that infants did not have Categorical Perception in a visual search task (colour pop-out), Davies & Franklin (2002) addressed their methodological concerns by testing adults on a target detection task in which participants showed Categorical Perception effects, concluding that infants might have colour categories after all. These conclusions lead Franklin to investigate Bornstein's work.

Bornstein Kessen & Weiskopf's (1976) results have been criticized due to the method adopted, in particular the wavelength coordinates (e.g., Werner and Wooten, 1985) and habituation technique (e.g., Franklin, Pilling and Davies, 2005). Franklin and Davies (2004) wanted to investigate the validity and reinstall reliability to these results and replicated this study using a paired-comparison novelty preference method. Infants were familiarized to a stimulus and when shown a novel colour, infants looked longer to the right or the left depending on which screen it appeared. Infants showed Categorical Perception for colour near the 'linguistic' category boundary, for 2 parts of the colour continua (blue-green, blue-purple), not showing any novel preferential responding for within-category trials even when the distance between stimuli was increased for the between-category trials, and when saturation and brightness were made to vary on trials (pink-red). Bornstein, Kessen & Weiskopf's (1976) results were thus replicated and extended to secondary colours refuting the differential treatment of colour such as might be predicted from Berlin and Kay's work (1969). Nevertheless, Franklin and colleagues concluded that despite the latter infant categories showed an innate, universal nature.

Taking a step further, a target detection study was carried out, using the same paradigm for infants and adults, in which evidence from eye movements showed Categorical Perception for both adults and infants. The target detection task comprised of a participant detecting a circle against a background of same or different colour. Franklin Pilling & Davies (2005) present this task as being advantageous for several reasons; the first being that this task could be administered to infants and adults so that direct comparison could take place; the second reason is that this task was free of a memory component, thus eliminating a possible use of colour terms that might be stored in memory (e.g. prototype effect in memory in Huttenlocher, Hedges, & Vevea, 2000). Their point was extended further in their results which showed shorter target fixation times on between-category trials than on within-category trials, and this effect was found for two chromatic separation sizes, showing that neither infants nor adults were sensitive to perceptual distance. Thus, this paradigm was applicable across populations and showed inherent colour categories. Seeing that Categorical Perception is innate, perhaps colour categories are also identical in older children and evidence of those should then be found across cultures.

Franklin et al. (2004) argued that 'colour term knowledge does not affect Categorical Perception of colour in toddlers'. Himba and English toddlers were tested on a 2 AFC task, adjusted to fit a 2-year-old's attention i.e. bear/boy wearing shirts. The English toddlers were found to produce cross-category accuracy effects for blue-green and blue-purple irrespective of colour term acquisition. Only tested on the blue-purple continuum, coupled with a lack of naming and comprehension data, the Himba were also found to show Categorical Perception effects identical to the English. Thus, from this cross-cultural developmental study colour categories also appear to be innate.

In conclusion, the infant and children's findings by Franklin and colleagues are challenging for the linguistic relativity hypothesis. However these results seem to be incompatible with several other cross-cultural investigations (Roberson et al., 2004) and also another study of which Franklin is the co-author in which older children and adults show Categorical Perception differences based on colour terms (Daoutis et al., 2006). Therefore Franklin suggests that it remains to be seen whether linguistic reorganization of colour categories when colour terms are acquired, is at the origin of the incompatibility between the infant work and the cross-cultural investigation of categorization.

2. The current study

2.1 Motivation for the current study

In practice, there are three experimental approaches to address the nature/nurture debate in the quest for the origins of colour categories: phylogenetic (e.g. Sandell, Gross & Bornstein, 1979; Matsuzawa, 1985; Matsuno, Kawai & Matsuzawa, 2004), ontogenetic (e.g. Bornstein, Kessen & Weiskoph, 1976; Pitchford & Mullen, 2001; Franklin & Davies, 2004), and cross-cultural (e.g. Berlin & Kay, 1969; Davidoff, Davies & Roberson, 1999; Kay & Regier, 2003). These approaches address different populations under the same hypothesis, testing the relation between language and categorization. If similar colour categories between two populations are found, the innate view would be favoured; a biological mechanism can be adopted as a valid explanation for Categorical Perception in both populations and the relation between language and categorization is denied. However, if differences are found, colour categories are acquired through learning, and a non-biological mechanism can be suggested. This non-perceptual mechanism affecting conceptual grouping could be argued to be language.

The proposed research contains a phylogenetic investigation questioning differences in colour categorization between human and non-human primates (baboons), and cross-lingual investigations of toddlers in an ontogenetic enquiry, and cross-cultural investigations of adults addressing possible category differences. From the literature it seems that there is a fairly shared amount of experimental evidence supporting different experimental outcomes. It is the aim to establish a novel link between comparative, cross-cultural, and ontogenetic approaches to assemble and expand the knowledge and research carried out on the origins and nature of categories.

2.2 The current set of experiments

2.2.1 Methodological considerations

Some methodological considerations have been made during the course of the studies of this thesis. The four paragraphs below assess methodologies for cross-cultural, phylogenetic and ontogenetic work, and also discuss the Munsell coloured stimuli briefly. More detail can be found in the various experiment chapters that are to follow (see method sections of chapters 3, 4, 5, 6, and 7) and are therefore not further elborated here.

From a cross-cultural point of view, Lucy (1992) reviews possible flaws in previous comparative research within the paradigm of colour cognition both from the period before 1969 and from that which followed. He concludes that multicultural comparisons shed light on the issue of linguistic relativity and puts forward a number of methodological recommendations for future work. He suggests that new methods of examination should include 'back-translation' (requiring at least two bilingual members of a speech community so that the experimenter can ask one individual to translate a question into their own language, and then ask the other member to translate it back again to ensure that the question being asked is the correct one). In the current set of investigations back-translation was possible and carried out with the translator and translator's colleague before carrying out the experiments with the Himba (Languages used were English to Himba and Himba back to English). Lucy also recommends that experiments should be designed so that precise behavioural predictions can be made about the influence of language on thought; the experiments in section 2-7 of the experimental work have been so designed. He suggests the use of materials and/or categories which are familiar to both the Western and traditional cultures being compared such as the Munsell colours presented to the Himba and English were, and in the colour experiments of this thesis in experiment 4 the Himba and English are both exposed to a part of the continua characterized by a firm linguistic boundary and another unfamiliar boundary to their language. The naming data of this section confirms the cross-cultural validity of the materials in that both cultures are equally well able to name the colours presented to them. Although the picture portrayed is that the English have more colour category boundaries and therefore have an advantage at recognizing colours along the spectrum, another experiment in the series of this thesis makes sure the Himba also have the opportunity of being the ones that have the larger vocabulary and thus of perhaps showing an advantage for recognizing differences on the animal pattern continuum. Thus, one is warned of Emics (Pike, 1954), i.e., exposing a culture to another culture's knowledge, yet this aspect is counterbalanced and perhaps not such a concern seeing that perceptual processes are being tested.

Taking a phylogenetic approach, methodology needs to be adequate in order to provide more information about the presence or absence of monkey categories and the data needs to be comparable to that of humans. Even though they are our closest relatives in the animal world apart from apes, comparative studies need to be sensitive to monkey's abilities to fulfil task demands. Monkeys usually have to be trained on experimental procedures in order for them to learn procedures that can then be carried over to test conditions (e.g., Fagot, Krushke, Dépy and Vauclair, 1998). For the case of the matching-to-sample method, monkeys have been shown to successfully learn via identity-matching and are just as good at carrying out this task as human primates are (e.g. Fagot and Deruelle, 1997). There is evidence that monkeys are capable of classifying by abstraction (Neiworth and Right, 1994) using transfer skills so that the conceptualizing is not ruled out by learning and transferring skills and makes the matching-to-sample and ideal and simple procedure to follow (McKone, Martini & Nakayama, 2001). Moreover, the evidence for identity concept learning in monkeys, although by no means uniformly positive (e.g. Premack, 1983; Fugita, 1982), seems stronger than it is for other members of the animal kingdom (e.g. birds, Holmes, 1979). Thus the experimental procedure chosen in this thesis for the comparative work seems adequate (D'Amato, Salmon, and Colombo, 1985). Another important methodological consideration is the monkey's work routine established in the animal laboratory. Monkeys are usually invited to work for no longer than an hour a day and are not taken out of their daily routines such as food-time etc. Monkeys lose their focus of attention quite rapidly and in order to keep their attention levels high while at work, they are reinforced during training and only partially reinforced (at random) during the test phase. This is a standard procedure used in most monkey studies and does not interfere with results obtained.

Taking an ontogenetic point of view, toddlers are of interest here because they can enlighten lifespan issues and more particularly can possibly point to the origins of colour categorization because they too have not acquired colour terms at a very young age, and provide another situation in which non-linguistic grouping can be tested. Thus such questions as put forward by Bornstein, Kessen & Weiskopf (1976) can be tackled by studying infants or very young children; 'Does the infant see colour? If so, what organization of the colour world does the child bring to language learning? Lastly, in what ways does the preverbal child's organization of colour resemble that of adults?' The method adapted to this area of study depends on how old the children are. Very young infants have been studied using the habituation technique (Franklin & Davies, 2004; Borstein, Kessen & Weiskopf, 1976), a method that induces the infant to get accustomed to seeing a particular colour, followed by a record of the infant's looking time to a novel colour (novel colour can be from the same or different category to the training colour). This diffuculty here is that this method has not been used with adults to make adequate comparisons. The closest to the habituation technique may be the often used two alternative-forced-choice or the detection method that currently replaces the habituation technique (Franklin, Pilling & Davies, 2005). However, with older children, methodology can be adapted to be closer to the ones applied to adults. Tests to assess development, such a language acquisition, are usually administered to children, so as to categorize them into various groups of different ability, and test them according to those abilities. The difficulty of the experiment needs to be controlled so as to make a methodology easy enough for children to follow, and fun enough to keep their attention at a high enough level to carry out the task. This, we believe, has been done following Franklin et al. (2005) procedure. From the assessment children can be expected to be at different stages of development, the first stage is comprised of children who don't know any colour terms (colours for purposes of test), the second stage is comprised of children who know some but not all colour terms, and the third stage is comprised of children who know all colour terms. Further children's performance at the first and third stages is compared as they provide two groups that differ on their level of colour term acquisition and that can be controlled i.e. their results subsequently interpreted.

To resume, the colour studies all used Munsell stimuli (although a set of color aid stimuli was also used for the toddler study). The reason for this being first that early work such as Rosch Heider's and more recent work by Roberson et al. (2004, 2005) used the same stimuli and these experiments were partly meant as an extension of their work. Second, some findings with an English population have shown that the use of these stimuli do provide comparable results to the only other widely used manual set of colour stimuli, the OSA system (Sturgess & Whitfield, 1995).). Third, use of the Munsell system does not presuppose that the organization of the Munsell colour space (or any other system) necessarily reflects the underlying organization of the representation of colour in the internal colour space but that these dimensions are sufficient, if not necessary, for the precise methodologies in this series of experiments using colour e.g. matching-to-sample (Jameson & Hurvich, 1989; Thompson, 1995). Also Indow (1988) found that 3-dimensional Euclidian map is a close-fitting representation of the perceptual colour space that does not lose significant amounts of information compared to models of greater dimensionality (7-dimensional). From this and other conclusions about the Munsell system, Munsell stimuli were chosen in the present set of cross-cultural experiments as an available set of calibrated, easily portable stimuli. Third, some of the studies by Bornstein and colleagues under direct scrutiny (e.g. Sandell, Gross & Bornstein, 1979) have also been addressed for their methodological flaws of colour illumination (monochromatic beam, lamp) using wavelength (nm) as the standard colour metric and instead the Munsell system has been found to be a reliable set of stimuli to use.

2.2.2 The participant populations of this thesis

The Himba

The Himba are semi-nomadic pastoralists who inhabit the Kunene Region in north-western Namibia. Unlike many indigenous groups in Africa, the Himba have managed to maintain much of their traditional lifestyle, perhaps owing to the fact that the land they occupy is so harsh and unyielding that it has been rarely coveted by the colonialists and commercial farmers that have affected so many other regions of the continent. They are monolingual and are not subject to western influence.

The Himba (together with the Herero) moved into present-day Namibia and Botswana as part of a larger migration of Bantu-speaking peoples from east Africa several hundred years ago. About 150 years ago, the group began to split and a large group that we know today as the Herero moved southward, while the ancestors of the present-day Himba remained.

The estimated 20,000 to 50,000 Himba, long among Africa's most prosperous herders, are resilient. In the 19th century those in Namibia survived cattle raids by marauding ethnic groups from the south. Most fled to Angola, joining with the Portuguese military and forming their own armies of raiders. Eventually many returned to Namibia. Starting in the 1920s, South African rulers confined them to a prescribed "homeland," officially forbidding them to trade, graze livestock freely, or garden and gather wild plants along the Kunene River. Yet they endured, even if at times life was extremely difficult.

When drought and war struck Namibia in the 1980s, it looked as if the culture of the indigenous Himba people might disintegrate. Ninety percent of Himba cattle, the centre of their economy and identity, died. Some families left to Angola. Lacking any other means of survival and desperate for cash, a number of men joined South Africa's army in its fight against guerrillas seeking Namibian independence. Unable to feed themselves, Himba flowed into the town of Opuwo for relief food, settling in slums of cardboard and plastic sacks.

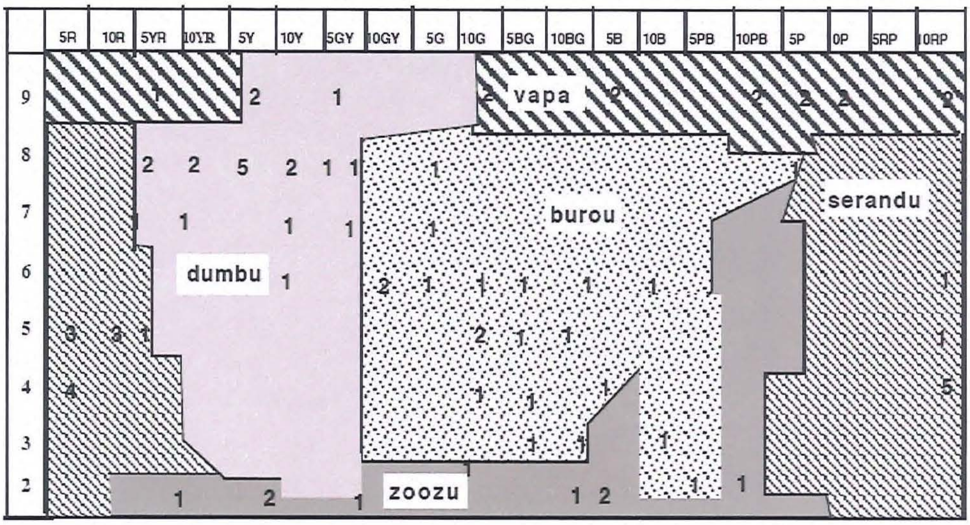
With the peace and good rains that came to Namibia in the 1990s, the Himba rebuilt their herds and, working with international activists, helped block a proposed hydro-electric dam that would have flooded ancestral lands along the Kunene. They also have benefited from new opportunities provided by the government of independent Namibia—mobile schools where Himba children are slowly learning English, and conservancies that give Himba control of wildlife. This does not withstand that the fortunes of the Himba are still very much tied to the earth and its weather as they continue to try and maintain there pastoral way of life.

Today, the Himba live as they have for centuries by following their herds of goats and cattle to new grazing areas and waterholes as existing areas become depleted. They are still affected by the threat of setting up a hydraulic dam at the Kunene River, but seem to live on despite upcoming pressures. They live in small villages or family groups throughout the Kunene Region.

The Himba language is a dialect of the Herero language, but cultural isolation has resulted in a variety of cultural and linguistic differences from Herero. The Herero culture is stable and broadly agricultural. Most villages now have schools and radios, and the people have adopted Western dress. Herero has acquired borrowed colour terms such as 'grine' and 'pinge' (green and pink) that Himba speakers do not use. Himba people are semi-nomadic and have a strong and distinctive traditional cultural identity. They have retained traditional clothing and lifestyles and usually are in little contact with other cultures.

With some of the Himba's history in mind, they proved to be an interesting population to study as they are monolingual and are not subject to the same abstractions that Western people have in the domain of colour. Previous field work has been carried out in this region by Roberson, Davidoff, Davies & Shapiro (2004, 2005) and also Roberson, Davidoff & Shapiro (2002). Interesting results have been reported about their colour term knowledge, of most interest for the purposes of this thesis. I have therefore chosen to include a summary of naming patterns distributions (figure 1.2), as reported in Roberson et al. (2005).

Himba Naming Distribution



*The colour terms that span the Himba's vocabulary are Vapa, indicating white, light and pastel colours, Serandu, indicating red, orange and pink, Zoozu, indicating black, and all dark colours, Dumbu, indicating mixtures of yellow and green with brown undertones, and burou, indicating mixtures of green and blue.

English Naming Distribution

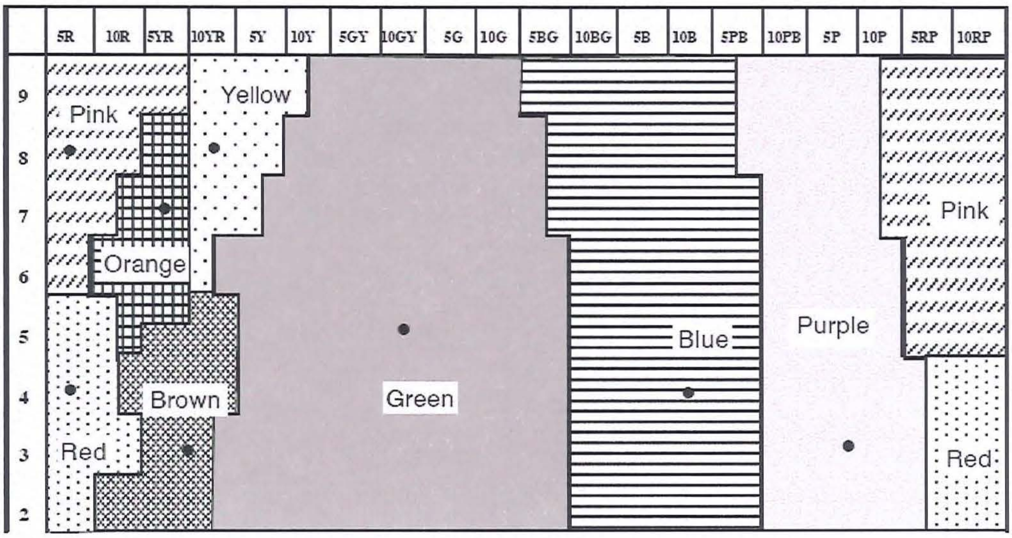


Fig. 1.2 Distribution of Himba naming and choices of best exemplars for the 160 chip saturated array (for 31 observers) compared to those English speakers for the same array. Numbers represent number of individuals choosing an exemplar as best example of the category. Munsell lightness values are indicated on the vertical axis; hue values are indicated on the horizontal axis.

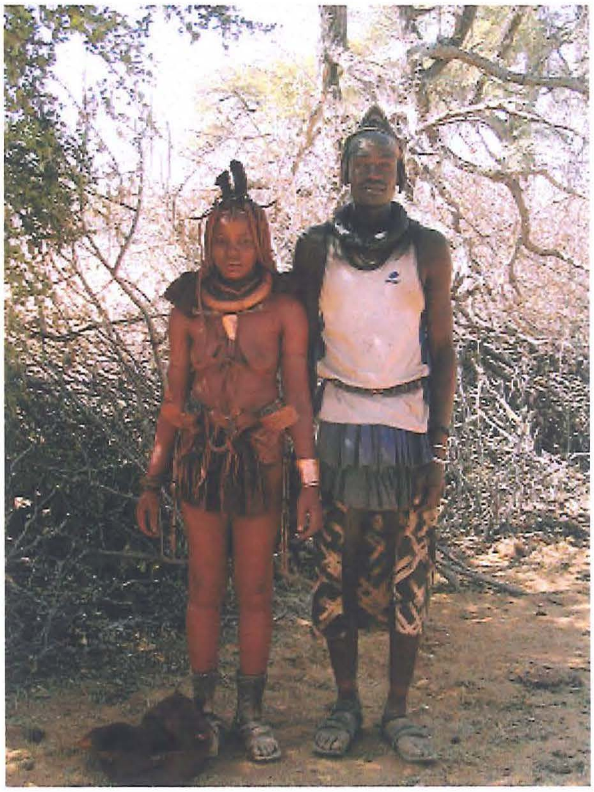
Figure 1.2 supports the idea that there are linguistic boundary differences (where the thick black lines intersect), lines that do not correspond to the predicted places of the 11 colour terms found in the English and also other western languages (Berlin & Kay, 1969). This tells us that the Himba have colour groupings that correspond to their colour naming.

Himba animal patterns were also investigated in this thesis for the first time. Although Himba might base their naming judgments on a combination of colour and shape, this thesis addresses animal patterns as a separate issue for which it is not easy to establish a continua so unfamiliar to the Western eye. The advances made are finding out some naming patterns to explore further possibilities of fieldwork experiments to support linguistic relativity. The main pattern names are illustrated in Eckl (2000), who carried out a more in depth investigation of herds and naming patterns some of which are illustrated in the table 1.1. Those patterns correspond to the findings in this thesis, and several animal patterns are used for further testing.

Table 1.1 Animal patterns used in Goldstein & Davidoff (in press) that also correspond to description found in Eckl (2000). These animal patterns have been collected from 32 Himba observers based on picture naming of various animals.

Name	Pattern
Omvahe	Light, yellow-brown
Ombahiona/ombahozu	Spotted on light-coloured background
Ombambi	Dark, red-brown
Ongwenje	Spotted and white stomach
Ongange	Speckled dark, white spotted
Ekwara	White neck
Onganja	Chestnut, lots of even tiny spots
Ongara	Coal-coloured, black
Ondamberawa	Coloured-back
Onzoozu	Pitch dark, black
Ondumbu	Beige, colour of dry soil
Ombotozu	Dark, with dark patches
Ekondo	Dark with white stripes
Ongwiti	Coloured head and neck (brown or black)
Onthemba	White spots on dark colour

Himba adults and their children (toddlers) participated in the experiments of this thesis. Although there are some mobile schools in some areas, the Himba that were sought out for participation were from remote areas, where mobile schools were not the norm. Due to rapid changes in this area and the invasion of beer traders, some Western cloth has become accessible to the Himbas. The cloth is mainly worn by young Himba men. Women have not adopted any of the invading culture and remain strongly set in their traditional ways of dressing e.g. mix of butter fat, ochre, and herbs on their skin, and shell hanging around their neck with a metal plate hanging from the back, heavy metal bracelets around their feet etc. The women are seen to preserve the culture a lot more than men (see picture below).



Himba woman and man

Children and men are often in charge of travelling with the herds to waterholes, located close to riverbeds that have usually dried up, and only fill up again during the rain season. Those have provided good locations for testing. Himba were always tested just outside of their settlement or at waterholes, and remunerated for participating with maize flour which is a produce they are familiar with; the Himba have their own maize fields, where they collect the seeds, and grind them to flour. Last but not least, the Himba have always been friendly, and very cooperative and happy to see us. They have received pictures from us and have always been very welcoming.

Baboons (Papio Papio)

The Guinea Baboon (Papio papio) is a baboon from the Old World monkey family, probably one of the most common non-human primate species after the chimpanzee which belongs to the ape family. Some (older) classifications list only two species in the genus Papio, this one and the Hamadryas Baboon. In those classifications, all other Papio species are considered subspecies of Papio Papio and are called the Savannah Baboon. The Guinea Baboon inhabits a small area in western Africa. Its range is from Guinea, Senegal, Gambia, southern Mauritania and western Mali. It has reddish brown hair and a hairless, dark-violet or black face with the typical dog-like face, which is surrounded by a small mane. The Guinea Baboon is the smallest baboon species. It is terrestrial, but sleeps in trees at night. The number of suitable sleeping trees limits the group size and the range. Although little is known about its social behaviour, it is likely to live in harems. Like all baboons it is omnivorous, eating fruits, buds, roots, insects and small mammals. Due to its small range and the loss of its habitat, the Guinea Baboon is classified as 'near threatened' by the IUCN (World Conservation Union).

The baboon population used in this study lives in the CNRS in Marseille, in the animal laboratory in enclosures that have an outside platform and inside chamber with several metal floors so that the monkey can sleep in an elevated place at night. They are fed once a day in the afternoon and monkeys do not work more than an hour to two hours a day prior to feeding time. In order for monkeys to work, monkeys are seized in an animal-box vehicle and transported to a different enclosure in the testing rooms. The monkey has been trained to use a joystick, and most psychological experiments are based on the use of this joystick (see picture below). The monkey is usually reinforced with flavoured fruit pellets during training by a feeder attached to the experimental enclosure. The experimenter controls the animal behaviour and behavioural responses on monitors that are located outside the experimental room. Once the monkey has finished the task they are returned to their living enclosures.

Male baboon (Papio Papio) carrying out an experiment

The baboon population consists of 3 female and 6 male monkeys. The dominant male monkey does not participate in experiments, probably out of fear of losing the dominant role in the group. Thus, eight baboons have been tested in the experiments of this thesis.

3. Experiments 1 and 2: A cross-species comparison of colour categorization in a matching-to-sample task.

3.1 General introduction

Herrnstein and Loveland (1964) were the first to demonstrate that an animal species (i.e., the pigeon) could learn to sort visual stimuli on a categorical basis. In their study, pigeons learned to provide behavioural responses consistent with the categorical structure of the stimuli and, even more impressively, generalized to novel, previously unseen, exemplars of the categories. Since Herrnstein and Loveland (1964), categorization has been demonstrated in a variety of species and with a variety of experimental procedures (D'Amato & van Sant, 1988; Dépy, Fagot & Vauclair, 1997; Fagot, Young and Wasserman, 2001; Martin-Malivel & Fagot, 2001).While these studies show that some forms of categorization are possible in animals, and consequently that language might not be a necessary condition for categorization, the cognitive procedures permitting those categorical responses still remain largely unclear. It is particularly unclear if animals solve the task by way of categorical procedures, such as feature analysis or prototype learning, implying an analysis of the physical dimensions of the stimuli, or rely on more abstract thought.

Among the possible reasons limiting the heuristic values of categorical studies in animals is the general use of non-continuous variables to define differences between categories. For example, studies have tested the capability of animals to discriminate between categories such as tree, monkey or people (e.g., Herrnstein & Loveland, 1964). Use of these types of categorical structures makes it very hard to identify the cognitive procedure employed by the animals to solve the task, and therefore to make conclusions about the level of abstractness achieved by the subjects.

Notable exceptions for the use of non continuous variable with animals species are the studies on categorical perception of auditory stimuli in chinchillas or monkeys (e.g., May, Moody, & Stebbins, 1989; Ohlemiller et al., 1999). Most of these authors consider the Voice Onset Time (VOT) as the critical variable in their studies. The VOT corresponds to the interval between consonant onset and the start of rhythmic vocal-cord vibrations. Perception of consonants such as /d/ and /t/ appears indeed categorical in humans, with a boundary lying at a VOT of 20-40 ms. Application of the VOT paradigm has suggested that the discrimination of such auditory stimuli is also categorical in animals (e.g., May, Moody, & Stebbins, 1989; Ohlemiller et al., 1999; Steinschneider, Fishman & Arezzo, 2003), like humans, and therefore that there is no need for language for this kind of categorical perception procedure to appear. This conclusion might however be premature for the following three reasons. First, it should be noted that the stimuli employed in VOT paradigms are not unidimensional, as they vary in a number of dimensions, such as spectral, temporal, or overall energy. It is therefore difficult to ascertain with this procedure the actual mechanism at the origin of categorical perception effects. Listeners in VOT studies were noticeably found to attend to different phonemic cues (Ohlemiller et al, 1999), therefore confirming that different cognitive procedures might be applied by participants with VOT procedures. Second, not all experiments could demonstrate categorical perception with the VOT procedure (eg., Hopp, Sinnot, Owren, & Petersen, 1992). Finally, this kind of procedure does not rule out the possibility that the stimuli are of special ecological value for the animals, and therefore hardly control for training effects.

Other evidence for categorical perception in animals has resulted from categorization studies using visual stimuli. Thus, Wilson and Debauche (1981) tested categorical perception of length, orientation and texture in macaques. After training with the extremities of a set of stimuli, category boundaries were identified for each dimension (50 choices over all comparisons). After training, normal macaques (in comparison with animals with inferotemporal lesions) had improved discrimination performance for between compared to within category discrimination.

Unfortunately, very few studies have assessed categorical perception of colours in animals, including in primates that are known to have the same wavelength discrimination function as humans (Bowmaker et al., 1991). A few attempts have been made. First, categorical perception of colours was obtained in Sandell, Gross & Bornstein (1979), but there are concerns regarding the stimuli used in their study. Because their study did not take into account the spectral sensitivity of the primate eye, the advantage reported by the authors for cross-category contrasts might be accounted for by a perceptually easier discrimination for cross-category than within-category contrasts.

Second, Yoshioka et al. (1996) claim that neurons in the visual cortex of macaques that are specific to colours is not a definitive demonstration of categorical perception of colours, and provides no evidence for the origin of this specificity. Third, the fact that a chimpanzee categorises colours as humans (Matsuzawa, 1985) after being trained to use symbols to name the so-called 11 basic colours provides no evidence that such process would be applicable in absence of specific training, as these categories do not seem to be well-formed in a chimpanzee with only two years of experience with colour symbols (Matsuno, Kawai & Matsuzawa, 2004). The comparative literature therefore provides some evidence for categorical perception in animals, and only controversial evidence supporting the hypothesis that animals have colour categories. Replication and extension of previous studies is therefore warranted. Therefore, this phylogenetic comparative investigation was aimed to examine discrimination thresholds and colour perception similarities and differences between monkeys and humans, using simple procedures inclined to elicit categorical responding and verify whether monkeys and humans show basic perceptual differences (McKone, Martini & Nakayama, 2001).

A matching-to-sample procedure with further modelling was used in Experiment 1 to ascertain the boundary behaviour of monkeys and humans for colours of the chosen continuum. If boundary differences are found between monkeys and humans, the reason may be human's use of language as a grouping device that seems to allow for more abstract perceptual processes and perhaps a clear division between colours. However, more evidence from biological and discrimination functioning are needed to reinforce those result if they are true. To confirm findings of categorical perception, a psychophysical indicator of biological functioning was ascertained in experiment 2a, and categorical perception was tested on this basis in experiment 2b. Discrimination thresholds were measured, by way of an adaptive staircase procedure (Levitt, 1971) leading to a rapid convergence near the asymptotic threshold level. Expectations are that monkeys and humans will have similar colour perception thresholds (see Bowmaker et al, 1991 above), but will show a difference in behaviour on a test that involves category decision over an above simple discrimination (two-alternative-forced-choice procedure). Thus, one expects monkeys and humans to categorize colour differently, but show very little differences in discrimination thresholds indicating that language is at the heart of categorization of colour rather than categorical perception being inherited biologically.

3.2 Experiment 1: Colour categorization in matching-to-sample

(Also a publication: Fagot, Goldstein, Davidoff, & Pickering (2006). Cross-species differences in colour categorization. *Psychonomic Bulletin and Review, 13 (2),* 275-280).

3.2.1 Method

3.2.1.1 Participants

They were 5 male and 3 female adult Guinea baboons (Papio Papio), who lived in the same social group at the CNRS-INCM institute. The baboons were already familiar with the matching-to-sample procedure (MTS ; Fagot & Deruelle, 1997; Fagot, Wasserman & Young, 2001) and were not food deprived. Five men and three women also served as subjects for payment. They all had normal colour vision as tested by Ishihara plates (Ishihara, 1998). The adults varied in age from 21 to 37 years (mean 25.7 years).

3.2.1.2 Apparatus

The experiment took place in a darkened room. The baboons were tested in an experimental enclosure (60x50x72 cm) facing an analogue joystick, a metal touch pad, and a 14 inch colour monitor driven by a Pentium 4 personal computer. On the front of the enclosure were a view-port, a hand-port, and a food dispenser delivering banana-flavoured food pellets into the enclosure. Manipulation of the joystick induced isomorphic displacements of a cursor on the monitor. The monkey viewed the screen through a viewing-port, maintaining the eye/screen distance at approximately 49 cm. For the human participants, the same monitor and joystick were placed on a table so that viewing distance remained approximately equal to 49 cm. Control and randomisation of conditions were achieved through purpose-made programs written in E-Prime V 1.0 (Psychology Software Tools, Inc).

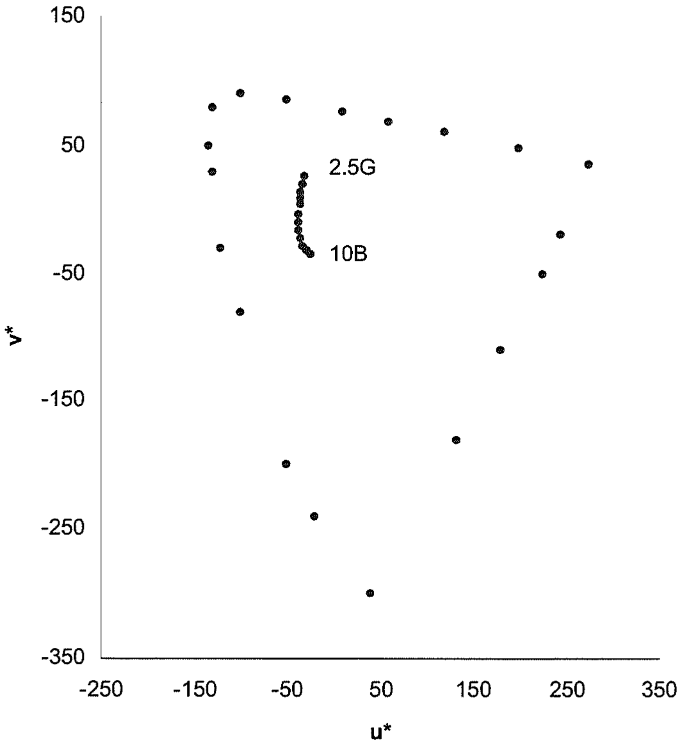
Reliable, consistent representation of highly specific colour samples (see stimuli) requires that a computer screen be precisely and reliably calibrated. This calibration process must be carried out repeatedly and always if the computer monitor has been switched off or exposed to daylight. The monitor should also be switched on for 45 minutes prior to calibration. The custom written software was used in conjunction with a Minolta CS100 colour gun camera. The customised software measures the white point for luminosity and colour along its variant intensities of red, green and blue allowing for the required gamma correction, for which each monitor has its own characteristics. The relationship between gamma and any presented colour is a non-linear one; in other words there would be a discrepancy between the actual value and the intended one. Calibration compensates for this non-linearity, the underlying formula thus matches actual image to desired image as presented via the computer's video card. The monitors used in this research were always calibrated prior to testing and also calibrated intermittently during the testing process to ensure reliability of presentation. The monitors were running Windows 98 and had a 14inch-monitor screen. The PCs were Pentium class processors with integrated graphics cards set to 32 bit colour, 'true colour mode', which ensured sufficient spectral resolution to guarantee reliable repeated presentation of specific colours.

3.2.1.3 Stimuli

The stimuli used in the program were derived from Munsell colours (Munsell, 1905). The stimuli are standardized so that adjacent stimulus pairs are perceptually equidistant by comparison to the distance between pairs of stimuli on the grey scale. Each Munsell stimulus is identified by a hue followed by two numbers divided by a slash mark. The first number (increases in steps of 1) indicates the sample's lightness (higher number means lighter), the second (increases in steps of 2) indicates saturation (higher number means more coloured). In this experiment, Munsell stimuli were represented on a computer screen (same one used for both participants groups) with the help of computer programs that produced colour patches from the Munsell coordinates. The computer was subsequently calibrated so as to verify that the colour presentation on screen matched the colour coordinates, and the correct coordinates on the computer (see apparatus for calibration details).

All colours in the present studies kept brightness (level 5) and saturation (level 6) constant. Twelve colours were chosen to be linearly spaced (with 2.5 Munsell hue units between each adjacent colour), going from 2.5G to 10B. Colour stimuli subtended 6.4 by 6.4 degrees visual angle. Those colours were chosen on the basis of a boundary found around 495 nm for spectral stimuli and close to 7.5BG when using the Munsell system (Davidoff et al., 1999). Contrary to some criticism (e.g., Ratner,1989), the Munsell preference ratings have been noted as carefully calibrated standard stimuli of which the range and foci of colour terms have been compared across a wide variety of cultures (Berlin, Kay & Merrifield, 1991; MacLaury, 1991, 1992). Nevertheless, a number of other colour systems have since been introduced which have attempted to address some problems of Munsell stimuli presentation e.g., controlled under Illuminant C lighting and presented against a grey background (Derefelt, 1991), but comparisons of different systems to the Munsell system have found them to be comparable (Sturges & Whitfield, 1995). Additionally since 1931 the C*I*E* (Commission Internationale de l'Eclairage) has produced specifications for a standard observer, under specified illuminant conditions, which allow translation of the position in the colour space of any stimulus from one system into another. 'Standard Observer' data (taken as the adjusted mean for a number of human observers) represents perceptual differences between two colours on three dimensions (hue, lightness and saturation) as co-ordinates three dimensions (L*, u*, v*) of Euclidean space. Thus, the exact position of a stimulus from one system can be compared to that for one from another system, in the same standardized colour space. Because of this widely acclaimed specification for a standard observer in psychology, Munsell glosses that have been used in these studies were also plotted in Euclidean space and translated into Y,x and y values in Figure 3.1.

Figure 3.1 Munsell stimuli plotted in CIE L*u*v* space.



3.2.1.4 Procedure

In each trial, the baboons placed one hand on the touch pad that produced a .5° circular green cursor along with a white .5°x.5° square-shaped stimulus on the screen, located 1.5° above or below the cursor. Baboons had to manipulate the joystick in order to place the green dot on the white square so as to initiate the MTS procedure; then a square-shaped sample stimulus appeared with 4.5° of lateral eccentricity on the right or the left of the screen. The sample stimulus was displayed for 500 ms and immediately replaced by a display comprising a cursor and two patches of colour for comparison with the preceding sample. In balanced order, one colour square appeared on the top, and the other on the bottom half of the screen. In response to the display, baboons had to point with the cursor to the comparison stimulus matching the sample. There was no time limit for responding. Response choices and response times were recorded. Baboons sometimes received a food pellet inside the enclosure, in accordance with the prevailing reinforcement contingencies.

The training procedure used only 2.5G and 10B stimuli as sample and comparison stimuli. Training sessions were composed of 96 randomly ordered and differentially reinforced identity matching trials, resulting from a completely balanced stimulus identity by stimulus position design. Participants from both species were required to reach a criterion level of 80% correct in training trials before being tested. Verbal instructions were given for humans, but they had to discover the matching rule by themselves.

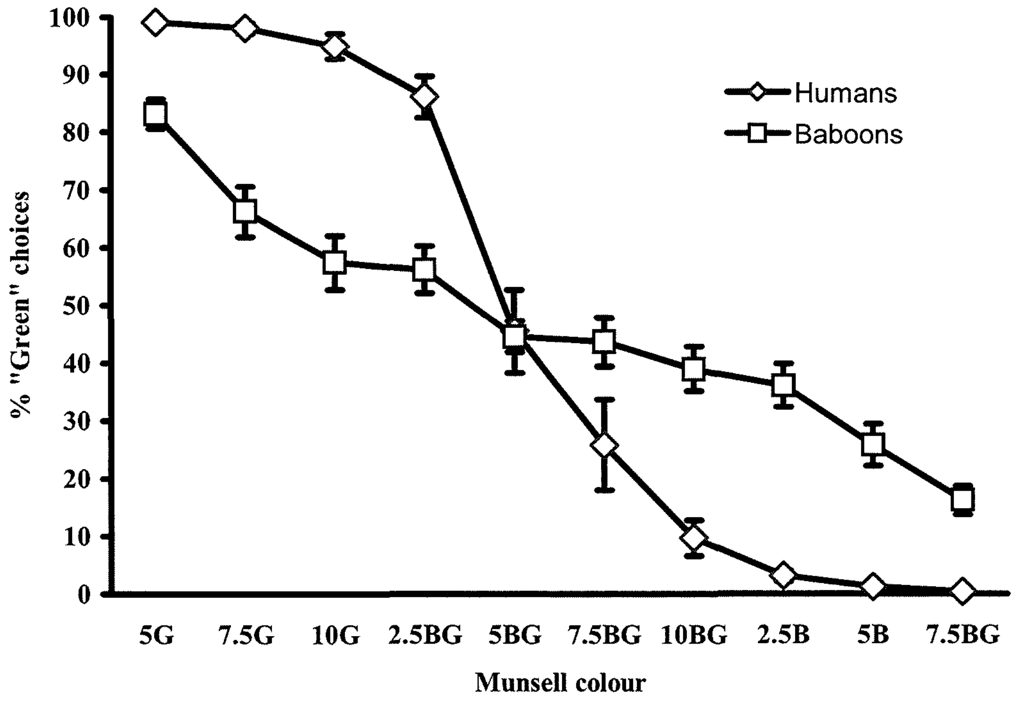
Experimental test sessions were composed of identity (baseline) and similarity (probe) matching-to-sample trials. In the identity trials, the target was identical to one of the comparison stimuli. In the similarity matching trials, the target was randomly chosen from the 10 intermediate colours, going from 5G to 7.5B. For all trials, the colour-comparison stimuli were 2.5G and 10B. All possible combinations of stimulus position were given equally often in a random order. The 2 stimuli used on identity matching trials were each shown 15 times, and the 10 stimuli used on similarity matching trials were each shown once, for each combination of sample and comparison stimulus position, resulting in a total of 160 trials per session. The proportion of identity and similarity matching trials (3:1) was chosen so as to maintain attention during the test. Ten test sessions were required for each participant and species. Baboons received a food pellet whenever a correct matching response was given in baseline trials. Independently of response, probe trials were randomly reinforced at a 80% rate. In cases of an incorrect baseline response, and for the remaining 20% of the probe trials, baboons received a 3-sec time-out period with a black screen. The unique procedural differences for humans were that they were allowed breaks on every occasion after completion of the first three test sessions and were not reinforced during this test.

3.2.2 Results

Training performance varied from one baboon to the next. The sessions needed for the MTS rule to be learned varied between 4 and 26, but overall an average of 10 sessions was needed for baboons to reach criterion. One session sufficed for all human participants to learn the rule.

Figure 3.2 illustrates the percentage of "green" responses for probe trials for each species. There are clearly different response profiles in baboons and humans across the 10 probe stimulus types, with the human profile being curved above the negative diagonal on the left (green) side and below the negative diagonal on the right (blue) side. Although generally more linear, the baboon graph shows the opposite pattern of curvature.

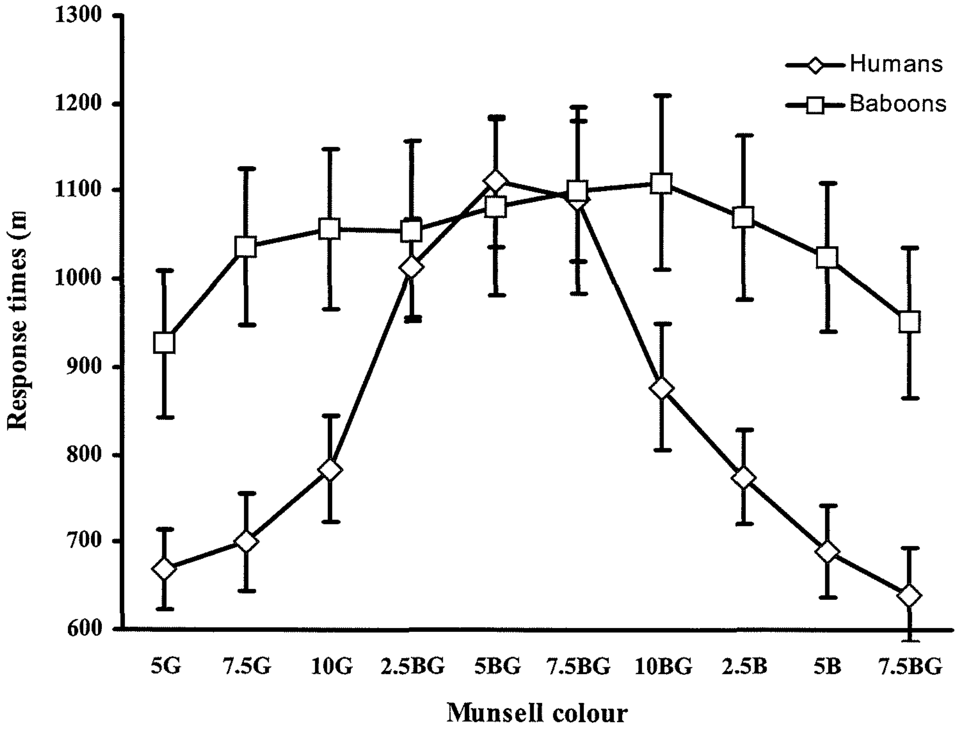
Figure 3.2 Responses choices (% of "green" responses) and standard errors for human and baboon participants.



The frequency of selection of the extreme green colour (i.e., 2.5G) was considered as the dependent variable in a species (human, baboon) by probe sample colour (10 possible colours from 5G to 7.5B) ANOVA. Findings indicated a significant main effect of probe colour, $F(9,126) = 177.9$, $p < .001$, as well as a significant probe colour by species interaction $F(9,126) = 34.5$, $p< .001$. Tukey honestly significant difference post-hoc tests (Tukey HSD, $p < .05$) were computed to understand this interaction. When the left side of the curve is considered, corresponding to colours in the green range, Tukey tests revealed a reliably greater number of "green" responses in humans than baboons for the 5G, 7.5G, 10G and 2.5BG colour patches. When the blue colours of the right side of the curve are considered, data indicated a reliably lower number of "green" responses for humans than baboons, for the 7.5BG, 10BG, 2.5B, 5B and 7.5B colours. There was by contrast no reliable difference for the 5BG colour.

The analysis of scores for probe trials was completed by an analysis of response times. Because there was no correct response in probe trials, all the trials were retained for the analysis, and log transformed, before being submitted to a species (human, baboon) by sample colour (10 colours) two-way ANOVA. The main effect of species emerged as significant, $F(1,14) = 2.79$, $p < .03$, humans being faster on average than baboons. The main effect of sample colour was also significant, $F(9,126) = 17.52$, $p < .001$, intermediate colours giving rise to longer response times than more extreme colours in the continuum. More importantly, there was a significant sample colour by species interaction which is illustrated in Figure 3.3, $F(9,126) = 8.2$, $p < .001$. Tukey HSD tests showed faster responses times in humans than baboons for 5G, 7.5G, 10G, 10BG, 2.5B, 5B, 7.5B. Note that there was no difference between humans and baboons for 2.5BG, 5BG and 7.5BG.

Figure 3.3 Observed latency data and standard errors for human and baboon participants.



Modelling green responses

A central question for this research is whether the participants' categorisation behaviour can be characterized by a "boundary" between subsets of the probe stimuli. If evidence for such boundaries is found, then an ancillary question is whether human and baboon participants use them in a similar way. Fortunately, categorization behaviour has, for many years, been analysed using formal models, which postulate the existence of such boundaries. These models are referred to as decision boundary (DB) models (Ashby, 1992) and they have proved extremely successful in accurately capturing human performance on a range of category learning tasks (Maddox & Ashby, 1993).

We constructed a hierarchy of simple DB models, which could be fit to the probe trial responses made by each participant. The models begin by assuming that the probe trial stimuli each create a perceptual response. We can denote this by p_i for the i^{th} probe stimulus (where i ranges from 1 to 10 for the ten probe stimuli from the most green to the most blue along the Munsell hue continuum). The perceptual response, p_i, across repeated presentations of the same probe stimulus, is assumed to vary due to perceptual noise. As such, it is assumed to be drawn from a random normal distribution with a mean x_i and variance σ_{pi}^2. In the simplest DB model, which can be applied to the current data, it is assumed that the values of the means ($x_1 - x_{10}$) are equally spaced, in keeping with the equal spacing, in Munsell hue terms, of the probe stimuli. In addition, we can arbitrarily set the mean values of x_1 to x_{10} to be 1 to 10, as the values are scaled by other (noise) parameters in the model. Furthermore, the variance of the noise is assumed to be independent of the mean, and to be the same for each probe stimulus. It can therefore be represented by a single parameter, σ_p^2.

The DB model then proposes that the participant sets a simple linear boundary at a particular value somewhere along the range of perceptual response values created by the probe stimuli. The perceptual response value corresponding to this boundary is denoted by k. For each probe trial, if the resulting perceptual response is greater than the value of the boundary, then the participant responds with the same behavioural response as was learned for the blue training stimulus: if the perceptual response is less than the value of the boundary then the participant emits the same behavioural response as was learned for the green training stimulus. The same boundary is employed for all probe stimuli but its position varies according to a zero-mean Gaussian noise term, σ_b^2. This noise can be collapsed together with the perceptual noise term (σ_p^2) into a single overall noise variance parameter, σ^2. It is straightforward to estimate the probability of green responses produced for each probe stimulus by using the properties of the normal distribution. The probability of a green response for probe stimulus i, $P(G|i)$, is given by,

$$P(G|i) = \Phi([k - x_i]/\sigma) \qquad \ldots (1)$$

where $\Phi(z)$ is the cumulative distribution function of the standard normal distribution. To fit this model to each participant's data we used an iterative fitting routine to obtain maximum likelihood estimates of the two parameters (σ^2 and k). The two-parameter model, given by equation (1), was fit to the probability of green responses made for each of the 10 types of probe stimulus.

In fact, we began the modelling with a more elaborate 4-parameter version of the above model. In this model, we assumed that participants adopted two boundaries, one closer to the perceptual response elicited to the green training stimulus (at perceptual response value, k_G), and one closer to the blue training stimulus (at perceptual response value, k_B). If a probe stimulus produces a perceptual response below k_G then the behavioural response produced is the same as that given to the green training stimulus; if the perceptual response is above k_B then the behavioural response produced is the same as that given to the blue training stimulus; a guess is produced if the perceptual response lies at or between the two boundaries. On the guessing trials, the probability of emitting the same behavioural response as that given to the green training stimulus is given by a parameter, g ($0 \leq g \leq 1$); the probability of emitting the same behavioural response as that given to the blue training stimulus is $(1 - g)$. Thus, the four-parameter model can be written formally as,

$$P(G|i) = (1 - \mathrm{g})^* \Phi([k_G - x_i]/\, \sigma) + g^*\, \Phi([k_B - x_i]/\, \sigma) \qquad \ldots (2)$$

The characteristic curve shape produced by the 2-parameter, single decision boundary model resembles the mean performance of the human participants in Figure 4 (convex then concave, from left to right). By contrast, the curve shape produced by the 4-parameter, double decision boundary model resembles the mean performance of the baboons in Figure 4 (concave then convex, from right to left). It was therefore anticipated that the 2-parameter model would better be able to account for the human performance whereas the 4-parameter, double boundary model would better be able to account for the baboon performance.

A likelihood ratio (LR) test statistic was computed in order to compare the fits of the two models (for a review of this standard approach to modelling response probabilities and comparing model fits, see Maddox & Ashby, 1993; Wickens, 1982). The 2-parameter model was considered the better fitting model when the LR test statistic was non-significant, i.e. when its value did not exceed the 5% value of the chi-squared distribution (with degrees of freedom (*df*) equal to the difference in numbers of parameters between the two models; $df = 2$ in this case). As a final test of fit, the better fitting model (with either two or four parameters) was then compared with a perfect saturated model. This saturated model used ten parameters (which were the observed probabilities of "green" responses for each of the ten probe stimulus types). Of course, such a model is (trivially) able to predict the observed probabilities perfectly. Another LR test statistic is computed to compare the fit of the two (or four) parameter model with this perfect model. If this LR test statistic is non-significant (below the 5% value of the chi-squared distribution, with $df = 6$ or $df = 8$ as appropriate) then the simpler model is not significantly worse than the perfect model, and can be deemed a well-fitting model.

The model parameters needed to be constrained to certain ranges in order to compare the two models appropriately. First, one can see from equations 1 and 2 above, that if g approaches 0 or 1 then the 4-parameter model reduces to the 2-parameter model. The value of g was therefore restricted to lie in the range 0.2 to 0.8. Second, if the noise parameter becomes very large under either model the curve shape tends towards linearity and so can produce a reasonable fit to either of the curve shapes illustrated in Table 3.1 Given the range of values used to code the stimulus parameters (1-10), it was decided to restrict the noise variance parameter, σ^2, to lie between 0 and 10 (most of the good fits were achieved with values between 1 and 3; see Table 3.1). Finally, the positions of the boundaries were also restricted: for the 2-parameter model the single boundary (k) was constrained to lie somewhere between the perceptual values corresponding to the two training stimuli (0 and 11 in the stimulus space used in the modelling). For the 4-parameter models the green boundary (k_G) was constrained to lie between the value of the green training stimulus (0) and the midpoint of the stimuli used (5.5); the blue boundary (k_B) was constrained to lie between the midpoint (5.5) and the value of the blue training stimulus (11).

Table 3.1 Best-fitting models and parameter values for individual human and baboon participants.

*The p-value given in parentheses is that associated with the likelihood ratio (LR) test comparing the 2- and 4- parameter models ($p>0.05$ means that the 2-parameter model is preferred);

**This LR statistic compares the better-fitting model with the perfect saturated model.

Participant	**Better-fitting model (2 or 4 parameters)***	k	k_B	k_G	g	σ^2	*LR***	*df*	*p**
Human #1	2 (0.28)	4.61	-	-	-	2.30	21.3	8	0.0064
Human #2	2 (0.95)	5.01	-	-	-	0.74	3.84	8	0.87
Human #3	2 (1.00)	5.34	-	-	-	4.27	6.34	8	0.61
Human #4	2 (0.08)	5.32	-	-	-	1.32	9.09	8	0.33
Human #5	2 (1.00)	4.49	-	-	-	1.56	21.0	8	0.0071
Human #6	2 (0.45)	5.73	-	-	-	1.38	6.00	8	0.65
Human #7	2 (0.67)	6.05	-	-	-	2.28	21.95	8	0.0050
Human #8	2 (0.08)	4.69	-	-	-	1.59	20.89	8	0.0074
Baboon #1	4 (0.01)	-	1.21	9.64	0.40	0.74	1.60	6	0.95
Baboon #2	4 (0.002)	-	3.20	9.37	0.28	10.00	1.96	6	0.92
Baboon #3	4 (0.02)	-	2.09	8.14	0.71	3.01	6.79	6	0.34
Baboon #4	4 ($<10^{-14}$)	-	1.20	10.41	0.51	3.13	10.11	6	0.12
Baboon #5	4 ($<10^{-12}$)	-	1.52	10.02	0.39	2.89	8.60	6	0.20
Baboon #6	4 (0.014)	-	2.31	8.80	0.41	1.56	5.82	6	0.44
Baboon #7	4 ($<10^{-14}$)	-	0.94	9.41	0.47	2.13	3.38	6	0.76
Baboon #8	4 (0.011)	-	2.50	10.57	0.32	2.99	2.90	6	0.82

Table 3.1 shows the model-fitting results for each baboon and human participant. Clearly, the 2-parameter model better fit the human data than the 4-parameter model (8 out of 8 humans); the 4-parameter model, by contrast, better fit the baboons, than the 2-parameter model (8 out of 8 baboons). For 4 of the human participants even the saturated model was not a better fit than the 2-parameter model. The baboon data, with the exception of baboon #2, were extremely well fit by the 4-parameter model. For baboon 2, the best-

fitting noise parameter was the maximum allowed (10.0), a much higher value than in any of the other model fits. As already noted, such a high noise value tends to remove the curvature from the curve produced under the model, and so allows the model to fit the data artefactually. However, the fit of the 2-parameter model was significantly worse than that of the 4-parameter model, and the best-fitting 2-parameter model also was obtained with the maximum noise parameter of 10.0. For this participant, it seems that neither model truly captured the data. For the remaining 7 baboons the fit was excellent; in all cases the fit of the 4-parameter model was not significantly inferior to that of the saturated model (which captures the data exactly). For the human participants, the single boundary was close to the mid-point of the probe stimuli used (range 4.49-6.05 c.f. the midpoint of 5.5).

3.2.3 Discussion

Experiment 1 shows that monkeys and humans perform differently on a matching-to-sample task. This is perceived in their category functions both for accuracy and latencies. Whereas humans show categorization effects for green responses and latencies at test, baboons show a somewhat gradual decline in their green-response scores, invoking matching behaviour, and a flatter latency function showing little difference in reaction to various foils on the blue-green continuum. However, behaviour for the middle point of the continuum did coincide; monkeys and humans showed similar accuracy and latency scores for colour 5BG. To further ascertain whether monkeys and humans were truly behaving differently these results were statistically modelled. In the modelling of green responses it was found that humans showed evidence of processing colours according to one boundary located in the middle of the continuum whereas baboons showed that they were processing colour according to two boundaries located close to the training foils, indicative of matching-to-endpoint behaviour. The latter is not characteristic of categorical perception and a possible explanation of similar behaviour is therefore excluded. It was therefore concluded that human and non-human primates show very different behaviour on a task requiring matching-to-sample of colours, with only humans showing evidence of Categorical Perception.

3.3 Experiment 2: Colour discrimination and categorization in 2 alternative-forced-choices

Experiment 2 (Fagot et al, 2006) casts doubt on whether monkeys show Categorical Perception as humans do for the blue-green part of the continuum. The results of the matching-to-sample task were interpreted as evidence against the proposal that colour categories are innate (Bornstein, Kessen & Weiskopf, 1976) and for the proposal that they are linguistic in origin (Davidoff, Davies & Roberson, 1999; Roberson, et al., 2005). However, similarity matching tasks are inherently open to subject interpretation (Roberson et al 1999). It could be that the human would just as easily mimic the monkey behavior if instructed to match only if the similarity was very close. We therefore decided to adopt a task where both humans and monkeys have to simply decide which of two stimuli is the same as the one previously presented. The two alternative-forced-choice task will allow category effects to be demonstrated because in this task it is harder to distinguish between a perceptually equidistant within-category foil than a perceptually equidistant cross-category foil (Harnad, 1987).

In changing the task from similarity matching-to-sample from Experiment 1 (Fagot et al., 2006) to one of an identity match, one opens the possibility that some colours might be inherently easier to discriminate than others due to the visual neurophysiology. For example, pigeons categorize colours differently from humans and these categories have been argued to be based on the avian wavelength discrimination functions (Wright & Cumming, 1971). Thus, we need to show that there is no enhanced discrimination at the colour boundary within our chosen colour range. It is known that baboons have the same retinal colour vision as humans (Adams, Bryan & Jones, 1968) but in Experiment 2a, we also reassured ourselves that baboons and humans show the same discrimination functions for the specific colour range that will be used in Experiment 2b (identity matching-to-sample or 2 alternative-forced-choice).

3.3.1 Experiment 2a: Discrimination data

There is general agreement on the boundary position between green and blue, at least in Western populations. It is around 495 nm for spectral stimuli and close to 7.5BG when using the Munsell system (Roberson, Davies & Davidoff, 2000). In this human/baboon experiment, we examined discrimination thresholds in the range of Munsell stimuli that were used in Experiment 1 (Fagot et al, 2006) and that crossed the human green/blue boundary.

3.3.1.1 Method

i. Participants

They were two Guinea baboons (*Papio Papio*) who lived at the CNRS (Marseille) and two young human female adults. The baboons had already taken part in many different experiments using the matching-to-sample procedure and joystick computerized systems (e.g. Experiment 1).

ii. Apparatus and Stimuli

The apparatus and stimuli were identical to Experiment 1.

iii. Procedure

Discrimination thresholds were measured with an adaptive staircase procedure (Levitt, 1971) leading to a rapid convergence near the asymptotic threshold level. Thresholds were obtained in separate blocks for the following reference (sample) stimuli: 7.5G-10G-2.5BG-5BG-7.5BG-10BG-2.5B-5B, avoiding the training stimuli used in Experiment 1 (Fagot et al (2006)). The same sample stimulus was systematically employed in all trials of a single session. All stimuli were of constant brightness (Value 5), saturation (Chroma 6) and of 6.4° by 6.4° visual angle.

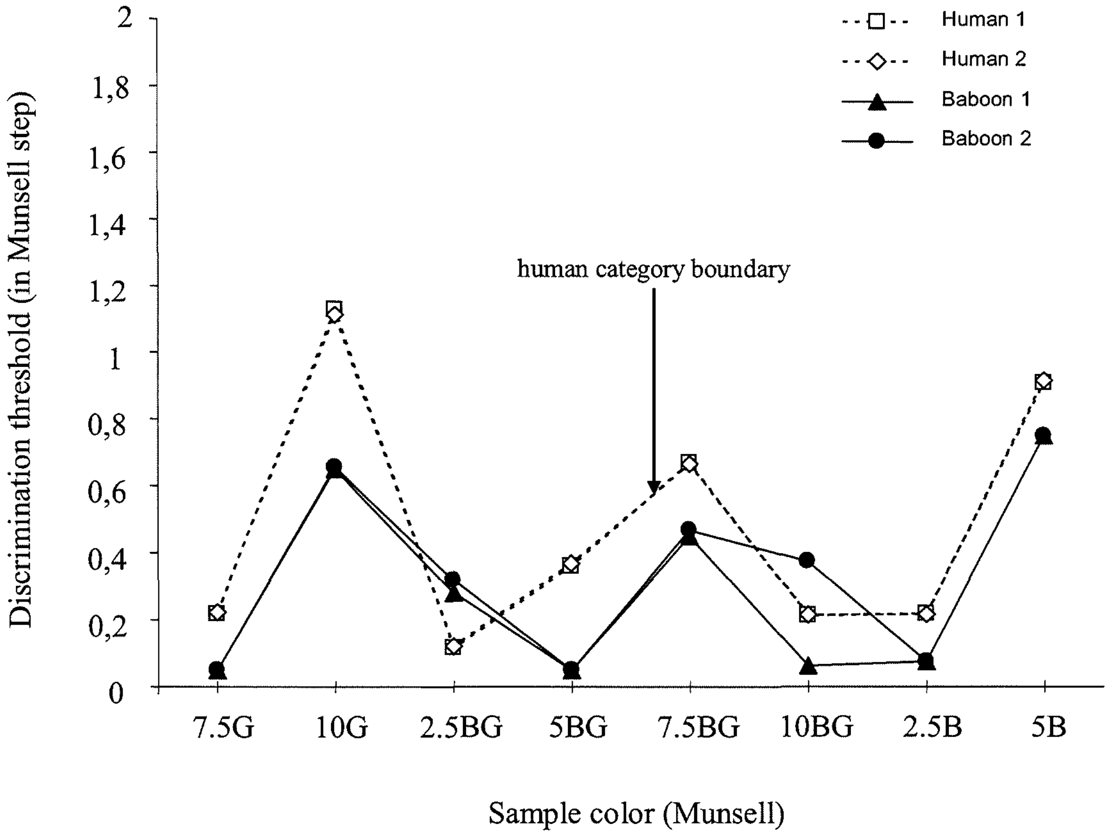
Each trial started with the display of a .5° circular green cursor along with a white .5°x.5° square-shaped stimulus, located 1.5° above or below the cursor. In response to this display, the participant had to place the green dot on the white square so as to initiate the trial. Once done, the sample stimulus was displayed at random on the left or right of the screen for 500 msec and was immediately followed by a pair of stimuli arranged vertically on the mid-line of the screen. One of the stimuli was identical to the sample. The task was to point with the cursor to the stimulus matching the sample. Baboons received a short inter-trial interval (ITI) of 1 second and a food pellet inside the enclosure in case of correct response. When their response choices were erroneous, responses were followed by a three second timeout during which the screen remained black. For humans, the feedback consisted of a short ITI for a correct response, and a three second timeout for an incorrect choice.

The session started with trials having a high probability of positive response. In these early trials, the foil was two Munsell steps different from the sample. When a run of three positive responses was obtained, the next trial reduced the match/foil colour difference by 1/20th of a Munsell step. When a single negative response was obtained, the colour difference was increased by 1/20th of a Munsell step. The procedure continued until 35 reversals (defined as either a colour difference increment or colour difference decrement) were obtained; this compares favorably to most other usage of the staircase procedure (Garcia-Pérez, 1998). Humans received three series of test sessions with each of the eight possible sample colours. Baboons received 6 series with each sample colour. The order of colour testing within a series was randomized. For baboons, the test sessions lasted approximately from 200 to 350 trials. Humans often reached the asymptotic discrimination level more quickly than baboons, and needed from 150 to 200 trials to perform a session. The total testing time was thus of approximately 12,000 trials for each baboon and 4,800 for humans.

3.3.1.2 Results

The increment thresholds were defined, for each sample colour, as the minimal average colour difference obtained in any run of 8 consecutive reversals in any session (Garcia-Pérez, 1998). In the context of very low thresholds for both species, there was generally superior performance from the baboons (see Figure 3.4). Thus, on average for the 8 sample colours, discrimination thresholds for the two baboons were of .29 (baboon 2) and .34 (baboon 1) Munsell steps. It was of .48 Munsell steps for each of the two human participants. Correlations between the thresholds were extremely high and significant within species (humans: $r = .99$; baboons: $r = .93$) and between individuals from the two species (range of *r*s: .81-.89). In addition, each baboon indicated a lower discrimination threshold than the humans (two-tailed pair t-tests, all ps < .02). Regardless of these similarities and differences, it is clear from Figure 3.4 that there is no evidence for superior sensitivity at the category boundary in the region of 7.5BG.

Figure 3.4 Discrimination thresholds obtained in humans (dash lines) and baboons (continuous lines) for 8 equally spaced Munsell colours in the green-blue range.



3.3.1.3 Discussion

Experiment 2a allowed two conclusions. First, in spite of lower discriminations thresholds in baboons than in humans, baboons show the same pattern of results as humans with similar threshold variations along the colour continuum. Second, there is no evidence for either human or baboon for an enhanced sensitivity at the known human green/blue colour boundary (i.e., 5BG-7.5BG). Therefore it seems safe to conclude that any categorical similarity in colour appearance can not derive from the differential sensitivity at the boundary. However, whether Categorical Perception of the blue-green continuum is unique to humans remains to be demonstrated. This will be investigated in Experiment 2b.

3.3.2 Experiment 2b: Establishing a cross-category advantage

Experiment 2b assessed Categorical Perception as defined by Harnad (1987). After having re-confirmed the category boundary of humans with our testing system, cross- and within-category trials were presented to both humans and baboons in a sequential identity matching task. There have been no other colour studies with monkeys that have used the 2AFC recognition procedure but Sandell, Gross & Bornstein (1979) reported Categorical Perception for macaques from a generalization paradigm that followed habituation to Munsell stimuli. The task involved training Monkeys to respond in the presence of one chromatic stimulus and were tested, in extinction, for generalization to others. Sandell, Gross & Bornstein (1979) found that the macaques were more likely to generalize within a category than across a category boundary. They therefore claimed that colour categories were to be found innately within the primate visual system. However, there are some concerns about the stimuli used in this study.

The colours that were mounted on a rotating disk were illuminated at three different luminance levels. Although this is worrying, as luminance levels may affect the viewed colour sample, the most worrying were the actual stimuli that were chosen to be used at test. The distances between the colour samples in 3 of the 4 experiments seemed to predict outcome; the Munsell-metric-distance was always smaller between the target and foils that were of the same category and the distance was always larger between the target and foils crossing a category boundary (ascertained by human colour naming). The

Munsell-metric-distance values are illustrated in table 3.2. As indicated by the Munsell distances, the within-category pairs always, without exception, span a smaller distance than the between-category pairs. In other words, the within-category colours are always easier to recognize at extinction because they are closer in Munsell distance to the target the monkeys have been trained at recognizing. Thus, it is no surprise that in extinction, the monkeys responded at similar and high levels to stimuli that fell in the same basic human hue category as the training stimulus and at similar and much lower levels to stimuli that fell in a different human hue category from the training stimulus. The conclusion that macaques and humans categorize the spectrum in a similar fashion is therefore not conclusive and rather shows that macaques can learn to match colours to a closely related target in a perceptual fashion based on samples that are most similar to the target rather than carry out a generalization task categorically.

Table 3.2 Munsell-metric-distance for experiment 1,2, and 3 between the target and foils for within- and between-category pairs of Sandell, Gross & Bornstein's (1979) study. In the experiments, monkeys were habituated to a target and made to generalize to a foil at test. Foils in experiment 1 and 2 were all equal in Saturation and Brightness (Munsell values 6 and 8 for experiment 1 and values 7 and 10 for experiment 2). Foils in experiment 3 and 4 varied in Saturation. There were three different trials in experiment 1 comprising of green targets, experiment 2 contained a green target as well, and experiment three and four had a red and yellow target respectively.

Experiment	Luminant wavelength (nm)	Human colour name (target)	Munsell designation (target)	Munsell designation (foil)*	Human colour name (foil)	Category pair (with target)	Munsell distance (steps from target)
1	497	green	2.5BG 6/8	5G 7.5G 2.5B 5PB	green green blue blue	within within between between	3 2 4 9
1	505	green	7.5G 6/8	5G 2.5BG 2.5B 5PB	green green blue blue	within within between between	1 2 6 11
1	513	green	5G 6/8	7.5G 2.5BG 2.5B 5PB	green green blue blue	within within between between	1 3 7 8
2	552	green	10GY 7/10	2.5R 2.5Y 5GY 2.5G	yellow yellow green green	between between within within	11 7 2 1
3	609	red	5R 6/10	2.5R 5/10 2.5R 6/10 10YR 6/10 2.5Y 7/10	red red yellow yellow	within within between between	1 1 6 7
4	582	yellow	10YR 6/10	2.5R 6/10 5R 6/10 2.5Y 7/10 7.5GY 6/10	red red yellow green	between between within between	7 7 6 7

It was therefore decided that an experiment would be carried out to verify whether monkeys have colour categories, with the prediction that although monkeys and humans share the same biological functioning, they do not perceive colour categorically as has been suggested by Sandell, Gross & Bornstein (1979).

3.3.2.1 Method

i. Participants

They were 5 male and 2 female adult Guinea baboons (*Papio Papio*) who lived in three social groups at the CNRS, Marseille. This group included the two subjects tested in Experiment 3a and was the same as tested 22 months previously in Experiment 1 (Fagot et al., 2006), with the exception of one baboon that was unavailable for testing. Eight young French adults also served as subjects for payment. They all had normal colour vision (Ishihara, 1998).

ii. Stimuli

In a preliminary testing, the green/blue category boundaries of 10 different French participants were determined in a naming task from 12 linearly spaced Munsell colours going from 2.5G to 10B. This boundary was established between 5BG and 7.5 BG, as found with English speakers (Roberson et al., 2000). Consequently, we used 8 stimuli in Experiment 2b (i.e., 5G, 10G, 2.5BG, 5BG, 7.5BG, 10BG, 2.5B, 7.5B) selected to have the boundary approximately in the centre of the continuum.

iii. Procedure

Each trial began with a fixation procedure identical to Experiment 1. Once the cursor has been placed on the fixation point, a square-shaped sample stimulus appeared with 4.5° of lateral eccentricity on the right or left of the screen. This stimulus was displayed for 500 ms and immediately replaced by a display comprising a cursor and two patches of colour for comparison with the preceding sample. In balanced order, one colour square appeared on the top and the other on the bottom half of the screen. During the response period, participants had to point with the cursor to the comparison stimulus matching the sample. There was no time limit for responding. Response choices and response times were recorded. Reinforcement contingencies were as in Experiment 1.

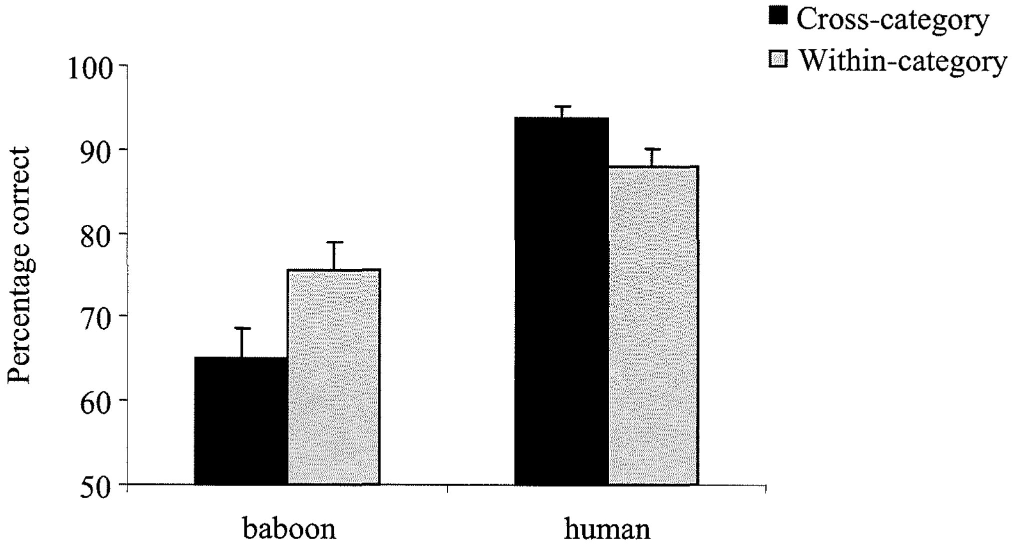
The testing consisted of three sessions of 64 randomly ordered trials each per participant, using the stimulus pairs 5G-10G and 2.5B-7.5B for within-category comparisons, and the stimulus pairs 2.5BG-7.5BG and 5BG-10BG for cross-category comparisons. All stimuli were used equally often as sample and comparison and their location on the screen was completely balanced. Prior to testing, the participants from the two species received 16 randomly ordered and differentially reinforced training trials using two colours (purple and grey) different from the test colours. These trials were repeated until subjects reached a criterion level of 80% correct.

3.3.2.2 Results

Trials associated with latencies less than 150 ms or greater than 4 sec (less than 3% of the total number of trials) were omitted from statistical analyses because they likely reflect anticipation or inattention. Baboons responded in 1167 ms on average, humans were faster with an average response time of 772 ms. Response times were log transformed prior to statistical analyses. Application of species (baboons, humans) by category (within, between) ANOVA on transformed response time revealed a significant effect of species, $F(1,13) = 18.75$, $p < .001$, and a significant species by category interaction, $F(1,13) = 63.95$, $p < .001$. Analysis of this interaction with a Tukey HSD test indicated that response times were reliably shorter for humans in cross- compared to within-category trials (mean of log RT for cross = 6.50, mean for within = 6.59, $p< .05$). Baboons showed an opposite effect corresponding to a reliable within-category advantage (mean log RT for within = 6.86 ms, mean cross = 7.0; $p < .05$).

Accuracy scores were also submitted to an ANOVA using the species (baboons, humans) as between-subject factor and the category (within, cross) as within-subject factor.

Figure 3.5 Percentage of correct responses for the two species, and for cross- and within-category pairs.



As confirmed in figure 3.5, there was a significant effect of species, F(1,13) = 38.5, $p < .001$, corresponding to greater performance on average for humans (90.8% correct) than for baboons (70.3%). The other reliable effect was the species by category interaction, F(1,13) = 18.55, $p < .001$, showing a numerically greater performance for cross- (93.7% correct) compared to within-category trials (88%), but ceiling effects prevented this difference reaching significance (Tukey HSD, $p > .05$). Baboons showed the opposite effect. For baboons, performance was greater for within- (75.56% correct on average) than for between-category trials (65.1% correct; Tukey HSD, $p < .05$, see figure 2). In brief, the analyses of accuracy converge on the pattern of performance for the latency data with no indication of a speed-accuracy tradeoff.

Both latency and accuracy analyses revealed an initially surprising within-category advantage only for baboons. There would be no *a priori* reason why baboons should show a within-category advantage unless it was a consequence of the previous training in Experiment 1 (Fagot et al, 2006). To examine the within-category advantage for baboons, we tested recognition performance for the 8 individual sample colours to verify if matching performance was uniform throughout the continuum. A one-way ANOVA on latencies revealed a reliable effect of sample stimulus, F(7,55) = 4.69, p < .001. Significant differences only emerged from pair-wise comparisons involving the two endpoints of the continuum (i.e., 5G and 7.5B). A similar analysis on accuracy also revealed a significant effect of sample colour, F(7,55) = 4.92, p < .001. Tukey HSD post hoc paired-comparisons between all colours again revealed that only contrasts involving the endpoints (5G and 7.5B) gave significant differences (p < .05). Not surprisingly a similar analysis on the human data gave different effects. For accuracy, there was an overall difference between colours, F(7,63) = 2.45, p =.029, but no post-hoc Tukey tests were significant. The comparisons less than p < .2 were all cross-category not within-category pairings. For latencies, there were no reliable effects of sample colour, F(7,63) = 1,53, p = .18.

3.3.2.3 Discussion

The results for accuracy and latencies repeated the cross-category advantage for humans found in previous studies (Davidoff, Davies & Roberson, 1999; Roberson & Davidoff, 2000). A quite different result was shown for baboons, who surprisingly gave a within-category advantage in the same task. However, further analyses showed that this within-category advantage for baboons was restricted to stimulus pairs involving either endpoint of the continuum of test stimuli. The most likely explanation for the superior performance for the endpoints is that our baboons had simply remembered these colours from the previous and first experiment (Fagot et al., 2006). That conclusion is only credible because baboons have prodigious visual memories (Fagot & Cook, 2006) allowing them to retain thousands of visual stimuli over extended periods (in the range of a year, Fagot & Cook, 2006). So it is quite reasonable that they were able to remember two highly trained particular colours for 22 months. The training stimuli in Experiment 1 (Fagot et al., 2006) were 2.5G and 10B. Modeling the matching performance of these baboons showed that they were prepared to match only the 5G and 7.5B stimuli to the training stimuli. In consequence, if the baboons were basing their recognition partly on their previous experience, then it would only be the endpoints for which we would expect superior performance, and this is precisely what has been found in our research.

3.4 General Discussion of experiments 1 and 2

Human and non-human primates show different results when asked to categorize colour; humans categorize showing effects of Categorical Perception and baboons do not categorize in a similar fashion, rather they match to endpoints as a result of learning. Thus, there is no indication of colour categories for non-human primates and suggests language has a role in the formation of colour categories. This view was first suggested in Experiment 2 and reinforced in experiments 2a and 2b. In Experiment 2a, we looked for a differential sensitivity between species at a colour category boundary but found no evidence for enhanced sensitivity in either species. This confirmed similar biological functioning and colour discrimination function as suggested by Bowmaker et al (1991). In Experiment 2b, only humans showed Categorical Perception for colours implying that it is only humans that see within-category colours as resembling each other more than equidistant cross-category colours. The results of Experiment 2b do not fit well with the view that colour categories derive from inborn discontinuities in colour discrimination functions (Bornstein, Kessen & Weiskopf, 1976) and confirm that only humans have the capacity to show Categorical Perception. Thus, our data do fit the view that colour categories are instantiated through top-down influences and, in particular, from the colour terms in a speaker's language (Davidoff, Davies & Roberson, 1999; Roberson et al., 2005). Recent computer simulations come to a similar conclusion because robots with human colour vision do not arrive at the same colour categories unless they communicate with each other (Steels & Belpaeme, 2005).

The data also speak to the differential effects of training for the two species. It is clear that the baboons benefit in our recognition task from their previous experience. In Experiment 1 (Fagot et al, 2006), the baboons were initially trained to match two specific blue or green colours, after which they had the opportunity to match a range of intermediary colours to the training samples. Here we found that, even twenty-two months later, superior recognition was achieved for colours close to those two trained stimuli. However, the learning outcome for the monkeys was simply to reinforce the trained exemplars as they only gained advantage for colours that were very similar to the training stimuli. Such a restriction in similarity matching would make human broad colour categories difficult to achieve. It would rather appear that some form of labeling is required to achieve the acquisition of new colour categories (see for example, Ozgen & Davies, 2002). For that reason, it could be that chimpanzees, unlike monkeys, are capable of demonstrating colour categories. Matsuzawa (1985) trained a chimpanzee (i.e., Ai) with extensive symbol training to associate symbols to human prototypical colours. The learning generalized to other colours and mimicked human colour categories though this did not appear clearly in another chimpanzee with only two years experience with colour symbols (Matsuno, Kawai & Matsuzawa, 2004). Thus, even the large number of trials required to establish categorization of multi-dimensional stimuli (Freedman et al., 2001) could be insufficient to produce Categorical Perception for colours if the species is incapable of symbol use.

In conclusion, these set of experiments have made a novel contribution towards trying to establish Categorical Perception in animals by using a uni-dimensional set of stimuli (Munsell colours with measurable factors) in a set of simple testing procedures that adds value to the more extensively researched area of categorization found in speech sounds (May, moody & Stebbins, 1989; Ohlemiller et al, 1999). The experiments also tried to address some of the pitfalls of the only other study showing colour categories in monkeys (Sandell, Gross & Bornstein, 1979) and found that possibilities of colour Categorical Perception in animals are minimal, if any, and that the ease with which Categorical Perception manifests itself in humans is probably due to what separates human and non-human primates, namely language.

4. Experiment 3: A cross-cultural comparison of colour categorization in matching-to-sample

4.1 Introduction

The second strand of the research studied Himba and English populations to ascertain whether language differences affect perception. The direct link between language and cognition has been proposed in several studies (e.g., Roberson et al., 2005a; Roberson, et al., 2002; Roberson et al., 2000) using similar or same cross-cultural contexts. More particularly, language-specific categorical effects have been found for the Himba, the Berinmo and the English using a range of procedures such as short-term and long-term recognition memory, similarity judgements, two-alternative forced-choice judgements, and paired-associate learning. In this experiment, a MTS procedure is used to be consistent with the Baboon-Human data, and to detail categorical findings further. Moreover, universal and cultural constraints have been identified in a grouping task across cultures (Roberson et al, 2005b) indicating some expectation that participants group (universal) and do so with some marked differences (relative). Consequentially, were a grouping measure also found to be a categorizing measure able to account for differences and similarities in categorical responses, it could be concluded from using an matching-to-sample task that the categorization process is perhaps not solely perceptual but also influenced by language. These results would be all the more compelling under the restraints of such a perceptual task that has also been used previously as a psychopysical measure of similarity (e.g., Sperling & Harwerth, 1971).

Nevertheless, evidence of grouping variations can be found among the colours under scrutiny that are suggested to be largely determined by non-perceptual, arguably linguistic, influences (green-blue vs. grue, in Roberson et al, 2005b) and are also expected under the conditions at test. The English have two colour names, blue and green, and a category boundary between them; this is shown in their categorical function of previous research when comparing monkeys and humans. The Himba, however, only have one colour name, *buru* (Roberson et al., 2005a) for both blue and green and therefore will predictably not show the exact same categorical function. Using the identical matching-to-sample colour paradigm that has been used in comparative work (see section 4), a quantitative difference between the Himba and the English is foreseen. The reason is that both cultures have acquired colour terms, and on this basis linguistic relativity would predict the use of those terms to categorize differently although using a universal grouping mechanism. So, having the terms increases the likelihood for marked differences. Thus category grouping on matching-to-sample tasks is predicted to be more pronounced for the English on the blue-green boundary and for the Himba on the dumbu/buru boundary. This prediction, if true, will also be visible by looking at the category functions, as typical categorization curves have been identified (e.g., Maddox and Ashby, 1993). Those categorical functions show steepness around category boundaries with a sharp drop in positive responding.

In sum, predictions are that the Himba will perform with more certainty in grouping and their response function will look steeper when using colour terms specific to their language. In comparison, their responses to the blue-green boundary should show a flatter, less determined categorical function. The English will perform with the reversed pattern, more determined in grouping in the green/blue boundary matching-to-sample task. This may lead to a perceived double dissociation in which differences in acquired linguistic terms sets English and Himba performance apart, not in terms of accuracy but in terms of conceptual grouping advantages.

4.2 Method

4.2.1 Participants

Different participants took part on the different matching-to-sample tests. For the Blue-green Set there were 12 English (4 men and 8 women, varying in age from 18 to 30 years, with a mean age of 21.5 years) and 12 Himba (eight woman and four men aged from 18 to 40 years) participants in this experiment. For the Dumbu-buru Set there were 12 English (5 men and 7 women, varying in age from 18 to 33 years, with a mean age of 20.9 years) and 12 Himba (4 men and 8 women, varying in age from 18 to 40 years) participants in this experiment. The English participants were students at Goldsmiths College and took part for course requirements. The Himba were monolingual, tested in Northern Namibia and paid in kind. All participants were screened for colour vision abnormalities with the City Color Vision Test (Fletcher, 1980).

4.2.2 Apparatus and Stimuli

Stimuli used were in the form of 1 inch square glossy finish Munsell chips that were mounted on 2 inch square pieces of white card, and the choice given in the MTS comprised of two colours being mounted on a white card in vertical alignment. Per Set, twelve Munsell colours were used covering eleven equally spaced steps in the Munsell continuum, for the Dumbu-buru Set going from 5Y to 2.5BG, covering the Himba boundary 7.5GY between dumbu and buru colours, and for the Blue-green Set going from 2.5G to 10B, covering the boundary between 5BG and 7.5BG. All colours had value of 5 and chroma of 6 following Munsell notation. The boundary locations were previously estimated and taken from Roberson et al (2004, 2005). The colours are also represented in CIE L*u*v* colour space (see figure 4.1 for Dumbu-buru and refer figure 3.1 for Blue-green). The experiments were carried out in controlled lighting conditions under a lamp with Illuminant C lighting. Same lighting conditions were aimed for with the Himba; the experiments were carried out manually under a shaded tree and colours were placed into a metal box when shown to the participant. The light conditions (verified with a Gossen Luna-star light meter) were at 6000 K (brightness) and 2000 Lux (Luminance) with minor fluctuations that did not seem to affect the level of responding. The experimenter registered responses on a scoring sheet.

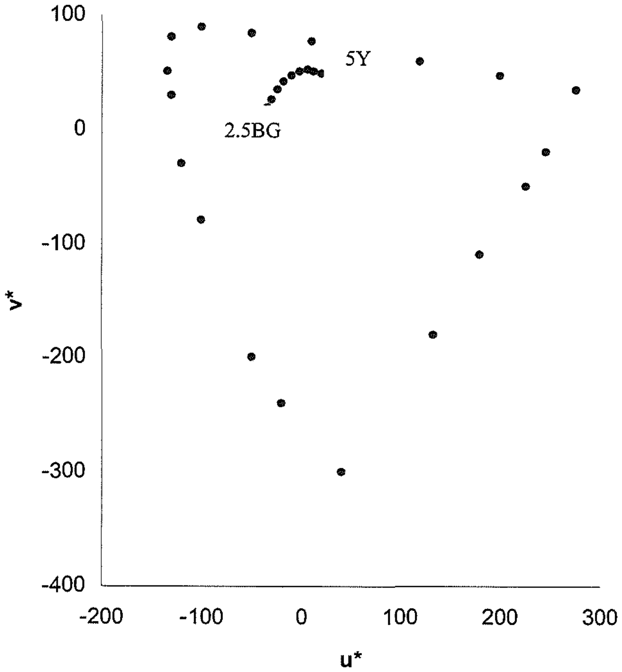


Figure 4.1 Equidistant (Saturation 5, Brightness 6) Munsell colours 5Y, 7.5Y, 10Y, 2.5GY, 5GY, 7.5GY, 10GY, 2.5G, 5G, 7.5G, 10G, 2.5BG represented in CIE L*u*v* colour space.

4.2.3 Procedure

Practice- Identity Matching Trials

For the Blue-green Set colours 2.5G and 10B, and similarly for the Dumbu-buru Set colours 5Y and 2.5BG, were used in the practice trials. They were shown individually, one at a time in random order and participants were asked to look at the colour carefully. The colour was then removed, and with an interval of a couple of seconds, it was followed by the administration of the choice comprising of the two colours, corresponding to the top and bottom colour, placed on a card, shown upright or upside down so that the presentation of the order was counterbalanced. The participant was then asked which colour he/she had seen. The participant was trained until he/she got all the practice trials right 4 times consecutively.

Test- Similarity Matching Trials + Identity Matching Trials

The probe trials were introduced and administered individually in equal random fashion to the training procedure. There were 10 alternative probes ranging from 5G to 7.5B for the blue-green Set, and 7.5Y to 10 BG for the Dumbu-buru Set. The choice given at matching-response time remained the same as in the practice trials. The similarity matching trials were intermixed with the identity matching trials for 5 experimental sessions comprising of 240 trials.

Naming

At the end of the experiment, for both sets, participants were shown the twelve stimuli, one at a time, in random order and asked, in English or Himba, for its colour name. As said previously, the colour terminology used by the Himba consisted of 5 colour terms (Roberson et al., 2005): vapa (light and pastel colours, white), serandu (red, orange and pink), zozu (dark colours and black), dumbu (beige, yellow and some light green), and buru (green and blue).

4.3 Results

4.3.1 Naming

Naming: Blue-green Probe colours

English participants used colour terms green and blue for this colour range. The mean % of colour terms used for every colour is illustrated in table 4.1A. A boundary was determined where the 2 colour terms were each used 50% of the time, and averaged across. For example in table 4.1A, for the colour green, one can see that the 50% mark would be between 7.5BG and 10BG. The boundary calculation substracted 50% from the actual value of colour term use for 7.5BG (91.7%), and the remaining figure was then divided by the difference between the 7.5BG's value and 10BG's value amounting to 66.7% difference. The value of the division was then multiplied by 2.5 to translate the number into Munsell value, which amounted to 1.56 Munsell steps. Thus the 50 % mark for using green was exactly at 7.5BG + 1.56BG i.e., 9.06BG. The same calculation procedure was used for the colour blue and numbers were averaged across for a naming boundary between green and blue. The three Colour terms buru, dumbu and zozu were used by the Himba for this colour range. The mean % of colour terms used for every colour is illustrated in table 4.1B. The colour 'Buru' was predominantly used to name most colours. As this was the only colour used at 50% or above, no reliable boundary was found.

Table 4.1A Mean % Colour Term Used for English in the naming task of green-blue colour range.

Colour Term Colour Range	**Green**	**Blue**
2.5G	100	0
5G	100	0
7.5G	100	0
10G	100	0
2.5BG	100	0
5BG	100	0
7.5BG	91.7~	8.3
10BG	25	75~
2.5G	0	100
5G	0	100
7.5G	0	100
10G	0	100

~ Boundary calculation: Green: 91.7 - 50 = 41.7; 41.7 / 66.7 = 0.625; 2.5 x 0.625 = 1.56 → 7.5BG + 1.56 BG = 9.06 BG; Blue: 75 - 50 = 25; 25 / (75 – 8.3) = 0.375; 2.5 x 0375 = 0.94 → 10BG – 0.94 BG = 9.06 BG; Average Green & Blue: 9.06 BG.

Table 4.1B Mean % Colour Term Used for Himba in the naming task of green-blue colour range.

Colour Term Colour Range	**Buru**	**Dumbu**	**Zozu**
2.5G	41.7	41.7	16.6
5G	58.3	25	16.7
7.5G	58.3	25	16.7
10G	66.6	16.7	16.7
2.5BG	66.6	16.7	16.7
5BG	50	25	25
7.5BG	58.3	16.7	25
10BG	58.3	8.4	33.3
2.5G	50	25	25
5G	58.3	16.7	25
7.5G	50	25	25
10G	66.7	8.3	25

*No reliable boundary

Naming: Dumbu-buru Probe colours

The 3 Colour terms buru, dumbu and zozu were used by the Himba for this colour range. The mean % of colour terms used for every colour is illustrated in Table 4.2A. Dumbu was predominantly used to name colours up to 7.5GY, after which buru was used most often; 7.5GY was the reliable boundary and 50% point see table 5.2A). Zozu was used very little as compared to other colour terms and was therefore not considered for boundary calculations.

English participants used colour terms brown, yellow, green and blue for this colour range. The mean % of colour terms used for every colour is illustrated in Table 4.2 B. 50% colour term use was found between 5Y and 7.5Y and also at 2.5BG, indicating that there are two boundaries for this colour range. Munsell boundary location was calculated as in tables 4.1A, 4.1B, and 4.2A and also further annotated below table 4.2B. This time a reliable boundary was found at 8Y (With terms Brown and Green) and 2.5BG (With terms Green and Blue).

Table 4.2A Mean % Colour Term Used for Himba in the naming task of dumbu-buru colour range.

Colour Term Colour Range	**Buru**	**Dumbu**	**Zozu**
5Y	8.3	83.4	8.3
7.5Y	8.3	83.4	8.3
10Y	8.3	83.4	8.3
2.5GY	8.3	83.4	8.3
5GY	33.4	58.3	8.3
7.5GY	50*	50*	0
10GY	75	8.3	16.7
2.5G	83.3	0	16.7
5G	83.3	0	16.7
7.5G	83.3	0	16.7
10G	83.3	0	16.7
2.5BG	83.3	0	16.7

* Boundary dumbu & buru : 7.5GY

Table 4.2B Mean % Colour Term Used for English in the naming task of dumbu-buru colour range.

Colour Range	**Brown**	**Yellow**	**Green**	**Blue**
5Y	50~	16.7	33.3	0
7.5Y	8.3	16.7	75~	0
10Y	8.3	0	91.7	0
2.5GY	8.3	0	91.7	0
5GY	8.3	0	91.7	0
7.5GY	0	0	100	0
10GY	0	0	100	0
2.5G	0	0	100	0
5G	0	0	75	25
7.5G	0	0	75	25
10G	0	0	58.3	42.7
2.5BG	0	0	50*	50*

* Boundary green & blue: 2.5 BG.

~ Boundary calculation: 75 –50 = 25; 25 / 41.7 = 0.6; 2.5 x 0.6 = 1.5 → 7.5Y - 1.5Y = 6Y; Average green & brown: (5Y+6Y) / 2 = 5.5Y

4.3.2 Test

Practice

All participants needed 16 identity matching trials to reach criterion for the test.

Test

A 2 (Culture: English and Himba) x 2 (Set: Blue-green and Dumbu-buru) x 10 (Probe Colours:1, 2, 3,..., 10) 3-way Anova, with repeated measures for the second and third factors, was carried out on the participants' performance (number of green or dumbu responses) in similarity matching trials. The significant effects were the 3-way interaction of Culture by Set by Probe colours (F(9,198) = 4.86, p = .007), in the context of 2-way interactions of Culture by Set (F(1,22) = 4.47, p = .046) and Probe by Set (F(9,198) = 5.36, p = .004). The Culture by Probe interaction was not significant (F(9.198) < 1). Main effects of Set (F(1,22) = 6.74, p = .016), and Probe colour (F(9,198) = 643.57, p < .001 are also reported with no main effect of Culture (F(1,22) < 1).The 3-way interaction was further explored with 2-way Anovas of Culture by Probe colour for each Set.

Blue-green Set: Post-hoc

As confirmed in table 4.3 and figure 4.2 (below), the Culture by Probe colour interaction was significant for the Blue-green Set (F(9,198) = 3.85, p < .001) in the context of a main effect of Probe colour (F(9,198) = 318.46, p < .001) and no main effect of Culture (F(1,22) = 2.12, p < .2), indicating that overall performance did not differ between Cultures for the Blue-green Set, but that overall performance did differ among Probe colours.

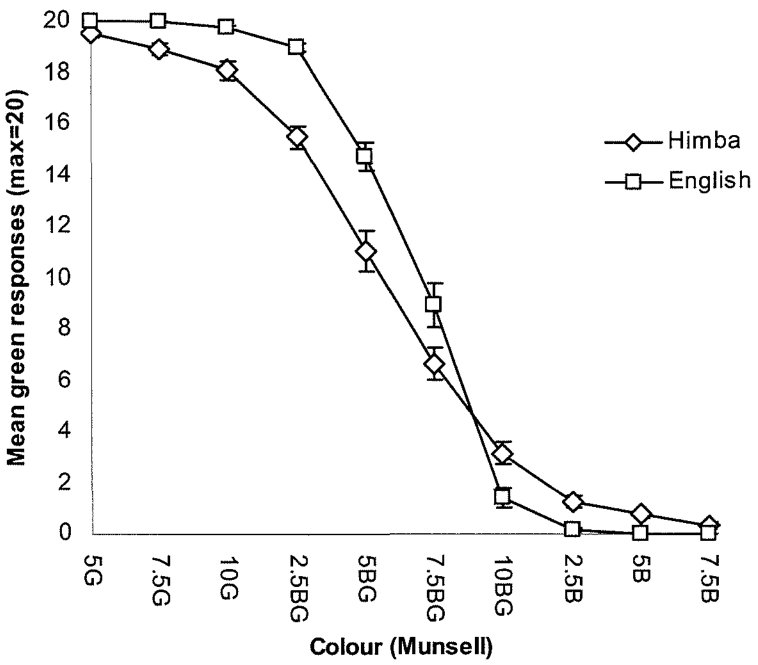
Tukey (HSD, p < .05) pair-wise comparisons were carried out to ascertain differences in treatment of Probe colours across and within Cultures. For 28 comparisons with 22 degrees of freedom the Tukey HSD value was 2.41. The comparison results show that the English gave significantly more green responses for probes 2.5BG and 5BG (and marginally so for 7.5BG). Even though Himba and English showed significantly different allocation (> 2.41) of green responses between probes 2.5BG and 5BG, 5BG and

7.5BG, and 7.5BG and 10BG with the exception of Himba showing an additional difference between probes 10G and 2.5BG, the English showed larger differences between those green response allocations on these probes indicating a sharper categorical function for more peripheral green and blue colours near the boundary. Trend analyses were carried out to assess the categorical function obtained from the Probe Colour response data. They showed a significant cubic trend by group interaction (F(1,22) = 17.2, p < .001) and upon inspection, the English show a stronger cubic trend indicative of a sharper categorical function with particular reference to colour near the boundary between green and blue. The boundary was assessed as follows: a t-test was carried out to assess the colour in the continuum where participants were at chance. The at-chance-performance for every participant was located on the continuum at 10 green responses given (from 20 as maximum performance). The difference between the English chance performances was not significantly different from the Himba chance performances (p > .05) indicating midpoints were not significantly different in location. Calculations were needed to convert chance score of 10 into Munsell steps as done before in Table 4.1A and 4.1B. For e.g., the Himba midpoint: 11.08 at 5BG, is 1.08 different from 10. 1.08 divided by the difference in score between 5 and 7.5 BG (4.41) is .244. This value is then converted into Munsell steps by multiplying it by 2.5 (= 1 Munsell setp). The end result is a score of 10 at 5BG + .61BG i.e., 5.61BG. The same was done for the English midpoint amounting to 7.03BG.

Table 4.3 Mean green responses (+/- 1 SE) for Himba and English participants for Similarity Matching trials.

Colour (Munsell)	**Himba (N = 12)**	**English (N = 12)**
5G	19.5 (0.23)	20 (0.00)
7.5G	18.92 (0.50)	20 (0.00)
10G	18.08 (0.65)	19.75 (0.18)
2.5BG	15.5 (0.81)	19 (0.35)
5BG	11.08 (1.61)	14.75 (1.14)
7.5BG	6.67 (1.36)	8.92 (1.70)
2.5B	1.25 (0.51)	0.17 (.11)
10BG	3.17 (0.85)	1.42 (0.82)
5B	0.75 (0.28)	0 (0.00)
7.5B	0.33 (0.26)	0 (0.00)

Fig. 4.2 Mean green responses (+/- 1 SE) for Himba and English participants for Similarity Matching trials.



Dumbu-buru Set: Post-hoc

As confirmed in table 4.4 and figure 4.3, the Culture by Probe colour interaction was significant for the Dumbu-buru Set (F(9,198) = 2.63, p = .01) in the context of a main effect of Probe colour (F(9,198) = 272,38, p < .001) and no main effect of Culture (F(1,22) = 3.31, p < .1), indicating that overall performance did not differ between Cultures for the Dumbu-buru Set, but that overall performance did differ among Probe colours.

Tukey (HSD, p < .05) pair-wise comparisons were carried out to ascertain differences in treatment of Probe colours across and within Cultures. For 28 comparisons with 22 degrees of freedom the Tukey HSD value was 2.67. The comparison results show that the Himba gave significantly more dumbu responses than the English for probes 2.5GY, 5GY, and 7.5GY. Even though Himba and English showed significantly different allocation (> 2.67) of dumbu responses between probes 5GY and 7.5GY, 7.5GY and 10GY, with the Himba additionally showing a response difference between 10GY and 2.5G in comparison to the marginal English difference, the Himba differences were larger than the English response, indicating a sharper categorical function for dumbu and buru colours near the boundary. Furthermore, English also showed a significant difference in response between 10Y and 2.5GY, and 2.5GY and 5GY, indicating a more gradual descending function with dumbu responses more spread across matching-to-sample trials than the Himba. Trend analyses were carried out to assess the categorical function obtained from the Probe Colour response data across Cultures. They showed a cubic trend by Culture interaction approaching significance (F(1,22) = 3.55, p = .073. From inspection, a slight cubic difference is noticeable as the Himba look to show a stronger cubic trend than the English. The Himba show a stronger cubic trend indicative of a sharper categorical function with particular reference to colours near the boundary between dumbu and buru. The boundary was assessed as follows: a t-test was carried out to assess the colour in the continuum where participants were at chance. The at-chance-performance for every participant was located on the continuum at 10 green responses given (from 20 as maximum performance). The difference between the English chance performances was not significantly different from the Himba chance performances (p >

.05). The midpoints (as calculated in table 5.2A and 5.2B) for the English was on average at 6.34GY, and the Himba midpoint was on average at 7.6GY.

Table 4.4 Mean dumbu responses (+/- 1 SE) for Himba and English participants on similarity matching trials.

Colour (Munsell)	**Himba (N = 12)**	**English (N = 12)**
7.5Y	20.00 (0.00)	20.00 (0.00)
10Y	19.58 (0.19)	19.92 (0.08)
2.5GY	19.75 (0.13)	16.58 (1.73)
5GY	17.83 (0.94)	13.58 (1.86)
7.5GY	10.25 (1.40)	7.00 (1.71)
10GY	3.75 (0.88)	2.83 (1.28)
2.5G	1.08 (0.47)	0.5 (0.20)
5G	0.25 (0.13)	0 (0.00)
7.5G	0.25 (0.13)	0 (0.00)
10G	0.08 (0.08)	0 (0.00)

Fig. 4.3 Mean dumbu responses (+/- 1 SE) for Himba and English participants on similarity matching trials.

4.4 Discussion

This experiment achieves evidence of universal human grouping at the same time as showing some differences across cultures that is continuum specific. Therefore results are also consistent with linguistic relativity (Davidoff et al., 1999; Roberson et al., 2000) and replicate conclusion drawn by Roberson et al. (2005b) from their sorting task that show linguistic influence in colour grouping with the additional value these effects have been found in a more constrained task.

The Sets used seem to affect responding depending on the Culture carrying out the task. For the blue-green continuum, compared to the Himba, the English made significantly more green responses on the "green" side of the boundary and fewer green responses on the "blue" side of the boundary (see Fig 4.2). The Himba also showed evidence of a linguistically driven difference for the dumbu-buru continuum (see Fig 4.3). Moreover, significantly different cubic functions, sharper for host cultures, support the claim that categorical functions (e.g., Maddox & Ashby, 1992) are markedly different for performance on both continua. The different cubic functions support linguistic relativity findings even further (e.g., Robertson et al, 2005a, b) under restrained task demands of matching by similarity. The origin of this general grouping propensity is both perceptual and conceptual in nature. On the one hand, it would seem that this propensity seems inevitably human as monkeys perform differently matching colours strictly to their endpoints rather than making broad perhaps abstract generalizations. On the other hand cross-cultural differences are not to be ruled out showing evidence of conceptual grouping.

A possible counter-argument may be that the oddity of the task has induced the differences found across cultures, for when cultures were exposed to unfamiliar boundaries they may have found the task more difficult. However, task difficulty is not reflected in performance levels that did not vary across the different sets and only varied across probe colours near the boundary. Thus, although there are some perceptual constraints that make the categorical behaviour of two different cultures with different colour terms look similar in matching to sample, there are differences of linguistic nature that should not be mistaken for noise. In more detail, the same boundary behaviour has been found for the host culture of both continua, at the 4th, 5th and 6th colours of the 10 probe colours, indicating there is something culturally specific about the differences found. The findings that boundaries are created from shared linguistic experience and communication seems to fit nicely with this result (Braisby & Dockrell, 1999; Steels & Belpaeme, 2005).

It is concluded that despite inherent grouping behaviour, linguistic differences are perceived across cultures for different contexts. The Himba perform more sharply in their own context and more loosely in an English termed context, and the English categorize more sharply in their own context and more loosely in the Himba termed context. This double dissociation shows that although categorizing involves inherent grouping behaviour, language seems to bring out differences to be found at the boundary between two colours.

5. Experiment 4 and 5: A developmental study of effects of colour term acquisition and a cross-cultural comparison of colour categorization in toddlers.

5.1 General introduction

Do children who have not yet acquired colour terms respond to colour categorically? If the universal perspective is adopted (Bornstein, Kessen & Weiskopf, 1976; Franklin & Davies, 2004) children at a very young age will show evidence of colour categories despite the tardy acquisition of colour terms. If language plays a role in determining colour categories (Roberson et al., 2004; Roberson, Davies & Davidoff, 2000), children are not expected to show these categories unless they have acquired colour terms. Two experiments (experiments 5 and 6) were undertaken to address the nature/nurture debate on the origin of colour categories in toddlers of 2- to 4-years-old. Experiment 5 particularly addressed colour term acquisition in English toddlers to ascertain whether the level of acquisition corresponds with the acquisition of CP. It attempts to replicate Franklin et al. (2005) where it is argued that there are panhuman colour categories that are not established by language but by biology (Franklin & Davies, 2004). A tight link between language and categorical perception would be in direct opposition to their universal approach to Categorical Perception. Experiment 6 considered the issue from a cross-lingual perspective.

Franklin et al. (2005) looked for a causal relationship between naming accuracy and Categorization effects. In their study, children showed Categorical Perception irrespective of their colour naming ability but we will argue that their assessment of colour term acquisition paid insufficient account of colour term comprehension. Franklin et al. (2005) only took comprehension into account in one part of their analyses, namely, when investigating the relationship (correlation) between colour term fluency and Categorical Perception. Moreover, their assessment used only 11 colours with just one measurement for naming and another for comprehension; this is not enough to be sure that the colour terms are understood. The present investigation rectifies some of those flaws by testing a larger sample of children and by using 22 colours to verify colour term acquisition. The meaning of colour terms is exoplored further by asking additional questions as to whether other samples would also be given the same colour name. Under these circumstances, one predicts that CP could apparently be achieved following Franklin et al.'s (2005) criteria, but that when stricter criteria are applied Categorical Perception will only be found for children that have properly acquired their colour terms needed at test (i.e., blue, green and purple).

5.2 Experiment 4: A developmental study of effects of colour term acquisition with English toddlers

5.2.1 Method

5.2.1.1 Participants

A total of 60 English-speaking children (37 girls and 23 boys) between 31 and 51 months of age (mean age = 42.24 months) were recruited from 4 different nurseries and a music group for toddlers in London, UK (Parkway Under 5s, Lingfield Nursery, Woodlands Nursery, Powerhall Infants School, and Hampstead Toddler Music Group). None of the children tested were colour blind (as ascertained by information given by the school).

5.2.1.2 Stimuli

Two identical bear cut-out figures (Bear A and Bear B) were used. The coloured stimuli in the form of sweaters were placed over the bear figures. All sweater cut-outs were mounted on cardboard and coloured either in ColourAid paper or Munsell coloured paper. The 22 colours used for naming and comprehension tasks (see Roberson et al., 2004) were made from Colour-Aid paper. These colours included the 11 basic colours used by Franklin et al. (2005) with an additional set of 11 non-focal exemplars as in Roberson et al. (2004), chosen to be intermediate between each of the chromatic categories (e.g., half way between pink and orange), given the Munsell arrangement of colour in a spherical three-dimensional space. The focal colours were (with Colour-Aid codes in parentheses): black (BLACK), white (WHITE), red (RO Hue), green (G Hue), yellow (Y Hue), blue (B Hue), brown (O S3), pink (R T4), purple (V Hue), orange (YO

Hue), and gray (GRAY 4). The non-focal colours were (with Colour-Aid codes in parentheses): blue (BV HUE), pink-purple (RVR HUE), yellow (YOY HUE), orange (O HUE), green (GBG HUE), pink-red (ORO T2), green (YGY T1), orange (OYO T4), orange-yellow (YOY T3), purple (RVR T3), pink (ROR T2). All coloured sweaters were designated with a number code in the back. Table 5.1 gives the designations and CIE L*u*v* co-ordinates for each colour.

Table 5.1 Designations of the 22 Color Aid sweaters in the CIE Lu*v* metric under Illuminant C as measured by Roberson et al (2005)

Tile code	*Color-Aid*	*Name*	*L*	*u**	*v**
1	Gray 4	Grey	71	1	1
4	White	White	100	0	0
6	O S3	Brown	44	13	8
23	YO HUE	Orange	82	101	62
28	V HUE	Purple	40	21	-31
29	BV HUE	Blue	36	-2	-23
33	RVR HUE	Pink-purple	57	113	1
42	Black	Black	35	5	2
45	Y HUE	Yellow	90	54	80
46	YOY HUE	Yellow	92	83	78
47	O HUE	Orange	75	146	44
49	RO HUE	Red	59	136	28
50	B HUE	Blue	49	-30	-69
53	GBG HUE	Green	51	-46	-14
54	G HUE	Green	57	-55	4
59	R T4	Pink	89	70	12
77	ORO T2	Pink-red	70	113	28
78	YGY T1	Green	74	-26	60
79	OYO T4	Orange	83	100	49
80	YOY T3	Orange-Yellow	87	78	57
81	RVR T3	Purple	63	81	-13

82	ROR T2	Pink	70	108	15

The test stimuli sweaters for the 2AFC task were made from Munsell Glossy paper. There were two sets of test stimuli, blue–green and blue–purple, with three stimuli per set. The Munsell notation of the stimuli for the blue-green set were 7.5G, 5BG, and 2.5B, all with lightness value 5 and saturation value 10. The stimuli for the blue-purple set were 10B, 7.5PB, and 5P, all with lightness value 3 and saturation value 10. The test stimuli are referred, as in Franklin et al. (2005) as A1, A2, and B, A1 and A2 belonging to one category and B being of another category. All test stimuli varied only in Munsell hue; Munsell Chroma (colourfulness, like saturation) and Munsell Value (lightness) were kept constant.

For each set, there was a within-category pair and a between-category pair, and the separations of within- and between-category pairs were equated in Munsell units. There were 3 Munsell steps between stimuli in a pair. Although the nature of the human colour space upon which the Munsell set of colours is based is non-Euclidian and completely uniform colour differences are impossible to represent in a three dimensional colour model (MacEvoy, 2005), Franklin et al. (2004) made an attempt to also equate colour differences in the CIE Lu*v* colour space. It was successfully found (see figure 6.2) that test foils were equally separated in colour space units (ΔE) lending further support to the equidistance between colours suggested by Munsell (1905) and other 3-dimensional colour representation that are found by some to be the best at approximating human colour space (Indow, 1988). The location of the category boundaries was taken from Franklin and Davies' (2004) adult naming and similarity judgments, as was done in Franklin et al. (2005). As presented by Franklin et al. (2005), the Munsell codes and categorical status of stimulus pairs are shown in figure 5.1 A and B. (Blue–green and blue–purple separations were around 30 units in CIE colour space (30 ΔE)).

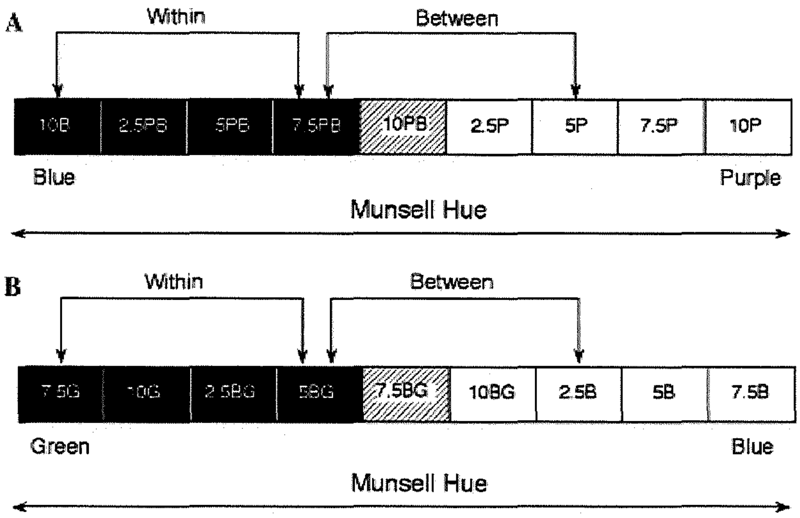


Figure 5.1 Munsell codes, categorical status, and Munsell distances of the stimuli of the category pairs used in Experiment 5 as in Franklin et al. (2005). The categorical relationships (within or between) of the category pairs are shown for blue–purple (A) and blue–green (B). In panel A, Chroma = 3 and Value = 10. In panel B, Chroma = 5 and Value = 10. (Taken from Franklin et al. (2005))

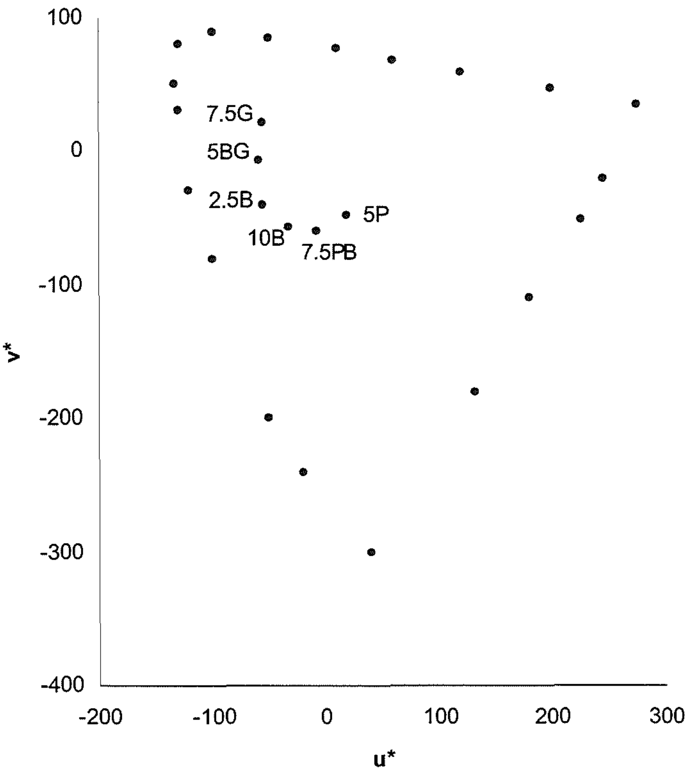


Figure 5.2 Munsell colours of figure 1 represented in CIE Lu*v* colour space. Same values used in Franklin et al. (2005).

5.2.1.3 Procedure

Overall Design

The design was identical to Franklin et al. (2005) except that test stimuli were reduced to two sets; blue-purple and blue-green sets were kept and the red-pink set was omitted. All children completed the training, naming, and comprehension task using the 22 coloured sweaters. For the 2AFC task, children were tested on both sets with a counterbalancing of task order; half the children got the blue-purple set first, and the other half got the blue-green set first. For each continuum, eight category-pairs were randomly presented until the four between- and four within-category pairs were exhausted. The 2-AFC task was always presented after naming and comprehension, and the naming task and comprehension task were presented in a randomized order.

Comprehension and naming of focals and non-focals

The 22 focal and non-focal sweaters were laid out in front of the child and the child was asked to point at a specific colour. The colour of the chosen sweater was recorded, with additionally asking the child if there were any other sweaters that corresponded to this specific colour (e.g., 'Are there any other red ones?'). The additional answers were recorded until the child said there were no more sweaters of that colour. This comprehension procedure was carried out for the 11 focal colour names. Each of the 22 sweaters was then presented individually and in random order, and the child was asked to name the colour of the sweater.

Training

Training was carried out as in Franklin et al. (2005). The aim of the training session was to show each child that when Bear A wore a particular sweater, Bear B also wore the identically coloured sweater. Bearing in mind that the meaning of 'same' and 'different' are not properly understood by 2-year-olds (Glucksberg, Hay & Danks, 1976), the term 'different' was not used in the procedure at all, and the training phase ensured that the child understood the meaning of the word 'same' as follows: The two bears were placed flat on the table in front of the child, who was encouraged to give each bear a name. The child was told that Bear A has lots of coloured sweaters, and the 11 focal sweaters were randomly laid out above Bear A. The child was then told that Bear B also has lots of coloured sweaters and another set of 11 focal sweaters were randomly laid out above Bear B. A sweater from Bear A's set was randomly chosen and placed on Bear A. The child was told that if Bear A wears this sweater, then Bear B also wears this sweater, and so the corresponding sweater was picked out of Bear B's set and placed on Bear B. This was repeated three times with different coloured sweaters. Another sweater was chosen and placed on Bear A but this time Bear B's set of sweaters was covered with white cardboard. The child was allowed to inspect the sweater, and after 5 s of stimulus presentation, Bear A and its sweater were covered. After a further 5-s delay, Bear B's set of sweaters were uncovered and the child was asked to find the same sweater for Bear B out of the set of 11 sweaters. After the child had made a choice, Bear A was uncovered and the child was encouraged to evaluate his or her response. If the choice was correct, then the child was praised; if the choice was incorrect, then the child was encouraged to amend his or her choice. This was repeated until a correct response was given three times and the experimenter was sure that the task was understood by the child. All of the children passed this training.

Two-alternative forced-choice task

Test stimuli sweaters were used in a 2-AFC task under standardized lighting conditions that simulate natural daylight with the use of a Gretag Macbeth lamp (D65, 6500 K, at 810–1880 lux). These lighting conditions are necessary for the uniformity of the Munsell system to be maintained (Davies & Franklin, 2002). The procedure and goal of finding the matching sweater for the other bear was the same as in the training task except that the child was given a choice of only two sweaters: an incorrect choice (foil) and a correct choice (target) identical to Bear A's sweater. The categorical relationship (between category / within category) of the incorrect and correct choices was manipulated (for stimulus pairs see panels A and B in Fig. 5.1). The procedure was conducted for both within- and between-category stimulus pairs four times each. For two of these four trials, one stimulus was the target and the other stimulus was the foil; for the remaining two trials, this target/foil allocation was reversed. Therefore, the child made a total of eight judgments, and the order was randomized for each child. The child was then presented with each of the stimuli individually and was asked to name the colour of the sweater.

5.2.2 Results

Four analyses were carried out on the English toddler's data; three analyses identical to Franklin et al.'s (2005) and one additional analysis (analysis 4). The first analysis considered, for all children, category effects for the blue-purple and blue-green continua in the 2 alternative-forced-choice task. The second analysis addressed effects of naming ability on the category effect in the first analysis. The third analysis addressed colour term fluency in relation to the categorical effects of the first analysis. The fourth (additional) analysis is a variation of the first analysis but assigns children to whether they know or do not know colour terms based on further probing of their comprehension of colour terms.

Analysis 1: Showing CP for blue-purple and blue-green

For each child, the number of correct identifications in the 2-AFC task was calculated when the choice was between two stimuli of the same category (within-category pairs) and when the choice was between stimuli of different categories (between-category pairs). The maximum number of correct identifications was four for within-category pairs and four for between-category pairs. Tables 5.2 and figure 5.3 give the accuracy scores for within and between pairs for the blue-purple and blue-green set. The between-category accuracy was higher than within-category accuracy for all conditions.

Table 5.2 Mean accuracy (±1 SE) for within- and between-category pairs for the blue-purple and blue-green sets.

Category Set	Within	Between	*Total*
Blue-purple	2.78 (.152)	3.43 (.105)	3.11 (.129)
Blue-green	2.58 (.123)	3.45 (.091)	3.02 (.107)
Total	2.68 (.138)	3.44 (.098)	3.07 (.118)

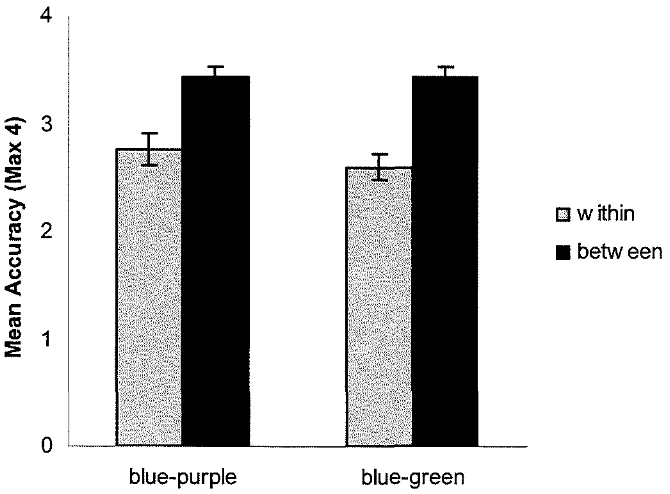


Figure 5.3 Mean accuracy (±1 SE) for within- and between-category pairs for the blue-purple and blue-green sets.

The results were supported by an analysis of variance (ANOVA) looking at the effects of Category (Between or Within), Set (Blue-purple or Blue-green), and Task Order (Group 1 (Blue-purple set first) or Group 2 (Blue-green set first)). Category and Set were repeated measures. The only significant effect was the main effect of Category. Accuracy was greater for Between-category pairs than for Within-category pairs, $F(1,58) = 59.48$, $p < .001$. Overall accuracy did not vary across Set, $F(1,58) < 1$ or Task Order, $F(1,58) < 1$. Moreover, Order did not vary by Set, $F(1,58) < 1$, or Category, $F(1,58) < 1$. Finally, the Category by Set interaction was not significant, $F(1,58) = 1.04$, $p > .3$, nor was the Category by Set by Order interaction, $F(1,58) < 1$. Accuracy was above chance (2/4) for both Sets; Blue-purple set, $t(59) = 9.94$, $p < .001$ and Blue-green set, $t(59) = 13.21$, $p < .001$, and a Bonferroni post-hoc test did not reveal a significant difference in accuracy between the Blue-green set and Blue-purple set, $t(59) < 1$. Therefore, as in Franklin et al.'s (2004), Categorical Perception was demonstrated equally for both Sets, regardless of Task Order.

Analysis 2: Effect of naming accuracy on the size of the category effect

Table 5.3 gives the naming frequencies for each of the two alternative-forced-choice stimuli for all three sets. Overall, the majority name is between 60%-85% in agreement with the intended linguistic boundary for the blue-green and blue-purple sets. Out of the three stimuli, A2 seems to be most ambiguous for both sets, probably because it is closer to the linguistic boundary than A1 and B. These dips are also perceived in Franklin et al.'s table 3, albeit to a lesser extent.

Table 5.3 Frequencies (percentages) of the colour terms offered for the stimuli of each set

Set	**Stimulus Type: Munsell code**	**Colour term**	**Percentage of children offering the term**
Blue-Purple	A1: 10B 3/10	***Blue***	83.6
		Purple	6.8
		Green	4.9
		Red	4.9
	A2: 7.5PB 3/10	***Blue***	59
		Purple	31.2
		Green	6.6
		Red	1.6
		Don't know	1.6
	B: 5P3/10	***Purple***	83.6
		Blue	6.6
		Red	3.3
		Yellow	3.3
		Green	1.6
		Pink	1.6
Green-Blue	A1: 7.5G 5/10	***Green***	85.3
		Blue	3.3
		Red	3.3
		Yellow	3.3
		Purple	1.6
		Pink	1.6
		Orange	1.6
	A2: 5BG 5/10	***Green***	65.6
		Blue	21.3
		Red	6.6
		Don't know	3.3
		Purple	1.6
		Yellow	1.6
	B: 2.5B 5/10	***Blue***	85.2
		Green	6.6
		Pink	3.3
		Red	3.3
		Yellow	1.6

As in Franklin et al. (2005), individuals' naming patterns were analyzed to investigate the effect of linguistic categorization on the category effect. If individuals gave A1 and A2 the same term and gave B a different term, then they were classified as having a between-category linguistic boundary (boundary). An ''I don't know'' response was also counted as a term. Therefore, two patterns of naming qualified: (A1: name 1; A2: name 1; B: name 2) and (A1: name 1; A2: name 1; B: ''I don't know''). If individuals gave A2 and B the same term and gave A1 a different term, then they were classified as having a within-category linguistic boundary or reverse boundary. Again, an ''I don't know'' response was counted as a term. Therefore, two patterns of naming qualified: (A1: name 2; A2: name 1; B: name 1) and (A1:''I don't know''; A2: name 1; B: name 1). If individuals gave A1, A2, and B the same term, they were classified as having no linguistic boundary. Again, ''I don't know'' was counted as a term. Therefore, two patterns of naming qualified: (A1: name 1; A2: name 1; B: name 1) and (A1: ''I don't know''; A2: ''I don't know''; B: ''I don't know''). In Franklin et al. (2004), naming accuracy for the different sets was combined because the sample sizes were not sufficient to look at the effect of language for each boundary separately. No further details were given by Franklin et al. (2004) about how this general naming pattern was inferred across the three data sets and there is a concern about those children that could name some colours correctly but not all of them. For example, if a child had a linguistic boundary for blue/green, a reverse linguistic boundary for blue/purple, and no boundary for pink/red, what naming pattern would this child qualify for if collapsing naming accuracy across all sets? In the present study, the matter was resolved by a sample size large enough to look at effects of language for each boundary separately. Thus, participants were grouped into a name boundary, reverse boundary, or no boundary group as in Franklin et al. (2005), although crucially for separate sets.

For the blue-purple set, 30 children had a name boundary, 20 had a reverse boundary and 10 had no name boundary. For the blue-green set, 38 children had a name boundary, 8 had a reverse boundary and 14 had no name boundary. Tables 5.4 and 5.5, and Figures 5.4 and 5.5 show within- and between-category accuracy for the three name boundary groups per set.

Table 5.4 Mean accuracy (±1 SE) of Name Boundary for the Blue-purple set on Within- and Between-category pairs.

Category **Name Boundary**	Within	Between	*Total*
Name boundary	3.13 (.22)	3.53 (.162)	3.33 (.156)
Reverse boundary	2.9 (.27)	3.05 (.198)	2.98 (.191)
No boundary	2.2 (.381)	2.6 (.28)	2.4 (.271)
Total	2.74 (.172)	3.06 (.127)	2.9 (.15)

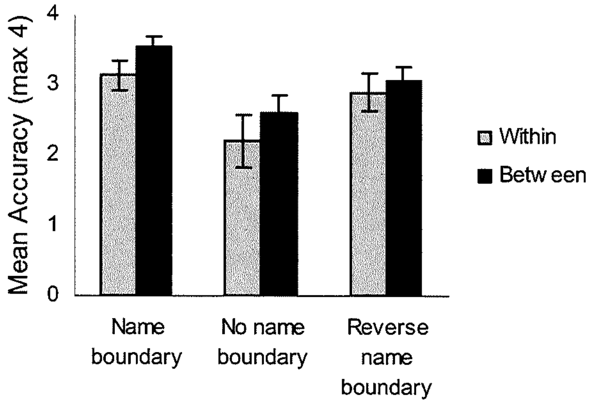


Figure 5.4 Mean accuracy (±1 SE) of Name Boundary for the Blue-purple set on Within- and Between-category pairs.

Table 5.5 Mean accuracy (±1 SE) of Name Boundary for the Blue-green set on Within- and Between-category pairs.

Category **Name Boundary**	Within	Between	*Total*
Name boundary	2.5 (.167)	3.47 (.130)	2.99 (.107)
Reverse boundary	2.38 (.363)	3.5 (.282)	2.93 (.233)
No boundary	2.71 (.275)	2.71 (.213)	2.71 (.176)
Total	2.53 (.162)	3.23 (.126)	2.88 (.144)

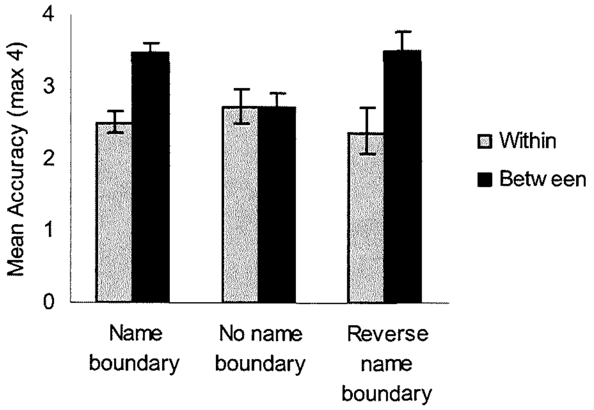


Figure 5.5 Mean accuracy (±1 SE) of Name Boundary for the Blue-green set on Within- and Between-category pairs.

The results are supported by 2 ANOVAs (for the 2 different sets) on accuracy scores with Category (Within or Between), and Name Boundary (Boundary, No Boundary or Reverse Boundary) as factors. Category was a repeated measures factor.

For the blue-purple set, only one effect was significant; accuracy varied with Name Boundary, F(2,57) = 4.59, p = .014 with post-hoc tests showing that the only significant difference in performance was between the Name Boundary and No Boundary groups (t(38) = 2.85, p = .007). A trend was also found for the main effect of Category, F(1,57) = 3.18, p = .08 with increased performance on Between-Category trials. The Category by Name Boundary effect was not significant, F(2,57) < 1.

For the blue-green set, the main effect of Category, F(1,57) = 12.01, p < .001, and the Category by Name Boundary interaction, F(2,57) = 3.31, p = .044, were significant.

Accuracy did not vary with Name Boundary, F(2,57) < 1. Bonferroni post-hoc tests revealed that the difference between Within- and Between-category pairs was only significant for the Boundary group, t(37) = 4.68, p < .001. The Reverse Boundary group just escaped a significant difference, t(7) = 2.55, p = .038, and the No Boundary group showed no significant difference, t(13) < 1.

It would seem that linguistic terminology does play a role for the blue-green set, and therefore the data are in contrast to Franklin et al.'s (2005) results of equal CP for all groups. The data for blue-purple are more in keeping with Franklin et al. (2004) though it is noted that there is only a trend towards an effect of Category.

Analysis 3: Colour term fluency and Categorical Perception

Franklin et al. (2004) analyzed colour term acquisition in children that were 2- to 4-years-old. These children were split into 4 age groups covering 6 months each (2-2.5, 2.5-3, 3-3.5, 3.5-4 years). Participants' age in our study spanned from 2.5 to 4.5 years. Colour term acquisition was also analyzed at four stages with the removal of the age group 2-2.5 years and with the additional group of 4-4.5 years. Thus, the groups amounted to 2.5 to 3 years, 3 and 3.5 years, 3.5 and 4 years, and 4 and 4.5 years of age. Overall, 12 children were between 2.5 and 3, 23 children between 3 and 3.5, 17 children between 3.5 and 4, and 8 children between 4 and 4.5 years of age. Tables 5.6 and 5.7 give the percentages of correct responses for each focal colour.

Table 5.6 Percentages correctly named focals in the naming task for each age band

Age **Colour**	2.5-3 years (n=12)	3-3.5 years (n=23)	3.5-4 years (n=17)	4-4.5 years (n=8)	*Total* (N=60)
Black	58.3%	60.9%	88.2%	88.9%	74.08%
White	33.3%	82.6%	88.2%	88.9%	73.25%
Red	75%	69.5%	94.1%	88.9%	81.88%
Green	50%	78.3%	94.1%	88.9%	77.83%
Yellow	75%	82.6%	88.2%	100%	86.45%
Blue	58.3%	78.3%	100%	100%	84.15%
Orange	58.3%	60.8%	88.2%	77.7%	71.25%
Pink	75%	73.9%	88.2%	88.9%	81.50%
Purple	58.3%	78.3%	82.3%	88.9%	76.95%
Brown	9.1%	43.4%	64.7%	66.6%	45.95%
Gray	33.3%	26.1%	41.2%	55.6%	39.05%
Total	53.08%	67.58%	83.40%	84.85%	72.42%

Table 5.7 Percentages correctly identified focals in the comprehension task for each age band using Franklin et al's criteria.

Age **Colour**	2.5-3 years (n=12)	3-3.5 years (n=23)	3.5-4 years (n=17)	4-4.5 years (n=8)	*Total* (N=60)
Black	66.7%	78.3%	88.2%	88.9%	80.53%
White	75%	82.6%	76.4%	77.8%	77.95%
Red	66.7%	73.9%	64.7%	77.8%	70.78%
Green	58.3%	65.2%	76.5%	66.7%	66.68%
Yellow	66.7%	78.3%	94.1%	77.8%	79.23%
Blue	58.3%	78.3%	94.1%	77.8%	77.13%
Orange	66.6%	60.9%	82.3%	55.6%	66.35%
Pink	75%	65.2%	88.2%	77.8%	76.55%
Purple	58.3%	73.9%	76.5%	77.8%	71.63%
Brown	41.7%	47.8%	58.8%	55.5%	50.95%
Gray	25%	69.6%	64.7%	66.6%	56.48%
Total	59.85%	70.36%	78.59%	72.74%	70.38%

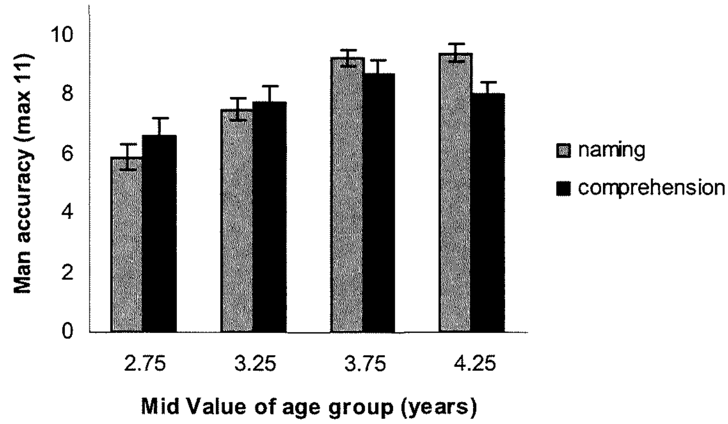


Figure 5.6 Mean numbers correct on naming and comprehension tasks for each age band. Bars represent ±1 SE.

Children's naming and comprehension of the 11 focal colours improved from a mean of approximately 6 terms to a mean of approximately 9.5 terms (see figure 5.6). This is slightly different from Franklin et al.'s (2005) colour term acquisition that improved from a mean of approximately 6 to 11 terms. Although, using similar naming and comprehension tasks, some researchers have found that toddlers acquire names faster than they comprehend colour terms (e.g., Soya, 1994). As in Franklin et al. (2004), there was no difference here between naming and comprehension, t(10) < 1 (mean naming = 72.42%; mean comprehension = 70.38%).

The relationship between colour term fluency (colours named/colours comprehended) and the extent of the category effect was explored as in Franklin et al. (2005). Two indices were calculated to this effect; a colour term fluency index, averaging the mean number of colours named and the mean number of colours identified (as in Franklin et al., 2004), and a categorical effect index, subtracting the within-category score from the between-category score (on the two alternative-forced-choice task) for each child. A score higher than 0 indicates a categorical effect; that is, between-category accuracy is greater than within-category accuracy. A score of 0 indicates no categorical effect; that is, within-category accuracy and between-category accuracy are equal. A score lower than 0 indicates a reversed categorical effect; that is, within-category accuracy is greater than between-category accuracy.

For the blue-purple set, looking at scores obtained from 60 children, the relation between the colour term fluency and categorical effect indices was not significant, r = -.13, p > .2. For the blue-green set, the relation between the colour term fluency and categorical effect indices was also not significant, r = .12, p > .3. When looking at naming and comprehension from this perspective, it seems that there is no effect of the acquisition of colour terms on obtaining categorical effects. This suggests that showing CP for blue, purple and green does not rely on how many other colours have been acquired (whether 6 or above 6 terms), and that perhaps a majority of children who have acquired as few as 6 terms, know all three colours (blue, purple and green) in our task. Indeed, Pitchford and Mullen (2005) found that purple (a secondary colour) is acquired at the same time as primary colours.

Franklin et al. (2005) found a small significant negative correlation signifying that the acquisition of more colour terms may have adverse effects on CP. Their result seems to indicate that CP will deteriorate with the acquisition of additional terms and will improve with the acquisition of fewer terms. This result seems counterintuitive and, indeed, Franklin et al. give no explanation of the negative correlation. Franklin et al. imply that a relativist account would predict a positive correlation between the extent of Categorical Perception for one colour and the number of other colour terms known by the child.

However, there is no reason to make that assumption. When children truly *know* a colour term this only implies that they do not confuse it with other nearby terms and not whether the children know fewer or more colours. Thus, there is no reason to believe that acquiring fewer secondary terms (e.g., not knowing orange and pink) would make a significant difference for Categorical Perception, as is suggested by Franklin et al. (2005). The present results show no relation between colour term fluency and categorization effects and point out the need for a different kind of test of the acquisition of terms for Categorical Perception effects to be observed. This point is addressed in Analysis 4.

Analysis 4: Categorical Perception with a revised assessment of the understanding of colour terms

Toddlers were assigned as knowing colours following Soja's procedure (1994), previously used with English and Himba toddlers in a study by Roberson et al. (2004). In Franklin et al. (2005), whether a child has a boundary is decided on the basis of naming test foils (Munsell colours used at test). Here we use a 22 coloured array to assess colour term acquisition with a comprehension measure that subsequently asks to point to any other colours that would be given the same name. This comprehension measure is also different from the one used to assess comprehension in colour term acquisition by Franklin et al. (2005); they used pointing to 11 focal colours. Thus, using our criteria before the test, children are classified as not knowing a colour name if they point to a focal colour incorrectly more than once and also if they name it incorrectly.

There is evidence in table 5.8 that children classified as in Franklin et al. (2005) as knowing colour terms did not properly comprehend them. Take the blue/green Linguistic Boundary group, 38 children qualified as in Franklin et al. (2005) but only 27 of them showed full naming and comprehension, hence leaving 11 children with an incomplete understanding of blue and green. However, the question of allocation to groups is of equal concern for the no boundary group where some children clearly had comprehension of colours. Of the 10 children classified in the No Boundary Group for Blue/purple, 4 knew the colours and of the 14 children classified to the No Boundary Group for blue/green, again 4 knew the colour terms. Thus, there is reason to believe that children have been mistakenly allocated as having or not having a linguistic boundary. The next analysis tries to get a more complete picture of colour term acquisition and its effects on Categorical Perception.

Table 5.8 Number of children who were allocated to Franklin et al.'s (2005) Linguistic Boundary, Reverse Linguistic Boundary and No Boundary groups according to naming and knew either blue and purple or blue and green according to prior naming and comprehension tests.

Set	Boundary Group	Franklin et al. (2005) Naming at test	Naming/Comprehension *(Children that knew…)*
Blue-purple	Linguistic Boundary	30	23
	Reverse Linguistic	20	10
	No Boundary	10	4
Blue-green	Linguistic Boundary	38	27
	Reverse Linguistic	8	3
	No Boundary	14	4

Following the naming and comprehension criteria in the current study, children who knew their colour terms were defined as those who named blue and purple or/and blue and green correctly (one measure for each colour), and pointed at these colours in the comprehension test no more than once. Children, who did not know their colour terms were defined as those who did not give the correct names for focal colours blue and green or blue and purple or pointed to those on more than one occasion (when probed for other colours). In sum, there were thirty children who knew blue, purple and green, and twelve children who did not know all these terms. Additionally, there were seven children who knew blue and purple and 2 children who only knew green. Four children who knew blue and green and 1 who only knew purple. The remaining 9 children for the blue-purple condition did not qualify because they either knew blue or purple. The remaining 13 children that did not qualify for the blue-green condition either knew blue or green. These children were excluded from the analysis because no definite conclusion can be drawn about their performance based on their colour understanding. Table 5.9 shows the allocation of children based on their naming and comprehension data.

Table 5.9 Number of children that qualified for full, partial or no colour term knowledge.

Colour term knowledge	**Number of children (N=60)**
Blue, green, purple	30
Blue, green	4
Blue, purple	7
Green, purple	1
Blue	3
Green	2
Purple	1
None of the three	12

Table 5.10A Mean accuracy (±1 SE) of Linguistic Terminology for Within- and Between-category pairs for the blue-purple set

Category Linguistic Terminology	Within	Between	*Linguistic Terminology*
Know BG	2.84 (.176)	3.60 (.127)	3.21 (.126)
Don't Know BG	2.71 (.285)	2.78 (.207)	2.75 (.204)
Category	2.77 (.168)	3.19 (.121)	

Table 5.10B. Mean accuracy (±1 SE) of Linguistic Terminology for Within- and Between-category pairs for the blue-green set

Category Linguistic Terminology	Within	Between	*Linguistic Terminology*
Know BG	2.56 (.160)	3.44 (.128)	3.00 (.115)
Don't Know BG	2.69 (.259)	2.69 (.207)	2.69 (.186)
Category	2.62 (.152)	3.07 (.122)	

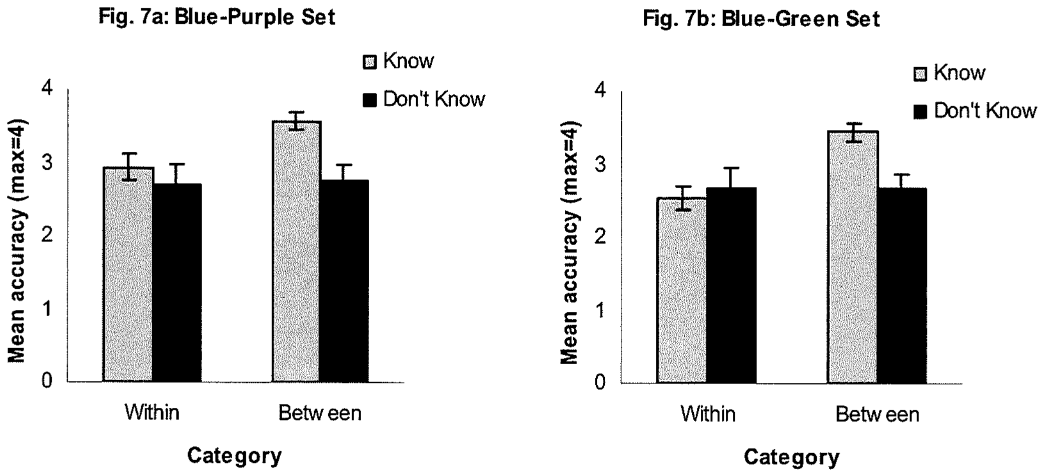


Figure 5.7A and 5.7B: Mean accuracy (±1 SE) of Linguistic Terminology for Within- and Between-category pairs for blue-purple and blue-green sets

When comparing the Know group and Don't Know group for both sets, the Know group showed a category effect in the predicted direction, whereas the Don't Know group did not show CP.

Tables 5.10A and 5.10B and figures 5.7A and 5.7B illustrate this finding that was supported by 2 two-way ANOVAs (separate analysis per set) with Category (Between or Within), as repeated measures, and Linguistic Terminology (Know, Don't Know) as factors. For the blue-purple set, two significant effects were found. The main effect of Category (F(1,49) = 6.12, p = .017) was significant with better performance on Between-category trials. The interaction was also significant (F(1,49) = 4.19, p = .046), showing CP only for children who Know their terminology (t(36) = 4.13, p < .001; Don't Know: t(13) < 1). The main effect of Linguistic Terminology was not significant (F(1,49) = 3.77, p < .1) showing no significant difference on overall performance between both groups. For the blue-green set, the same two effects were significant; there was a main effect of Category (F(1,45) = 6.93, p = .012) with better performance on between-category trials and the interaction (F(1,45) = 6.93, .012) with Categorical Perception only for the children who Know their terminology (t(33) = 4.96, p < .001; Don't Know: t(12) = 0). The main effect of linguistic terminology was not significant (F(1,45) = 1.97, p < .2) showing no difference in overall performance between both groups. These results critically indicate that when colour term knowledge is taken into account using this procedure for both sets, Categorical Perception effects are only found for children who Know their colour terms.

5.2.3 Discussion

Four sets of analyses were carried out in the replication of Franklin et al. (2005). The first analysis focused on Categorical Perception effects regardless of age and colour term acquisition as in Franklin et al. (2005); their result was repeated here with sixty toddlers who showed Categorical Perception effects across both the blue-purple and blue-green boundaries. The second analysis that looked at naming measures taken for the Munsell colours used at test were only somewhat in agreement with Franklin et al.'s (2005) Categorical Perception effects. The data did not fully replicate findings especially for the blue-green continuum where the only group that consistently showed Categorical Perception for the blue-green continua were the toddlers that showed a boundary through correct naming of test foils. The third analysis addressed colour term acquisition by including comprehension measures, with a range that covered slightly older children than in Franklin et al. (2005) between the ages of 2.5 and 4.5. It was found that children of that age acquire 6 to the 9.5 basic colour terms across that age span. This finding did not exactly match Franklin et al.'s (2005) findings as they found children at the age of 4 had acquired all 11 basic colour terms. Nevertheless, Franklin et al. (2005) found that colour term fluency was negatively related to Categorical Perception which seems odd when taking the first analysis into account where overall effects of categorization were found regardless of colour term acquisition. From results of the first analysis, one would expect fluency of colour terms not to matter. Moreover, from a universalist's point of view there should be no such relation because colour categories are expected to be innate and Categorical Perception a by-product of ontogeny. From a relativist's point of view the relation of interest is the precise knowledge of colour terms needed at test. Thus, the effects of knowing other terms than blue, green and purple are not of primary interest, because knowing more terms than those that were required (such as orange or pink) would certainly not deteriorate effects of Categorical Perception. This prediction was further supported in the present study's fourth analysis where effect of colour term acquisition from this point of view (whether a child knows blue and green or/and blue and purple) was assessed. It was found that knowledge of those particular colour terms is most important to obtain colour categories.

5.3 Experiment 5: A cross-cultural comparison of colour categorization in toddlers.

Experiment 4 indicated a categorical effect only for those English children who know their colour terms. Experiment 5 will assess whether this finding can be reinforced by studying Himba toddlers that are unlikely to have acquired their own colour terms (Androulaki, 2003, in Franklin et al., 2004). Franklin et al. (2004) found Categorical Perception for Himba toddlers who, whether or not they have their own colour terms, clearly would not have had the English terms blue and purple used in their study. Their result fits with a nativist rather than a relativist account. However, the decision to use blue-purple as the only continuum seems rather an odd choice. So, in addition to replicating their study we also carried out the test with a continuum between primary colours (green-blue). Testing took place on two different field trips.

5.3.1 Method

5.3.1.1 Participants

As closely as allowed, the first Himba group of 36 toddlers (21 boys and 15 girls ranged between 2 and 4 1/2 years of age), resulting from the first field trip, and the second Himba group of 24 toddlers (10 boys and 14 girls also ranging between 2 and 4.5 years of age) resulting from the second field trip, were matched in age to the group of 60 English toddlers from Experiment 6. As in Franklin et al. (2005), the age of the Himba toddlers had to be estimated as the Himba do not keep track of their age in the same way Westerners do. Ear touching (Gabriel, 2000), physical height and any other information the parents could give us (e.g., age in relation to sibling) were used to estimate the age of a child.

5.3.1.2 Stimuli, design, and procedure

Stimuli, design and procedure followed experiment 4 very closely, although bears were replaced by boy-cut-outs as in Franklin et al. (2005). Naming and comprehension were not collected by Franklin et al. (2005) because from an initial check the children were found not to have colour terms. The same applies to the present data. A comprehensive account of colour term acquisition is given in Roberson et al.'s (2004) study with Himba children, but time constraints made it impossible to replicate that with the present sample. The same blue-purple set was chosen to replicate the Franklin et al. (2005) findings but, in addition, the blue-green set from experiment 4 was also used in this within-subjects design. As in Franklin et al., testing was in the shade of a tree, and not in direct sunlight for the first Himba toddler group. With the second Himba group, a light box was used to control for any light variations that may have affected the appearance of the coloured shirts. A Namibian translator helped with instructions, and mothers sometimes came to sit with their children but were refrained from commenting and helping the child, so that every child was tested individually, and free from distraction.

5.3.2 Results

As with the English children, for each child, the number of correct identifications on the two-Alternative Forced-Choice task was calculated when the choice was between two stimuli from the same category (within-category pairs) and when the choice was between stimuli from different categories (between-category pairs). The maximum number of correct identifications was four for within-category pairs and four for between-category pairs. The Himba children carried out the task for both color sets balanced for order. In the critical analyses of Category, it will be necessary to exclude children who are at floor or ceiling. To make sure that order effects will be unimportant for that analysis, the accuracy data were first analysed summing over the two sets in a 2: Task Order (blue-purple set first vs. blue-green set first) x 2: Category (Between vs. Within) x 2: Lighting (Natural daylight vs. Lightbox) mixed design ANOVA with repeated measures over the factor Category. No effects were significant (all ps > .13).

For the blue-purple set, 17 children were at ceiling and 4 at or below chance leaving 39 children for analysis of Category effects. The data from the 39 children were analysed in a 2: Category (Between vs.Within) x 2: Lighting (Natural daylight vs. Lightbox) mixed design ANOVA with repeated measures over the first factor. The only significant effect was the main effect of Category [$F(1,37) = 7.71$, $p < .011$, $h_p^2 = .162$] with better performance on Between-category trials (see Figure 6). There was no main effect of Lighting [$F(1,37) < 1$] and no significant interaction between Lighting and Category [$F(1,37) = 1.55$, $p < .3$, $h_p^2 = .04$].

For the blue-green set, 27 children at or below chance were removed, and the remaining 33 children were submitted to a similar analysis. Neither main effect was significant [$F(1,31) < 1$], and the interaction was also not significant [$F(1,31) = 2.45$, $p < .2$, $h_p^2 = .073$]. (see Figure 6)

5.3.3 Discussion

Our results replicate those of Franklin et al. (2005) for the blue-purple set while giving a quite different outcome for the blue-green set. The performance of Himba children in the blue-green set resembled that of the Don't Know group of English toddlers, with no evidence of Categorical Perception. .Importantly, since a relativist view of colour cognition would predict a null result, the toddlers were well above chance at the task. The results of the blue-green set therefore further support the view that acquisition of colour terms is required to show Categorical Perception.

Both Franklin et al. (2005) and the present study showed Categorical Perception for Himba toddlers using the blue-purple set of colors. We wondered whether testing under natural daylight conditions in Franklin et al. (2005) could have produced the category effect for the blue-purple set but this seems unlikely as performance under controlled lighting was essentially similar to those replicating the conditions in their study. We would rather suggest that category effects for the blue-purple set arise because of some knowledge of the Himba color terms *zoozu* and *burou* (see Appendix I), or because they are better acquainted with cattle terms than color terms (Goldstein & Davidoff, 2008). In considering the Himba names for colors, Roberson et al. (2004, 2005) found extensive use of terms describing animal skin patterns (combinations of color, pattern and texture) for several colours, and confirmed that the preferred term used by adults for colours in the blue-purple range was not always a color term. Indeed, for the three colors used in the blue-purple continuum in Experiment 2, the purple color was named equally often with a cattle term as with the dominant color term. However, the two blue colors were called by the dominant color term around four times more frequently. Thus, if cattle terms were known by the children, they would likely have been used only for the purple color and hence set up an artifactual boundary.

5.4 General Discussion

A replication of Franklin et al.'s study (2005) was carried out with additional analyses that might help examine the role of toddlers' color naming and comprehension in color Categorical Perception. Despite some small procedural differences in the replication from the original study (within-group rather than between-group, naming undertaken before rather than after the two-alternative forced-choice, exclusion of the youngest age group) the results were initially very similar to those of Franklin et al. (2005). Of these small changes, it might have been thought that prior testing of naming before the two-Alternative Forced-Choice task might be the most likely to have changed the outcome. It might have encouraged color naming and artefactually produced the between-category superiority in the two-Alternative Forced-Choice task (Munnich & Landau, 2003), but the data show otherwise. Indeed, despite the remarkable findings in Roberson and Davidoff (2000), now replicated in many laboratories, that perceptual similarity changes with verbal load, we know from the procedures introduced in Franklin, Pilling and Davies (2005) that overt naming is not the origin of Categorical Perception.

Experiment 4 found an overall effect of category for both color sets for the sample of 60 children (analysis 1). However, with such a large age range, a different performance by a relatively small number of children that did not know color terms might not be revealed in the group analysis. Hence, as in Franklin et al. (2005), children were allocated according to their color naming ability for the stimuli used in the two-Alternative Forced-Choice (analysis 2). Using their procedure, we showed substantial agreement with Franklin et al.'s data for the blue-purple range though we did not fully replicate their findings of equivalent Categorical Perception independent of the accuracy of color naming for the blue-green range,. More striking, a different assessment of the child's understanding of color terms produced a quite different outcome. Critical here is how we define the understanding of a color term. The procedure adopted by Franklin et al. (2005) of naming the colors used in the two-Alternative Forced-Choice may appear to be a reasonable way to assess the knowledge of a color term. However, we argue that for young children it is more important to have some estimate of the range of colors that the child finds acceptable for the color term. Only then can we be confident that colors from within the same category are perceived as being more similar to each other than would be assumed from their separation in L*u*v* space. Most rigorously, one might use a signal detection procedure (Mullen & Pitchford, 2002, 2003) but the simpler procedure from Soja (1994) was sufficient to cause a rather different interpretation of the data. In the present study, color term acquisition was re-analyzed using a method that split children into Know and Don't Know (analysis 4). Now, the only group that showed category effects was the group of children who truly knew their colour names.

Within our analysis 2, we replicated Franklin et al.'s (2005) intriguing finding that a substantial number of children would be classified as having a reverse boundary, because their name boundary between green and blue (and also between purple and blue) was not yet aligned to that of adults. Franklin et al. make little comment on this reverse boundary group. Clearly, both their and our data find that Categorical Perception does not simply align with actual naming but there is more than one possible interpretation. It might be interpreted to show that color term acquisition is not necessary for the establishment of Categorical Perception. However, we suggest that the names reported by the children do not properly reflect their underlying colour knowledge (see also Pitchford & Mullen, 2003). Table 3 shows that almost half of the children classified as having a reverse boundary should actually be classified as knowing the color terms, which supports the second interpretation.

With respect to analysis 3, Franklin et al. (2005) suggested that, if color term acquisition was needed to establish Categorical Perception, then there could be a positive relation between color term fluency and Categorical Perception. This argument presumes that a relativist account would predict a positive correlation between the extent of Categorical Perception for any particular color and the number of other color terms known by the child. In fact, they found an unexplained small but nevertheless significant negative correlation. However, there is no reason to assume a relationship between Categorical Perception for a particular boundary and the total number of terms known. When a child definitely knows a color term this implies only that they will not confuse that color with other nearby terms, but says nothing about their knowledge of other regions of color space. There is no reason to suppose that acquiring more secondary terms (e.g., knowing orange and pink) would make a significant difference for Categorical Perception to green, blue or purple and the present data found no such correlation.

Considering the Himba data, Experiment 5 replicated the cross-category advantage found by Franklin et al. (2005) for the blue-purple range but did not find the same for the additional blue-green range of colours. Yet, the boundary between blue and green is one that is still regarded as potentially universal (Kay & Regier, 2003) while that between blue and purple is not. The lack of Categorical Perception for the blue-green range is not due to the Himba children misunderstanding the task, as their performance was well above chance. We further examined whether testing under natural daylight conditions in Franklin et al. (2005) could explain the differences between the two sets of results, but this seems unlikely, as the pattern of results under controlled lighting were essentially similar to those replicating the conditions in their study. Above, we proposed that the use of non-color words for the purple color could have produced a boundary for the Himba toddlers.

In summary, though our results looked at first to be very similar to those of Franklin et al (2005) they end up not supporting their conclusions that there is one particular set of universal color categories (those named in English) that are independent of language acquisition (Franklin & Davies, 2004; Bornstein et al, 1976). The counter-argument here points to the importance of color term acquisition by children of Himba and English cultures in establishing both color categories and hence Categorical Perception. These findings reflect differences across cultures in accord with other color categorization studies (Davidoff et al, 1999; Roberson et al, 2004, 2005; Winawer et al., 2007;

Roberson et al., 2007) and among toddlers of the same culture separated by language acquisition (Roberson et al, 2004; O'Hanlon & Roberson, 2006).

6. Experiment 6: A cross-cultural comparison of animal pattern categorization in two-alternative-forced-choices

(Also a publication: Goldstein, J., & Davidoff, J. (2007). Categorical perception of animal patterns. *British Journal of Psychology, In Press*).

6.1 Introduction

As part of the more general issue of whether culture can affect perception (Nisbett & Miyamoto, 2005), the present paper addresses the Whorfian question of whether the language available to describe perceptual experience can influence the experience itself (Whorf, 1956). There has been a considerable revival of interest in the Whorfian hypothesis (Lucy, 1992; Davidoff, Davies & Roberson, 1999; Boroditsky, 2001; Özgen & Davies, 2002; Saunders & van Brakel, 2002). The revival has provided evidence for the Whorfian view from systematic investigations of the relationship between language and thought. Differences between languages in their grammatical structure and vocabulary have been associated with perceptual differences of the same experience in the following domains: grammatical gender (Sera, Berge & Pintado, 1994; Sera et al, 2002, Boroditsky, Schmidt & Phillips, 2003); colour (Kay & Kempton, 1984; Davidoff et al, 1999; Roberson, Davidoff, Davies & Shapiro, 2004); material and shape classification (Lucy, 1992); spatial relations (Levinson, 1996; Bowerman & Choi, 2001), number systems (Gumpertz & Levinson, 1996, Gordon, 2004), artifact categories (Malt & Johnson, 1998); modes of motion (Gennari, Sloman, Malt & Fitch, 2000); time (Boroditsky, 2001) and shape (Roberson, Davidoff & Shapiro, 2002). However, other recent studies have argued against the influence of linguistic differences on mathematical abilities (Gelman & Butterworth, 2005), colour (Kay & Regier, 2003), and perceptual classification, both at the level of terminology (Munnich & Landau, 2003; Malt et al, 1999) and grammatical structure (Karmiloff-Smith, 1979; Pérez-Pereira, 1991). The present study seeks to provide further evidence in favour of an effect of vocabulary on perceptual classification by the study of the Himba, a remote culture which has a rich vocabulary of pattern terms.

Eckl (2000), and Turton (1980) for an Ethiopian group of cattle herders with similar life-style to that of the Himba, outlines cultural interests and reasons why colour and patterns are used to recognise individual animals. Therefore, it is not surprising that the Herero and Himba have many names for cattle patterns (Eckl, 2000). The present study will assume that each of these names refers to a category of patterns and will argue that these categories should obey the principles of Categorical Perception (CP). Harnad (1987) provides a comprehensive discussion of CP across a range of auditory and visual categories. He argued that, with CP, a physical continuum is perceived as qualitatively discontinuous. Items from different categories appear more different than items from the same category despite being equated for physical difference (e.g., by numbers of just-noticeable-differences). CP is thereby revealed by faster and more accurate discriminations between stimuli that cross a category boundary than between two stimuli that are both good exemplars of the same category (Bornstein, 1987). Goldstone (1994) further argued that this between-category expansion might also be accompanied by a within-category compression where items within a category become to look more similar. As a technique, CP has been used to consider debates in face (Etcoff & Magee, 1992) and colour categorisation (Roberson, Davidoff & Braisby, 1999).

In cross-cultural studies, CP has been used to show precise differences based on colour terms in the speaker's language (Roberson, Davidoff, Davies & Shapiro, 2005). However, there has been considerable debate as to whether the origin of colour CP is linguistic (Roberson et al, 2004, 2005) or at least is partially based on universal neurophysiological constraints in colour vision (Bornstein, 1975; Franklin & Davies, 2004). No other visual type of stimulus has been used in cross-cultural investigations of CP. The present paper by extending the CP methodology to animal patterns does so clearly to categorisation that is not innately given. It will use the two-alternative forced-choice (2AFC) methodology as in the previous studies (Roberson, Davies & Davidoff, 2000) to examine whether a target is better recognised when contrasted with a foil from a different category compared to when the foil is from the same category. In summary, the aim of this study was to provide a different type of stimulus to confirm previous findings with colour that linked perceptual and cognitive differences to labelling differences.

6.2 Production of the animal continua

Stage 1: Preliminary naming

Two individuals from different Himba villages, (man, age 45 and woman, age 35) helped in the picture-taking of animals at two different waterholes. With their advice, 128 pictures were taken with a digital camera; these were transferred to a laptop pc and shown to 10 Himba men (age 23-45), from two separate groups. Like all the subsequent participants, they were rural inhabitants of Northern parts of Namibia (Kaokoveld); mainly herders taking their animals to waters sources, often having to travel quite far by foot. They were monolingual, and paid in kind. All participants were screened for colour vision abnormalities with the City Colour Vision Test (Fletcher, 1980). Every participant was tested individually in a darkened vehicle. The experimenter showed one picture at a time on a computer, and the translator asked the participant to give a name per picture and recorded the answers in a notebook.

Stage 2: Obtaining the morphed stimuli

Eighteen pictures were selected on the basis of results from naming stage 1; for these pictures, there was 90% naming agreement. The animals were isolated from their context/background and were manipulated in Adobe Photoshop and given a light yellow background, approximating the colour of the earth/sand of the Himba territory. Next, pictures were paired by the experimenter and manipulated in the Meditor morphing package to produce a continuum. Every continuum was based on two pictures (endpoints) from the same species (cow, goat or sheep) and resembled each other to increase the likelihood of a 'natural' continuum after morphing (see figure 6.1 A-E for endpoints used in the 2AFC Test). However, to ensure that the animal shape was identical for both end-points, the same silhouette was used for both endpoints. The pattern was lifted from one end-point in Photoshop and transferred to the outline of the other end-point. A continuum consisted of 21 pictures in total as there was a 5% difference between each morphed picture going from one endpoint to the other. Nine continua were created following this procedure, resulting in 189 pictures (21 pictures x 9 continua) that were printed and prepared for use in Stage 3.

Figure 6.1 A-E. Endpoint pictures of continua used for 2AFC test.

Figure 6.1A Continuum 1

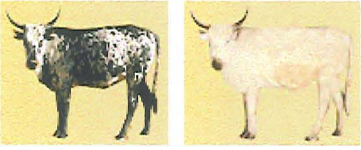


Figure 6.1B Continuum 2

Figure 6.1C Continuum 3

Figure 6.1D Continuum 4

Figure 6.1E Continuum 5

Stage 3: Naming the morphed continua

For the second phase of naming, laminated pictures were placed on a table, either in a sequential or random order. The sequential order (1-21 in Fig 2) was simpler to administer but as this could have led to perseveration in naming, we added a random order condition. Fifteen different Himba from those in Stage 1 (12 men and 3 women, varying in age from 17 to 45 years) participated in the sequential naming, and 15 different Himba from those in Stage 1 and those who did the sequential order (12 men and 3 women varying in age from 19 and 50 years) participated in the naming in random order. Every Himba participant was asked to give the name of the animal that the experimenter pointed at, or was asked to name every picture individually (depending on the condition). In the sequential phase, four blocks were needed to assess 189 pictures, as only 50 pictures fitted the table. For individual picture naming, the Himba participant named 189 pictures placed on the table in random order. All answers were recorded on an answer sheet. There was some disagreement about the names given in phase 2 to those in phase 1. Our interpreter advised us that this was to be expected as the Himba participants came from a large area and that there was some naming variation between different groups.

Stage 4: Deciding on the continua

The fourth stage in the production of the animal continua was to ensure that they would be suitable for use in a two alternative-forced-choice experiment. In one respect, all continua were suitable. In all continua (see, for example, figs 6.2 a-e) there was a sharp drop where the name given to one endpoint changes to the name given to the other endpoint. As these intermediary points represent "imaginary" animals, one might ask why there is not a gradual change from one name to the next rather than a sudden change. The most likely answer is that we have chosen prototypical animals for the endpoints. Animals with the same name would bear a strong family resemblance to the prototype, and the morphing procedure would move, in a few images, from pictures that would easily be named as the endpoint to ones that would not be so named. Of course, there was an area on every continuum that some of the Himba were not prepared to give either endpoint name and gave other names. However, for a different reason, not all nine continua could be used. Four of the nine continua consisted of boundaries (name changes) that were very close to either one of the endpoints. It was therefore not possible to create a two alternative-forced-choice procedure for these continua as the test requires a minimum of a boundary location at 5 pictures away from the endpoints.

The remaining 5 continua were suitable for use in the two alternative-forced-choice test. The agreement on endpoint names and our interpreter's definition of names were as follows: Continuum 1 Cow (Ongange–Omvahe), 90% consensus for Ongange (white spots on body and face), and 56.66 % for Omvahe (cream colour); Continuum 2 Goat (Orupera-Ongwiti) 53.33% agreement for Orupera (brown body with white circling rim around the body), and 86.66% for Ongwiti (white body and brown or black head and neck); Continuum 3, Goat (Ombundu-Onganga), 36.66% consensus for Ombundu (light red body) and 50% for Onganga (tiny dots all over a red body); Continuum 4, Goat (Othaona-Ongara), 40% consensus for Othaona (golden-brown colour) and 96.66% for Ongara (black background with golden-beige patches); Continuum 5, Sheep (Ondondo-Ombotozu) 40% consensus for Ondondo (white with black around the body and ears) and 43.33% for Ombotozu (brown/black with white on the stomach). To ensure that the two naming techniques (sequential or random) from Stage 3 did not cause a difference in naming, t-tests were performed on the naming scores from each technique comparing each of the 21 positions for each continuum. There were no significant differences. All t values < 1 except Continuum 3(Onganga, $t(20) = 1.2$, $p = .25$) and Continuum 4 (Ongara, $t(20) = 1.48$, $p = .15$).

The boundary at which one name changed to the other was assessed by eye from the name agreement at all 21 positions (see fig 6.2 A-E). For Continuum 1, the boundary was between pictures 12 and 13. For Continuum 2, the boundary was between pictures 15 and 16. For continuum 3, the boundary was between pictures 6 and 7. For continuum 4, the boundary was between pictures 5 and 6, and for continuum 5, the boundary was between pictures 12 and 13.

Figure 6.2 A-E Naming agreement from 30 Himba participants on five continua for animal patterns. Each continuum consisted of 21 pictures (19 equally spaced morphed images from the two endpoints). Continuum 1 (a) was for cow patterns, Continua 2 (b), 3(c) and 4(d) were goat patterns and Continuum 5(e) was for sheep patterns. The two curves for each continuum represent the number of times the endpoint names were given to each picture.

Figure 6.2A Continuum 1

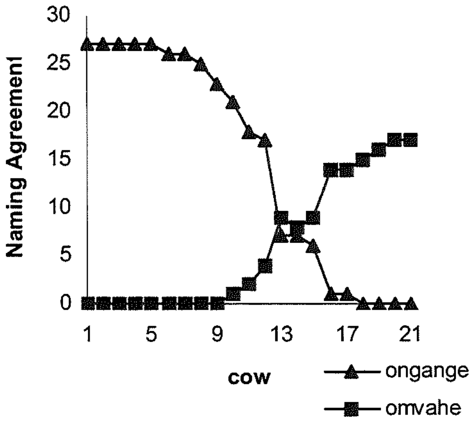


Figure 6.2B Continuum 2

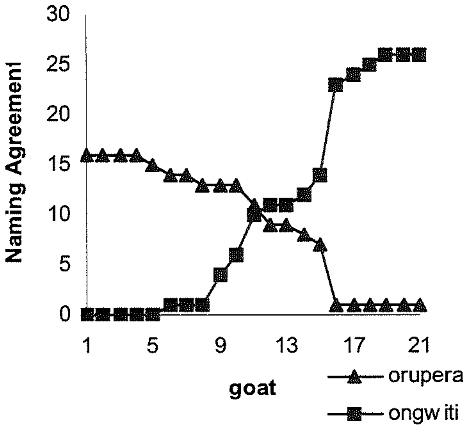


Figure 6.2C Continuum 3

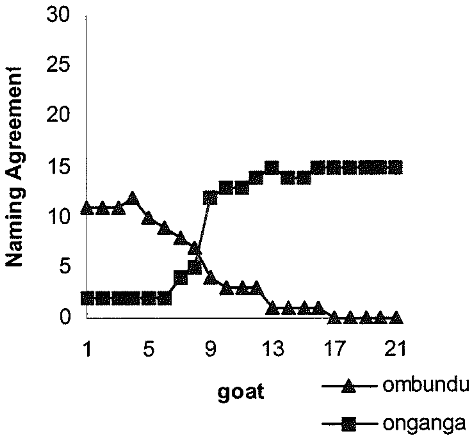


Figure 6.2D Continuum 4

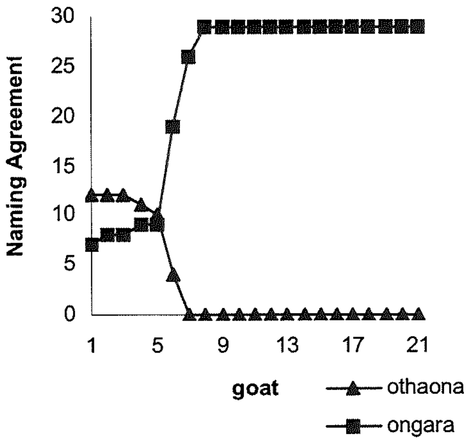
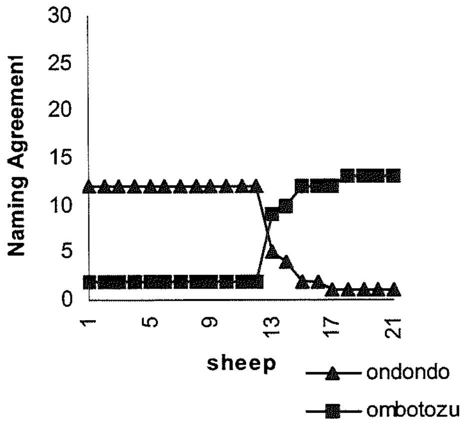


Figure 6.2E Continuum 5



6.3 Two alternative-forced-choice test

6.3.1 Method

6.3.1.1 Participants

Thirty different Himba (17 women and 13 men, varying in age from approximately 18 to 45 years) and 30 English (17 women and 13 men varying in age from approximately 20 to 43 years) took part in the 2AFC Test. The English participants were students at Goldsmiths' College and participated under course requirements. As for the production of the animal continua, the Himba were monolingual, rural inhabitants of Northern parts of Namibia (Kaokoveld), tested in those areas and paid in kind. All participants were screened for colour vision abnormalities with the City University ColourVision Test (Fletcher, 1980).

6.3.1.2 Stimuli

Pictures for the between-category and within-category items were selected from the five continua (see Stage 4 of preceding section). There was always 10% morph length between two pictures in each pair. Fig 3 a-e shows the section of each continuum used to select pictures for the 2AFC range. It also shows the actual pictures used for each type of trial. Note that though the same labels (a-j) are used in figure 6.3 to denote the section of each continuum used in the 2AFC, these came (see Stage 4 above) from different positions on the 21 step morphed continua.

Figure 6.3 A-E Subsets of the 5 continua containing the pictures used in the 2AFC test. The boundary for each continuum was between e and f. Within-category pairs were pictures a-c and h-j (|-----|). Between-category pairs were d-f and e-g (|……|). (next page)

Figure 6.3A Continuum 1

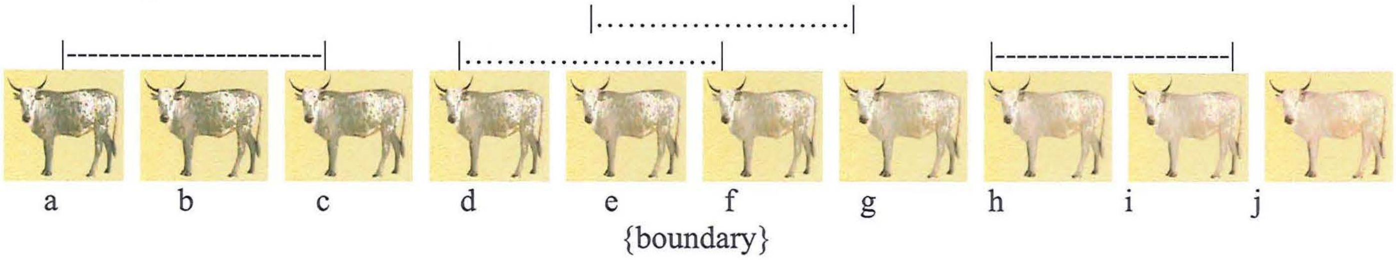


Figure 6.3B Continuum 2

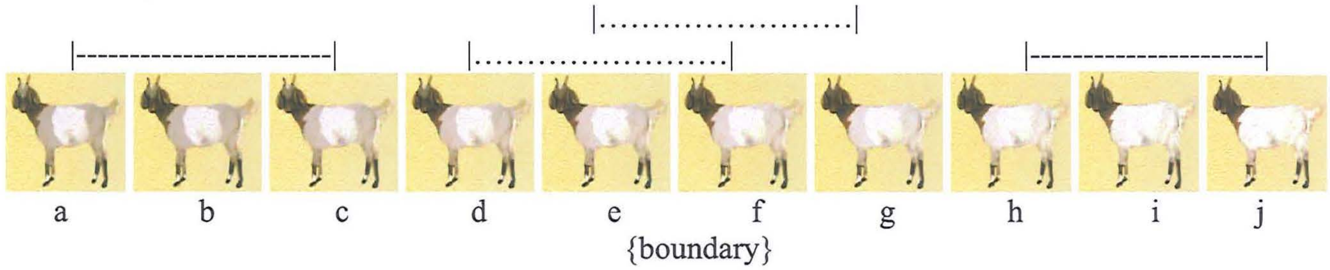


Figure 6.3C Continuum 3

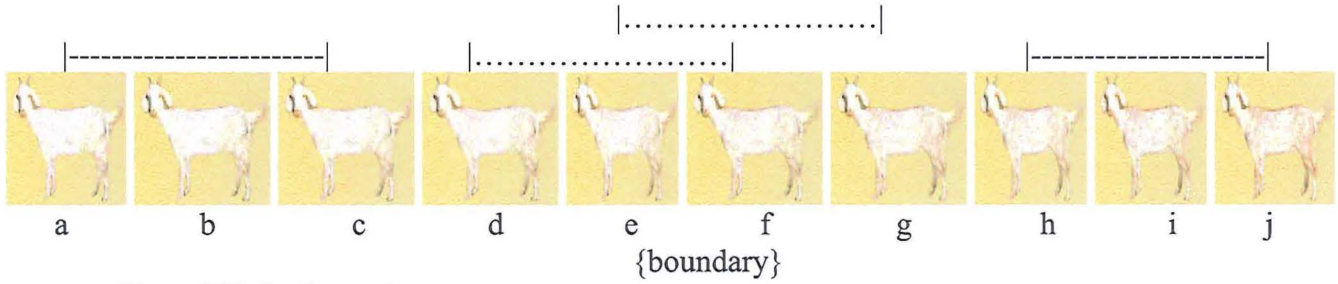


Figure 6.3D Continuum 4

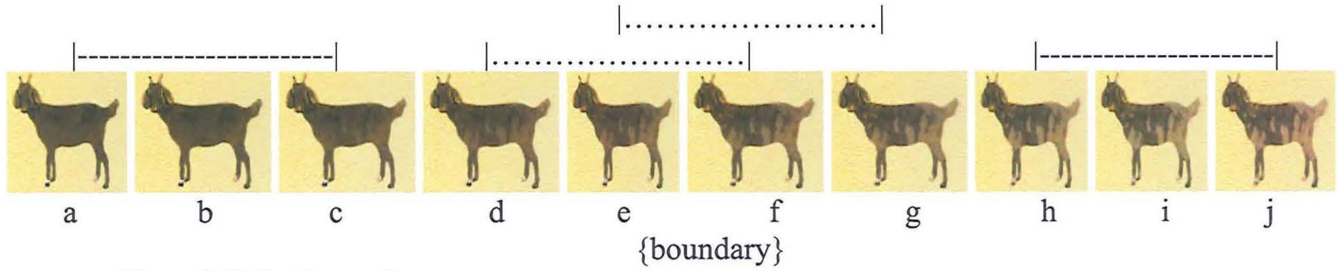
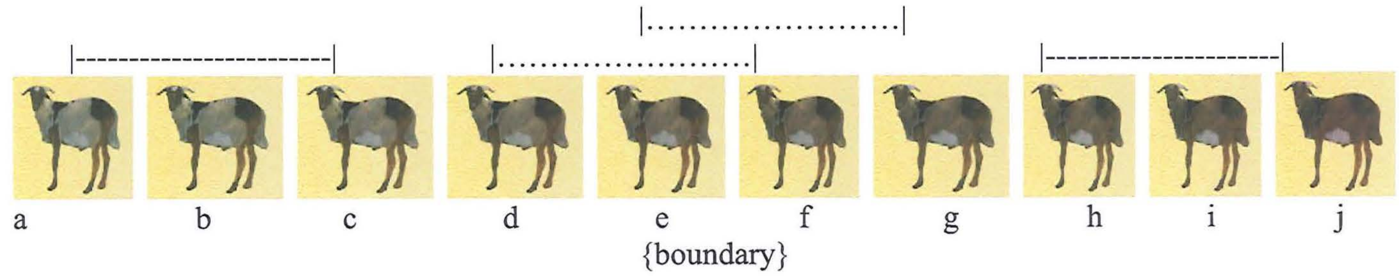


Figure 6.3E Continuum 5



6.3.1.3 Procedure

Practice

The task was explained to the Himba via an interpreter. He told the participant that he would show a picture of an animal and that they had to remember what they had seen. After the first picture, there would be immediately two pictures one of which was identical to the first. The participant was asked to point to the one that was identical. Practice trials (maximum of 10) with feedback were given using animals not included in the Test until there were three consecutively correct answers. Exactly the same procedure was adopted for the English participants except that the stimuli were shown on a computer (programmed in E-prime) with a 5 sec interval between target and 2AFC. The participants were asked to respond by pressing key 'v' for the left picture and 'n' for the right picture when presented with the two alternatives.

Test

Procedure at Test was similar to that at practice. The experiment was carried out manually in Namibia in a shaded area and answers were recorded on a scoring sheet. The test consisted of 40 2AFC trials per continuum (20 cross-category and 20 within-category trials), summing to 200 trials per participant. Half of the trials for each continuum had the correct answer on the left and half the correct answer on the right. A different random order of the 200 trials was administered to each participant with trials for all the continua intermingled. To ensure smooth running of the experiment, for the Himba, the trial cards were assembled prior to the start of the experiment.

Naming

At the end of the test, Himba participants were shown all pictures of the test one at a time and asked to name them.

6.3.2 Results

Naming

Naming agreement was assessed for the 8 pictures used for each continuum. Each picture was classified as in agreement if the same name was given as in the production of the animal continua (see stage 4 of previous section). For each continuum, some Himba participants were in complete agreement but others showed disagreement on one name or more. The numbers of Himba with naming agreement for the continua was as follows: 27 for continuum 1; 22 for continuum 2; 10 for continuum 3 (and therefore 20 participants in naming disagreement); 29 for continuum 4; 19 for continuum 5 (and 11 participants in naming disagreement).

Two alternative-forced-choice

For the initial analysis, all Himba and English participants were included to allow all continua to be included in the same analysis. A 2 (Group: Himba vs. English) x 2 (Trial Types: Between-Category vs. Within-Category) x 5 (Continuum: Continua 1 vs. 2 vs. 3 vs. 4 vs. 5) Anova, with repeated measures over the second and third factors was carried out on the accuracy data. All main effects were significant. The Himba were more accurate than the English ($F(1,58) = 7.84$, $p = .007$), Between Category trials were more accurate than Within-Category Trials ($F(1,58) = 11.36$, $p = .001$) and Continua differed in their recognition accuracy ($F(4, 232) = 11.37$, $p < .001$, with highest performance on Continuum 1 and 4 and lowest scores on Continuum 2 and 5).

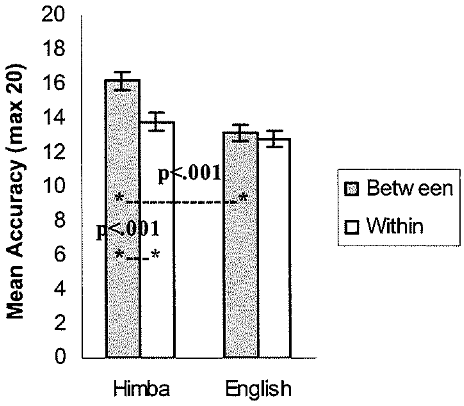
In the context of these main effects, the analysis showed three significant two-way interactions. Most important, there was a significant Group x Trial Types interaction ($F(1,58) = 32.62$, $p < .001$), revealing on post-hoc analysis (Bonferroni correction, $p < .013$), higher scores for the Himba on Between-Category trials than Within-Category trials, $t(29) = 5.34$, $p < .001$, whereas English performance did not reach reliable levels between Trial Types, $t(29) = 2.22$, $p = .034$ which, in any case was in the direction of greater accuracy for Within-Category trials. The Himba were more accurate than the English on Between-Category trials, $t(29) = 5.62$, $p < .001$ but the Groups did not differ on Within-Category trials, $t(29) < 1$.

There were also reliable effects that involved Continuum. The Category x Continuum interaction gave F(4,232) = 10.2, p < .001 and there was also a weakly significant Group x Continuum interaction, F(4,232) = 2.43, p = .048. In light of these effects that involved differences between Continua and also because Himba were included in the analysis who did not agree on the names for the animal patterns, further analyses were carried out on individual continua. Analyses were carried out on the accuracy data for continua 1, 2 and 4 on only the participants who were in naming agreement. However, in Continua 3 and 5 there was substantially more naming disagreement than for other continua, so Himba naming disagreement groups were also considered in the data analyses of these continua.

Continuum 1

A 2 (Group: Himba vs English) x 2 (Trial Types: Between-Category vs. Within-Category) Anova, with repeated measures on the second factor was carried out on the accuracy data.

Figure 6.4 Mean accuracy scores (max = 20) and standard errors for English and Himba participants on continuum 1 for Between-Category and Within-Category trials in the 2AFC test. Comparisons giving p < .001 are indicated.

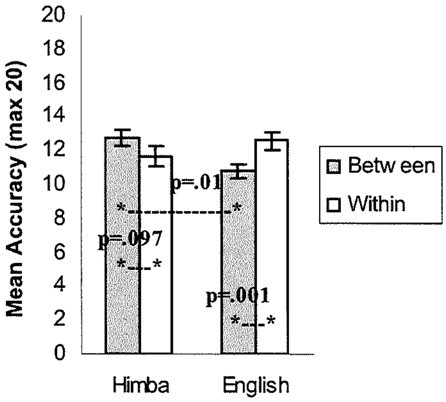


As confirmed in figure 6.4, the analysis showed a significant interaction (F(1,55) = 11.18, p = .001) accompanied by main effects of Trial Type (F(1,55) = 21.87, p < .001) with better performance for Between-Category Trials, and Group (F(1,55) = 10.28, p = .002) with better performance for the Himba (see fig 6.4). Analysis of the interaction revealed higher scores for the Himba on Between-Category trials (t(55) = 4.6, p < .001) but not on Within-Category trials (t(55) = 1.39, p > .1). However, the Himba (t(26) = 5.58, p < .001) have a Between-Category advantage over Within-Category trials in their own language whereas the English do not (t < 1). The lowest mean accuracy (M = 12.80, SD = 2.34) was for the English on Within-Category trials, and this score was above chance (t(29) = 6.55, p < .001).

Continuum 2

A 2 (Group: Himba vs. English) x 2 (Trial Types: Between-Category vs. Within-Category Trials) Anova, with repeated measures on the second factor, was carried out on the accuracy data.

Figure 6.5 Mean accuracy scores (max = 20) and standard errors for English and Himba participants on continuum 2 for Between-Category and Within-Category trials in the 2AFC test. Comparisons giving p < .1 are indicated.

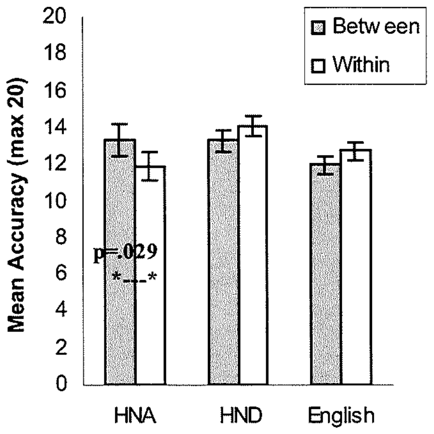


As confirmed by figure 6.5, the analysis only showed a significant interaction (F(1,50) = 13.41, p = .001). Both Main effects (Trial Type and Group) gave F < 1 (see fig 6.5). Analysis of the interaction revealed higher scores for the Himba on Between-Category trials (t(50) = 2.57, p = .013), but not on Within-Category trials (t(50) = 1.37, p > .1). However, unlike for Continuum 1, the Himba only showed a trend for superior performance on Between Category trials (t(21) = 1.74 , p = .097) whereas the English (t(29) = 3.17, p = .001) were more accurate on Within-Category trials. A further analysis showed that much higher accuracy was obtained for English participants on one of the Within-Category pairs (t(29) = 3.88, p = .001). It was for the Within-Category pair h-j in fig 3b. As one may see, the contrast in picture j is low and could be remembered as having a much larger white area than the other. The lowest mean accuracy (M = 10.80, SD = 2.54) was for the English on Between-Category trials; this score was not above chance (t(29) = 1.73, p > .05). The next lowest mean accuracy (M = 11.68, SD= 2.38) was for the Himba on Within-Category trials; this score was above chance (t(21) = 3.32, p = .003).

Continuum 3

A 3 (Group: Himba Naming Agreement (HNA), vs. Himba Naming Disagreement (HND) vs. English) x 2 (Trial Types: Between-Category vs. Within-Category) Anova, with repeated measures on the second factor was carried out on the accuracy data.

Figure 6.6 Mean accuracy scores (max = 20) and standard errors for English and Himba participants on continuum 3 for Between-Category and Within-Category trials in the 2AFC test. Himba participants are divided into those with naming agreement (HNA) and those in disagreement (HND). Comparisons giving p < .05 are indicated.

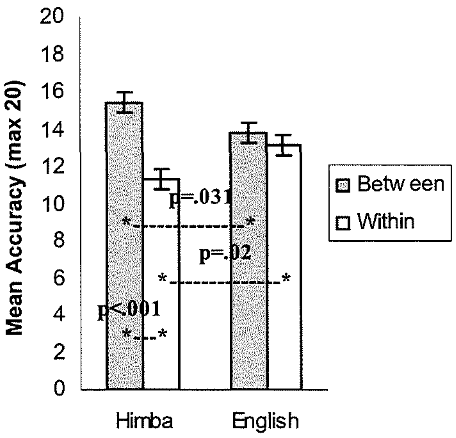


As confirmed from figure 6.6, the main effect of Trial Type gave F < 1, no effect of Group (F(2,57) = 2.30, p > .1) was found, and the interaction showed a trend (F(2,57) = 2.49, p = .092) (see fig 6.6). This trend was further explored in post hoc tests. When comparing trial types within groups, the only significant value found was for HNA with a between-category advantage (t(9) = 2.58, p = .029). A similar comparison for HND gave t(19) = -1.67, p > .1, and for English gave t(29) = 1.18, p > .2. When comparing groups across a single trial type condition, the groups failed to show significantly different performances from each other on Between-Category trials (F(2,57) = 2.05, p > .1), or Within-Category trials (F(2,57) = 2.61, p > .05). The lowest mean accuracy (M = 11.90, SD = 1.52) was for HNA on Within-Category trials; this score was above chance (t(9) = 3.94, p = .003).

Continuum 4

A 2 (Group: Himba vs English) x 2 (Trial Types: Between-Category vs Within-Category Trials) Anova, with repeated measures on the second factor was carried out on the accuracy data.

Figure 6.7 Mean accuracy scores (max = 20) and standard errors for English and Himba participants on continuum 4 for Between-Category and Within-Category trials in the 2AFC test. Comparisons giving p < .05 are indicated.



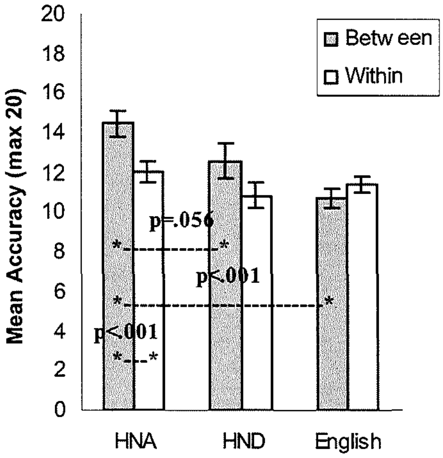
As confirmed from figure 6.7, a significant interaction (F(1,57) = 17.29, p < .001) was found in the context of a main effect of Trial Type (F(1,57) = 30.15, p < .001) with superior performance on Between-Category trials but no main effect of Group (F < 1) (see fig 6.7). The interaction showed that the English were more accurate on Within-Category trials (t(57) = 2.38, p = .02) and the Himba more accurate on Between-

Category trials (t(57) = 2.21, p = .031). However, English participants did not distinguish between the Trial Types (t(29) = 1.22, p > .2), whereas the Himba did (t(28) = 5.71, p < .001) scoring more accurately on Between-Category trials than Within-Category trials. The lowest mean accuracy (M = 11.38, SD = 3.61) was for Himba on Within-Category trials; this score was above chance (t(28) = 2.05, p = .05).

Continuum 5

A 3 (Group: Himba Naming Agreement (HNA) vs Himba Naming Disagreement (HND) vs. English) x 2 (Trial Types: Between-Category vs Within-Category Trials) Anova, with repeated measures on the second factor, was carried out on the accuracy data.

Figure 6.8 Mean accuracy scores (max = 20) and standard errors for English and Himba participants on continuum 5 for Between-Category and Within-Category trials in the 2AFC test. Himba participants are divided into those with naming agreement (HNA) and those in disagreement (HND). Comparisons giving p < .06 are indicated.



As confirmed from figure 6.8, the analysis showed a significant interaction (F(2,57) = 8.26, p < .001) in the context of a main effect of Trial Type (F(2,57) = 8.41, p = .005) with superior performance on Between-Category trials and of Group (F(2,57) = 5.96, p = .004) with better performance for the Himba (see fig 6.8). Analyses of the interaction showed differences between Groups only for Between-Category trials (F(2,59) = 15.68, p < .001) and not Within-Category trials (F < 1). The largest difference was found between the HNA and English Groups (Tukey HSD, p < .001), with a trend towards a significant difference between the HND and English Groups (Tukey HSD, p = .056). When looking at performance differences according to Trial Types, the only significant difference was found for HNA (t(18) = 4.42, p < .001) with an advantage for Between-Category Trials; HND (t(10) = 1.51, p > .1) and English (t(29) = -1.53, p > .1) groups performed equally well on both Trial Types. The lowest mean accuracy (M = 10.73, SD = 1.64) was for Himba on Within-Category trials; this score was above chance (t(29) = 2.45, p = .021).

6.4 Discussion

The present study examined CP with a type of stimulus not previously used to assess the effect of labels on perceptual judgements. In cross-cultural studies of colour CP, it has been found that colour categories correspond to the terms in the speaker's language (Roberson, Davidoff & Davies, 2000; Riberson et al., 2004, 2005). Indeed, Roberson et al (2005) showed that colour CP for the Himba was different to CP in another language with the same number of colour terms. For the cattle patterns, the Whorfian view predicted that it would only be the Himba who showed superiority for cross-category decisions as only they have the appropriate labels. The Whorfian view was largely upheld from the recognition accuracy for these animal stimuli.

Continua 1 and 4 gave the clearest results. For these continua, the Himba participants showed strong name agreement to those given at production of the continua. For Continua 1 and 4, the Himba showed greater recognition for cross-category trials whereas the English found the two types of trial equally easy. The pattern of results was, in Goldstone's (1994) terms one of between-category expansion in that it would seem that, to the Himba, animals from different categories look more different. One can also see, for example in Continuum 4, evidence of within-category compression in that the Himba performed worse than the English on these trials. Similar trends are in the other continua but they are difficult to be confident of in the context of the generally superior accuracy for the Himba with these animal patterns. The same pattern of cross-category advantage was seen for Continua 3 and 5 though the data were complicated by the fact that not all the Himba agreed on the names given by those who helped make those continua. However, if we considered only those Himba who showed name agreement, exactly the same pattern was found. The Himba who showed name disagreement showed category effects identical to the English participants. The clear pattern in the other four continua was less marked in Continuum 2 that only produced a trend in the same direction of the Himba showing superior performance for cross-category trials. We offer no certain explanation for the weaker effect for this continuum. However, as noted above, there are effects of contrast that could prompt the use of additional labels. We also note that the only chance score (English, Between-Category) was found on Continuum 2 and one of the within-category pairs appeared to offer easier recognition for the English participants.

It would seem that the CP found for the Himba derives from their possession of pattern categories not found in the Western participants. However, alternative explanations might be considered. Stimuli were presented to the Himba manually and it is possible that this different procedure altered overall performance. In fact, despite it being an artificial task, the Himba were generally better at the recognition of the animals than the English participants. In any case, the critical aspect of the data is not the overall level of the performance but the superiority of between-category trials and this cannot be explained by the pictures being shown as photographs rather than on the computer. We can also rule out the possibility that the task was simply too difficult for the English observers. While there is significantly lower overall performance for the English observers for several of the continua, this is not always the case and performance levels were above chance for English participants on almost all the continua.

Another potential artifact is that of explicit naming by the Himba whereby, it could be argued, superior recognition memory could be achieved as a result of matching by name in the 2AFC. The 2AFC task is that most commonly used in cross-cultural studies of CP (Roberson et al, 2000, 2004, 2005) but it is perhaps not the best. However, the same argument about labelling could have been made if the 2AFC task had been replaced by similarity judgements to determine CP (see Roberson et al, 2000). We could have more clearly ruled out explicit naming by use of interference procedures (Roberson & Davidoff, 2000) but these were not possible with a Himba population. Probably a better technique would have been to employ rapid presentations that would make naming difficult (Gilbert, Regier, Kay & Ivry, 2006) though even then some limited naming might be possible.

In fact, there was no evidence to the experimenter of any explicit naming but it would be difficult to rule out some implicit naming from this first study of animal pattern CP. Indeed, one might not want to rule it out; activation of a perceptual difference linked to a linguistic network (i.e, implicit naming) could be critical to all CP (Roberson & Davidoff, 2000). An indication that categorical representations are activated implicitly was shown in the neuropsychological case study of Roberson et al (1999). The patient demonstrated intact CP (colours and facial expressions) when using the 2AFC procedure as given to the Himba. The limited role that explicit labelling played in the task was clear because the patient was unable to name colours or facial expressions correctly so any such attempt by him would have been detrimental to his performance.

The neuropsychological data (Roberson et al, 1999) are also consonant with the view that networks activated in colour CP are not even implicitly linguistically driven but are simply perceptual (Pilling, Wiggett, Ozgen & Davies 2003). Indeed, the introspection of observers (Kay & Kempton, 1984) is that discriminating between colours from the same or different categories is a perceptual decision. The issue concerning the underlying perceptual or linguistic mechanism for CP has to date been examined by use of interference techniques. In some cases, verbal interference has resulted in complete abolition of CP (Roberson et al, 2000; Gilbert et al, 2006) and in others less than complete abolition (Pilling et al, 2003). For the case of animal pattern CP, as there is no likelihood of innately given categories, the perceptual codes underlying CP would presumably be linked to a linguistic network but that would need verification in future studies.

CP for animal patterns we regard as highly likely for the Himba. Colour categories have very limited functional significance for the Himba; this is not the case for animal patterns. However, the CP could arise from ways other than from labels. A different explanation for the origin of animal pattern CPs could be that they derived from the function of an animal with a particular type of pattern though of course these functions are going to be highly correlated with labels. Indeed, by adulthood, it is clear from our HNA groups that the classification is intrinsically linked to the names but a distinction between the two potential origins might be possible in future testing. Even very young children look after animals in the Himba culture so one might be able to carry out the same sort of longitudinal study as in Roberson et al (2004) to investigate the development of animal pattern CP. In Roberson et al (2004), the acquisition of colour CP was linked closely to the acquisition of colour terms. It might be different for animal patterns though one must stress it would be hard to distinguish between the two explanations as pattern names might be acquired very early given their cultural significance.

In summary, the data clearly imply that labels have affected the Himba perception of animal patterns. Indeed, one could simply point to the continua in figures 6.2 A-E and be surprised that what appears to us to be a smooth progression could be marked with an abrupt change to any observer. The proposal that it is the acquisition of vocabulary in the language that causes the abrupt change is rather nicely confirmed by the fortuitous result from the HND groups in Continua 3 and 5 who gave no evidence for cross-category superiority. Thus, though one cannot rule out alternative explanations, the consideration of animal patterns has given further evidence in favour of linguistic relativity with respect to categorical perception.

7. General Discussion- From grouping behaviour in perception to conceptual grouping in categorization

The discussion of the current results focuses on a progression from findings of conceptual grouping that has been found to take place under the influence of language. First of all, perceptual grouping is identified as a mechanism made use of by both human and non-human primates even though this kind of grouping has been questioned in monkeys (e.g. D'Amato, Salmon & Colombo, 1985). Monkeys show evidence of matching by similarity in experiments 1 and 2; they match to training colours and show an advantage for within-category matching. Human primates rather show categorical grouping preference as perceived in experiments 1, 2, 3, 4, 5, and 6 that has identified an involvement of culture and language setting humans and monkeys apart as seen from experiment 1 and 2. The discussion further consolidates ontogenetic, phylogenetic and cross-cultural findings in a context that describes the implications of this research to universality and relativity of categorization behaviour. The discussion then further leads to a three-pronged investigation with the inclusion of phylogenetic, ontogenetic, and cross-cultural approaches to address the nature and origin of grouping behaviour. In comparison to former cross-sample investigations (e.g., Bornstein, 1980), the overall findings indicate that sometimes universal perceptual grouping trends are identified, but that categorical grouping for the case of colour and animal patterns are always relative to language and culture.

7.1 Cross-species grouping differences

The first achievement of this thesis found cross-species differences in a set of experiments that involves colour categories (experiment 1 and 2). Cross-species differences have been found that may indicate language is involved in the findings. Perhaps this may be due to the fact colour has been identified as a special case of conceptual rather than perceptual grouping (Braisby & Dockrell, 1999). Although previous research shows similar colour apparatus for monkeys and humans (e.g., Bowmaker et al., 1991) and some researchers believe monkeys do show the ability to categorize (e.g., May et al., 1989; Ohlemiller et al., 1999; Thompson & Oden, 2000), it does not necessarily follow that monkeys conceptualize colour as humans do

(Matsuzawa, 1985; Matsuno, Kawai, & Matsuzawa, 2004). In fact, we achieve cross-species differences on categorization of colour in matching-to-sample and two-alternative forced choice tasks that lead us to believe monkeys do not categorize as humans do. The results of experiment 2 and 3 are in contradiction to the only other comparative colour categorization study carried out almost thirty years ago by Sandell, Gross & Bornstein (1979) that showed identical colour categories across species.

The cross-species grouping behaviours perceived in MTS and 2 AFC tasks coincide with previous findings that have shown that global precedence is specific to human beings even though common biological mechanism have been identified across species (Fagot & Deruelle, 1998). In so far as categorization may be viewed as a form of global grouping behaviour, human primates are the only group to show this behaviour in experiments 2 and 3, due to their response to the boundary that exists between blue and green and their subsequent cross-category advantage. Despite evidence of monkey's excellent memory (Fagot & Cook, 2006) that may have allowed them to learn about the green and blue endpoints in experiment 2 and as noted for results of monkey's behaviour in experiment 3, monkeys show grouping locally by similarity, but do not show abstraction for categorical grouping. Moreover, monkey's learning and memory skills should not be confused with the presence of colour categories, because although they can memorize colour they categorize colour quite differently to humans. Indeed, monkeys were found to show carry-over memory effects from matching in experiment 3 in which learning of a local nature was identified by showing extensive within-category advantages related to the endpoints of the continuum used in both the matching-to-sample and two alternative-forced-choice.

Contrary to human colour category learning (Ozgen & Davies, 2002) which has been shown to be influenced by language, there is evidence of extensive and long-term symbol learning needed for colour categories to emerge in non-human primates (Matsuzawa, 1985). These results indicate that monkeys in the current experiments may not have had well-informed colour categories and may have reflected that at test, as has also been shown for a chimpanzee with only two years experience of colour-symbol training (Matsuno et al., 2004). This is an indicator of how much previous experience monkeys may need for conceptual grouping to occur as probed for in experiments 2 and 3, showing that language is for them perhaps not as natural an abstraction as it is for human-primates. Thus, the current cross-species differences indicate that conceptual coding involving the use of symbols such as may be helped by acquiring linguistic terms is needed for colour categorization to take precedence over matching and learning by perceptual similarity.

7.2 Some cross-cultural linguistic influences in colour grouping

Using matching-to-sample as with the comparative study, results of similar grouping patterns have been found between the Himba and English that are in sharp contrast to the grouping behaviour of monkeys. Nevertheless, some small variations have also been found between the two cultures that are likely to be caused by language. The result of experiment 3 are in agreement with other findings of colour grouping on restrained (Davidoff, Davies, & Roberson, 1999; Roberson et al., 2005a) and less restrained tasks like colour sorting (Davies & Corbett, 1997; Davies, 1998; Roberson et al., 2005b) in some of which the Himba and English took part. All the more interesting is perhaps that matching-to-sample may be on the other side of the continuum in terms of task restraints than colour sorting, seeing that matching-to-sample has also been used as a psychophysical procedure (e.g. Sperling & Harwerth, 1971) probing perceptual mechanisms, and yet one still finds significant differences related to language.

Relating to similarities between the Himba and English response functions and matching-to-sample tasks also being used as psychophysical procedures, Helson (1964) has suggested an adaptation level theory which may also be applied to the current response pattern similarities of experiment 3. Adaptation-level theory suggests behaviour can be predicted from a stable context if one looks at the frames of reference given in an experiment. In the matching-to-sample of experiment 3 frames of reference were the middle of blue and green categories (2.5G and 10 B) and dumbu and buru categories (5Y and 2.5BG). These endpoints were the reference points which participants used to split the continua into two groups; one can see Himba and English participants have adapted to matching-to-sample levels of grouping and split the continuum into two in quite a similar way. Although these similarities may be accounted for by the adapting behaviour of participants in this task, there were significant differences based on conceptual levels of responding that can be perceived at the category boundaries.

Just as found for the cross-species difference in categorization, the differences are related to language because it is found for colours near category boundaries that have been found to be markers of conceptual grouping under the influence of language (Braisby & Dockrell, 1999; Roberson, Davidoff & Braisby, 2000). These regions have been argued to be salient in categorization with particular attention to categorical perception of colour (Bornstein & Korda, 1984) and in agreement, experiment 4 also shows a sharper decrease in responses for both cultures at the intersection between two colours that are identified by their names; for the English the sharper decrease is found at the blue-green intersection and for the Himba the same is found at the dumbu-buru colour intersection. It is interesting that monkeys showed a response function induced by two boundaries in their adaptation to this task versus humans showing typical categorical functions characterized by one boundary. This perhaps indicated that perceptual grouping behaviour as can be predicted from adaptation levels may also be influenced by language, resulting in a more typical categorical function induced by one boundary that characterizes a type of conceptual grouping as well.

Thus, even in a constrained task as is matching-to-sample versus other tasks that may be less demanding (Roberson et al., 2005b), one still finds evidence of linguistic difference for colours near the boundary in experiment 3 between the Himba and the English, and this evidence is all the more compelling seeing that adaptation behaviour would predict perceptual rather than conceptual behaviour.

7.3 Conceptual grouping influenced by language in developmental work on colour.

In the developmental work of experiment 4 and 5, categorical grouping behaviour has only been found among children who have colour terms needed to establish a blue-green and blue-purple boundaried two alternative-forced-choice task. These results were unlike published work by Franklin et al. (2005) that found colour categories for all children irrespective of their culture and colour term acquisition. The differences between the studies were marked by the different criteria used to determine whether children knew their colour terms with particular attention to the comprehension measure in experiment 4, and also by using two boundaries rather than one for the cross-cultural comparison in experiment 5. Moreover, the current results of experiment 4 and 5, showing no parsing of the continuum unless a toddler has acquired colour terms (as also suggested for the blue-purple continuum for the Himba toddlers), is also not in agreement with previous evidence of infant colour categories suggesting innate English categories (Franklin and Davies, 2004; Franklin, Pilling & Davies, 2005; Bornstein, Kessen & Weiskopf, 1976). With particular relevance to Franklin et al. (2005), the supporting evidence for innate categories has not been achieved due to our partial non-replication in experiment 4 and experiment 5.

Experiment 4 also repeated findings of research (e.g., Carey, 1985; Gelman & Markman, 1986; Keil, 1989; Rips, 1989; Soja, 1994; Sandhoffer & Smith, 1999) showing that categorization is a special matter that evokes the relations between attributes and more abstract knowledge rather than perceptual knowledge. This conclusion is made in the context of categorization being of particular difficulty that suggests biological/perceptual mechanisms are not sufficient to categorize. For instance, tardy acquisition of colour terms has been shown (Shatz et al., 1996; Soja, 1994; Andrick & Tager-Flusberg, 1986; Rice, 1980; Carey, 1978) compared to other categories such as for example shape (Clark, 1987; Baldwin, 1989). Experiment 4 also shows that colour term acquisition begins from the age of 3, and 9 terms are achieved at the age of 4.5. As to what causes the delay of full colour term acquisition of 11 terms, some researchers (Sandhoffer & Smith, 1999, Soja, 1994) have argued a lack of being able to abstract or filter the necessary information required to make the colour word meaningful and that this comes slowly because one needs to learn deeper relational properties than perceptually similar attributes (Medin, Wattenmaker & Hampson, 1987) suggesting a conceptual grouping problem. Elsewhere, the meaning of colour terms has also been pointed out as being a special case because the boundaries between colours are not as sharp and definite as they are for other categories of natural kinds indicating that those are the consequences of semantic processing triggered by colour names involved in colour categorization (Braisby & Dockrell, 1999).

The results of experiment 4 agree that unless a child comprehends their colour terms they will not respond more accurately to colour crossing a category boundary indicating that a semantic level of naming and comprehension is required to do so. The latter is further supported by previous findings of cross-cultural differences and colour category acquisition linked to colour names (Roberson et al, 2004). Those Himba children's behaviour concurred with findings of experiment 5 for Himba children that did not express a differential response for between cross- and within-category trials on the bue-green continuum. These results further agree with linguistic relativity for the same cross-cultural comparison with adults (Roberson et al., 2005) and for other areas of research that have tested children's spatial cognition (Brown & Levinson, 2000). Not in the least, experiment 4 and 5 are also compatible with the comparative work in experiments 1 and 2 because they further emphasize the crucial role of language for colour categories.

7.4 Cross-cultural findings of conceptual grouping of animal patterns influenced by language.

For the first time, we report categorical perception depending on having animal pattern terms. The Himba in comparison to the English, were the only culture to show categorical grouping of animal patterns. This result further supported linguistic relativity because the Himba, who had not acquired animal pattern terms in line with naming data, behaved just like the English using a perceptual mechanism to do the task.

In agreement with cultural significance of having animal patterns (Turton, 1980; Eckl, 2000) for e.g., animal patterns being indicators of wealth and ceremonial use, and previous reports of linguistic relativity (Roberson et al, 2005), one can say that Himba animal pattern categories are a regularity for them in their everyday environment with a strong need to communicate about them that leads to strengthening boundary areas in this continuum, as shown from evidence of categorization with robots and their need to communicate for optimal colour categories to occur, where shared regularities in the environment have not coincided in shared colour categories (Steels & Belpaeme, 2005; Introduction, section 1.5.4). Our results would seem to model discriminative and communicative success achieved through language i.e., communicative interaction is needed over and above learning from a shared environment just as the robots show in their communicative intentions to achieve colour categorization that is apparent in humans. Himba's discriminative and communicative success also points to The Paradox of The Heap for further explanation of what is meant by the importance of communication (Introduction, section 1.3.3). Thus, without communication it is to have more than a vague notion of animal pattern categories. This is consistent English participant's perception of animal categories, but also with the Himb vague and more robust animal pattern categories as found in experiment 6.

We know from the animal pattern results that the setting of boundaries resultin process of communication shows strong conceptual grouping in accordan categorical grouping hallmarks (Bornstein & Korda, 1984). Moreover, in experi the boundaries have been established through naming, and had they been established one might expect the Himba to be as confused as the English were in alternative-forced-choice tasks that followed. The fact that most Himba w confused about the boundaries and showed categorical grouping, indicates that na essentially important (Braisby & Dockrell, 1999; Roberson, Davidoff & Braisby Roberson et al., 2005). Perceptual grouping, in this case seen in the English and the Himba for some continua, involves perceptual similarity or local grouping these categories, which shows some parallels with monkey behaviour of experi and children who have not acquired their colour terms in experiment 4. It is the safe conclusion to suspect categorical grouping is influenced by language fo patterns as it is for colour.

7.5 Implications for universality

Universalists have shown that language is not a necessary pre-condition for ca and suggest categories are innate rather than learned (Franklin & Davies, 2004, Pilling & Davies, 2005), despite cultural differences (Berlin & Kay, 1969; Rosch Kay & Regier, 2003, 2006b), and irrespective of species-differences (Sandell, Bornstein, 1979).

Common colour categories across cultures have been found (Rosch-Heider, 1 Rosch-Heider & Olivier, 1972) that have pioneered a universal approach to categories based on prototypes. However the research has now been mar misrepresentation of some differences among cultures that were interpreted in acc with colour universals (Saunders & Van Brakel, 1997; Roberson, Davies & Da 2000). For example, the Dani showed extremely poor performance on their rec for English foci and to conclude that there are universal foci seems to discount and Dani performance and perceptual differences of these foci. Had this issu addressed, one may also have expected further comparison for colours for which tl may have been expected to have readily available labels with the prediction that tl may have performed better in this case. Recent evidence in colour research has s shift from absolute universality, showing the same colour categories across cultur for the latter part complex statistical similarities, to a more relaxed form of u grouping (Kay & Regier, 2006a) opening possibilities for acknowledgment differences among cultures as is seen by for e.g., pointing out 70 % commun cultures of 5-colour terms, yet leaving 30% worth of cultural differences leadin more cautious conclusions about universality. This thesis accommodates some universal findings in that it partially agrees that our biology plays a role in our per grouping behaviours because there are biological constraints of perceptual simil the grouping behaviour of experiment 1 and the adaptation behaviour of experime both Himba and English cultures. However, significant differences are found cultures in experiment 4 for colours near the boundary, and in experiment 7 for pattern categories that do not support universal categories. Moreover, in all the cultural experiments of this thesis, Himba and English showed significant differ performance levels from chance, indicating genuine relativity that is not confour task difficulty and the animal pattern task of experiment 7 gives another approaching cross-cultural research that looks at a comparison from the perspecti experimenters' foreign language as has been previously done for colour (Roberso 2005a).

Universality is nevertheless still strongly advocated by researchers that have evidence of innate categories for human- and non-human primates (Bornstein, Ke Weiskopf, 1976; Sandell, Gross & Bornstein, 1979; Franklin, Pilling & Davies, Although these pioneering experiments in the 1970s have been re-addresse conclusions replicated (e.g., Franklin & Davies, 2004), the same has not been d the comparative work. Thus although the same cannot be said of monkey categor infant abilities have been found to be largely similar to those of adults and have n reinforcement from other areas of research such as numerical abilities (Spelke &

2007). However, this thesis does not find evidence pointing in this direction. Exp
4 does not find the same infant abilities in toddlers that were similarly not able t
their colour categories. A particularly sharp statement against innateness is als
when not finding colour categories in monkeys in experiments 1 and 2. Even
cross-cultural colour categories in toddlers do not fully correspond to English
categories (blue-green), and one wonders about innate colour categories (Franklin,
& Davies, 2005) being universal. Therefore it seems that universality is not accoun
by innate colour categories because if it were one would expect the same p
categories across and within cultures, and also across species. One would also not
from innate colour categories, which are shown to be the same in infancy as in adu
for those to be lateralised in the brain in relation to language processing (Gilber
2006). Language would not be needed if colour categories were innate and un
Perhaps one can say colour categories have evolutionary and therefore cultural
(McNamara, 2004) that make language important to their evolution; colour cat
may be influenced by culture in an evolutionary sense and our knowledge may
across time under that influence. Perhaps an evolutionary explanation, tending t
increased social intelligence needed in a more populated environment, makes la
all the more important (Dunbar, 1996). This could also explain the strange
Darwin's (1977; in Bornstein, 1987) reports of children not having acquired categ
the age of 7 years, when now children acquire their categories between the ages o
6 years (Roberson et al., 2004) in accordance with experiment 4. Elsewhere, t
separating us from other non-human primate species in terms of different evolu
paths due to differing cognitive adaptations (Bartow, Cosmides & Tooby, 1
Cartwright, 2000) can also be inferred in experiment 1 and 2 due to findings o
species differences.

In conclusion, this thesis takes the view of a limited universal grouping behavi
may have biological origins for perceptual mechanisms and rather more evolu
origins influenced by language and culture for conceptual mechanisms that woul
infant categories, cross-species differences, and cross-cultural differences to e
More precisely, language skills that affect colour categories are human skills that

found in other species, nor are they identical for all toddlers or adults within and a cultures, and this thesis disagrees with those that conclude otherwise.

7.6 Implications for relativity

For relativity, conceptual grouping is of particular interest because this behavio influenced by language and therefore not innate as shown by the boundaries that mar categories seen in all experiment of this thesis. Therefore results of this thesis a agreement with previous cross-cultural findings (Davidoff, Davies & Roberson, 1 Roberson et al., 2004, 2005a) and coincide with other relativist findings (Levinson, 1 Boroditsky, 2001) that categories are conceptually formed with the help of language. This thesis could endorse a strong view of the Whorfian hypothesis (Whorf, 1956; 1 1992) seeing that where conceptual grouping is found, linguistic influence is also i evident in its most absolute form. It is partially the case that this thesis finds lang influences grouping behaviours even when under a perceptual constrained tasks su matching-to-sample (experiments 1 and 3). Moreover, monkeys do not show Who behaviour in this task indicating that matching-to-sample requires human abiliti show a categorical function that includes one boundary rather than two. Additior differences can also be strictly attributed to language in experiment 4 and 6, becau both colour and animal pattern studies there are differences found among infor belonging to the same culture who know and don't know their terms. In the same f of investigation as the latter, cross-cultural differences between Himba and En categories are found in experiment 5 and 6 and additionally those effects are also f in experiment 3. Thus, this thesis shows no doubt that language is involved in the pr of concept formation and production for colour and animal patterns.

There are also reasons for abstaining from a strong Whorfian view i.e. that lang influences category formation but it is not the only influence. Clearly, experiment 3 shows similarities in grouping of a non-Whorfian kind; humans use similar grou strategies clearly not permeated by language. Experiment 5 also shows some grou similarities on the blue-purple continuum, and although this seems a coincidence d possible animal pattern naming and blue and purple designated English colour terms is not sure of the coinciding categories' nature.

Nevertheless, other influences have been noted that make grouping relative to experie
and are not any different from the implication and suggestions linguistic relativity ma
about categories being nurtured and moulded by experience. Therefore when diffe
cultural experiences are noted, different linguistic expressions are also found to be a
origin of the variation, as seen in the examples of the next paragraph that an
accordance with experiment 3, 4, 5, and 6 of this thesis.

Levinson suggests spatial cognition is affected by frames of references in space that
relative to expression of orientation in different cultures e.g., differences found betw
Tzeltal Maya with an Absolute frame of reference, and the Dutch found to have a rela
frame of reference whereby cross-cultural differences are found on a spatial alignm
task (Majid et al., 2004). Boroditsky (2001) and colleagues (e.g., Cassasanto
Boroditsky, 2003; Cassasanto et al., 2004) showed that grammar affects memory
different objects in different cultures and also suggested that time may be perce
differently due to different cultural experience and in particular the linguistically diffe
expressions found in Mandarin where time is perceived vertically and in English w
time is perceived horizontally (Boroditsky, 2001). Findings of experiment 6 may als
in agreement with the influence of culture on cognition as suggested by Nisbett
colleagues (Nisbett et al., 2001; Nisbett & Miyamoto, 2005) who have studied
differences in attentional patterns and their link to the origin of afforded cult
differences. The origin has mainly been argued as one of socialisation and indivi
difference. For example, the way cultures talk about self may impact on the way
plays attention to other objects and individuals (Markus & Kitayama, 1991). This ma
well with acculturation of the Himba and the relation they have to their herds.
example, a lot of Himba are given cattle names such as Gongoro meaning 'spotted li
leopard' (Eckl, 2000) indicating the spotted animal pattern, or Vahekonde, meaning 'l
coloured and speckled' (Eckl, 2000), and in this case indicating a combination
pattern of off-white colour and speckled. This may lead to a close relationship betw
self and animal patterns, not in the least because wealth is described in terms of
number of animals a particular family has. The latter may also be a clue to a n
thorough outlook on animal patterns that is indicative of the need for expertise in
area. Cultural practices have also been found to extend to differences in the environ

that may lead to an attention pattern only found in the host culture; Americans have been found to be analytic or field independent and Asians are described as holistic or field dependent in various cognitive tasks such as the Framed-line Test (Kitayama et al., 2003). Perhaps it is the case that the closer relationship to nature/animals found in the Himba way of life makes the Himba consider their context more closely (depending on their herds for survival) as the Asians do in showing holistic or field dependent attention due to their strong sense of interdependence. This type of acculturated attention can also be characterised as grouping behaviour of a particular kind that allows one to move from perceptual similarity to conceptual grouping as found in the discrete grouping tendencies of the Himba in experiment 6. Although findings fit a cultural theory, it is argued, that findings point to culture closely linked to language too, because experiment 6 also shows evidence of linguistic influence between Himba who know and don't know their colour terms emphasising the importance of language affecting category boundaries within the same culture. An identical finding has been reported for colour cognition in English toddlers. Thus without language one cannot reinforce the shared and acculturated attention one has for the environment (see section 7.4).

Furthermore research that supports linguistic relativity in other areas such as time (Boroditsky, 2001) and colour cognition (Roberson et al., 2004, 2005a) also indicate the tight bond between cultural and linguistic experience, and the fact that this tight bond may also be affected among people of the same culture indicating communication is so important (e.g., colour cognition: Agrillo & Roberson, 2007). Thus, categorization is concluded to be relative to culture and language in nature and origin rather than a mechanism adhering to fixed and innate grouping patterns.

7.6.1 A three-pronged investigation's implications for relativity versus conclusions of previous two-pronged investigations

The thesis contains a three-pronged investigation to address the nature of colour categories. This type of investigation has only ever been approached using a two-pronged enquiry. The current results' implications for linguistic relativity are held in comparison to these previous two-pronged investigations. The first compared ontogenetic and phylogenetic work (Bornstein, 1980), and found common colour categories across species indicating a common innate origin for monkeys and humans. This thesis does no have the same conclusions because differences between species have been found (experiment 1 and 2), and evidence of blue and green infant colour categories (Bornstein et al., 1976) would clearly not be in line with monkey behaviour found in this thesis. Nor would these findings be in line with evidence from children that have not acquired their colour terms (experiments 4 and 5). The second two-pronged investigation compared ontogenetic and cross-cultural work (Franklin, 2003) and again found common colour categories, indicating colour categories are innate. From children (experiment 4) we find colour categories are not present unless names are learnt from language, and the same can be said from cross-cultural work (experiments 3, 5 and 6), indicating language is a cultural and conceptual mechanism ensuring categorical grouping with no signs of innate colour categories. The third kind of two-pronged approach is in accordance with current results and is suggested in papers addressing cross-cultural and developmental differences (e.g., Levinson, 1996; Brown & Levinson, 2003; Roberson et al., 2004 2005a). The results of these studies agree with current developmental and adult work showing differences in cognition that are under linguistic influence. Importantly, the current reported findings of adult colour categories (experiments 1, 2 and 3) culturally correspond to those found for children who have acquired their colour terms (experiments 4 and 5). Thus, this three-pronged investigation has implications for linguistic relativity and from additional information obtained from cross-species categorization differences does not conclude with a similar argument to the unanimous universal and innate categories found from previous two-pronged investigations.

7.7 Future research

Suggestions are made for future improvement of the current findings and conclusions First, the three-pronged investigation could be ameliorated by adopting an identical task for all populations to make comparisons more effective. For instance the matching-to-sample could be adopted with young children. This remains to be investigated with the positive consequence that three-way i.e. phylogenetic, ontogenetic and cross-cultura comparisons might bring fruitful additional information by comparing three populations that may show a lack of colour terms or show none at all like with the monkeys. It is interesting that Himba show a weak cubic function just like the English on parts of the continuum that are unfamiliar, and if grouping is inherent to humans, as it is not to monkeys, one wonders at what age this kind of grouping behaviour is established. The outcome of children at a young age on matching-to-sample might give further evidence of the origins of the grouping behaviour perceived in this task.

Second, one might want to pursue grouping behaviour in matching-to-sample to find ou whether the same grouping i.e. adaptation behaviour is evident when changing the context (Helson, 1964, see section 7.2). The context of colours could be changed to spar 3 categories so that category compression as part of categorical perception for within category trials may also be predicted and further conclusions may be made abou language's influence under this context. The line of investigation is also worth pursuing with the idea of further investigating how perceptual grouping may be flexible in this kind of task.

Third, one might adopt tasks that minimize language use even further than in this thesis where the 2AFC method is used, such as visual search (e.g. Nagy & Sanchez, 1990 Wolfe, 1998; Daoutis, Pilling & Davies, 2006) with minimal response time, so that one can further ensure language's influence on cognition by minimizing direct effects o language (e.g., Witthoft et al., 2003). In the animal pattern task one could argue that the Himba made access to pattern names and therefore showed linguistic effects. Although this is highly unlikely due to short response time given in the 2 alternative-forced-choice this possibility cannot be ruled out completely. Regardless, whether language can ever be ruled out completely is a philosophical matter that for some researchers does not seem possible (Dedrick, 1996; Saunders &Van Brakel, 1997; Braisby & Dockrell, 1999 indicating that this avenue may perhaps be futile seeing that there are non-trivial constraints such as language on colour categories that leave their traces in memory by for e.g. affecting the boundaries as discussed previously. Although this thesis agrees with non-trivial constraints on colour categories that might extend to other categorization instances, this suggestion is fitting in the sense that it might be in agreement with the results of experiment 7 and further support linguistic relativity.

Fourth, a gap in the nature and origin of grouping behaviour is cross-cultural infant work, presumably because infant work is a lengthy process that requires patient experimental procedure, and from a relativist point of view one might be chasing null results. However, null results may serve a purpose in questioning Franklin Pilling & Davies's (2005) evidence of infant categories and are worth carrying out. Related to cross-cultural investigation, further work with different populations (other 5-colour term populations as in Kay & Regier, 2006a) might continuously elaborate on cross-cultural differences. So far in the area of colour, the Dani of Papua New Guinea, and the Himba of Namibia have participated in cross-cultural investigation that have shown linguistic relativity between those two cultures as well as in comparison to the English (Roberson et al., 2005a) . It would be interesting to replicate the latter findings by further research of other cultures that Kay & Regier (2006a) point universal patterns for, such as the Yaminahua of Peru and Jicaque of Honduras, to perhaps find further differences between those cultures.

Last but not least, testing bilingual Himba who lead a more secular lifestyle might provide us with more information about language's role in the formation of categories. For example, evidence can be sought of primary and secondary linguistic influence (Boroditsky, 2001; with Mandarin and English) and the outcomes of learning at different ages to find out whether there is anything special about the tardy more uniform acquisition found in the West (when a child / adolescent started going to school might vary a lot among the Himba).

8. Conclusion

Having suggestions for future research does not outweigh the fact that for the first time, we have brought together a three-pronged investigation of grouping behaviour that has evidenced the use of perceptual and conceptual mechanisms in different situations. Rather than adhering to absolute linguistic relativity (Lucy, 1992) as initially suggested by Whorf (1956), a weaker form of the Whorfian hypothesis (Davidoff, Davies & Roberson, 1999; Roberson, Davidoff & Braisby, 2000) fits with findings of universal perceptual grouping tendencies and conceptual grouping influenced by language.

In conclusion:

1. The phylogenetic work shows cross-species differences that set humans and monkeys apart in their grouping of colour (experiment 1 and 2) and further shows conceptual grouping is a human skill (experiments 1 and 3).
2. The ontogenetic work shows language influences colour grouping, and does so for children of the same culture (experiment 4) and for children from different cultures (experiment 5).
3. Some cross-cultural work shows inherent grouping behaviour (experiment 3, 5). However, those findings also reveal more specific evidence of category expansion and suppression that are the hallmarks of linguistic influence on categorization (experiment 3, 4, 5 and 6).

References

Abramov, I. & Gordon, J. (1994). Color Appearance: On seeing red- or yellow, or green, or blue. *Annual Review of Psychology, 45*, 451-485.

Adams CK, Bryan AH, Jones AE (1968) Electroretinographic determination of the spectral sensitivity of the baboon (Papio anubis). *Vision Research, 8*, 1399-1405

Agrillo, C. & Roberson, D. (2007). Colour memory and communication in English speakers: Are focal colours easier to name and remember*? International Council of Psychologist* Proceedings, 2007.

D'Amato, M., Salmon, D. & Colombo, M. (1985). Extent and limits of the matching concept in monkeys (cebus paella). *Journal of Experimental Psychology: Animal Behavior Processes, 11(1),* 35-51.

D'Amato, M.R. & van Sant, P. (1988). The person concept in monkeys (Cebus apella). *Journal of Experimental Psychology: Animal Behavior Processes, 14,* 43-55.

Andrick, G.R. & Tager-Flusberg, H. (1986) The acquisition of colour terms. *Journal of Child Language 13*, 119-134.

Ashby, F.G. (1992). Multidimensional models of categorization. In F.G.Ashby (Ed.), *Multi-dimensional models of perception and cognition* (pp. 449-483). Hillsdale, NJ: Erlbaum.

Baldwin, D.A. (1989). Establishing word-object relations: A first step. *Child Development, 60,* 381-398

Belpaeme, T. (2001). Reaching coherent color categories through communication. In Kröse, B. et al. (eds.), *Proceedings of the 13th Belgium-Netherlands Conference on Artificial Intelligence*, Amsterdam: Netherlands. p 41-48.

Berlin, B. & Kay, P. (1969). *Basic color terms: Their universality and evolution.* Berkeley and Los Angeles: University of California Press.

Berlin, B., Kay, P. & Merrifield W. R. (1991). Biocultural implications of systems of colour naming. *Journal of Linguistic Anthropology, 1(1),* 12-25.

Boas, F. (1940). *Race, Language, and Culture.* Chicago: University of Chicago Press.

Bornstein, M. H. (1975). Hue is an absolute code for young children. *Nature, 256*, 309-310.

Bornstein, M. H. (1976). Infants' recognition memory of hue. *Developmental Psychology*, 12(3), 185-191.

Bornstein, M. H. (1980). *Comparative methods in psychology*. New Jersey: Hillsdale.

Bornstein, M. H. (1987). *Perceptual Categories in vision and audition*. In Harnad, S. R., Categorical perception: The groundwork of cognition. Cambridge: Cambridge University Press.

Bornstein, M. H., Kessen, W. & Weiskopf, S. (1976). Color Vision and hue categorization in young human infants. *Journal of Experimental Psychology: Memory, Perception and Performance*, 2(1), 115-129.

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

Bornstein, M. H. & Korda, N. O. (1985).Identification and adaptation of hue: parallels in the operation of mechanisms that underlie categorical perception in vision and audition. *Psychological Research*, 47, 1-17.

Bornstein, M. H., & Monroe, M. D. (1980). Chromatic information processing: rate depends on stimulus location in the category and psychological complexity. *Psychological Research, 42,* 213-22

Boroditsky, L. (2000). Metaphoric structuring: understanding time through spatial metaphors. *Cognition*, 75, 1-28.

Boroditsky, L. (2001). Does language shape thought?: Mandarin and English speakers' conceptions of time. *Cognitive Psychology*, 43, 1-22.

Boroditsky, L., Ramscar, M., & Frank, M. R. (2001). The roles of body and mind in abstract thought. *Proceedings of the 23^{d} Annual Meeting of the Cognitive Science Society.*

Boroditsky, L., Schmidt, L.A., & Phillips, W. (2003). Sex, syntax and semantics. In Gentner, D. & Goldin-Meadow, S. Eds., *Language in Mind.* Cambridge, London: MIT Press.

Bowerman, M. (2000). Where do chidren's word meanings come from? Rethinking the role of cognition in early semantic development. In Nucci, L. P., Saxe, G. B., & Turiel, E. Eds. *Culture Thought and development.* New Jersey, London: Lawrence Erlbaum Assosiates Publishers.

Bowerman, M. & Choi, S. (2001). Shaping meanings for language: universal and language-specific in the acquisition of spatial semantic categories. In Bowerman, M, & Levinson S.C. Eds. *Language Acquisition and Conceptual Development.* Cambridge: Cambridge University Press.

Bowmaker, J. K., Astell, S., Hunt, D. M., & Mollon, J. D. (1991). Photosensitive and photostable pigments in the retinae of Old World monkeys. *Journal of experimental Biology, 156*, 1-19.

Boynton, R. M. (1997). Insights gained from naming the OSA colors. In Hardin, C. L., & Maffi, L. (Eds.) *Color Categories in Thought and Language*. Cambridge: Cambridge University Press.

Braisby, N. & Dockrell, J. (1999). Why is colour naming difficult? *Journal of Child Language, 26,* 23-47.

Brown, R. (1976). Reference in memorial tribute to Eric Lenneberg. *Cognition*, 4, 125-153.

Brown, R. & Lenneberg, E. H. (1954). A study in language and cognition. *Journal of Abnormal Social Psychology*, 49, 454-462.

Brown, P. & Levinson, S.C. (1993). Uphill and downhill in Tzeltal. Journal of *Linguistic Anthropology,* 3(1), 46-74.

Brown, P. & Levinson, S.C. (2000). Frames of spatial reference and their acquisition in Tenejapan Tzeltal. In Nucci, L. P., Saxe, G. B., & Turiel, E. Eds. *Culture Thought and Development.* New Jersey, London: Lawrence Erlbaum Assosiates Publishers.

Cangelosi, A. & Harnad, S. (2001) The Adaptive Advantage of Symbolic Theft Over Sensorimotor Toil: Grounding Language in Perceptual Categories. Evolution of Communication 4(1) 117-142

Carey, S. (1978) The child as word learner. In Bresnan, Miller & Halle (eds.), *Linguistic theory and psychological reality.* Cambridge, MA: MIT Press, 264-293.

Carey, S. (1985). *Conceptual change in childhood.* Cambridge, MA: MIT Press.

Carey, S. (2000). The origin of concepts. *Journal of Cognition and Development, 1*, 37-41.

Cartwright, J. (2000). *Evolution and human behaviour.* MA: MIT Press

Cassasanto, D. & Boroditsky, L. (2003). Do we think about time in terms of space? *Proceedings of the 25th Annual Meeting of the Cognitive Science Society.*

Cassasanto, D., Boroditsky, L., Phillips, W., Greene, J., Fotokopoulo, O., Pita, R. & Gil, D. (2004). How deep are effects of language on thought? Time estimation in speakers of English, Indonesian, Greek, and Spanish. *Proceedings of the 26th Annual Meeting of the Cognitive Science Society.*

Cassirer, E. (1946). *Language and myth.* NY: Dover Publications Inc.

Choi, S. & Bowerman, M. (1991). Learning lo express motion events in English and Korean: The influence of language-specific lexicalization patterns.*Cognition, 42,* 83-121.

Choi, S., McDonough, L., Bowerman, M. & Mandler, J. M. (1999). Early sensitivity to language-specific spatial categories in English and Korean. *Cognitive Development,* 14, 241-268.

Choplin, J. M., & Medin, D. L. (1999). Similarity of the perimeters in the Ebbinghaus illusion. *Perception & Psychophysics, 61,* 3–12.

Coren, S., & Enns, J. T. (1993). Size contrast as a function of conceptual similarity between test and inducers. *Perception & Psychophysics, 54,* 579–588.

Coren, S., & Miller, J. (1974). Size contrast as a function of figural similarity. *Perception & Psychophysics, 16,* 355–357.

Clark, E. V. (1987). The principle of contrast: A constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition.* London: Erlbaum

Cook, R. S., Kay, P. & Regier, T. (2005). The world color survey database: history and use. In Cohen, H. & Lefevbre, C. Eds. *Handbook of categorization in cognitive science.* Elsevier.

Cowey, A, Heywood, C.A. & Irving-Bell, L. (2001). The regional cortical basis of achrmatopsia: a study on macaque monkeys and an achromatopsic patient. *European Journal of Neuroscience, 14,* 1555-1566.

Crandall, D. P. (2000). The place of stunted ironwood trees: A year in the lives of the cattle herding Himba of Namibia. New York: Continuum-Academy.

Daoutis, C. A., Franklin, A., Riddett, A., Clifford, A. & Davies, I. (2006). Categorical effects in chidren's colour search: A cross-linguistic comparison. British Journal of *Developmental Psychology*, 24, 373-400.

Daoutis, C. A., Pilling, M. & Davies, I. R. L. (2006). Categorical effects in visual search for colour. *Visual Cognition, 14(2),* 217-240.

Davidoff, J. (1975). *Differences in visual perception.* London: Academic Press

Davidoff, J. (1997). The neuropsychology of colour. In Hardin, C. L. & Maffi, L. (eds.) *Colour categories in thought and language*. Cambridge: Cambridge University Press.

Davidoff, J. (2001). Language and perceptual categorization. *Trends in Cognitive Sciences,* 5(9), 382-387.

Davidoff, J., Davies, I. & Roberson, D (1999). Colour categories in a stone-age tribe. *Nature,* 398, 203-204.

Davies, I.R. (1998).A study of colour grouping in thee languages: A test of the linguistic relativity hypothesis. *British Journal of Psychology*, 89, 433-452.

Davies, I.R.& Corbett, G.G. (1997). A cross-cultural study of colour grouping: evidence for weak linguistic relativity. *British Journal of Psychology, 88(3),* 493-517.

Davies, I. & Franklin, A. (2002). Categorical similarity may affect colour pop-out of infants after all. *British Journal of Developmental Psychology*, 20, 185-203.

Depy, D., Fagot, J. & Vauclair, J. (1998). Comparative assessment of distance processing and hemispheric specialization in humans and baboons (*Papio papio*). *Brain and Cognition, 38*, 165-182.

Derefelt, G. (1991). Colour appearance systems. In Gouras, P. (ed.). *The Perception of Colour.* Basingstoke: Macmillan Press Ltd.

Deruelle, C. & Fagot, J. (1998). Visual search for global/local stimulus features in humans and baboons. *Psychonomic Bulletin & Review, 5(3),* 476-481.

Dummett, M. (1975). Wang's Paradox. *Synthese, 30*, 301-324.

Eckl, A.E. (2000). Language, culture and environment. The conceptualization of Herero cattle terms. In Bollig, M. & Gewald, J-B, (Eds.) *People, Cattle and Land. Transformatiom of a pastoral society in Southwestern Africa* (History, Cultural Traditions and Innovations in Southern Africa, Vol. 13). Cologne: Ruediger Koeppe, 401-431.

Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition*, 44(3), 227-240.

Evans-Pritchard, E. E. (1940). *The Nuer.* Oxford, England: Oxford University Press.

Fabre-Thorpe M., Richard G., Thorpe S. (1998) Rapid categorization of natural images by rhesus monkeys. *Neuroreport, 9* , 303-308

Fagot, J. & Cook, R. G. (2006). Evidence for large long-term memory capacities in baboons and pigeons and its implications for learning and the evolution of cognition. *Proceedings of the National Academy of Sciences of the United States of America, 103(46),* 17564–17567.

Fagot, J. & Deruelle, C. (1997). Processing of global and local visual information and hemispheric specialization in humans (Homo sapiens) and baboons (Papio papio). *Journal of Experimental Psychology: Human Perception and Performance,* 23(2) 429-442.

Fagot, J., Kruschke, J. K., Dépy, D. & Vauclair, J. (1998). Associative learning in baboons (Papio papio) and humans (Homo sapiens): species differences in learned attention to visual features. *Animal Cognition, 1(2),* 123-133.

Fagot, J., & Tomonaga, M. (1999) Global-local processing in humans (Homo sapiens) and chimpanzees (*Pan troglodytes*): Use of a visual search task with compound stimuli. *Journal of Comparative Psychology, 113*, 3-12.

Fagot, J., Wasserman, E. & Young, M. (2001).Discriminating the relation between relations: The role of entropy in abstract conceptualization by baboons and humans. *Journal of Experimental Psychology : Animal Behavior Processes, 27*, 4, 316-328.

Fahle M (2004) Perceptual learning: a case of early selection. Journal of Vision, 4, 879-890

Farroni, T., Valenza, E., Simion, F. and Umiltà, C. (2000). Configural processing at birth: Evidence for perceptual organisation. *Perception, 29,* 355–372

Fletcher, R. (1980*). The City University Colour Vision Test.* Windsor, Berks: Keeler, Ltd.

Fobes, J.L. & King, J.E. (1982). *Primate Behavior.* New York: Academic Press.

Franklin, A. (2003*). The origin and nature of Categorical Perception of colour.* PhD thesis, University of Surrey

Franklin, A. (2006). Constraints on children's color term acquisition. *Journal of Experimental Child Psychology. In Press.*

Franklin, A., Clifford, A., Williamson, E. & Davies, I. (2005). Colour term knowledge does not affect categorical perception of colour in toddlers. *Journal of Experimental Child Psychology, 90,* 114-141.

Franklin, A. & Davies, I. R. L. (2004). New evidence for infant colour categories. *British Journal of Developmental Psychology, 22,* 349-377.

Franklin, A., Pilling, M. & Davies, I. (2005). The nature of infant color categorization: Evidence from eye movements on a target detection task. *Journal of Experimental Child Psychology,* 91, 227-248.

Fujita, K. (1982). An analysis of stimulus control in two-colour matching-to-sample behaviors of Japanese monkeys (Macaca Fuscata). *Japanese Psychological Research, 24,* 124-135.

Gabriel, D. J. (2001). *The spirit of childhood.* Cambridge, UK: RSC Press.

Garcia-Pérez MA (1998) Forced-choice staircases with fixed step sizes : asymptotic and small-sample properties. *Visual Research, 38,* 1861-1881.

Gegenfurtner, K.R. & Kiper, D.C. (2003). Color Vision. *Annual Review of Neuroscience, 26,* 181-206.

Gelman, R. & Butterworth, B. (2005). Number and language: how are they related? *Trends in Cognitive Sciences, 9,* 6-10.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cogntion, 23,* 183-209.

Gennari, S. P., Sloman, S., Malt, B., & Fitch, W. T. (2000). Motion events in language and cognition. Cognition, 83, 49-79.

Gerhardstein, P., Renner, P. & Rovee-Collier, C. (1999). The roles of perceptual and categorical similarity in colour pop-out in infants. *British Journal of Developmental Psychology,* 17, 403-420.

Gilbert, A. L., Regier, T., Kay, P. & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Science,* 103(2), 489-494.

Glucksberg, S., Hay, A., & Danks, J. H. (1976). Words in utterance contexts: Young children do not confuse the meaning of same and different. *Child Development, 47,* 737–741.

Goldstein, K. (1948). *Language and Language Disturbance.* Grune & Stratton: New York.

Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology, General*, 123(2), 178-200.

Goldstone, R. (1998).Perceptual learning. *Annual Review of Psychology,* 49, 585-612.

Goldstone, R. L. & Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition*, 65, 231-262.

Goodale M.A.,& Milner A. D. (1992). Separate visual pathways for perception and action. *Trends in Neuroscience, 15*, 20–25.

Goodman, N. (1972). Seven strictures on similarity. In Goodman, N. (Ed.) Problems and projects. New York: Bobbs-Merrill

Gordon, P. (2004). Numerical Cognition without Words: Evidence from Amazonia. *Science,* 306, 496-499.

Gumpertz & Levinson, S. C. (Eds.) (1996). *Rethinking linguistic relativity*. Cambridge, UK: Cambridge University Press.

Haffenden, A. M. & Goodale, M. A. (2000). Independent effects of pictorial displays on perception and action. *Vision Research, 40*,1597–1607

Hampton, J.A. (1998). Similarity-based categorization and fuzziness of natural categories, *Cognition, 65,* 137-165.

Hampton, J.A. (2005). Rules and similarity - a false dichotomy. *Behavioral and Brain Sciences, 28,* 26-27.

Hanazawa, A. Komatsu, H. & Murakami, I. (2000). Neural selectivity for hue and saturation of colour in the primary visual cortex of the monkey. *European Journal of Neuroscience, 12,* 1753-1763.

Harnad, S. (1987). Psychophysical and cognitive aspects of categorical perception: A critical overview. In S. Harnad (Eds), *Categorical Perception: The groundwork of cognition.* Cambridge: Cambridge University Press.

Helson, H. (1964). *Adaptation- level theory.* NY; Harper & Row.

Herrnstein R. J., Loveland D. H. (1964). Complex Visual Concept in the Pigeon. *Science, 146*, 549-551.

Heywood, C.A., Cowey, A. & Newcombe, F. (1991). Chromatic discrimination in a cortically color blind observer. *European Journal of Neuroscience, 3*, 808-812.

Holmes, P. W. (1979). Tranfer of matching performance in pigeons. *Journal of Experimental Analysis of Behavior, 31*, 103-114.

Hopp, S. L., Sinnott, J. M., Owren, M. J., & Petersen, M. R. (1992). Differential sensitivity of Japanese macaques (Macaca fuscata) and humans (Homo sapiens) to peak position along a synthetic coo call continuum. *Journal of Comparative Psychology, 106*, 128-136.

Horner, v.; Bonnie, K., & De Waal, F. B. M. (2005). Identifying the motivations of chimpanzees: Culture and collaboration. *Behavioral and Brain Sciences, 28,* 704-704.

Horner, V. & Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (*Pan troglodytes*) and children (*Homo sapiens*). Animal Cognition, 8(3), 164-181.

Huttenlocher J, Hedges LV, Vevea JL (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General, 129(2),* 220-41.

Humboldt, W. (1836). *On language: The diversity of human language-structure and its influence on the mental development of mankind.* In Gentner, D. & Goldwin-Meadow, S. (Eds.) (2003). Language in mind: advances in the study of language and thought. Massachusetts: MIT Press.

Imai, M., & Gentner, D. (1997). A cross-linguistic study of early word meaning: universal ontology and linguistic influence. Cognition, 62(2), 169-200.

Imai, M. & Mazuka, E. (1997). Learning Proper Nouns and Common Nouns without Clues from Syntax. *Child Development*, 72(3), 787-802.

Indow, T. (1988). Multidimensional studies of the Munsell color solid. *Psychological Review, 95*, 456-470.

Ishihara S (1998) *Ishihara's Tests for Color Blindness*. Hodder & Stoughton Educational, London,UK

Jackendoff, R. (1990). *Semantic structures*. MIT Press: Cambridge, MA.

Jameson, K.A. (2005). Culture and Cognition: What is Universal about the Representation of Color Experience? *Journal of Cognition & Culture, 5*, 293-347.

Jameson, D. & Hurvich, L. M. (1989) Essay concerning colour constancy. *Annual Review of Psychology, 40,* 1-22.

Johnson, E. G. (1977). The development of color knowledge in preschool children. *Child Development, 48,* 308-311.

Karmiloff-Smith, A. (1979). *A functional approach to child language: A study of determiners and reference.* Cambridge, UK: Cambridge University Press.

Kay, P., Berlin, B. & Merifield, W. R. (1991). Biocultural implications of systems of color naming. *Journal of Linguistic Anthropology, 1(1),* 12-25.

Kay, P., Berlin, B., Maffi, L., & Merifield, W. R. (1997). Color naming across languages. In Hardin, C. L., & Maffi. L. (Eds.) Color categories in thought and language. Cambridge: Cambridge University Press.

Kay, P., Cook, R.S., & Regier, T. (2005). Focal colours are universal after all. *Proceedings of the National Academy of Science.* 102 (23), 8386-8391.

Kay, P & Kempton, W. (1984). What is the Sapir-Whorf hypothesis? *American Anthropologist,* 86(1), 65-79.

Kay, P., & Maffi, L. (1999). Color appearance and the emergence and evolution of basic color lexicons. *American Anthropologist, 101,* 743-760.

Kay, P. & McDaniel, C. (1978). The linguistic significance of the meaning of basic colour terms. *Language,* 54(3), 610-646.

Kay, P. & Regier, T. (2003). Resolving the question of colour naming universals. *Proceedings of the National Academy of Science.* 100 (15), 9085-9089.

Kay, P. & Regier, T. (2006a). Colour naming universals: The case of the Berinmo. *Cognition.* In press.

Kay, P. & Regier, T. (2006b). Language though and colour: Recent developments. *Trends in Cognitive Sciences.* In press.

Keil, F. C. (1989). Spiders in the web of belief: The tangles relations between concepts and theories. *Mind and Language, 4,* 43–50

Kimchi, R. (1988). Selective attention to global and local levels in the comparison of hierarchical patterns. *Perception & Psychophysics, 43,* 189-198.

Kitayama, S., Duffy, S., Kawamura, T. & Larsen, J. (2003). Perceiving an object and its context in different cultures: A cultural look at New Look. Psychological Science, 14(3), 201-206.

Lamberts, K. (1997). Process models of categorization. In: Lamberts, K. & Shanks, D. (Ed.). *Knowledge, concepts, and categories*. Psychology Press.

Lammens, J. M. (1994) A computational model of color perception and color naming. PhD thesis, State University of New York, Buffalo.

Lantz, D. & Streffle, V. (1964). Language and cognition revisited. *Journal of Abnormal Social Psychology*, 69, 472-481.

Lenneberg, E. H. (1957). A probalistic approach to language learning. *Behavioural Science,* 2, 1-12.

Lenneberg, E. H. (1967). *Biological foundations of language*. New York, NY: Wiley.

Lenneberg, E. H. & Roberts, J. M. (1956). The language of experience: A study in methodology. *International Journal of American Linguistics* (Memoir no 13).

Levinson, S. C. (1996a). Language and space. *Annual Review of Anthropology*, 25, 353-382.

Levinson, S. C. (1996b). (1996). Frames of reference and Molyneaux's question: Cross-cultural evidence. In P. Bloom, M. Peterson, L. Nadel & M. Garrett (eds.), *Language and Space* . MIT Press.

Levinson, S. C. (1997). Language and cognition: the cognitive consequences of spatial description in Guugu Yimithirr. *Journal of Linguistic Anthropology*, 7(1), 98-131.

Levinson, S. C. (2003). *Space in language and cognition: Explorations in cognitive diversity.* Cambridge University Press: Cambridge, UK.

Levinson, S. C., & Schmitt, B. (1993). *Animals in a row.* In Cognition and Space Kit Version 1.0 (pp. 65–69).Nijmegen: Cognitive Anthropology Research Group at the Max Planck Institute for Psycholinguistics.

Levinson, S. C. & Wilkins, D. P. (2006). *The background to the study of the language of space.* In S. C. Levinson and D. P. Wilkins, Grammars of space: Explorations in cognitive diversity. Cambridge University Press: New York, NY.

Levinson, S. C., Kita, S., Haun, D. B. M., Rasch, B. H. (2002). Returning the tables: language effects spatial reasoning. *Cognition,* 84, 155-188.

Levitt H (1971) Transformed up-down methods in psychoacoustics. *Journal of the Acoustic Society of America A,* 49, 467-477

Lewis, H. (2001). Boas, Darwin, Science and Anthropology. *Current Anthropology, 42(3),* 381-406.

Lucy, J. (1992). *Language diversity and thought: A reformulation of the linguistic relativity hypothesis*. Cambridge: Cambride University Press.

Lucy, J. (1997). The linguistics of color. In Hardin, C. L. & Maffi, L. (Eds.) *Color Categories in Thought and Language*. Cambridge: Cambridge University Press.

Lucy, J.A. and Gaskins, S. (2001). Grammatical categories and the development of classification preferences: a comparative approach. In S.C. Levinson and M. Bowerman (eds.), *Language Acquisition and Conceptual Development*, 257-283. Cambridge: Cambridge University Press.

Lucy, J. and Schweder, R. A. (1979) "Whorf and his critics: Lingusitic and non-linguistic influences on color memory." *American Anthropologist, 81,* 581-615.

MacEvoy, B. (2005). Modern Color Models – Munsell Color System. *Color Vision.* http://www.handprint.com/HP/WCL/color7.html#MUNSELL

Maclaury, R. E. (1991). Social and cognitive motivations of change. Measuring variability in colour semantics. *Language, 67,* 34-61

Maclaury, R. E. (1992). From brightness to hue: An explanatory model of color-category evolution. *Current Anthropology, 33*, 137-186.

Maclaury, R. E. (1997). *Colour cognition in Mesoamerica: Constructing categories as vantages*. Austin: University of Texas Press.

Maddox, W.T., & Ashby, F.G. (1993). Comparing decision bound and exemplar models of categorisation. *Perception and Psychophysics, 53*, 49-70.

Martin, P.R., Gruenert, U., Chan, T.L., & Ghosh, K.K. (2001). Retinal pathways for colour vision: Studies of short-wavelength sensitive ("blue") cones and their connections in primate retina. *Color Research and Application, 26*, S112-117.

Matsuno, T., Kawai, N. & Matsuzawa, T. (2004) Color classification by chimpanzees (*Pan troglodytes*) in a matching to sample task. *Behavioural Brain Research, 148(1-2),* 157-165.

Matsuzawa, T. (1985) Color naming and classification in a chimpanzee (*Pan troglodytes*). *Journal of Human Evolution, 14*, 283-291.

Majid, A., Bowerman, M., Kita, S., Haun, D. B. M. & Levinson, S.C. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences*, 8(3), 108-114.

Malt, B. C. & Johnson, E. C. (1998). Artifact category membership and the intentional-historical theory. *Cognition*, 66, 79-85.

Malt, B. C., Sloman, S., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language*, 40, 230-262.

Markus, H., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98,* 224-253

Martin-Malivel, J., & Fagot, J. (2001). Cross-modal integration and conceptual categorization in baboons. Behavioral and Brain Research, 122, 209-213.

Matsuno, T., Kawai, N. & Matzusawa, T. (2004). Color classification by chimpanzees (Pan troglodytes) in a matching-to-sample task. *Behavioral Brain Research, 148,* 157-165.

Matsuzawa, T. (1985). Colour naming and classification in a chimpanzee. *Journal of Human Evolution, 14,* 283-291.

May, B. J., Moody, D. B., & Stebbins, W. C. (1989) Categorical perception of conspecific communication sounds by Japanese macaques (Macaca fuscata). *Journal of Acoustic Society of America, 85,* 837–847.

McKone, E., Martini, P., Nakayama, K. (2001). Categorical perception of face identity in noise isolates configural processing. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 573-599.

McNamara, T. E. (2004). *Evolution, Culture and Conciousness*. NY: University Press of America.

McNeil, J. E., & Warrington, E. K. (1993). Prosopagnosia: A face-specific disorder. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 46*(A), 1–10.

Medin, D. L., & Ross, B. H. (1997). *Cognitive psychology*. Florida: Harbourt Brace & Company.

Medin, D., Wattenmaker, W. D. & Hamson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology, 19(2),* 242-279.

Munnich, E. & Landau, B. (2003). The effects of spatial language on spatial representation: Setting some boundaries. In Gentner, D. & Goldin-Meadow, S. (Eds), *Language in mind: Advances in the study of language and thought.* Cambridge, MA: MIT Press.

Munsell, A. H. (1905*). A color notation.* Boston: Ellis

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289-316.

Nagy, A. L., & Sanchez, R. R. (1990). Critical color differences determined with a visual search task. *Journal of the optical Society of America, 7,* 1209-1217.

Namibian Government statistics (2004). Avaliable from : <www.grnnet.gov.na/>.

Navon (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9,* 353-383

Neiworth, J. J. & Wright, A. A. (1994). Monkeys (Macaca mulatta) learn category matching in a nonidentical same-different task. *Journal of Experimental Psychology: Animal Behavioral Processes, 20(4),* 429-35.

Nisbett, R. E. & Miyamoto, Y. (2005). The influence of culture: holistic versus analytic perception. *Trends in Cognitive Sciences*, 9(10), 467-473.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic vs. analytic cognition. *Psychological Review,* 108, 291-310.

Norenzayan, A., Smith, E. E., Kim, B. J. & Nisbett, R. E. (2002). Cultural preferences for formal versus intuitive reasoning. *Cognitive Science, 26(5),* 653-684.

Notman L, Sowden PT, Özgen E (2005) The nature of learned categorical perception effects: a psychophysical approach. *Cognition, 95,* 1-14.

Ohlemiller, K.K., Jones, L. B., Heidbreder, A. F., Clark, W. W. & Miller, J. D. (1999). Voicing judgments by chinchillas trained with a reward paradigm. *Behavioral Brain Research, 100,* 185-195.

Okajima, K. Robertson, A.R. & Fielder, G.H. (2002). A quantative network model for color categorization. *Color Research and Application, 27,* 225-232

Ozgen, 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(4), 477-493.

Pederson, E., & Schmitt, B. (1993). *Eric's maze task.* In Cognition and Space Kit Version 1.0 (pp. 73–76).Nijmegen: Cognitive Anthropology Research Group at the Max Planck Institute for Psycholinguistics.

Perez-Pereira, M. (1991). The acquisition of gender: What Spanish children tell us. *Journal of Child Language,* 18, 571-590.

Perkins, & Deregowski, J. B., (1982). A cross-cultural comparison of the use of Gestalt perceptual strategy. *Perception, 11,* 279-286.

Pike, K. L. (1954). *Language in Relation to a Unified Theory of the Structure of Human Behavior, Vol. 1.* CA: Summer Institute of Linguistics.

Piling, M., Wiggett, A., Ozgen, E. & Davies, I. R. L. (2003). Is color "categorical perception" really perceptual? *Memory and Cognition,* 31(4), 538-551.

Pisoni, D. B. (1975). Auditory short term memory and vowel perception. *Memory & Cognition,* 3, 7-18.

Pitchford, N.J. & Mullen, K.T. (2001) Conceptualization of perceptual attributes: A special case for color? *Journal of Experimental Child Psychology, 80,* 289-314.

Pitchford, N. J. & Mullen, K.T. (2002). Is the acquisition of basic-colour terms in young children constrained? *Perception, 31,* 1349-1370.

Pitchford, N.J. & Mullen, K.T. (2005). The role of perception, language, and preference in the developmental acquisition of basic color terms. *Journal of Experimental Child Psychology, 90(4),* 275-302.

Posner, M.I. (1969). Abstraction and the process of recognition. In G. Bower (ed*.),* *Advances in Learning,* New York: Academic Press.

Premack, D. (1983). Animal cognition. *Annual Review of Psychology, 34,* 351-362.

Ratliff, F. (1976). On the psychophysical basis of universal color terms. *Proeedings of the American Philoophy Society, 120,* 311-330.

Ratner, C. (1989). A sociohistorical critique of naturalistic theories of color perception. *Journal of Mind and Behavior, 10,* 361-373

Rice, M. (1980) *Cognition to Language: Categories, word meanings, and training.*

Baltimore: University Park Press.

Rips, L. J. (1989). Similarity, typicality, and categorization. In Vosniadou, S. & Ortony, A. (Eds.), *Similarity and analogical reasoning* (pp.21-59). Cambridge, England: Cambridge University Press.

Roberson, D. & Davidoff, J. (2000). The categorical perception of colour and facial expressions: the effect of verbal interference. *Memory & Cognition*, 28, 977-986

Roberson, D., Davidoff, J. & Braisby, N. (1999). Similarity and categorization: neurophysiological evidence for a dissociation in explicit categorization tasks. *Cognition*, 71, 1-42.

Roberson, D., Davidoff, J., Davies, I. & Shapiro, L.R. (2004). The development of color categories in two languages: A longitudinal study. *Journal of Experimental Psychology: General*. 133(4), 554-571.

Roberson, D., Davidoff, J., Davies, I. & Shapiro, L.R. (2005a). Color Categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology*, 50, 378-441.

Roberson, D., Davidoff, J. & Shapiro, L.R. (2002). Squaring the circle: The cultural relativity of 'good' shape. *Journal of Cognition and Culture*, 2, 29-51.

Roberson, D., Davies, I. R. L., Corbett, G. G. & Vandervijver, M. (2005b). Freesorting of colors across cultures: Are there universal grounds for grouping? *Cognition & Culture*, 5, 86-124.

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

Roorda, A., Mehta, A.B., Lennie, P. & Williams, D.R. (2001). Packing arrangements of the three cone classes in the primate retina. *Vision Research, 41*, 1291-1306.

Rosch, E. (1971). Focal areas and the development of colour names. *Developmental Psychology, 4,* 447-455.

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, 4, 328-350.

Rosch, E. (1975a). Cognitive reference points. *Cognitive Psychology*, 7, 532-547.

Rosch, E. (1975b). The nature of mental codes for colour categories. *Journal of Experimental Psychology: Human Perception and Performance, 1,* 303-322.

Rosch Heider, E. (1972a). Probabilities, sampling, and ethnographic method: the case of Dani colour names. *Man,* 7(3), 448-466.

Rosch Heider, E. (1972b). Universals in colour naming and memory. *Journal of Experimental Psychology*, 93(1), 10-20.

Rosch Heider, E. & Olivier, D.C. (1972) The structure of the colour space in naming and memory for two languages. *Cognitive Psychology*, 3, 337-354.

Rosch, E. & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*, 573-605.

Sandell, J. H., Gross, C. G., & Bornstein, M. H. (1979). Color categories in Macaques. *Journal of Comparative and Pysiological Psychology*, 93(4), 626-635.

Sandhoffer, C. M. & Smith, L. B. (1999). Learning color words involves learning a system of mappings. *Developmental Psychology, 35,* 668-679.

Sauders, B. & Van Brakel, J. (1997).Are there non-trivial constraints on colour categorization? Behavioral & Brain Sciences, 20, 167-178.

Saunders, B. & Van Brakel, J. (2002). *Theories, technologies, instrumentalities of color: Anthropological and historical perspectives*. Lanham, MD: University Press of America.

Segall, M. H., Dasen, P. R., Berry, J. W. & Poortinga, Y. H. (1999). *Human behavior in global perspective.* Boston: Allyn & Bacon.

Serpell, R. (1979). How specific are perceptual skills? A cross-cultural study of pattern reproduction. *British Journal of Psychology, 54*, 276-183.

Sera, M. D., Berge, C. & Del Castillo Pintado, J. (1994). Grammatical and conceptual forces in the attribution of gender by English and Spanish speakers. *Cognitive Development*, *9*, 261-292.

Sera, M. D., Elieff, C., Forbes, J., Burch, M. C., Rodriguez, W., & Dubois, D. P. (2002). When language affects cognition and when it does not: An analysis of grammatical gender and classification. *Journal of Experimental Psychology: General*, 131, 377-397.

Shatz, M., Behrend, D., Gelman, S.A. & Ebeling, K.S. (1996) Colour term knowledge in two-year-olds: evidence for early competence. *Journal of Child Language, 23*, 177-199.

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

Slobin, D. I. (1996). From 'thought and language' to 'thinking for speaking'. In Gumpertz, J.J. & Levinson, S. C. (eds.). *Rethinking linguistic relativity*. Cambridge: Cambridge University Press.

Soja, N. N. (1994). Young children's concept of colour and its relation to the acquisition of color words. *Child Development, 65*, 918-937.

Spelke, E.S. & Kinzler, K. D. (2007). Core knowledge. *Developmental Science 10*, 89-96

Sperling, H. G. & Harwerth, H. S. (1971). Red-green cone interactions in the Increment-threshold spectral sensitivity of primates. *Science, 172,* 180-184.

Spinozzi, G., De Lillo, C., & Salvi, V. (2006). Local advantage in the visual processing of hierarchical stimuli following manipulations of stimulus size and element numerosity in monkeys (*Cebus apella*). *Behavioural Brain Research, 166*, 45-54.

Steels, L. (1997). The synthetic modeling of language origins. *Evolution of Communication, 1(1),*1–35.

Steels, L. (1998). *Synthesising the origins of languageand meaning using co-evolution, self-organisation and level formation.* In J. R. Hurford, M. Studdert-Kennedy, and C. Knight (eds.) Approachesto the Evolution of Language. Cambridge University Press.

Steels, L. & Belpaeme, T. (2005) Coordinating Perceptually Grounded Categories through Language. A Case Study for Colour. *Behavioral and Brain Sciences*. 28(4):469-489.

Steinschneider, M.., Fishman, Y. & Arezzo, J. C. (2003). Representation of the voice onset time (VOT) speech parameter in population responses within primary auditory cortex of the awake monkey. Journal of the Acoustical Society of America, 114 (1), 307-321.

Sturges, J. & Whitfield, T. W. A. (1995). Locating basic colours in the Munsell space. *Colour Research and Application, 20*, 364-376.

Subramania, I. (1983). *Vakyapadiya of Bhartrhari Kanda II. Containing the Tika of Punyaraja and the Ancient Vrtti*. Delhi: Bharatiya Vidya Prakashan.

Svorou, S. (1994). *The grammar of space*. John Benjamins: Amsterdam.

Thompson, R. & Oden, D. (2000). Categorical Perception and Conceptual Judgments by Nonhuman Primates: The Paleological Monkey and the Analogical Ape. *Cognitive Science, 24 (3),* 363-396.

Tomasello, M. (1999). *The Cultural Origins of Human Cognition.* Cambridge, MA: Harvard University Press.

Tomasello, M. & Call, J. (1997). *Primate Cognition.* Oxford: Oxford University Press.

Tomonaga, M. (1995). Transfer of odd-item search performance in a chimpanzee (*Pan troglodytes*). *Perceptual and Motor Skills, 80*, 35-42.

Turton, D. (1980). There's no such beast: Cattle and colour naming among the Mursi. *Man*, 15(2), 320-338.

Tversky, A. & Gati, I. (1978). *Studies of similarity.* In Rosch, E. & Lloyd, B. (Eds.), Cognition and Categorization (chapter 4). Erlbaum: Hillsdale (NJ).

Vauclair, J. & Fagot, J. (1996). Categorization of alphanumeric characters by baboons (Papio papio): Within and between class stimulus discrimination. *Current Psychology of Cognition, 15,* 449-462.

De Valois, R. L. , Abramov, I., & Jacobs, G. H. (1966).Analysis of Response Patterns of LGN Cells. *Journal of the Optical Society of America A, 56,* 966- 976.

De Valois, R. L. & Jacobs, G. H. (1968). Primate Color Vision. *Science, 162,* 533-540.

Watanabe S., Sakamoto J., & Wakita M. (1995) Pigeons' discrimination of paintings by Monet and Picasso. *Journal of Experimental Animal Behavior, 63*, 165-174

Webster, M.A. & Mollon, J.D. (1991). Changes in colour appearance following post-receptoral adaptation. Nature, 349, 235-238.

Werner, J.S. and Wooten, B.R. (1985). Unsettled issues in infant color vision. *Infant Behavior and Development, 8,* 99-107.

Webster M. A., Miyahara E., Malkoc G., & Raker V. E. (2000). Variations in normal color vision: Unique hues. Journal of the Optical Society of America A, 17(9),1545-1555.

Wertheimer, M., & King, D. (2004). Max Wertheimer and Gestalt theory. Transaction Publishers: Boston, MA.

Whiten, A., Custance, D. M., Gomez, J. C., Teixidor, P., & Bard, K. A. (1996). Imitative learning of artificial fruit processing in children (Homo sapiens) and chimpanzees (Pan troglodytes). *Journal of Comparative Psychology, 117(3),* 325-336.

Whiten, A., Goodall, J., McGrew, W. C., Nishida. T., Reynolds, V., Sugiyama, Y., Tutin, E. G., Wrangham, R. W., & Boesch, C. (1999). Cultures in Chimpanzees. *Nature, 399,* 682-685.28

Whorf, B.L. (1956). Language, Thought and Reality. In Carroll, J.B. (Ed.) *Language, Thought and reality*. Cambridge, MA: MIT Press.

Wilson M., & DeBauche B. A. (1981). Inferotemporal cortex and categorical perception of visual stimuli by monkeys. *Neuropsychologia, 19,* 29-41.

Wilson, T. P. & Wilson, M. (2003). Perception-action links and the evolution of human speech exchange. *Behavioral and Brain Sciences, 25,* 47-48

Witthoft, N., Winawer, J., Frank, M., Wade, A., & Boroditsky, L. (2003). Effects of language on color discriminability. *Proceedings of the 25th Annual Meeting of the Cognitive Science Society.*

Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science, 9,* 33-39.

Wright, C. (1975). On the coherence of vague predicates. *Synthese, 30,* 325-365.

Wright, A. A. & Cumming, W. W. (1971). Color-naming functions for the pigeon. *Journal of Experimental Psychology: Animal Behaviour Processes, 15(1),* 7-17.

Yoshioka. T., Dow, B.M. & Vautin, R.G. (1996). Neural mechanisms of colour categorization areas V1, V2 and V4 of macaque monkey cortex. *Behavioral Brain Research, 76,* 51-70.

Zeki, S. (1983). Color coding in the cerebral cortex: the responses of wavelength-selective and color-coded cells in monkey visual cortex to changes in wavelength composition. *Neuroscience, 9,* 767-781.