Colour Word and Colour Category Learning in Infants and Toddlers

Samuel H. Forbes
Brasenose College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Michaelmas 2017

Colour Word and Colour Category Learning in Infants and Toddlers

Samuel H. Forbes

Brasenose College, University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2017

This thesis examines how and when infants learn colour words, and how the knowledge of colour words affects their comprehension of colour categories. Over the course of seven experimental chapters, the ability of infants and toddlers to learn colour words, use colour words to process colours, and the role that colour words play in learning to perceive colour are all assessed. Chapters 2 and 3 assess claims that colour words are learned late using parental report and eye-tracking methods, finding that colour words are learned as early as 19 months. In contrast to this, Chapter 4 demonstrates that toddlers do not learn to modify colour words as dark or light until much later. Chapter 5 demonstrates that colour words are a crucial component for processing the colours of objects, showing that infants do not look to a colour-matched object unless they comprehend the colour word. Chapters 6 and 7 employ novel paradigms to explore categorical processing of colour, finding that infants have a preference for within-category colours, but that this has no effect on their attention to dynamic coloured stimuli. In Chapter 8, a prototype for an infant colour vision test is shown, demonstrating that the second year of life is crucial for development of visual closure. The generalisability of these results to infant perception and word learning is also discussed.

Acknowledgements

It goes without saying that writing a thesis and completing a body of research takes great support from many different sources. I am indebted to more people than I can name here, but I will attempt to name some key contributors nonetheless. I am fortunate to have a family supportive of my pursuits, and I am extremely grateful to my parents, and to my sisters, Brigette and Erin, for that.

I am most grateful to my primary supervisor, Professor Kim Plunkett, for the support and mentorship of these past few years, and for the faith shown in both me, and in the project. I would also like to thank the members of the BabyLab team, both past and present, whose advice, help and company have been a major reason that any work got finished at all. Working in the lab has been a great joy, and I am additionally appreciative to Amanda, Irina, Jelena, Janette, and Marlene, along with many others, for their help with participants. I would also like to name Dr Olivia Guest for her advice, and particularly for her encouragement to delve into open science and coding early in my DPhil. I am grateful also to the advice of many members of the department, who along the way have offered advice and support.

I would also like to thank the Rhodes Trust, for their continued support (both moral and financial) over three years, and to the department for continuing to fund my research thereafter. I am in tremendous gratitude to Brasenose College, who have supported several conference visits, and provided me with a wonderful common room full of people. I would particularly like to name my college advisor for much of the time, Dr David Popplewell and Tutor for Graduates Professor Elspeth Garman, for being thoughtful and inspiring academic role models in college.

This research could not have taken place without the generous visits of participants and their caregivers to the lab, for which I am most grateful. I would also like to acknowledge Dr Caroline Floccia and the Plymouth Baby-Lab team for sharing the CDI data in Chapter 2, and to the huge numbers of researchers who have donated data to both the CHILDES databases, and the MB-CDI database, which have also formed a crucial component of the research in Chapter 2.

Finally, much of my DPhil life has been made all the more enjoyable due to the presence of wonderful friendships and circles of people, and I am thankful to all the (extended) members of Hertford College Choir and Hertford College Orchestra, for tolerating my puns, singing, and conducting for so long. I would especially like to single out Ali for kind support these few years, and Rafa, whose loss I still feel most keenly. Last, and most certainly not least, I would like to thank Marianne, for being both a rock and an inspiration to me above all else.

Publications Arising From This Thesis

The results presented in Chapters 2, 3 and 5 of this thesis are currently under review for publication as the following:

Forbes, S.H. & Plunkett, K. (Under Review). Linguistic and cultural variation in early colour word learning. *Child Development*.

Forbes, S.H. & Plunkett, K. (Under Review). Infants show early comprehension of basic colour words. *Developmental Psychology*.

Forbes, S.H. & Plunkett, K. (Under Review). The role of colour labels in mediating infant visual attention. *Cognition*.

I confirm that the work contained herein is my own, under the supervision of Professor Kim Plunkett.

Contents

1 Introduction **1**

1.1 Background 1

1.2 The onset of colour word learning 3

1.2.1 Constraints on colour word learning in infants 5

1.2.2 The timing of colour word learning 9

1.2.3 The role of colour words in mediating visual attention in infants 13

1.3 The development of colour categories 15

1.3.1 Categorical perception of colour 16

1.3.2 Categorical perception in infants 17

1.3.3 Colour vision tests in infants 20

1.4 Outline of thesis 21

2 Linguistic and Cultural Variation in Early Colour Word Learning **23**

2.1 Introduction 23

2.2 Study 1 28

2.2.1 Methods 28

2.2.1.1 Participants 28

2.2.1.2 Materials 28

2.2.1.3 Analysis 29

2.2.2 Results 30
2.3 Study 2 34
2.3.1 Method 34
2.3.1.1 Participants 34
2.3.1.2 Materials 35
2.3.1.3 Analysis 36
2.3.2 Results 37
2.4 Study 3 40
2.4.1 Methods 40
2.4.1.1 Materials 40
2.4.1.2 Analysis 43
2.4.2 Results 44
2.5 Discussion 46

3 Early Colour Word Learning in British Infants 52
3.1 Introduction 52
3.2 Methods 55
3.2.1 Participants 55
3.2.2 Materials 56
3.2.3 Procedure 57
3.2.4 Analysis 58
3.3 Results 59
3.3.1 Pre-naming Phase 59
3.3.2 Post-naming Phase 61
3.3.3 Comparison to parental reports 66
3.3.4 Sentence type analysis 67
3.4 Discussion 70

4 Infants Display an Early Preference for Dark Colours 74

4.1 Introduction 74

4.2 Methods 77

4.2.1 Participants 77

4.2.2 Materials 77

4.2.3 Procedure 78

4.2.4 Analysis 79

4.3 Results 80

4.3.1 Post-naming phase 80

4.3.2 Pre-naming phase 83

4.4 Discussion 85

5 The Role of Colour Labels in Mediating Infant Visual Attention 89

5.1 Introduction 89

5.1.1 Comprehending Colour Words 92

5.1.2 The Present Study 93

5.2 Methods 94

5.2.1 Participants 94

5.2.2 Design 94

5.2.3 Stimuli 96

5.2.4 Procedure 97

5.2.5 Analysis 97

5.3 Results 98

5.3.1 Overall Results 98

5.3.2 Colour results by age 100

5.3.3 Colour results by colour word knowledge 103

5.3.4 Results at 19 Months 104

5.3.5 The Role of Total Vocabulary 105

5.4 Discussion 107

6 Preferential Looking in Categorical Perception of Colour 111

6.1 Introduction 111

6.2 Methods 115

6.2.1 Participants 115

6.2.2 Materials 116

6.2.3 Design 116

6.2.4 Procedure 118

6.2.5 Analysis 119

6.3 Results 120

6.3.1 Overall analysis 120

6.3.2 Colour word knowledge analysis 122

6.4 Discussion 124

7 Visual Attention and Categorical Perception of Colour 129

7.1 Introduction 129

7.2 Methods 132

7.2.1 Participants 132

7.2.2 Materials 132

7.2.3 Design 133

7.2.4 Procedure 134

7.2.5 Analysis 135

7.3 Results 136

7.3.1 Overall analyses 136

7.3.2 Individual effects 138

7.3.3 Looking patterns 140

7.3.4 First trial analysis 142

7.4 Discussion 144

8 The development of visual closure in infancy: Implications for colour vision testing 148

8.1 Introduction 148

8.2 Experiment 1 151

8.2.1 Methods 151

8.2.1.1 Participants 151

8.2.1.2 Visual stimuli design 152

8.2.1.3 Plate selection 153

8.2.1.4 Auditory stimuli 153

8.2.1.5 Procedure 154

8.2.1.6 Analysis 155

8.2.2 Results 156

8.3 Experiment 2 158

8.3.1 Methods 158

8.3.1.1 Participants 158

8.3.1.2 Stimuli 158

8.3.1.3 Procedure 158

8.3.1.4 Analysis 158

8.3.2 Results 159

8.4 Discussion 160

9 General Discussion and Conclusions 162

9.1 Introduction 162

9.2 Timing of colour word learning 163

9.3 Constraints on colour word learning 166

9.4 Mediating visual attention with colour words 168

9.5 Categorical perception of colour 169
9.6 Colour vision testing and visual closure 173
9.7 Concluding Remarks 175

References **177**

A Supplementary Information for Chapter 2 **199**

B Supplementary Information for Chapter 5 **211**

List of Figures

1.1 A timeline of when colour words were thought to be effectively learned by toddlers. 11

2.1 Recruitment information for Study 1. 28

2.2 Results of two fitted Bayesian models to the Oxford CDI comprehension and production data. Dotted lines indicate comprehension, solid lines indicate production. Narrow bands around each line indicate the credible interval of the mean. 31

2.3 Comprehension and production data from a subset of Oxford CDI participants, fitted with binomial curves. The trends for some colour terms (e.g. production of *orange*), directly overlap with other terms. 95% CI are not displayed here for ease of viewing. 34

2.4 MacArthur-Bates CDI production data modelled with a binomial Bayesian model, separated by language. Bands around each line indicate confidence in the mean. 41

2.5 Input frequency of occurrence of colour words for infants up to 3 years of age, by language. 42

3.1 Time course of a typical trial. 57

3.2 The proportions of looking to any target colour (Colour 1) over any other colour (Colour 2). Red indicates looking to that colour above 50%, yellow indicates looking below 50%. Proportions are listed in each box. 60
3.3 Looking proportions to the target for each age group. 61
3.4 Looking proportion to the target as trial progresses. 0ms is when the target is named. 63
3.5 Looking proportion to the target for different sentence types. . 68

4.1 Time course of a typical trial. 79
4.2 Looking proportions to dark and light in the post-naming phase, by age group. The dotted line indicates chance looking. 80
4.3 Looking proportions to dark and light in the post-naming phase, by age group. The dotted line indicates chance looking. 81
4.4 Looking proportions to dark and light in the pre-naming phase. The dotted line indicates chance looking. 83
4.5 Looking proportions to dark and light for each hue and age group in the pre-naming phase. Plotted are means and standard errors. By design, light and dark preferences are necessarily symmetrical. 84

5.1 Time-line in seconds of eye-tracking trials: this is a typical Colour Discrimination trial, where the intended target would be the yellow brick. 98
5.2 Overall results, collapsing over age, from target word onset. Related trials include colour related, semantically related, and colour discrimination trials. The dotted line indicates chance looking at 50%. 99

5.3 Overall results, comparing colour-related, colour discrimination, and semantically-related trials. 100
5.4 Colour trials vs semantic trials, fitted with a model by age in the top frame, and by knowledge of the relevant colour word, in the bottom frame. 102
5.5 Model fit on nineteen month-old participants, by colour word knowledge 105
5.6 Proportions of looking to related distractor in the 1500 - 2500ms time bin when each of the words in the Oxford CDI is known. The dotted line indicates the proportion when the relevant colour word is known. 107

6.1 Stimuli selections for each condition. The dashed line indicates an approximate category boundary. 118
6.2 A sample trial. Borders around each square are for ease of viewing in this figure, and were not included in the experiment. . . 119
6.3 Model fit of participant looking to the between-category colour, by condition. In the Green and Far conditions, squares are presented on a green background, in the Blue condition, squares are presented on a blue background. 122
6.4 Model fit of participant looking to the between-category colour, by participant age. 123
6.5 Model fit of 19 month-olds looking to the between-category stimulus, by colour word knowledge. 125

7.1 A typical trial. The patch of colour would alternate between 2 shades until the trial ends, every 500ms. 135

7.2 Numbers of participants who completed each trial, separated by age group. 137
7.3 Model fit to overall looking proportion data in the first 6 trials. The bars indicate standard errors of the mean. 138
7.4 Looking proportions in each condition. 139
7.5 Looking proportions to the target in each age group, on the left panel, and separated by colour word knowledge, on the right panel. 140
7.6 Looking proportions to the target over time within each trial type. 143
7.7 Looking proportions to the between-category pair or the within-category pair in the first trial. 144

8.1 A sample of a standard plate used in an Ishihara plate test. From http://www.color-blindness.com/ 150
8.2 An example plate, depicting a cat, used in the study. 152
8.3 The timecourse of a typical trial. 155
8.4 The fitted models from the two experiments, showing proportions of looking to the target, with Experiment 1 in the top panel, and Experiment 2 in the bottom panel. 157

A.1 Participant information for all data used from the Wordbank Repository, separated by age and gender. Grey indicates that there was no gender information available. 199
A.2 Fitted final model from Study 3, for comparison to Study 2. The results show *blue* further behind than in Study 2, indicating other possible factors. 200

List of Tables

2.1 Results of model on comprehension of colour words in the Oxford CDI. Colours are compared to *Blue*. 32
2.2 Results of model on production of colour words in the Oxford CDI. 32
2.3 Ages of participants who completed supplementary Oxford CDI data. 33
2.4 Comprehension model for the supplementary data. 35
2.5 Production model for the supplementary data to the Oxford CDI. 36
2.6 Numbers of participants in each language of the MacArthur-Bates CDI surveys. 37
2.7 List of words used in Study 2, based on MB-CDIs. 37
2.8 Main population-level effects on fitted model of MB-CDI data in 11 languages. Colours are as compared to black. Interaction terms and group-level terms are not included. 38
2.9 Population-level effects of final predictive model using both frequency and syllabic complexity as predictors of colour word learning. 45

3.1 Descriptive statistics for participants included in study. 56

3.2 Model coefficients for pre-naming phase. Colours are compared to black. Results for the second colour are identical but reversed, due to the nature of the data. 60
3.3 Full list of effects for logistic mixed-effects regression on the post-naming data. “ot” refers to the orthogonal time terms. Colours are compared to black. 64
3.4 Mean lengths and standard deviations for each stimulus type following target word onset. 67
3.5 Model coefficients of model fit comparing the type of sentence. Stimulus Type refers to the sentence type, and are compared to sentence type 1. 69

4.1 Descriptive statistics for participants included in the present study. 77
4.2 Model coefficients of binomial model on post-naming phase data. “ot” refers in turn to each of the polynomial time terms. . 82
4.3 Model coefficients of GLM examining effects of hue and age group on looking to the dark object. Hues are compared to blue. 85

5.1 Model coefficients of age-based model. “ot” refers to the orthogonal time terms. 101
5.2 Coefficients of knowledge-based model. 104
5.3 Model coefficients of knowledge-based model run only with 19 month-olds 106

6.1 Chromaticity coordinates (CIE 1931) for the three conditions in the present study. In the *blue* condition, the first 3 are used; in the *green* condition, the second to fourth stimuli are used. The three denoted with “far” are used in the *far* condition. 117

6.2 Coefficients of model on participant looking behaviour. Condition was compared to *blue*. 121
6.3 Output of the model on looking mediated by colour word knowledge in 19 month-olds. 124

7.1 Chromaticity coordinates (CIE 1931) for the three conditions in the present study. 133
7.2 Model output of multilevel model on first 6 trials. Conditions are compared to the *between* condition. 137
7.3 Output of model examining looking patterns within trials. . . 142

8.1 Stimulus ratings for all stimuli included in the experiment. Identification is the proportion of adults who correctly identified the stimulus, while the clarity score is the rating out of 100, divided by 100. 153
8.2 Output of generalised linear model. 156
8.3 Model output for Experiment 2. 159

A.1 Corpora used in analysing frequency of occurrence for each colour word. 201
A.2 Full frequency calculations for each Childes corpus used in Study 3. *Bernstein-Ratner corpus did not include the “interview” section as there was little parent data. **Szagun corpus only included the Normal-Hearing group. 202

B.1 Complete list of stimuli used during eye-tracking experiment. 211

Chapter 1

Introduction

1.1 Background

This thesis is concerned with how infants learn about colour categories, and how they learn the words that correspond to those colour categories. In essence, this thesis can be summarized in two main parts: When and how are colour words learned; and what role (if any) do colour words play in shaping perception. These two parts correspond loosely to the two halves of this thesis, respectively.

Past research had focussed on the errors that young children make when producing colour words, finding that they consistently use colour terms incorrectly (Franklin, 2006; Pitchford & Mullen, 2003; Sandhofer & Smith, 1999; Shatz, Behrend, Gelman, & Ebeling, 1996; Soja, 1994). The field reached a vague consensus that colour words were thus hard to master, and learned later than other classes of words. The present research challenges these assertions with a series of experiments designed to test when infants learn colour words.

This thesis also contributes to debates about the role that object properties, such as colour, play in processing of objects, and how labels of those properties contribute to that. There had been a great deal of research that suggested that the processing of the colour of an object could take an indirect route: when hearing "*frog*," it would activate the concept *green*, without needing to activate the word "*green*" (Huettig & Altmann, 2005; Huettig & McQueen, 2007; Huettig & Altmann, 2011; Johnson, McQueen, & Huettig, 2011; Johnson & Huettig, 2011). The present research critically evaluates these claims, extending the past research to include toddlers at various stages of learning colour words.

The current body of research also contributes to ongoing debate surrounding categorical perception of colour. Colour had been found to be perceived categorically (Franklin & Davies, 2004; Franklin, Clifford, Williamson, & Davies, 2005; Franklin, Drivonikou, Clifford, et al., 2008; Franklin, Drivonikou, Bevis, et al., 2008; Gilbert, Regier, Kay, & Ivry, 2006; Holmes & Wolff, 2012; Regier & Kay, 2009), with the effect being found in one visual field, and thus specialised in one hemisphere, more than the other. Slowly, concerns both theoretical and methodological, began to emerge about that body of research (A. M. Brown, Lindsey, & Guckes, 2011; Witzel & Gegenfurtner, 2011). These are addressed in the current thesis, examining the role that colour words play in categorical perception of colour with new methodology.

Finally, this thesis presents new research into the development of visual closure in toddlers, and the applications of that to a prospective colour vision test for infants. There is currently no test of colour vision for young infants and toddlers. As such, the present research provides a possible starting point to developments in that field.

In the remainder of this chapter, the literature in these fields is discussed and reviewed, highlighting the direction of the studies in this thesis. In Chapters 2 and 3, a series of experiments relating to a timeline of colour word learning and how colour words are learned, are presented. In Chapter 4, that body of work is expanded on, examining when toddlers learn to modify basic colour words with words such as "*light*" and "*dark*," and whether they have any preferences for different shades of colour. In Chapter 5, the role that colour words play in processing of colour related concepts is examined. Chapters 6 and 7 discuss the effect that colour words may have on the categorical perception of colour, while Chapter 8 examines visual closure in toddlers, with a view to a prospective colour vision testing paradigm. In Chapter 9, The overall findings and contributions of the thesis are examined, and the implications of the research discussed.

1.2 The onset of colour word learning

Word learning, in many respect, is a defining milestone in any child's development. The nature of the first words, whence they are learned, and the accuracy of their application, can often define the early development of the child. Word learning is, however, an enormous and never-ending task. The sheer enormity of the task can be imagined when considering that a child often learns over 60,000 words by adolescence (Bloom, 2000). That multiple words can refer to the same object, or different objects (Quine, 1960) only heightens the difficulty. It is remarkable, then, that despite all this, infants can and do frequently learn words despite them being presented without an obvious link to their meanings (Tomasello, Strosberg, & Akhtar, 1996).

To date, the vast majority of research into infant word learning has examined their acquisition of object labels. These studies range from examining parental reports of how many words infants might know at any given age

(E. Bates et al., 1994; Dale & Fenson, 1996; Fenson et al., 1994; Hamilton, Plunkett, & Schafer, 2000; Mayor & Plunkett, 2011), to the factors that allow them to be learned (Goodman, Dale, & Li, 2008), and even experimental manipulations of object labels and their role in assisting the learning of objects themselves (Plunkett, Hu, & Cohen, 2008; Robinson & Sloutsky, 2004; Waxman, 1999; Waxman & Booth, 2003; Younger, 1985). These studies highlight the prodigious growth in infant vocabulary.

As object labels make up the vast majority of infant vocabulary (Fenson et al., 1994), and are seemingly easier to learn than other classes of words (Childers & Tomasello, 2006), they have accounted for much of the research on infant word learning. Yet each of these labels map to an object that has context and properties, the labels of which all contribute to descriptions of the object. This contextual information, such as size, shape, number, or colour, make up crucial information for an infant to understand the world around them.

In the case of colour, this mapping has always been considered to be difficult (e.g. Franklin, 2006; Pitchford & Mullen, 2003; Rice, 1980; Sandhofer & Smith, 1999; Soja, 1994; Wagner, Dobkins, & Barner, 2013). It has long been thought that infants learn colour words late compared to other classes of words (Heider, 1971; Shatz et al., 1996), and that when infants do apply colour terms, they do so randomly, often resulting in randomly applying colour words to any shade of colour, rather than the correct one. Research to date had suggested that colour words were not learned until around the seventh year of development (Heider, 1971), with more recent research suggesting it might be around 4;0 that colour words are learned (Bornstein, 1985), with some surprising success in comprehending colour words in the third year of life (Mervis, Bertrand, & Pani, 1995).

1.2.1 Constraints on colour word learning in infants

Why, then, has colour word learning been perceived to be so difficult? There have been myriad explanations posited as to the constraints that cause colour words to be learned late, and why they are produced so haphazardly and randomly (Pitchford & Mullen, 2003). Much of the debate surrounding these difficulties are centred around whether the difficulty is a cognitive or conceptual one, or whether it is linguistic. Recent work has suggested that the errors involved in colour word learning may not be random, suggesting the possibility that colour word learning is not as difficult as has been suggested.

An early perspective on the difficulty with learning colour words was a neuromaturational account. Bornstein (1985) suggested the difficulty was a biological one, and that the ability to associate verbal properties with visual ones requires cortical structures that are not in place until around the fourth birthday (p. 83). However, other experiments around that time already showed that toddlers below that age show some comprehension of colour words (Heibeck & Markman, 1987; Macario, 1991; Smith, 1984), ruling out that hypothesis on an empirical basis. Despite this being ruled out, some support was found for colour word learning being conceptually difficult in the longitudinal work of Shatz et al. (1996). Shatz et al. found that children as young as 2;6 knew some basic colour words, but learned them possibly on the basis of warm-cool boundaries (see Gibson et al., 2017; Kay & Regier, 2007, for a discussion of the impact of this on colour word naming). On this basis it was concluded that the cortical development proposed by Bornstein may be a telling factor, but that it was in place during the third year of life.

Even without attributing blame to specific cortical structures, the conceptual ability to abstract colour was found to be a difficulty. Early reports suggested that it may be the conceptual difficulty of knowing how to divide the continuous spectrum of colour into discrete chunks that was the cause of the difficulty (Andrick & Tager-Flusberg, 1986). There is considerable evidence that, in fact, infants and toddlers who do not know colour words can categorize colours effectively (Franklin & Davies, 2004; Franklin, Clifford, et al., 2005; Franklin, Pilling, & Davies, 2005; Skelton, Catchpole, Abbott, & Franklin, 2017), but there was sufficient evidence from early studies to point to colour word learning being conceptually difficult (Soja, 1994).

Kowalski and Zimiles (2006) argued instead that the primary difficulty for children was in their ability to represent colour as an abstract domain prior to learning colour words. The conceptual constraint of being unable to determine that colour is a domain and abstract it from context was tested in two experiments, showing that the ability to represent colour and the knowledge of colour terms, were strongly associated. Thus Kowalski and Zimiles believed that because of their inability to abstract colour as a domain, infants will map novel words to objects, not abstracted properties – a stance that garnered additional support from separate word-learning studies (Booth & Waxman, 2003, 2009).

However, there was also considerable support for the difficulties in colour word learning being linguistic mapping, rather than conceptual. In a series of tests of toddler behaviour relating to colour word comprehension, Soja (1994) used a basic colour comprehension test to separate participants into two groups – those who knew colour words and those who did not. The task was one of colour matching to a word; they had to find, for example, a red crayon, when asked, *"can you get me the red crayon?"* The results of the experiments showed that those who did not know colour words were still able to map names onto colours, were able to judge similarity of objects based on colour, and were able to make colour-based inferences, but that many participants did not know colour words in their third year. These results were concluded to support the theory that the difficulty with learning colour words may be a linguistic one, rather than a conceptual one.

Sandhofer and Smith (1999) examined the series of mappings required to learn colour words, theorising that there may be levels of mapping. First, the toddler needs to determine that the domain of relevance is colour (that which is answerable by *"red"* or *"green"*), and then they need to map a given colour word to a given colour (*"red"* to red objects). The toddler then needs to combine these two mappings to respond correctly to colour tasks. It is only after all this that they were thought to be able to match by colour in a non-colour-related task. Using six test objects, as opposed to the two used by Soja (1994), Sandhofer and Smith found that colour words were learned with great difficulty. It was concluded that this was not just due to the task being harder than that of Soja, as the toddlers were able to complete the task when the relevant domain was size (although the same authors later concluded this may be related to the type of reinforcement learning that occurs for each domain; Sandhofer & Smith, 2001). While the comparative difficulty of the task came into question (Pitchford & Mullen, 2001, 2002), these findings reinforced the idea that due to the series of mappings required, colour words were much harder to learn than other words denoting object properties (Heider, 1971, 1972).

The idea that the main constraints on infant colour word learning may be linguistic and attentional, rather than conceptual were further tested in three experiments on Italian and English by O'Hanlon and Roberson (2006). O'Hanlon and Roberson used a computer-based task to teach three-year-olds three very low-frequency colour words (beige, crimson, and teal). The participants were assessed on comprehension of the words, as well as during a follow-up two weeks after the initial assessment. One of the aims of that study was to assess the interaction between the linguistic labels of the colour words, and the way they are presented; thus participants in some conditions just heard the target colour labels referred to, and in other conditions participants heard the target colour, and were explicitly told it was not any any of the distractor colours (i.e. *"This is my crimson star, not my yellow star or my blue star."* Across the experiments it was found that infants were most successful at learning colour words when the attentional aspects and the linguistic aspects of the novel colour word came together, so that they could learn with both labels and contrast, rather than when they were presented with only one of the two factors. This suggested that while the mappings were difficult, attention could be a factor, playing into the idea that frequency of exposure to colour words could also play a part (Pitchford & Mullen, 2005), as the attention in the task presented by O'Hanlon and Roberson appears mediated by linguistic information.

Thus while the causes, linguistic or conceptual, were debated, there was agreement in several senses. First, colour word learning occurred late and was difficult (although there was thought given to this being a task difficulty Pitchford & Mullen, 2001). Second, colour words were learned in an unusual fashion, with production in many cases preceding comprehension, leading to random and widespread errors in colour word usage (Shatz, Tare, Nguyen, & Young, 2010). These assertions were tested in a series of studies designed to test if colour word production errors really were as random as previously thought (Wagner et al., 2013; Wagner, Jergens, & Barner, 2014). Wagner et al. instead found that many of the production errors by infants of colour words were to the proximal colour. Rather than applying colour terms randomly and haphazardly, it was found that toddlers overextend the colour words, but accurately use the ones that they know for typical examples of that colour. Thus as they learn more colour terms, the categories shrink, until they approach the understanding that we possess as adults.

The assertions made by Wagner et al. raise the possibility that if, as suggested, a partial colour word comprehension precedes production of any sort, then colour word learning may occur earlier than thought previously, a claim tested in Chapters 2 and 3 of this thesis. In addition, if that is true, it is possible that the constraints may not be so severe, as colour word learning might not be as difficult as had been thought (e.g. Franklin, 2006).

1.2.2 The timing of colour word learning

With the constraints mentioned above, there have been various suggestions regarding the timing of colour word learning in infants. The timing of colour word learning as observed by researchers has changed over time, becoming earlier. There are many possible reasons for this, such as an increase in coloured plastic toys that increase children's exposure to colour words, or a change in research methodology (Franklin, 2006).

Due to the myriad aforementioned reasons, colour word learning was first described as not occurring successfully until as late as 7 years of age (Heider, 1971, 1972). While this seems incredibly late for successful colour word learning, it is likely that this refers to when colour words are comprehended and produced in an adult-like manner (Wagner et al., 2013). The consistent errors in producing colour words would likely have been regarded as random, rather than simply an overextension of the terms, as infants frequently do with other classes of words (Yurovsky, Wagner, Barner, & Frank, 2015).

Bornstein (1985) described a process (above) wherein once the correct cortical structures were in place, by around 4 years of age, then colour words could be correctly mapped to the colour categories. Again, the fact that colour words by this explanation are learned so late seems to suggest an underestimation of infant word learning. This account paid no heed to the fact that infants can produce colour words, albeit sometimes incorrectly, before that age.

One of the seminal studies on word learning around that time used a diary study to closely analyse the changes in an infant's vocabulary over time (Mervis et al., 1995). The results of that study suggested that colour words could be learned early in the third year of life, again earlier than previously found. Mervis et al. did not note how “adult-like” the comprehension of the colour words were, so it is possible that infants were found to regularly produce colour terms correctly (similar assumptions were made in reanalysis of diary studies, Mayor & Plunkett, 2009, 2011). Mervis et al. also noted that, while colour word learning occurred late, when colour word learning began, it happened quickly, with a close contiguity in the learning of multiple colour words. This finding appeared to lend additional weight to a mechanistic understanding of colour word learning, only occurring once a certain milestone had been achieved.

Later accounts demonstrate the early production of colour words overlooked by Bornstein (1985). Studies by Soja (1994); Pitchford and Mullen (2003); Sandhofer and Smith (1999) all demonstrated signs that infants were comprehending colour words earlier than previously shown, as early as 30 months in some cases. The early production of colour terms by infants was observed, and thus the theory on colour word learning was that they were learned unlike other classes of words. Infants were thought to produce colour words early, but without any comprehension of the term, producing them “haphazardly” (Pitchford & Mullen, 2003). By this account, production pre-

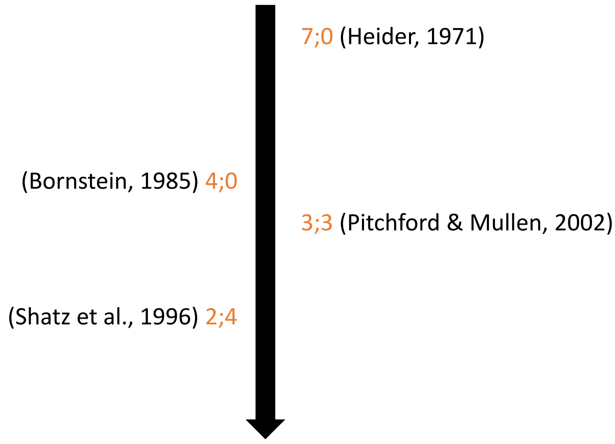

Figure 1.1: A timeline of when colour words were thought to be effectively learned by toddlers.

ceded comprehension, unlike other classes of words, where comprehension is partially comprehended, and then updated over time (E. Bates, Bretherton, & Snyder, 1988; Fenson et al., 1994).

An alternative account to the fast-mapped account discussed above, was proposed by Wagner et al. (2013). Following a behavioural study on infant production Wagner et al., and a further comprehension analysis based on eye-tracking (Wagner et al., 2014), some evidence was found that a partial comprehension preceded production. By this account colour words were slow-mapped, where infants only possessed a basic comprehension when producing the term. Wagner et al. demonstrated that the errors made by children in production of colour terms were not haphazard and random, but rather were over-extensions of the terms. A reanalysis of this data by Yurovsky et al. (2015), found further support for these claims. If colour word comprehension at least partially precedes production, as claimed by these analyses, then colour word learning may occur in the same way as other classes of words (Hidaka & Smith, 2010), and additionally may be earlier than previously claimed. One focus of Chapters 2 and 3 of this thesis is to examine the claims about when colour words are comprehended and produced by toddlers, and to examine the onset of comprehension, and whether it is earlier or later than production. In Chapter 4, terms such as "*dark*" and "*light*" which moderate colours on the periphery of each colour category are also examined.

Investigations into colour word learning have also investigated the internal timing of colour word comprehension and production, that is, whether there is a universal order in which colour words are learned. Berlin and Kay (1969) examined colour categories and colour words cross-culturally, establishing patterns of colour word allocations in each language (see also Kay, Berlin, Maffi, Merrifield, & Cook, 2011, for similar investigations into pre-industrialised languages). The authors discovered that languages consist of 11 basic colour terms (*blue*, *red*, *green*, *yellow*, *black*, *white*, *purple*, *pink*, *orange*, *brown*, and *grey*)[1]. In addition, when colour terms evolve in each language, they do so in a regular pattern. Any language that has colour words has a black (dark) and white(light) distinction, and then red is developed, following that, a language develops green and yellow in either order, then blue, then brown, and then the remaining colours are developed last.

Berlin and Kay (1969) hypothesised that as these colours were developed in languages in that order due to their natural perceptual states, this order would also be the developmental order in which infants acquire the colour words. Pitchford and Mullen (2002) investigated this claim in infants. While Pitchford and Mullen found no evidence of a strict colour word order as hypothesised by Berlin and Kay, they still believed that the primary colours would be learned before other colour words. There have been no other cross-linguistic, comparative developmental studies to examine the universality of any colour word orders, another main aim of Chapter 2 of this thesis.

[1]Later examinations discovered languages that also contain an additional blue, or a blue-green, such as Russian (Lucy, 1997).

1.2.3 The role of colour words in mediating visual attention in infants

Having examined the difficulties that children might possess in learning colour words in the previous subsection, it is necessary to also examine the role that colour words might play in influencing aspects of infant processing. While the ability of infants to use object labels to implicitly and explicitly process information about objects (Delle Luche, Durrant, Floccia, & Plunkett, 2014; Houston-Price, Plunkett, & Harris, 2005; Mani & Plunkett, 2010; Styles, Plunkett, & Duta, 2015), and about adults' ability to use feature and object labels to process information about features (Allopenna, Magnuson, & Tanenhaus, 1998; Chow, Aimola Davies, & Plunkett, 2017; Cooper, 1974; Huettig & Altmann, 2011), much less is known about how infants use feature labels, such as colour words, to process information about those features, the aim of Chapter 5.

In an investigation into adult online processing of object features, Huettig and Altmann (2011) employed three eye-tracking experiments to investigate which features drive shifts in attention. The authors found that the surface colour of the object, as well as the stored information about the typical colour of that object both mediated adult shifts in attention. Additionally, they found that taxonomic information was a more powerful mediator of shifts in attention than was colour knowledge, a finding that was later replicated in toddlers (Mani, Johnson, McQueen, & Huettig, 2013). This finding by Huettig and Altmann demonstrated that adults will fixate a green blouse, for example, if hearing the name of a typically green object, such as spinach, due to the information adults have stored about the typical colours of objects. The finding prompted questions about the role of the label in the processing that causes the switches in attention.

Discussions about label-mediated switches in attention lead to investigations of the same phenomenon in toddlers (Johnson et al., 2011; Johnson & Huettig, 2011). Johnson and Huettig (2011) used a similar method to that employed by Huettig and Altmann (2011) to demonstrate that toddlers at 36 months systematically fixate on colour-matched objects when hearing the label of a typically-coloured object in much the same way as adults. Lexical processing and visual attention are clearly linked from early in development, and toddlers clearly store contextual information, for example colour, from an early age. However, 36 month-old toddlers know a great deal of words, when compared even to a toddler 12 months younger, and thus while this work demonstrates that contextual information about objects is stored from an early age, it does not answer questions about the depth of lexical knowledge that is necessary.

Johnson et al. (2011) attempted a similar investigation with toddlers, this time at 24 months of age. The aim was to determine whether *feature labels*, rather than object labels, play a role in the ability to switch attention in a language-mediated visual search paradigm. The participants were 24 month-old toddlers who did not comprehend or produce colour words according to the results from a naming task, meaning that if they could still systematically attend to the colour-related object, as in the above examples, then colour words were not necessary for colour-mediated shifts in attention. To describe it another way, knowledge of the word "*green*" would not be necessary to fixate a green plate over a red plate when hearing the word "*frog*." The results demonstrated that the 24 month-olds tested in the study were still able to fixate the colour-related target, demonstrating the importance of the concept of colour over the colour words. Johnson et al. (2011) can be described as testing two alternative possible theories. One possibility is that of *direct acti-*

vation, where a colour concept can be activated without activating the associated colour label; the other is *label-mediated activation*, where the colour word is activated, and in turn activates the abstract representation of the colour. The results of the Johnson et al. study indicate strong support for the *direct activation* hypothesis.

However, with claims of ever-decreasing ages of learning colour words (Franklin, 2006), and with suggestions discussed above that a very early and basic comprehension of colour words might pre-date production and adult-like comprehension (Wagner et al., 2013, 2014), the claims that the 24 month-old toddlers tested in that study have no real knowledge of colour words may be questioned. A key aim of Chapter 5 of this thesis is to conceptually replicate and extend the findings of Johnson et al., and discover the role that the colour word has to play in colour-mediated shifts in attention.

1.3 The development of colour categories

For infants to learn about colour words, as discussed above, a series of word mappings are required, such that they learn to map colour words onto the abstract colour categories, and abstract colour categories onto real world examples of coloured objects, complete with nuances and shadings. For infants and toddlers to do this as described, some key assumptions are in place, namely that infants are in possession of perceptual colour categories onto which they can map colour words; and that infants have sufficient colour vision, allowing them to perceive colours in much the same way as others. These basic, albeit important, assumptions are considered with in Chapters 6, 7 and 8.

1.3.1 Categorical perception of colour

Infant colour perception has always been found to develop reasonably early. The basic psychophysical ability to distinguish colours has been shown to be in place by around 5 months of age (e.g. Morrone, Burr, & Fiorentini, 1993; Peeples & Teller, 1970, 1978; Teller, 1979, 1998). However, adults have been shown to perceive colour categorically (Bornstein & Korda, 1984; Daoutis, Franklin, Riddett, Clifford, & Davies, 2006; Pilling, Wiggett, Ozgen, & Davies, 2003; Roberson, Davies, & Davidoff, 2000; Roberson & Davidoff, 2000), that is to say that they differentiate colours faster or more accurately when they belong to different categories than when they belong to the same category of colour. But is this trait determined by language, or is it a basic, biological function of category boundaries? Chapters 6 and 7 of this thesis address these claims with infants who are yet to learn colour words, or have just started learning them to investigate discrepancies between the two groups, and determine whether language plays a top-down role in colour perception.

The basic idea that colour is perceived categorically was established since the 1980s (Bornstein & Korda, 1984; Kay & Kempton, 1984). However, in a later study (Gilbert et al., 2006), language was determined to have a major top-down influence. Gilbert et al. demonstrated with a reaction time task, that adults show categorical perception (CP) of colour in the right visual field (RVF), equating to the left hemisphere of the brain. The fact that CP was found only in the hemisphere of the brain that mediated language use was taken as evidence that CP of colour was linguistically mediated. The results of the study brought to light the possibility of powerful Whorfian influences of language on perception (Regier & Kay, 2009; Regier & Xu, 2017).

However, despite a large number of replications of the work with varying populations (Al-Rasheed, Franklin, Drivonikou, & Davies, 2014; Al-Rasheed, 2015a, 2015b, 2016; Drivonikou et al., 2007; Franklin, Catherwood, Alvarez, & Axelsson, 2010; Gilbert, Regier, Kay, & Ivry, 2008; Goldstone & Hendrickson, 2009), concerns surrounding the set up of the study by Gilbert et al. (2006) remain. Can evidence of a faster reaction time in one hemisphere be taken to mean that CP occurs in that hemisphere? Can those same effects be replicated with carefully-controlled stimuli (Witzel & Gegenfurtner, 2011)? Is reaction time even an appropriate measure for CP of colour (A. M. Brown et al., 2011)?

Despite the considerable success of both the original study and of subsequent investigations, there have also been considerable failures to replicate the findings of Gilbert et al. (A. M. Brown, Lindsey, Rambeau, & Shamp, 2009; Jraissati, 2012; Lindsey & Brown, 2009; Ocelak, 2016; Witzel & Gegenfurtner, 2013, 2016). Many claimed that the lateralisation of CP was a result of imperfections in the original studies, with some claiming that CP of colour itself may be in doubt, at least as measured with reaction times (A. M. Brown et al., 2011). Thus there is considerable cause to re-evaluate the claims of CP using measurements other than reaction times, a key aim of Chapters 6 and 7.

1.3.2 Categorical perception in infants

The considerable debate surrounding CP of colour was split between those claiming it was language-mediated (e.g. Roberson et al., 2000), and those claiming it was biological (e.g. Bornstein, Kessen, & Weiskopf, 1976). Thus testing infants and toddlers with varying levels of language acquisition became an important field of research. Categorical perception in infants is equally well-established. In a series of experiments with English and Himba[2] tod-

[2]An indigenous northern Namibian tribe, who also live in Angola. Himba has only 5 basic colour terms, as opposed to the 11 existent in English (Franklin, Wright, & Davies, 2009).

dlers, it was established that toddlers also display CP of colour, in much the same way as adults (Franklin, Clifford, et al., 2005). This same CP of colour was also demonstrated with infants as young as 4 months of age (Franklin, Pilling, & Davies, 2005), while similar effects were found with children of different language groups, up to around 7 years of age (Daoutis et al., 2006).

However, in a replication of earlier studies, Goldstein, Davidoff, and Roberson (2009) demonstrated that only toddlers with a fuller knowledge of the colour categories showed CP, both for Himba and English toddlers. The primary difference was that Goldstein et al. employed a much stricter definition of colour word knowledge than that used earlier by Franklin, Clifford, et al. (2005). In addition, Goldstein et al. demonstrated that Himba toddlers did not demonstrate CP over boundaries where English has a colour boundary but Himba does not, lending further weight to the linguistic determinism originally espoused by Roberson and Davidoff (2000). The contrasting theories of cognitive versus linguistic determinism were thus still unreconciled, even in infant research.

A crucial point came when the same type of hemispheric effects as shown by Gilbert et al. (2006) were demonstrated with pre-linguistic infants. Using an eye-tracking paradigm, Franklin, Drivonikou, Clifford, et al. (2008) demonstrated that infants have a faster latency to the first look for the between-category colour. However, unlike the findings of Gilbert et al., infants in the study by Franklin, Drivonikou, Clifford, et al. (2008) demonstrated CP of colour in the left visual field, rather than the right, reversing the hemispheric lateralisation seen in adults. Similar findings were made when comparing toddlers who did not know colour words to those who do; the hemispheric lateralisation of CP was found to reverse with the learning of colour terms (Franklin, Drivonikou, Bevis, et al., 2008). Thus it appeared evident that infants possess a different kind of CP of colour to those of adults, exchanging early, biological colour categories (Skelton et al., 2017) for linguistically-mediated CP with the learning of colour words.

As with the original Gilbert et al. (2006) study, the findings of Franklin, Drivonikou, Clifford, et al. (2008) did not pass without considerable scrutiny. The results were based on only 26 pre-linguistic infants, exactly half of whom were removed for fussiness. A result based on 13 infants may not be replicable, similarly a task with a 50% drop-out rate raises questions as to the reliability of those not excluded (Hanley & Roberson, 2008). In addition, the quality of the colour vision of infants only 26 weeks old may not be suitable for such a task (Roberson & Hanley, 2009). Finally, the reaction time criticism levelled at Gilbert et al. still applies in this case (A. M. Brown et al., 2011), perhaps more so, as an eye-tracker measuring at 50Hz may not be fast enough to capture the subtleties of reaction times using saccades. Franklin, Drivonikou, Bevis, et al. (2008) further examined the phenomenon of lateralisation of CP changing with colour term acquisition with toddlers, who apparently were on either side of colour term learning. However, in that sample, the mean age of participants was over 3 years old, an age when even many previous studies expected colour word learning was already well under way (Mervis et al., 1995; Pitchford & Mullen, 2002; Sandhofer & Smith, 1999).

Chapters 6 and 7 of this thesis aim to examine CP of colour in infants while avoiding the pitfalls found in previous studies. By employing novel sets of tasks that do not rely on reaction times to distinguish CP, various aspects of CP of colour in infants can be examined, and the presence of a top-down effect of language can be assessed.

1.3.3 Colour vision tests in infants

The experiments run for this thesis, and elsewhere, on topics relating to colour with infants, generally have a major caveat: it cannot be categorically ruled out that some participants may have colour vision defects. Here, and elsewhere (e.g. Maule, Witzel, & Franklin, 2014; Maule, Stanworth, Pellicano, & Franklin, 2017; Skelton et al., 2017) a probabilistic approach is often taken, ruling out from analysis any participant who reports a family history that suggests that the participant has a certain probability of colour vision problems.

Probabilistic approaches to anything genetic-related do come with significant drawbacks. However, to date, there are no reported colour vision tests that are applicable to toddlers below an age where they can communicate either with gesture or with word, with some degree of reliability. Due to the propensity of young toddlers to employ colour words incorrectly, or to overextend them (e.g. Wagner et al., 2013), parents often fear that the mislabelling may be due to colour vision problems, particularly when there is a family history.[3]

There is no shortage of available colour vision tests for adults. Many of them involve a response from the participant: in the case of pseudoisochromatic plates, it requires the participant to verbally respond to what was seen or to trace it with the finger (Ishihara, 1917), or in the case of other colour tests to match the plates or tiles by colour by physically moving them in to place. More recent dynamic tests require a participant to say when they are no longer able to see a moving stimulus (Barbur, Harlow, & Plant, 1994). These tests are problematic for infants who are unable to communicate reliably with a strong degree of accuracy, or who cannot yet be relied upon to dynamically respond

[3]This is purely anecdotal, but a very common claim in caregivers of participants tested for this thesis. Discussions with other researchers have yielded similar fears from caregivers.

to a stimulus.

In Chapter 8 of this thesis, a model prototype for a colour vision test is discussed, and possible limitations of its application with infants are investigated. The test discussed in Chapter 8 does not require a response from the participant other than looking, which infants can do with a considerable degree of success from a young age (Golinkoff, Mervis, & Hirsh-Pasek, 1994).

1.4 Outline of thesis

Below, the structure of the remainder of the thesis is outlined. In Chapter 2, both the Oxford CDI (Hamilton et al., 2000), and CDIs from 11 different languages (Frank, Braginsky, Yurovsky, & Marchman, 2016) are modelled with Bayesian growth curves to examine the timeline and patterns of colour word learning both individually for each language, and cross-linguistically. These are then compared with a frequency-based approach, to examine predictions of when colour words would be learned in different languages.

In Chapter 3, the veracity of the claims made in Chapter 2 is cross-examined with a large-scale eye-tracking task. Participant responses to the individual colours both before and after they have been named, and overall approaches to colour word learning are examined. The pre-naming element of the study can be used to examine infant colour preferences of focal colours, while the post-naming element examines colour word learning. In addition, Chapter 3 examines infant expectations that colour words should appear in the adjectival position.

In Chapter 4, claims that colours closer to the periphery of colour categories are hard to learn (Wagner et al., 2013) are tested with an examination of light and dark colours, as well as the words "*light*" and "*dark*." Preferences and word learning in Chapter 4 are examined in the same way as in Chapter 3.

Chapter 5 attempts to replicate and extend previous work suggesting that colour words are not necessary for colour word mediated shifts in attention (Johnson et al., 2011). In that chapter, the claim that infants systematically look to a red plate when they hear the word *"strawberry,"* even without knowing the word *"red,"* is tested by way of an eye-tracking task.

Chapter 6 examines categorical perception of colour across the blue-green boundary in infants. They way that the colour category boundary may effect infant looking preferences to the between-category colour or the within-category colour are examined. In addition, the role that learning a colour word has on the preferences is also examined.

Chapter 7 tests a similar question to that of Chapter 6, but on this occasion measures the effect of the category boundary on the attention of infants to a stimulus switching between the two colours – a novel paradigm for this type of study. As with Chapter 6, in Chapter 7 the effect of learning colour words on the attention paid in this task is also examined.

In Chapter 8, a novel paradigm for an infant colour vision test is explored, by examining the feasibility of such a task with respect to infant visual closure. The effect that vocabulary size has on this task and the general limitations of colour vision testing at young ages are also discussed.

In Chapter 9, I summarise the main findings of this thesis and make conclusions based on the results of the experiments. Future directions based on this work are also discussed.

Chapter 2

Linguistic and Cultural Variation in Early Colour Word Learning

2.1 Introduction

The domain of colour perception and categorization has played a central role in furthering our understanding of the impact of language on cognition, and of cognition on language for over 60 years (as discussed in Chapter 1, R. W. Brown & Lenneberg, 1954; Berlin & Kay, 1969; Heider, 1972; Cohen, Chaput, & Cashon, 2002; Roberson et al., 2000; Roberson, Davidoff, Davies, & Shapiro, 2005). In normal discourse, adult speakers treat colours categorically by grouping them into blocks linguistically, despite their continuous nature. With mounting evidence of emergent colour categories in infants (e.g. Franklin, Drivonikou, Clifford, et al., 2008; Skelton et al., 2017) that may adapt as colour terms are learned (Franklin, Drivonikou, Bevis, et al., 2008), it appears colour categories have a strong biological component. However, some of the questions discussed in the previous chapter are still without consensus. What brings about the change from infant biological colour categories to adult categories? What contribution remains for linguistic and cultural components in setting boundary conditions on the learning process? Does developing from a universal colour category imply a universal order in the learning of colour categories?

Explanations of the formation of colour categories has been the source of much debate. On the one hand, evidence from cross-linguistic differences in the perception of colour have shown how categorical perception can differ by language group (Roberson, Pak, & Hanley, 2008; Roberson, Hanley, & Pak, 2009), giving weight to the idea that colour categories are formed culturally (Roberson et al., 2000). On the other, analyses of World colour Survey (WCS) data (Kay et al., 2011) have shown universal similarities in colour naming across different languages (Abbott, Griffiths, & Regier, 2016; Kay, 2003; Regier, Kay, & Cook, 2005; Regier, Kay, & Khetarpal, 2007).

Evidence from infant experiments have provided a fresh perspective on this debate. Infants have been found to possess categorical perception (CP) of colour in the right hemisphere (Franklin, Drivonikou, Clifford, et al., 2008), as opposed to the left hemisphere in adults (Gilbert et al., 2006). Similar results were found comparing toddlers who had not learned colour terms, and therefore behaved like infants, with those who had, thus behaving like adults (Franklin, Drivonikou, Bevis, et al., 2008).

Recently, strong evidence has been reported for biological, pre-linguistic colour categories in a novelty-preference task, suggesting the presence of infant colour categories (Skelton et al., 2017). Skelton et al. found that when the results were plotted in a colour space representative of the retinogeniculate pathways that make up colour vision, most of the categorical distinctions made by infants were separated by the axes in that colour space. The results of that study therefore suggested that there is a strong association between the cardinal mechanisms of colour vision, and the way in which infants categorize colours. The infant colour categories were found to be similar to the category centroids of non-industrialized languages, suggesting some commonality with adult colour categories. However, there is great diversity in how the colour spectrum is divided across languages, suggesting that at some point, language and culture intervene and change the way colour is categorized from the original, biological infant categories, to make it more relevant to the language and culture in question. The diversity between languages in the number of colour words used may also lead to variability in the timing of acquisition of these terms.

The fact that there is a necessary transition from infant colour categories to adult colour categories (Skelton et al., 2017) may also provide further insights into why colour word learning is perceived to be difficult (Franklin, 2006; Johnson & Huettig, 2011; Mervis et al., 1995; Pitchford & Mullen, 2003; Soja, 1994; Wagner et al., 2013). Researchers have argued that infants learn colour words relatively late compared to other classes of words (Heider, 1971; Shatz et al., 1996; Soja, 1994), and that early colour-word usage is riddled with haphazard, random usage (Pitchford & Mullen, 2003; Sandhofer & Smith, 1999). As discussed in detail in the previous chapter, explanations have focused on the need to learn to categorise the continuous spectrum of colour (Kowalski & Zimiles, 2006), or the dominance of shape over colour as a salient dimension (Sandhofer & Smith, 1999). Despite this speculation, the reported age of acquisition of colour terms seems to have dropped dramatically in recent decades (Shatz et al., 1996; Franklin, 2006). That colour word learning really is difficult is thus worthy of reassessment.

The claim that colour categories have a biological root in infancy (Skelton et al., 2017), as well as the claim that colour categories in adults possess a common root in infant colour categories might suggest a universal order of colour term learning based on this biological root (see Bornstein, 1985; O'Hanlon & Roberson, 2006; Pitchford & Mullen, 2002). While Berlin and Kay (1969) found a general order in which colour words were developed by languages, no substantive evidence has been found to suggest that this might be mirrored in the way that infants learn colour categories. Evidence for a systematic, universal order of colour word learning has been equivocal (Andrick & Tager-Flusberg, 1986; Pitchford & Mullen, 2002; Shatz et al., 1996), suggesting that there have been some trends observed, but no overarching pattern. Most of these studies did not consider a wide range of languages, thus also limiting the ability to test whether there are emergent patterns in colour word learning.

In light of these findings, the current chapter explores some linguistic and cultural determinants for learning early colour words and colour categories in different languages, and analyses the developmental profile in colour word learning in order to examine the presence or absence of a universal order of colour word learning. To address these issues, data was employed from existing parental surveys of children's word learning, also known as Communicative Development Inventories (CDIs, Fenson et al., 1994). CDIs are generally considered valid and reliable indicators of infant word-learning (E. Bates et al., 1988; Dale, Bates, Reznick, & Morisset, 1989; Dale, 1991; Mills, Coffey-Corina, & Neville, 1993, 1997; Fenson et al., 1994; Styles & Plunkett, 2009b). CDIs have the advantage of being able to measure infant vocabulary on a large scale, with parents often asked to assess one or both of comprehension and production for each word. There has been considerable debate about the reliability of this measurement however, particularly when measuring comprehension, especially for terms as abstract as colour words (Houston-Price, Mather, & Sakkalou, 2007; Tomasello & Mervis, 1994). Studies have consistently found that CDIs measure word learning quite consistently when compared to normative developmental scales (Dale, 1991), and have been confirmed with measurement from ERPs (Mills et al., 1993, 1997). More recently, CDI validity of comprehension was tested against behavioural measurement (Styles & Plunkett, 2009b), finding that parents are quite conservative in their determinations of whether the infant comprehends a term, and that comprehension, as measured by parental report, is quite an accurate measurement.

In Study 1 of this chapter, I investigate toddler's comprehension and production of basic colour words in British English. If indeed colour word learning occurs late, as previous studies have suggested, we would expect to find that many of the participants will not have comprehended the colour terms by 2;6. Study 2 in this chapter extends this exploration to 11 other languages but for production only. Study 2 allows us to test whether colour words are learned following a universal order from their biological roots. If so, the overall order in which colour words are learned should show little variability across languages. Finally, Study 3 of this chapter examines the impact of colour word frequency and syllabic complexity of the different colour word forms across these languages, in attempt to identify potential sources of variation. Similar to the goals of Study 2, if colour words are learned purely based on their biological foundations, then frequency and complexity would not be expected to be strong predictors; alternatively if cultural and linguistic factors shape colour word learning, then frequency and complexity should be important factors.

2.2 Study 1

2.2.1 Methods

2.2.1.1 Participants

2962 8- to 30-month old participant's details were filled out by parents, either on paper or online before a visit to testing facilities either at the Plymouth Babylab or the Oxford BabyLab. Participant information that was lacking in either age or gender information was not included in the analysis. The majority of infants visited the laboratories only once, making these analyses cross-sectional, rather than longitudinal in character. A small number of participants visited more than once, giving a total of 3413 completed CDIs (1653 female).

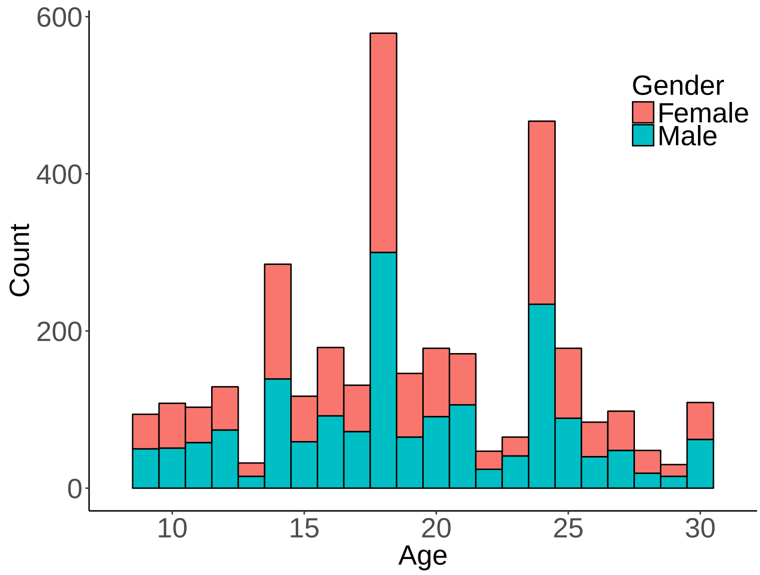

Figure 2.1: Recruitment information for Study 1.

2.2.1.2 Materials

In Study 1, previously collected data from the Oxford Communicative Development Inventory (Hamilton et al., 2000) was used to examine colour word comprehension and production. The Oxford CDI is a British adaptation of the MacArthur-Bates CDIs (MB-CDIs Fenson et al., 1994, 2007), measuring comprehension and production in 416 terms, and used from the earliest stages of word learning, up until around 30 months of age. The Oxford CDI contains 4 colour terms: *red*, *blue*, *green*, and *yellow*.

2.2.1.3 Analysis

In this, and the following analyses, parental report data is modelled with Bayesian binomial models. The objective of this analysis is to fit a curve to data that is binomially distributed (yes/no data), and in doing so be able to calculate a developmental trajectory (see Mirman, 2014; Bürkner, 2017, for general frequentist examples of modelling, along with some Bayes-specific examples). In addition to being able to view these trajectories, the population-level coefficients of the model provide information as to the effects that shape the model. In the approach used here, it can be considered strong evidence for a coefficient being an important factor in the model if the 95% Credible Interval of the coefficient does not intersect with 0 (Kruschke, 2013). This kind of model could be fitted with either a frequentist Generalized Linear Model, or a Bayesian model. The choice of Bayesian analysis was made in order to make inference-based analysis on the model output, while simultaneously avoiding shortcomings associated with some frequentist models (Cumming, 2014). In addition, the Bayesian method allows for greater flexibility in modelling, and the ability to fit complex models that maximum likelihood methods can fail to capture (Bürkner, 2017).

The 4 colour terms were isolated for each participant, and modelled with two separate Bayesian binomial models, each with 4 chains of 12000 iterations, of which 2000 were a warm-up. The chains were thinned by 2, to allow minimal autocorrelation. Both models included Age, Gender and the colour word in question as population-level (fixed) effects, as well as an interaction between Age and Gender to allow the possibility of different slopes for each gender. Colour and Gender, both categorical variables, were treatment coded, comparing to Blue in the case of colour, and to Female in the case of Gender. Age was treated for these analyses as a continuous, numeric variable. The data was modelled using brms (Bürkner, 2015), running in rstan (Stan Development Team, 2016).

An identical model was run on the production data. Priors were largely uninformative Student t-distributions with 10 degrees of freedom, a mean of 0, and deviance of 1 for all of the population-level effects. Both models were checked for proper convergence with no divergent steps, and a $\hat{R}$ of less than 1.1, with an effective sample size of greater than 10% of the total sample size, and a Monte Carlo Standard Error of less than 10% of the posterior standard deviation.

2.2.2 Results

The fit of the binomial curves can be seen in Figure 2.2. Overall there is a clear difference between comprehension and production for each of the colours, such that approximately 50% of infants are reported to comprehend, for example, blue at age 21 months, but only 25% of infants produce it at the same age. A similar difference between comprehension and production exists for each of the colour words tested. Below the results of the comprehension data and production data are discussed separately, in detail.

The results of the fitted model to comprehension data in the Oxford CDI suggest that each of the colour words are comprehended by around 50% of the infants at 21 months. From examining the graph, *blue* appears to be the

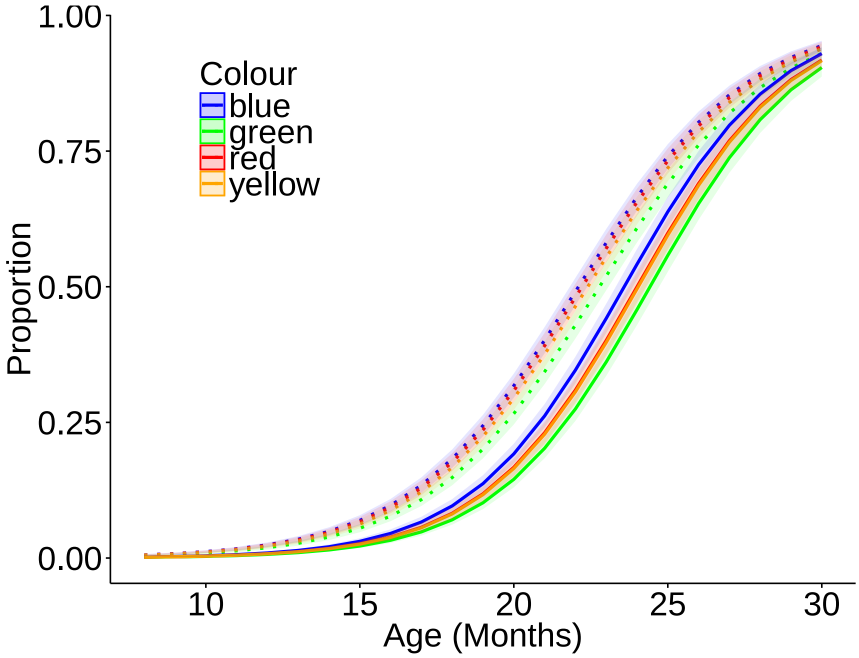

Figure 2.2: Results of two fitted Bayesian models to the Oxford CDI comprehension and production data. Dotted lines indicate comprehension, solid lines indicate production. Narrow bands around each line indicate the credible interval of the mean.

first of the colour words comprehended, by a small margin, while there is strong visual evidence that *green* is the last of the four colour words to be comprehended. These findings are supported by the results of the model as seen in Table 2.1, which show that *green* is learned later than *blue*, which is the reference colour in this model. The "Est." column gives the estimate of the means of the posterior distribution, while the two columns under "95% CI" give the upper and lower 95% Credible Interval around the estimate, and the "Err." column denotes the standard error on the estimate. "Samples" denotes the number of samples gathered for each individual parameter. Thus there is evidence for a difference between *blue* and *green*, where the 95% Credible Interval does not overlap with 0.

The results of the model also suggest a gap in comprehension between male and female participants. Notably, while there appears to be a possible

Table 2.1: Results of model on comprehension of colour words in the Oxford CDI. Colours are compared to *Blue*.

	Est.	Err.	95% CI		Samples
Intercept	-8.40	0.19	-8.78	-8.03	14753
Age	0.39	0.01	0.37	0.41	14760
Green	-0.25	0.07	-0.38	-0.12	16302
Red	-0.04	0.07	-0.17	0.08	15952
Yellow	-0.11	0.06	-0.24	0.02	16173
Male	0.61	0.26	0.11	1.12	14082
Age:Male	-0.04	0.01	-0.07	-0.02	14145

early advantage for male infants comprehending colour words, the interaction with age provides strong evidence of a shallower slope in learning for male infants than for female infants, indicating an overall advantage for females.

Table 2.2: Results of model on production of colour words in the Oxford CDI.

	Est.	Err.	95% CI		Samples
Intercept	-9.28	0.23	-9.72	-8.84	14587
Age	0.40	0.01	0.38	0.42	15227
Green	-0.34	0.07	-0.48	-0.20	17519
Red	-0.17	0.07	-0.31	-0.03	15388
Yellow	-0.18	0.07	-0.32	-0.04	16410
Male	-0.54	0.32	-1.16	0.07	14385
Age:Male	0.01	0.01	-0.02	0.03	14326

The model of production of colour terms in the Oxford CDI (Table 2.2) reveals a similar profile of learning to produce the terms to that of the comprehension data. Again, there is evidence that *green* is produced slightly later than the other colours, most notably *blue*. For production, the 95% credible interval for the difference between *blue* and each of the remaining three colour words does not intersect with 0. There is no strong evidence for a difference in either baseline or slope between the two genders in the case of production.

Finally in this study, a subset of these participants' caregivers were asked to fill out supplementary information to the Oxford CDI. This information asked caregivers to confirm whether the child comprehended or comprehended and produced all 11 basic colour terms (*red, blue, green, yellow, black, white, pink, orange, purple, brown,* and *grey*). A total of 256 participants completed the supplementary information. Table 2.3 shows the information by age group.

Table 2.3: Ages of participants who completed supplementary Oxford CDI data.

Age	*N*
12	52
16	59
18	4
19	72
24	39
26	5
28	5
30	20

Identical models for comprehension and production to the previous models were fit to the data, in order to examine trajectories of all the 11 basic colour words. The fit of the binomial curves can be seen in Figure 2.3. While the sample size for this data is limited, both models fit the trends for comprehension and production from the full Oxford CDI data, and show the close contiguity from the first four terms (*red, green, yellow,* and *blue*).

The supplementary data to the Oxford CDI shows that parents report *grey* to be comprehended and produced last, and brown to be learned moderately late, along with *black* and *white*. *Purple, pink,* and *orange* are produced and comprehended between the first four colours, and the later four. Overall the timing for learning these colour terms seems to agree with the overall Oxford CDI data. Model coefficients can be seen in Table 2.4 and Table 2.5, discussing the comprehension data and production data respectively.

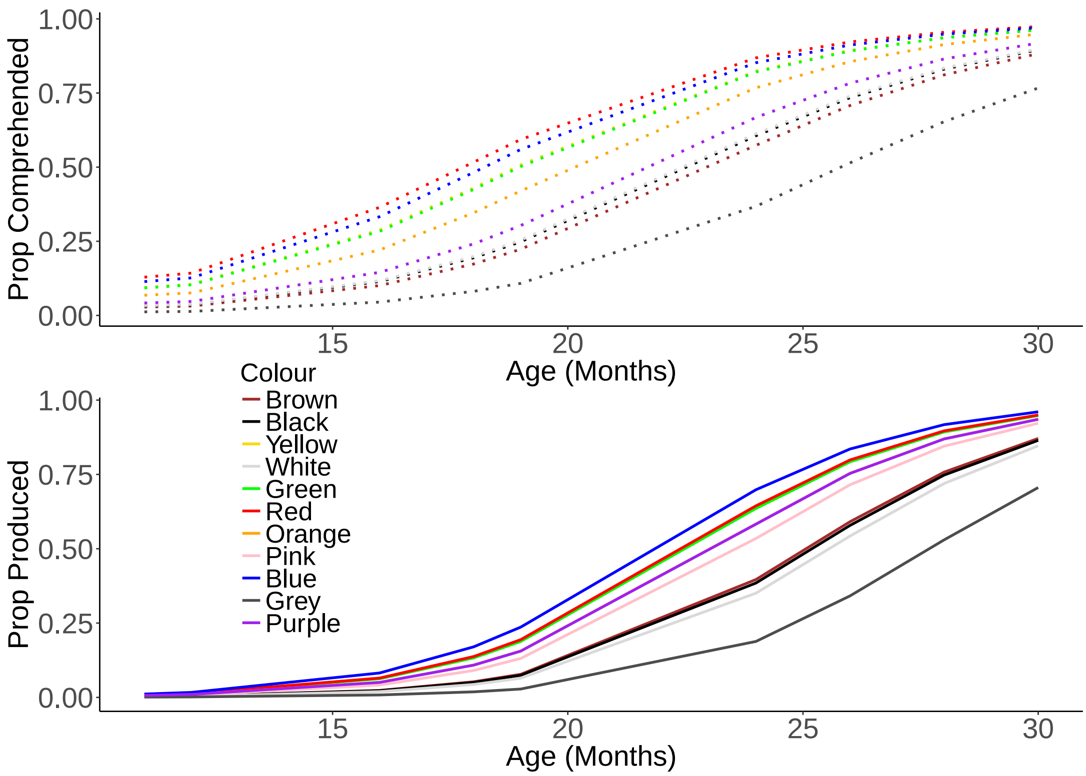

Figure 2.3: Comprehension and production data from a subset of Oxford CDI participants, fitted with binomial curves. The trends for some colour terms (e.g. production of *orange*), directly overlap with other terms. 95% CI are not displayed here for ease of viewing.

2.3 Study 2

2.3.1 Method

2.3.1.1 Participants

For Study 2 of this chapter, data from 22,642 participants was downloaded from the Wordbank database http://wordbank.stanford.edu/ (Frank et al., 2016) on 18/11/2016. Data was downloaded for 11 languages based on two selection criteria: first, the data needed enough participants to make it a generalizable sample, for the purposes of this experiment that was 600 participants. Second, the CDI data for that language needed to contain each of the 6 colour terms being examined (the four in Study 1 of this chapter, plus *black*

Table 2.4: Comprehension model for the supplementary data.

	Est.	Err.	95% CI		Samples
Intercept	-7.92	0.42	-8.76	-7.11	10661
Age	0.36	0.02	0.32	0.4	11445
Black	0.13	0.23	-0.31	0.57	10919
Yellow	1.29	0.22	0.87	1.72	10070
White	0.16	0.22	-0.27	0.6	10659
Green	1.26	0.22	0.84	1.69	9278
Red	1.64	0.22	1.2	2.07	10049
Orange	0.93	0.22	0.51	1.35	9970
Pink	0.42	0.22	-0.01	0.85	10519
Blue	1.5	0.22	1.08	1.91	9835
Grey	-0.88	0.25	-1.38	-0.39	12232
Purple	0.42	0.22	-0.02	0.84	10340
Male	1.4	0.45	0.52	2.31	10978
Age:Male	-0.09	0.02	-0.14	-0.05	11013

and *white*). Participants older than 2;6 were excluded from the analysis. Final participant numbers for each of the language groups can be seen in Table 2.6.

2.3.1.2 Materials

In each data set downloaded, participants' guardians had filled out the MB-CDI in the first language being developed by the child. Foreign language adaptations of the CDIs are often not direct translations of each of the terms, but instead are adapted to account for the differences in language. Each different language CDI set contained production data from participants between 1;4 and 2;6, with the exception of Russian, German, and Italian which began at 1;6, and Swedish which contained data from participants at three-month intervals from 1;4 to 2;4. In English, the terms used were *red*, *green*, *yellow*, *blue*, *black*, and *white*. All the languages used in this study had corresponding words for these terms. Where multiple terms exist for one of these colours, the most common one was used. In Cantonese, the standard modern equivalents for these terms were used, while Mandarin was already translated into

Table 2.5: Production model for the supplementary data to the Oxford CDI.

	Est.	Err.	95% CI		Samples
Intercept	-10.74	0.54	-11.82	-9.7	11728
Age	0.45	0.02	0.4	0.49	12514
Black	-0.06	0.28	-0.62	0.49	11573
Yellow	1.06	0.27	0.54	1.58	10087
White	-0.21	0.29	-0.78	0.36	11494
Green	1.02	0.27	0.5	1.55	10098
Red	1.06	0.27	0.54	1.58	10630
Orange	0.78	0.27	0.26	1.32	10513
Pink	0.58	0.27	0.06	1.12	11427
Blue	1.31	0.26	0.8	1.84	10083
Grey	-1.1	0.32	-1.73	-0.48	11980
Purple	0.79	0.27	0.26	1.32	10442
Male	0.69	0.58	-0.43	1.84	11458
Age:Male	-0.06	0.03	-0.11	-0.01	11453

English, it is expected that the terms used were the same as used in Cantonese. In Russian, the word *siniy* was used for *blue*, as *goluboy* was not included in the CDI. The full list of words used in the analysis can be viewed in Table 2.7.

2.3.1.3 Analysis

The data was modelled in a similar fashion to that of Study 1 of this chapter, except that the number of iterations was increased to 20,000, of which 4000 remained as a warmup, to allow for the larger dataset. Age, Gender, and Colour were again population-level effects, also including an interaction between Age and Colour. Age and Colour were both nested within the group-level effect of Language. Priors for population-level effects were as in Study 1, priors on group-level standard deviations were default half t-distributions with 3 degrees of freedom, while priors on group-level correlations were default Cholesky factors (Bürkner, 2015).

Table 2.6: Numbers of participants in each language of the MacArthur-Bates CDI surveys.

Language	*N*
Cantonese	987
Danish	2863
English(US)	5450
French(Quebec)	827
German	1183
Italian	639
Mandarin	1056
Norwegian	6931
Russian	712
Spanish	1094
Swedish	900

Table 2.7: List of words used in Study 2, based on MB-CDIs.

Language	Terms				
English	red	green	yellow	blue	black	white
Cantonese	hung	luk	wong	laam	hak	baak
Danish	rød	grøn	gul	blå	sort	hvid
French	rouge	vert	jaune	bleu	noir	blanc
German	rot	grün	gelb	blau	schwarz	weiß
Italian	rosso	verde	giallo	blu	nero	bianco
Mandarin*	red	green	yellow	blue	black	white
Norwegian	rød	grønn	gul	blå	svart	hvit
Russian	krasnyy	zelenyy	zheltyy	siniy	chernyy	belyy
Spanish	rojo	verde	amarillo	azul	negro	blanco
Swedish	röd	grön	gul	blå	svart	vit

NB: Mandarin CDI data was made available in English.

2.3.2 Results

Figure 2.4 shows the different trends in producing colour words, dependent on the language being learned. The model again converged with $\hat{R} = 1$ and no divergent transitions. The model coefficients Table 2.8 show very strong evidence for an effect of Age, as well as strong evidence for an effect of Gender, with male participants being generally behind female participants.

With the exception of *white*, the model suggests there is convincing evidence for the other colours to be produced ahead of *black*, with the 95% CI on

Table 2.8: Main population-level effects on fitted model of MB-CDI data in 11 languages. Colours are as compared to black. Interaction terms and group-level terms are not included.

	Est.	Err.	95% CI		Sample
Intercept	-8.84	0.47	-9.76	-7.91	21363
Age	0.34	0.02	0.3	0.38	21851
Male	-0.4	0.01	-0.43	-0.37	24205
Blue	0.98	0.24	0.5	1.45	25532
Green	0.48	0.21	0.07	0.89	26713
Red	0.83	0.2	0.44	1.22	27062
White	-0.11	0.26	-0.62	0.41	27915
Yellow	0.88	0.21	0.47	1.3	26174

the difference between them and *black* not including 0. This is consistent with the graphs depicted in Figure 2.4, where the four primary colours are produced before *black* and *white* in many of the languages examined. Overall, the general trend is a close contiguity with the four chromatic colour words, and then again a close contiguity between the two achromatic words. In many cases *red* or *blue* are the first terms learned, consistently above *yellow* and *green*.

In the five Germanic languages examined (English, German, Danish, Norwegian, and Swedish), there is a close contiguity in the production of the four primary colours, with *green* possibly the last of those four colours, except in the case of German. In these languages, each of the colour terms has been produced by at least 75% of infants at 30 months.

Figure 2.4 shows that the time-course of colour word production in Romance (French, Italian, and Spanish) languages is not as uniform as it is in the Germanic languages. In French, colour word learning happens relatively early, with all six colour terms known by around 75% of infants by 30 months. In Italian, colour word learning is slightly later than it is in French, while in Spanish, colour word learning happens much later, with each colour word only produced by around 50% of participants tested at 30 months. The most notable feature of the Romance languages examined here is that although *black* and *white* tend to follow behind the four primary colours, as with the Germanic languages, they do not do so by as large a margin as in the Germanic languages. *Black* is consistently the last word produced in these three languages, albeit by a small margin. The other consistent aspect in the Romance languages is that *green* tends to be produced after the other three primary colours, closer in timing to *white*.

Within the two Sinitic languages (Mandarin and Cantonese), there are distinct differences in the timing and order of colour words. While a dominant feature of the two Chinese languages is very early production of the word for *red*, in Mandarin this is matched by *white* as one of the first colour words produced, which is not the case in Cantonese. In Cantonese, the other five colour words are produced at essentially the same rate, whereas in Mandarin, *blue* is produced around a month later than the remaining three colour terms. There is a large difference in the rate of colour word learning between these two languages as well. In Mandarin, the parents report that almost all of the six colour terms are produced and understood by almost all infants by 30 months of age. In contrast, in Cantonese, most colour words are produced by about 60% of infants by 30 months, with the exception of *red* which is known by around 75% of infants.

In Russian, the overall pattern of colour word learning is not dissimilar to that of the Romance languages. The majority of colour words, excluding *white*, are produced by around 70% of infants by 30 months, and *black* is produced after the primary colours, although not by the amount seen in the Germanic languages. Where Russian differs greatly is that *white* is produced long after the other colour terms, with a gap of around 3 months. By 30 months of age, only around 30% of Russian infants are reported to produce the word *white*.

In Russian, the last of the primary colours to be produced is *yellow*, reflecting another possible difference in ordering.

2.4 Study 3

2.4.1 Methods

2.4.1.1 Materials

For Study 3, two sets of data are included. The first is the same set of MB-CDI data used in Study 2 of this chapter, while the second is data on the frequency of occurrence of words. Frequency data was obtained by downloading CHILDES (MacWhinney, 2000) CHAT transcripts from the website http://childes.talkbank.org/ on 19/12/2016 and examining the frequency of colour word appearance in each of the languages tested. Frequencies were then tallied up for each colour term, and divided by the total number of words to yield a proportion (see Figure 2.5). In line with Goodman et al. (2008), frequency of input was approximated by counting the frequency of appearance of each colour word in each language in CHILDES when spoken by the mother. Frequency varies substantially between languages and colours.

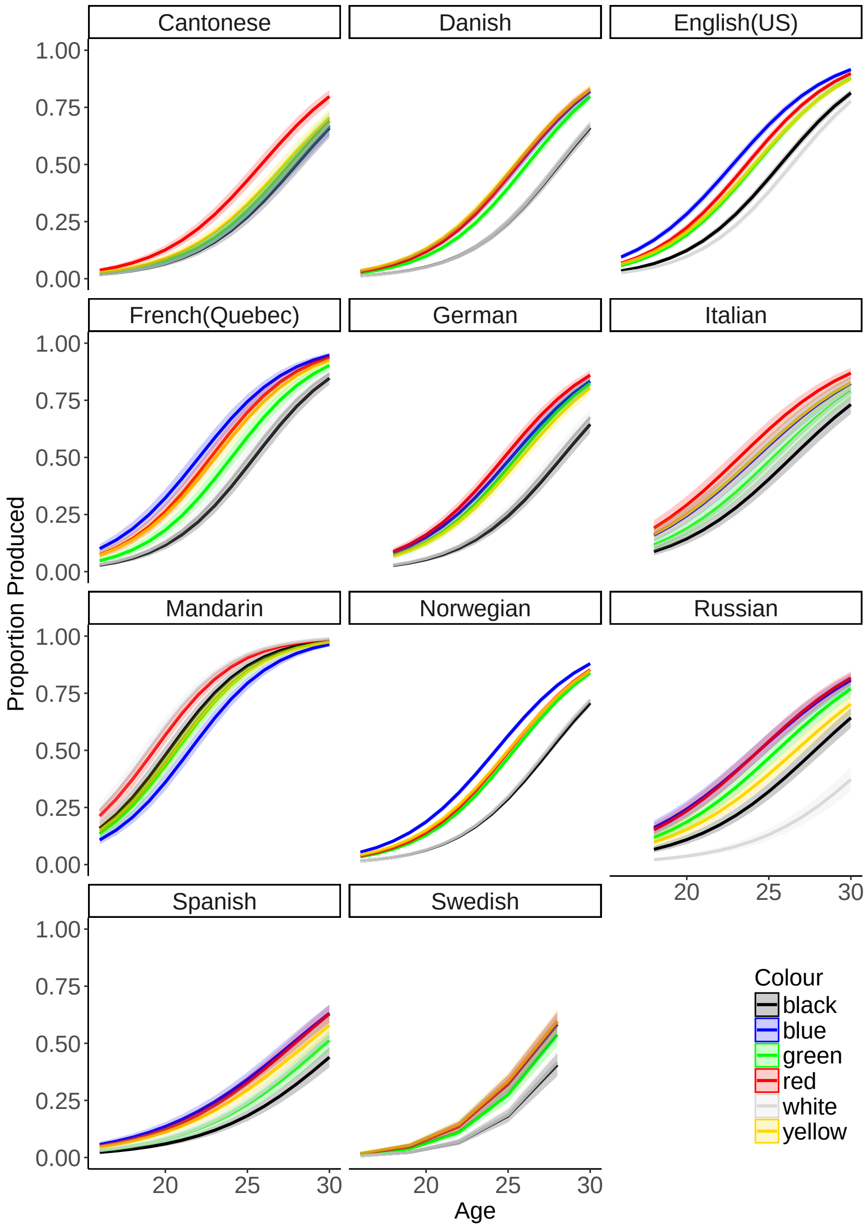

Figure 2.4: MacArthur-Bates CDI production data modelled with a binomial Bayesian model, separated by language. Bands around each line indicate confidence in the mean.

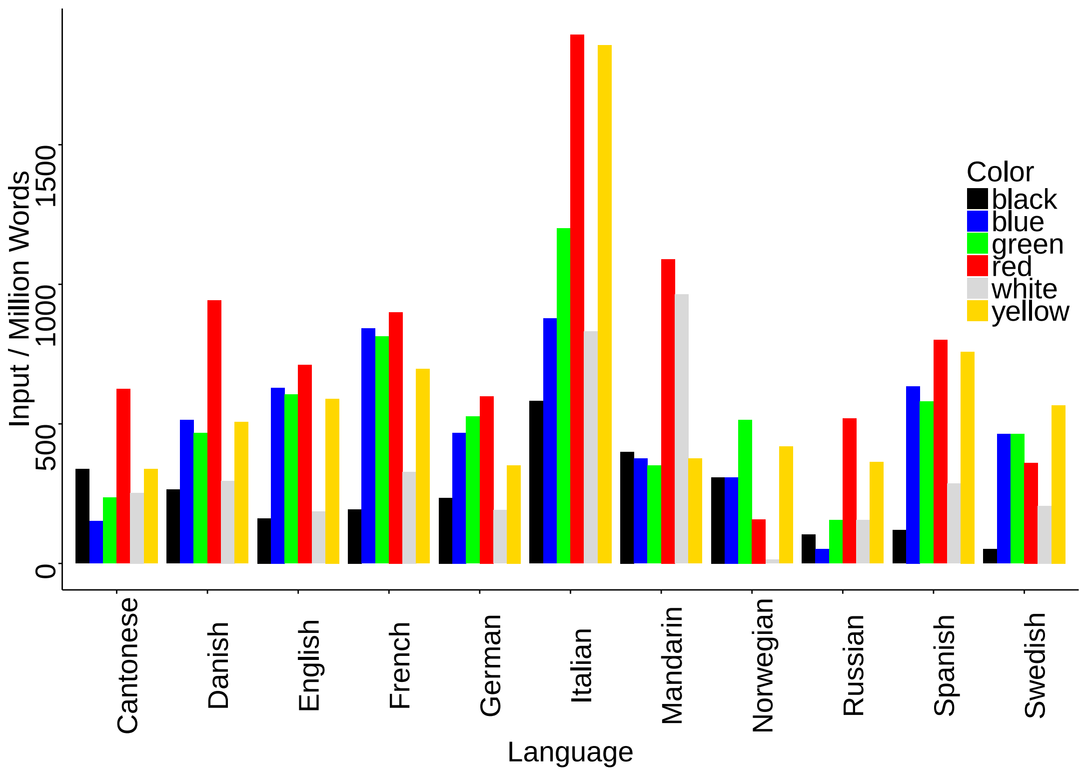

Figure 2.5: Input frequency of occurrence of colour words for infants up to 3 years of age, by language.

It has previously been shown that frequency, category size (i.e. how many shades are encompassed by the single term) and perceptual salience (i.e. the Euclidean distance from grey at the centre of the space) can predict precise colour word learning in a behavioural task (Yurovsky et al., 2015). In the present study, only frequency is used as a predictor, of those three possible options. While it is extremely likely that category size and perceptual salience are useful predictors, data on this for all of the languages included in this study is not currently available.

In each language, only transcripts in which the infants were three years or younger were used, and only occurrences in which the meaning of the colour term could be understood by native speakers to refer to the term, and not idiomatic expressions were accepted. For many of the European languages, care was taken to ensure that colour terms in each separate gender were included. Compound words, where the noun is made up of a colour word and another word (e.g. blueberry), were not included in the count, particularly in the case of the Sinitic languages, where the compound word may not require an understanding of the meaning of the colour words involved. As an example, the term *hong lu deng* (lit. red green light, meaning traffic light) was not included as one could understand the meaning of the term without necessarily understanding the words referring to *red* and *green*.

In two cases, because the colour term was a homonym with another commonly-used term, or transcribed the same way (German white and know), the frequency had to be predicted. In the German case, the frequency of Weiss was calculated by working out the ratio of masculine to feminine endings of each of the colour terms, and multiplying the average of that by the amount of times the feminine Weisse was used. In both Mandarin and Cantonese, care was taken to ensure that appearances transcribed in both pinyin and characters were included.

2.4.1.2 Analysis

In this study, Bayesian models were constructed in a similar fashion to the previous studies, with the same priors. In contrast to Study 2 of this chapter, the categorical variable of colour was replaced with the numeric variable of frequency, which was multiplied by one thousand to appear on the same scale as the other variables. An additional variable of syllabic complexity was also added, which was calculated as the number of syllables of each colour term. Because frequency is nested within language, each language will be affected differently by the coefficient of frequency, which allows for greater flexibility in fitting the model, but also means the model fit is less affected by the discrepancy in the overall frequency numbers in each language. The model used the log of the frequency, due to evidence that frequency of input should be log-transformed (Anderson & Schooler, 1991; Yurovsky et al., 2015).

Three models were run, each with the same specifications as in Study 2, except for increased iterations to 24000 and warmup of 8000 iterations, in order to allow for the more complex structure of the model. Running three separate models allowed the assessment of the addition of each term into the first, base model.

2.4.2 Results

The three models in this study were analysed separately, and then compared with Leave-One-Out Information Criteria (LOOIC). The first model analysed colour word production using only age and gender as population-level effects, with a group-level effect of age varying by language (LOOIC 12808.44, SE 264.22). In this first, base model, there was strong evidence for an effect of Age (95% CI 0.29 – 0.36) and for an advantage to female participants (95% CI -0.41 – -0.36).

The second model added frequency of the appearance of each colour word in each language in the CHILDES database as a predictor. Using frequency, age and gender to model the MacArthur-Bates CDI colour word data resulted in a dramatic improvement over the basic model containing only predictors of age and gender (LOOIC difference = 1871.41, LOOIC SE of difference = 157.37). There was strong evidence that frequency (95% CI 0.30 – 0.76), age (95% CI 0.29 – 0.37), and the gender difference (95% CI -0.42 – -0.37) all predicted word learning, with the 95% credible interval not intersecting 0 for any of those predictors.

In the final model, colour words in each language were assessed on the

Table 2.9: Population-level effects of final predictive model using both frequency and syllabic complexity as predictors of colour word learning.

	Est.	Err.	95% CI		Sample
Intercept	-7.68	0.39	-8.45	-6.91	4747
Age	0.33	0.02	0.29	0.37	4251
Male	-0.39	0.01	-0.42	-0.37	22618
log(Frequency)	0.54	0.12	0.3	0.77	6026
Complexity	-0.06	0.03	-0.11	0	22695

number of syllables each possessed, and added as a population-level effect into the previous model. Syllabic complexity was again found to be a potential, but weak factor in predicting colour word learning (95% CI -0.11 – 0.00), while frequency, age, and gender continued to have strong predictive power (Table 2.9). This final model proved to be arguably a slightly better fit than the frequency-only model (LOOIC difference 1.06, SE difference 5.17). The final model was successful at capturing much of the variance of colour word learning in different languages, but was not as optimal as the original descriptive model of the data which used colour terms as categorical variables in Study 2 of this chapter (LOOIC difference 1682.46, SE difference 177.24). In addition, the presence of positive evidence for slope of frequency to differ in each language (95% CI 0.22 – 0.64) suggests that the effect that frequency has differs greatly in each language.

One of the main points of difference between the predicted model and the original descriptive model was that the learning of *blue* was constantly underestimated in the final predicted model. Languages such as English, where the data suggests a clear advantage to *blue*, are instead modelled to show *blue* coming in behind other languages. A graph for comparison to Study 2 can be found in Appendix A. In addition the model predicts a closer contiguity between the colour terms than is realized in the data. These factors suggest that as well as the obvious strong effect of frequency, there are other elements as well, such as the category size or perceptual salience of the colour (Yurovsky et al., 2015).

2.5 Discussion

In this chapter, I have utilised measures of colour word learning from parental reports to assess the time line and trajectory of colour word learning from around 15 to 30 months of age. Recent research has substantiated the existence of a biological component in colour category formation in infancy (Skelton et al., 2017), but raises the question as to what causes the shift to adult colour categories? This strong biological contribution to early colour categories also reignites discussion of a potential universal order in which colour words are learned (Pitchford & Mullen, 2002), analogous to that in which colour words are proposed to emerge historically in languages (Kay et al., 2011), as discussed in Chapter 1.

The results of this chapter demonstrate that the order in which colour words are produced varies greatly between languages. The results from Study 2 in this chapter provide strong evidence for differences in colour word learning around the world, albeit with many similarities between groups that share a similar language and culture. Study 1 demonstrates that the onset of comprehension of colour terms follows a very similar trajectory to that of production, even for colour terms learned later, such as *brown* and *grey*. While Study 2 only measures production, comprehension can be assumed to precede it by a similar margin as shown in Study 1. Finally, in Study 3 of this chapter, the results of Study 2 are successfully approximated by modelling using the frequency of input and the lexical complexity of the colour term. Based on the evidence of Study 3, it appears that the frequency with which infants hear a word is a strong predictor of the timing of colour word production, and that syllabic complexity accounts for some further variance. This suggests that the timing of colour word learning is very much a linguistic and culturally mediated process.

That colour word learning is not universal, despite the biological foundations of colour categories, suggests a change in process in the understanding of colour categories. The visual colour categories evidenced in pre-linguistic infants (Franklin, Clifford, et al., 2005; Skelton et al., 2017) must adapt with the slow comprehension of colour words and their meaning (Franklin, Drivonikou, Clifford, et al., 2008). This process is not universal; the scope of the category for each word varies by language (Roberson et al., 2008), as does the timing of learning the word. Infants, in the learning of a colour term, are taught by frequent exposure to the term, as seen in Study 3 of this chapter. Thus, their understanding shifts as they slowly grasp a full comprehension of the meaning of the term. This may come about earlier or later, depending on how often they are exposed to the term. Wagner et al. (2013) found that when infants first comprehend a colour word, they comprehend the category centre, but over-extend it to include other colours. As the infant comprehends more terms, the additional category centroids force the infant to update their understanding of the original colour category, shrinking the category boundary with the addition of more terms. Thus a partial comprehension of the colour word precedes production, but is slow to mature. The comprehension data presented here, showing that colour words are learned differently due to different cultural and linguistic settings, captures the earliest part of that process – the basic comprehension of the focal colour term.

This study points to colour word learning in general occurring much earlier than previously reported (Heider, 1971; Mervis et al., 1995; Pitchford &

Mullen, 2002; Shatz et al., 1996; Soja, 1994). While it is possible that this may be part of a general trend of children learning colour words earlier than they used to (Franklin, 2006), there are two other possible considerations. One major consideration is that by asking parents to record whether their children can produce these terms, a larger-scale picture of word production that may be more sensitive than laboratory studies has been obtained, partly through the size of the samples used. The other consideration is that we may be measuring an earlier process, as it is possible that parents are able to report an early comprehension or production that is not yet consistent, or that the child cannot yet confidently reproduce in front of a stranger in a laboratory. The model of the Oxford CDI data (Study 1) demonstrates that toddlers' efforts to understand the meaning of colour words takes place earlier again, possibly during the second year of life in many languages. This again suggests that colour word learning may not be as difficult as previously thought (Andrick & Tager-Flusberg, 1986; Kowalski & Zimiles, 2006; Soja, 1994). While this may, in part, be due to environmental factors that promote the usage of colour words with young children, such as a focus on colour terms by the parents, or an increase in coloured plastic toys around the home, it is more likely reflective of a difference in measurement sensitivity. Parents are very sensitive to infants understanding and production of specific words (Hidaka & Smith, 2010), and have the opportunity to see them comprehend and produce words in a variety of contexts. By contrast, previous behavioural experiments lack the opportunity to assess colour term comprehension in a variety of environmental and grammatical contexts, leading to early understandings being ignored (Ramscar, Thorpe, & Denny, 2007). It is thus likely that colour words are learned in the same manner as other classes of words, where infants have a basic comprehension, which is then refined after they start producing the word. It may simply be that frequency affects learning of colour words in the same manner as other classes of words.

In Study 1 of this chapter, the gap between comprehension and production is shown to be quite small, and can be assumed to be similar for Study 2. This does, however, raise the question of measuring comprehension using parental report. Parental reports of production data are likely to be largely accurate. Asking parents to assess comprehension of a word, however, has been criticized (Houston-Price et al., 2007; Tomasello & Mervis, 1994) as an accurate method, although other findings have shown that CDI comprehension measures are useful estimates, at least in the case of concrete nouns (E. Bates et al., 1988; Mills et al., 1993, 1997). The short gap between comprehension and production was shown again for all 11 colours in the subset of participants who completed the additional colour word survey. Further suggestions that, if anything, comprehension estimates in CDI studies are an underestimation (Styles & Plunkett, 2009b), suggest that the short gap between comprehension and production indicates the possibility of earlier comprehension than reported here.

In Study 3 of this chapter, it is demonstrated that much of the variance in the difference in timing of the acquisition of colour words occurs due to input frequency and syllabic complexity. The variance in input frequency Figure 2.5 is incredibly large, a factor that may be cultural, or a peculiarity of the data. The prevalence of red in the Sinitic languages, for example, is likely to be cultural, given the associations between that colour and luck and fortune. While this is likely true in many cases, the recordings of child-parent interactions that make up CHILDES are limited, and could be biased by, in some instances, an infant playing with a toy of a particular colour, or a colour-based game. The variation that can be seen in the frequency of input data from CHILDES attests to this. Italian showed a much higher frequency for most colours than the other languages, suggesting a possible activity bias in one or more of the datasets. Frequency of input is a powerful predictor for colour word learning, but does not account for all of the variance. Syllabic complexity appears to account for some further variance, suggesting that the length of a word may make it harder for infants to learn. One possibility is that other predictors, such as the visual salience of the colour and the size of the colour word category could account for some of the remaining variance (Yurovsky et al., 2015). In this sense, it is likely that the absence or present of additional colour terms (such as the additional term for *blue* in Russian) may play an important role in the timing of colour word learning, as they change the category size for each of the surrounding colours.

It should be stressed that what is being measured here is not necessarily an adult-like understanding of colour words by young toddlers, but rather the beginning of a slow process of establishing the contents of a colour word category. While they may still be prone to errors in applying those terms correctly (Pitchford & Mullen, 2003), they may have understood that the colour term refers at least to the focal area of that colour word category. Infants clearly begin to understand colour words much earlier than first thought, and they do so with great variety, depending on both the individual and the language which they speak.

This chapter provides strong evidence for cultural and linguistic variation in the formation of colour categories, through analysis of parental surveys of British English children, and matched by parental surveys from around the world. Colour word learning follows no universal pattern or timeline, but instead varies with differing languages and cultures. In this sense, and in the sense that a partial comprehension seems to precede production (Wagner et al., 2013), colour words seem to be learned in much the same way as any other class of word. The results also suggest that colour word learning may occur much earlier than previously seen, thus suggesting that perhaps colour word learning is not as uniquely difficult as had previously been assumed (Franklin, 2006; Soja, 1994). Colour word learning in this chapter was measured with parental report, and further behavioural investigations into colour word production and comprehension, such as those in the following chapter, will be crucial to fully understanding this topic.

Infant colour categories appear to possess a biological, universal foundation (Skelton et al., 2017). However the infant colour categories change into adult-like understandings of the colour terms, a process that, from the evidence presented here, is determined by the nature of the language which they learn. Despite the universal, biological origin of colour categories, there is still an undeniable place for the cultural and linguistic.

Chapter 3

Early Colour Word Learning in British Infants

3.1 Introduction

As discussed in the previous two chapters, the nature and timing of colour word learning has been a topic of much debate. Early reports indicated that colour words were produced correctly as late as 7 years of age (Heider, 1971). Later evidence suggested more precocious knowledge of colour terms before 4 years (Bornstein, 1985; Franklin, 2006; Pitchford & Mullen, 2002), and some level of production during the second year of life (Mervis et al., 1995; Shatz et al., 1996). Even when the colour labels themselves are acquired, their usage was thought to be riddled with errors. Children regularly have more difficulty with some colours than others, particularly non-focal colours (Andrick & Tager-Flusberg, 1986; O'Hanlon & Roberson, 2006; Pitchford & Mullen, 2001, 2005), and having learned them they apply the colour terms inconsistently (e.g. Kowalski & Zimiles, 2006; Pitchford & Mullen, 2003; Rice, 1980; Roberson, Davidoff, Davies, & Shapiro, 2004; Sandhofer & Smith, 1999; Soja, 1994).

Studies to date have therefore found it difficult to establish a clear timeline for when colour words are learned, which has lead to various theories about why they might be so difficult for children to learn. Explanations have varied from children's inability to abstract the category boundaries in order to map the categorical colour words (Andrick & Tager-Flusberg, 1986), to infants lacking a conceptual representation of colour (Kowalski & Zimiles, 2006), as well as linguistic and attentional constraints (O'Hanlon & Roberson, 2006), as discussed in detail in Chapter 1. These arguments all gave credence to the idea that not only were colour words difficult to learn, but learning them may involve different mechanisms to other classes of words.

More recent work has found evidence that colour words are learned in much the same way as nouns are learned, with a partial production preceding comprehension, and that comprehension is slowly refined as the infants learn more about the category (Wagner et al., 2013, 2014). These claims find further support in studies of cross-linguistic parental report data from Chapter 2, which found that in 11 different languages, parents reported colour word comprehension prior to production. One of the aims of the present study is to examine this theory further with experimental, behavioural data.

In Chapter 2, I have also reported evidence for colour word comprehension beginning much earlier than previously found, with around 50% of infants comprehending the four basic colours by 21 months of age. Yet despite evidence verifying parental report as a reliable estimation of children's word learning (e.g. Dale, 1991; Mills et al., 1993, 1997), doubts remain over the ability of parental report to accurately capture comprehension of a word, especially for more abstract categories of words, such as colour (Houston-Price et al., 2007; Tomasello & Mervis, 1994). The findings reported in Chapter 2 of this thesis were also in stark contrast to previously collected behavioural data (e.g. Pitchford & Mullen, 2002; Sandhofer & Smith, 1999), making it unclear whether their findings are a result of the methods used, or consistent with children's real comprehension of colour words, a question conflated by the fact that many previous behavioural studies used production as a measure, or required the child to interact with the experimenter.

Measuring colour word comprehension with a behavioural task is complicated as it raises the question of colour preferences. In a controlled trial, if an infant reaches for the red shape having been prompted to "*find the red one*," that may be as much due to red being a colour of interest as to their possible comprehension of the word "*red*" (Pitchford, Davis, & Scerif, 2009). Colour preferences in infants have been well-documented, finding that even in pre-linguistic infants, infants look longer at red hues, as opposed to green hues (Franklin, Bevis, Ling, & Hurlbert, 2010), and that the preference for red is consistent across context (Franklin, Gibbons, Chittenden, Alvarez, & Taylor, 2012). Despite these findings, behavioural measures have yet to attempt to measure comprehension separate from colour preferences, one of the key aims of this study.

In controlled experimental conditions, infants often have great difficulty mapping adjectives to object properties (Mintz & Gleitman, 2002; Waxman & Markow, 1995). Colour is no exception to this. Children finding mapping a novel adjective on to a colour to be a very difficult task (e.g. Booth & Waxman, 2009). Additional linguistic context may make the mapping process easier, such that a child may find it easier to attend to an object property when a specific noun is provided. For example, the child may affix their gaze on the red car more readily when hearing, "*look at the red car*," than when hearing "*look at the red one*" (Mintz & Gleitman, 2002).

The aim of this chapter was to address each of the above questions by measuring colour word comprehension using Intermodal Preferential Looking (IPL) procedures (Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987). Each IPL trial can be examined in two steps: the pre-naming phase, which provides information about the baseline preferences infants have for one colour over another; and the post-naming phase, which measures their responses to auditory prompts. Based on the research presented in the previous chapter, it was hypothesised that the four chromatic colour words would be learned by the 24 month-old mark, earlier than previous behavioural experiments have shown. In addition, this chapter aimed to examine how the provision of different types of sentential information affects infants' comprehension of a colour word, by using three structures that differentially highlight the adjectival status of the colour word. It was predicted that the infants would look more reliably to the target when the colour word was embedded in a prototypical adjectival position, in line with the findings of Mintz and Gleitman (2002).

3.2 Methods

3.2.1 Participants

Participants were recruited in 5 age groups: 30 participants at 1;0 were recruited for a baseline no-comprehension control, as they were unlikely to understand colour words at that age. 28 participants ranging from 3;0 – 4;0 upwards were also recruited as the comprehension control group, as they were likely to all comprehend the colour words by that age. In between these groups, 29 participants at 1;4, 31 participants at 1;7, and 28 participants at 2;0 were recruited as the main experimental groups, for a total of $N = 146$ participants. An additional 23 participants were excluded for fussiness or parental interference with the task, while an additional 5 participants were excluded for failing to complete at least one trial with each colour as both distractor and target. Participant information can be found in Table 3.1.

Table 3.1: Descriptive statistics for participants included in study.

Age Group	*N*	Mean Age (months)	SD (months)
12	30	11.84	0.70
16	29	15.96	0.70
19	31	19.69	0.73
24	28	24.30	0.36
48	28	53.46	18.78

All participants were contacted after recruitment at the local maternity ward or online. Participants with one parent or grandparent with colour vision problems were not tested for this study. All participants were monolingual, learning English as their first language.

3.2.2 Materials

Auditory stimuli were recorded by a native female speaker of Southern British English (SBE), speaking slowly and clearly in an infant-directed manner. The auditory stimuli consisted of three different sentence types: sparse (*"Look, red!"*), general (*"Look at the red one!"*), and informative (*"Look at the red car!"*). Note that in all cases, attention to the colour label alone is sufficient to succeed in identifying the target. Both the colour and the named object varied depending on what was shown on the screen.

Visual stimuli were all objects that should be familiar to infants in daily life, such as vehicles, items of clothing, or furniture. Each object was chosen to be an object without a typical colour, and that could be easily recoloured. In each trial, the same object was presented on both the left and the right of the screen, varying only in the colour. Objects could be any one of six colours:

red, blue, green, yellow, black, or *white*, and each colour was selected to be a typical example of the colour category, and confirmed to be so by independent observers as well as during pilot testing. Where necessary, objects were recoloured in the GNU Image Manipulation Program (GIMP, www.gimp.org).

Participants saw each colour three times as a target, one corresponding to each of the sentence types, for a total of 18 trials. Participants were randomly assigned to different lists, in order to counter-balance which target colours appeared against which distractor colours, as well as counterbalancing which colours appeared with which objects. Trials were left-right randomised. All trials were run using a custom script in MATLAB, and recorded using a Tobii TX300 eye-tracker, recording at 120Hz.

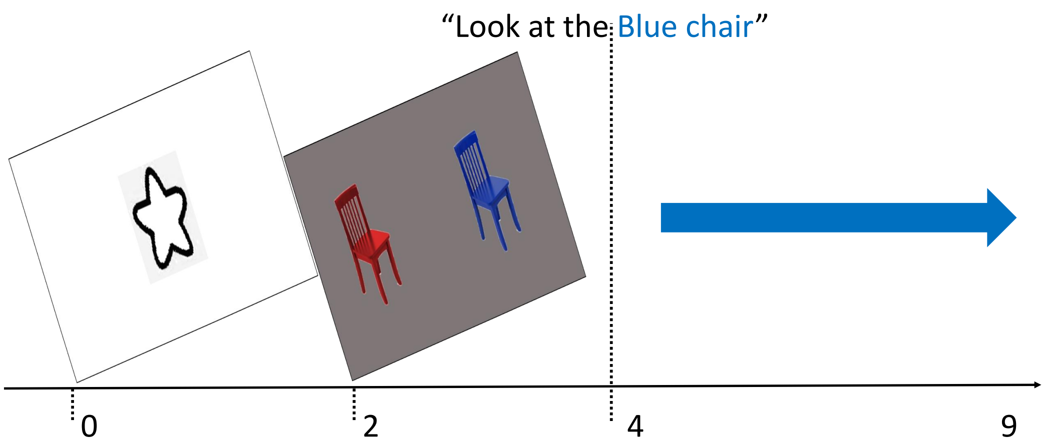

Figure 3.1: Time course of a typical trial.

3.2.3 Procedure

On arrival at the lab, participants and caregivers were shown to a playroom to allow the participant to familiarize themselves with the laboratory settings.

During this time, caregivers were asked to fill out consent forms, as well as a parental report, asking whether their child comprehends, or comprehends and says each of the 11 basic colour words (for details see Study 1 of Chapter 2). After this warming-up period, participants were seated on the lap of the caregiver, roughly 75cm from the eye-tracker and presentation screen.

The experiment commenced with a nine-point calibration sequence, which was repeated until at least 7 of the nine points were calibrated successfully, after which the trials commenced. Each trial lasted for nine seconds, the first two seconds of which was an attractive attention getter designed to orient the participant's attention to the centre of the screen. Immediately after, the two images appeared on the screen, on a neutral grey background. The auditory stimuli were presented so that the onset of the target colour word occurred exactly 2 seconds after the images appeared. Trials continued for another 5 seconds after the onset of the target colour word. For each participant, target/distractor colour pairings were counterbalanced, e.g., if a participant saw a blue chair as a target against a red chair as a distractor (Figure 3.1), then they would also see the red chair as a target against the blue chair as a distractor.

3.2.4 Analysis

Data for infant fixations were extracted with a custom MATLAB script. A fixation was defined as a stable gaze in one location for 100ms, allowing for a small amount of dispersion to account for the unsteadiness of the infant gaze. The area around each image was expanded slightly to allow for the same unsteadiness, so that the borders of each image were expanded by 25%. Trials were removed if greater than 60% of the trial was lost due to the infant focussing attention away from the screen. For each analysis, the variable of colour was dummy coded, while age and time elapsed during the trial were treated as continuous numeric variables.

Analysis was completed in R, using the MASS package (Ripley et al., 2017) and eyetrackingR (Dink & Ferguson, 2015). In the pre-naming phase, the data were aggregated so that for each participant, each colour could be compared to every other colour. In the post-naming phase, rather than modelling the proportion of looks to the target, which would allow the colour preferences of each infant to bias the result, trials were aggregated in such a way that for each colour tested, the looking to the target was compared instead to the looking to the distractor when that same colour was the distractor. In other words, for a given participant i and colour j:

$$Proportion_{\mathrm{ij}} = \frac{Target_{\mathrm{ij}}}{Target_{\mathrm{ij}} + Distractor_{\mathrm{ij}}}$$

3.3 Results

3.3.1 Pre-naming Phase

In the pre-naming phase, the data can be used to analyse the overall baseline colour preferences of the participants. The proportions of looking to each colour against each other colour can be seen in Figure 3.2. The figure is a matrix of preference for each colour against each other colour, where red suggests a preference for looking to that colour, and yellow suggests a preference for looking away from that colour. Visual inspection of the figure indicates a strong preference to look to red over most colours, and a strong preference to look to any other colour, in contrast with white.

The pre-naming phase data were fitted with a multilevel linear regression using the package lme4 (D. Bates et al., 2017). The model included Colour 1 and Colour 2 as fixed effects (see Figure 3.2), and varied the intercept for each participant, to allow for individual variance. The model coefficients (Table 3.2) reinforce the pattern depicted graphically in Figure 3.2, demonstrating strong evidence for looking toward red over black (the reference colour), and for looking away from white compared to black. While these are only compared to black in the model, they reinforce the pattern that can seen in Figure 3.2.

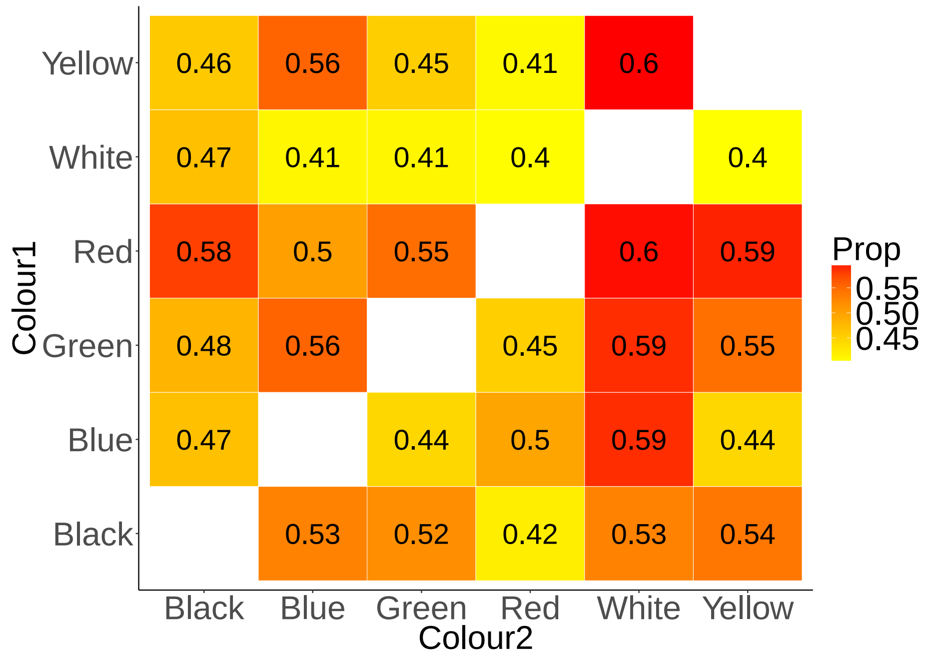

Figure 3.2: The proportions of looking to any target colour (Colour 1) over any other colour (Colour 2). Red indicates looking to that colour above 50%, yellow indicates looking below 50%. Proportions are listed in each box.

Table 3.2: Model coefficients for pre-naming phase. Colours are compared to black. Results for the second colour are identical but reversed, due to the nature of the data.

	Estimate	Std. Err	df	t value	Pr($> t$)
(Intercept)	0.50	0.02	2484	26.09	<0.001
Colour1Blue	-0.02	0.02	2484	-1.11	0.269
Colour1Green	0.01	0.02	2484	0.47	0.637
Colour1Red	0.04	0.02	2484	2.26	0.024
Colour1White	-0.08	0.02	2484	-4.13	<0.001
Colour1Yellow	-0.02	0.02	2484	-0.86	0.389

The results reinforce the need to correct for colour preference in the analysis of the post-naming data, as infants show a strong preference for red, and a strong preference for any colour over white. The results also suggest a preference for green over blue and yellow, for black over blue, green, and yellow, and for yellow over blue. The results of this analysis are consistent with previous reports of infants preferring colours on the red–green scale (e.g. Franklin, Bevis, et al., 2010; Franklin et al., 2012).

3.3.2 Post-naming Phase

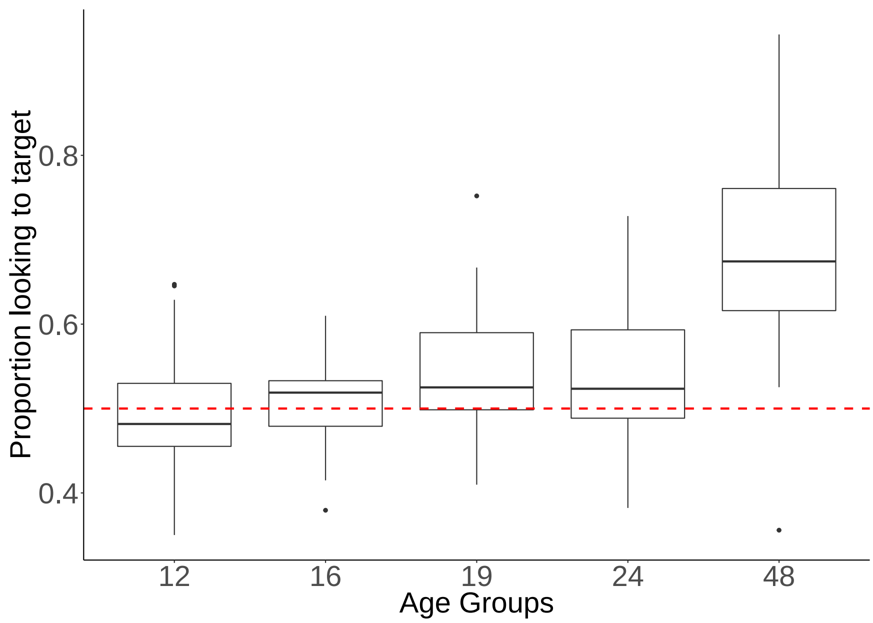

Figure 3.3: Looking proportions to the target for each age group.

In the post-naming phase, looking was first aggregated across the trial time to calculate whether colour word responses improved as the trial time increased. Participants were assessed on their looking to the target after the colour word was named; consistent looking to the target when prompted would suggest comprehension of the target colour word. For the post-naming phase, data from 0-3000ms after the target word onset were used, as looking after that time was deemed unlikely to be due to an effect of the target word. The proportion of looking to the target at each age group was compared to chance looking ($\mu = 0.5$) with One Sample t-tests. At 12 months, there was no evidence that the infants comprehended colour words ($t(29) = -0.430$, $p = 0.670$), nor was there any evidence of colour word comprehension at 16 months ($t(28) = 0.572$, $p = 0.572$). However, for the subsequent three age groups, there was strong evidence that participants comprehended colour words, looking consistently to the target after it was named: at 19 months ($t(30) = 3.229$, $p = 0.003$), 24 months ($t(27) = 2.361$, $p = 0.026$), and at 48 months ($t(27) = 7.710$, $p < 0.001$). A linear regression of proportional looking to the target by age confirms that colour word comprehension improves in each age group, with strong evidence for the intercept ($t = 28.030$, $p < 0.001$) and for the effect of age ($t = 9.619$, $p < 0.001$). Overall looking proportions in each age group can be viewed in Figure 3.3.

In order to further analyse the differences in colours and looking over the trial time, proportions of looks to the target were modelled with a binomial logistic mixed-effects regression, using the function glmmPQL in R. The regression was fitted with quartic orthogonal polynomials of the time elapsed after target word onset (Mirman, 2014). The numeric variable of participant age and the categorical variable of colour were included in the model. In addition, both the intercept and the slope of colour were allowed to vary for each participant, in order to allow for the fact that participant's understanding of different colours may vary greatly between individuals. The full list of effects can be viewed in Table 3.3.

The regression analysis demonstrated strong evidence for an effect of both the linear time term and the cubic time term, as well as for an interaction between both of those time terms and the age of the participant. This evidence

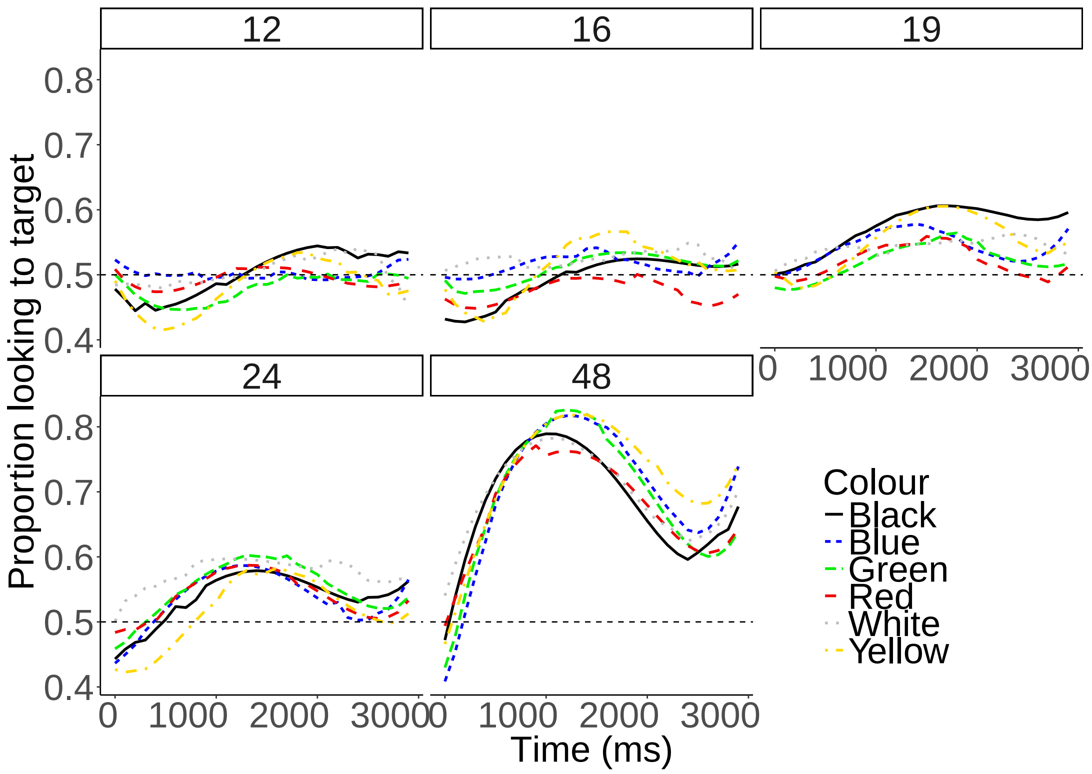

Figure 3.4: Looking proportion to the target as trial progresses. 0ms is when the target is named.

suggests that the looking patterns varied greatly across age groups. There was no strong evidence for an effect of colour, although there was evidence for interaction effects between some of the time terms and red and blue, suggesting that overall looking proportions for each of the colours may not have varied much, although there were some differences in the looking patterns for each of them.

The model fit (Figure 3.4) shows looking proportions for all colours at around chance at both 12 and 16 months, then at 19 months there is consistent looking to the target above chance, which becomes slightly more consistent at 24 months. The 48 month-olds consistently look to the target. The model fit also demonstrates very little difference between the colours, with looking to the target largely at chance for all six colours at 16 months, and above chance for the age groups thereafter.

Table 3.3: Full list of effects for logistic mixed-effects regression on the post-naming data. "ot" refers to the orthogonal time terms. Colours are compared to black.

	Value	Std.Error	DF	*t*-value	*p*-value
(Intercept)	1.226	1.946	25131	0.630	0.529
ot1	-915.378	194.255	25131	-4.712	<0.001
ot2	512.535	664.883	25131	0.771	0.441
ot3	-566.541	103.023	25131	-5.499	<0.001
ot4	153.179	241.368	25131	0.635	0.526
Age	0.220	0.909	144	0.242	0.809
ColourBlue	-0.630	2.702	25131	-0.233	0.816
ColourGreen	-0.339	2.687	25131	-0.126	0.900
ColourRed	0.803	2.661	25131	0.302	0.763
ColourWhite	-4.525	2.766	25131	-1.636	0.102
ColourYellow	1.336	2.764	25131	0.483	0.629
ot1:Age	646.923	90.756	25131	7.128	<0.001
ot2:Age	3.144	310.536	25131	0.010	0.992
ot3:Age	370.680	48.103	25131	7.706	<0.001
ot4:Age	50.328	112.735	25131	0.446	0.655
ot1:ColourBlue	419.477	270.791	25131	1.549	0.121
ot1:ColourGreen	10.708	268.538	25131	0.040	0.968
ot1:ColourRed	446.353	265.935	25131	1.678	0.093
ot1:ColourWhite	120.443	276.990	25131	0.435	0.664
ot1:ColourYellow	-202.788	276.776	25131	-0.733	0.464

ot2:ColourBlue	-236.221	924.158	25131	-0.256	0.798
ot2:ColourGreen	-129.556	919.177	25131	-0.141	0.888
ot2:ColourRed	305.124	910.192	25131	0.335	0.737
ot2:ColourWhite	-1554.771	945.458	25131	-1.644	0.100
ot2:ColourYellow	537.384	945.263	25131	0.569	0.570
ot3:ColourBlue	312.015	143.905	25131	2.168	0.030
ot3:ColourGreen	60.671	142.695	25131	0.425	0.671
ot3:ColourRed	313.148	141.338	25131	2.216	0.027
ot3:ColourWhite	114.232	147.156	25131	0.776	0.438
ot3:ColourYellow	-55.933	147.234	25131	-0.380	0.704
ot4:ColourBlue	-137.079	335.924	25131	-0.408	0.683
ot4:ColourGreen	-104.083	334.257	25131	-0.311	0.756
ot4:ColourRed	108.681	330.963	25131	0.328	0.743
ot4:ColourWhite	-575.569	343.558	25131	-1.675	0.094
ot4:ColourYellow	202.749	343.961	25131	0.589	0.556
Age:ColourBlue	0.941	1.259	25131	0.747	0.455
Age:ColourGreen	0.268	1.255	25131	0.214	0.831
Age:ColourRed	-0.007	1.238	25131	-0.006	0.996
Age:ColourWhite	1.289	1.277	25131	1.010	0.313
Age:ColourYellow	0.349	1.280	25131	0.272	0.785
ot1:Age:ColourBlue	-116.466	126.186	25131	-0.923	0.356
ot1:Age:ColourGreen	-67.787	125.037	25131	-0.542	0.588
ot1:Age:ColourRed	-285.077	123.930	25131	-2.300	0.021
ot1:Age:ColourWhite	-132.281	127.462	25131	-1.038	0.299
ot1:Age:ColourYellow	-47.353	128.453	25131	-0.369	0.712
ot2:Age:ColourBlue	332.262	430.750	25131	0.771	0.441
ot2:Age:ColourGreen	109.984	429.409	25131	0.256	0.798

ot2:Age:ColourRed	-1.086	423.693	25131	-0.003	0.998
ot2:Age:ColourWhite	426.918	436.488	25131	0.978	0.328
ot2:Age:ColourYellow	89.917	437.829	25131	0.205	0.837
ot3:Age:ColourBlue	-99.832	67.069	25131	-1.489	0.137
ot3:Age:ColourGreen	-56.833	66.443	25131	-0.855	0.392
ot3:Age:ColourRed	-176.324	65.916	25131	-2.675	0.007
ot3:Age:ColourWhite	-84.713	67.725	25131	-1.251	0.211
ot3:Age:ColourYellow	-55.821	68.412	25131	-0.816	0.415
ot4:Age:ColourBlue	144.359	156.679	25131	0.921	0.357
ot4:Age:ColourGreen	72.962	156.288	25131	0.467	0.641
ot4:Age:ColourRed	-0.810	154.151	25131	-0.005	0.996
ot4:Age:ColourWhite	146.084	158.712	25131	0.920	0.357
ot4:Age:ColourYellow	39.653	159.518	25131	0.249	0.804

3.3.3 Comparison to parental reports

Participants' performance in the eye-tracking task was compared with parental reports of the participants' understanding of these colour terms. The parental report data were derived from the reports they were asked to fill out when arriving at the lab. Participants were marked as comprehending the colour words in the eye-tracking task if their looking proportion to the target across all trials for that colour exceeded 0.55. Then collapsing across age, a Chi-squared test was performed on whether or not they looked to the target versus whether they were judged to have comprehended the colour term according to their parents. A strong association was found between parental report data and eye-tracking data for the colour word comprehension ($\chi^2(1) = 44.207$, $p < 0.001$).

The proportions of looking to the correct-coloured target were also regressed separately against comprehension according to parental reports. There was strong evidence for an effect of both the intercept ($t = 62.263$, $p < 0.001$) and whether they were judged by parents to have comprehended the colour term ($t = 6.460$, $p < 0.001$). Participant performance in this colour discrimination eye-tracking task is predicted by whether they are judged as comprehending the term according to their caregivers.

3.3.4 Sentence type analysis

There were three types of utterance used to introduce the colour terms:

1. Sparse: *"look, red!"*
2. General: *"look at the red one!"*
3. Informative: *"look at the red chair!"*

The effect of sentence structure on participant performance was analysed using a binomial mixed-effects regression, with quartic polynomials as above. The age of the participants and the sentence type were included as fixed effects, and the intercept was allowed to vary for each participant. Lengths of each stimulus from target word onset until auditory stimulus offset are recorded in Table 3.4.

Table 3.4: Mean lengths and standard deviations for each stimulus type following target word onset.

Stimulus	Mean	SD
Sparse	0.716	0.068
General	0.901	0.034
Informative	1.200	0.151

The model coefficients (Table 3.5) suggest strong evidence for a difference between the stimulus types (especially the difference between *sparse* and *in-*

formative, $t = 2.747$, $p = 0.006$), and also that these stimulus types vary greatly in their interactions with the polynomial time terms, where there is strong evidence for the stimulus types varying on the linear time term, as well as some evidence for them varying on the quadratic time term. There was also strong evidence for an interaction between the differences in stimulus type and age ($p = 0.005$ and $p = 0.002$, respectively).

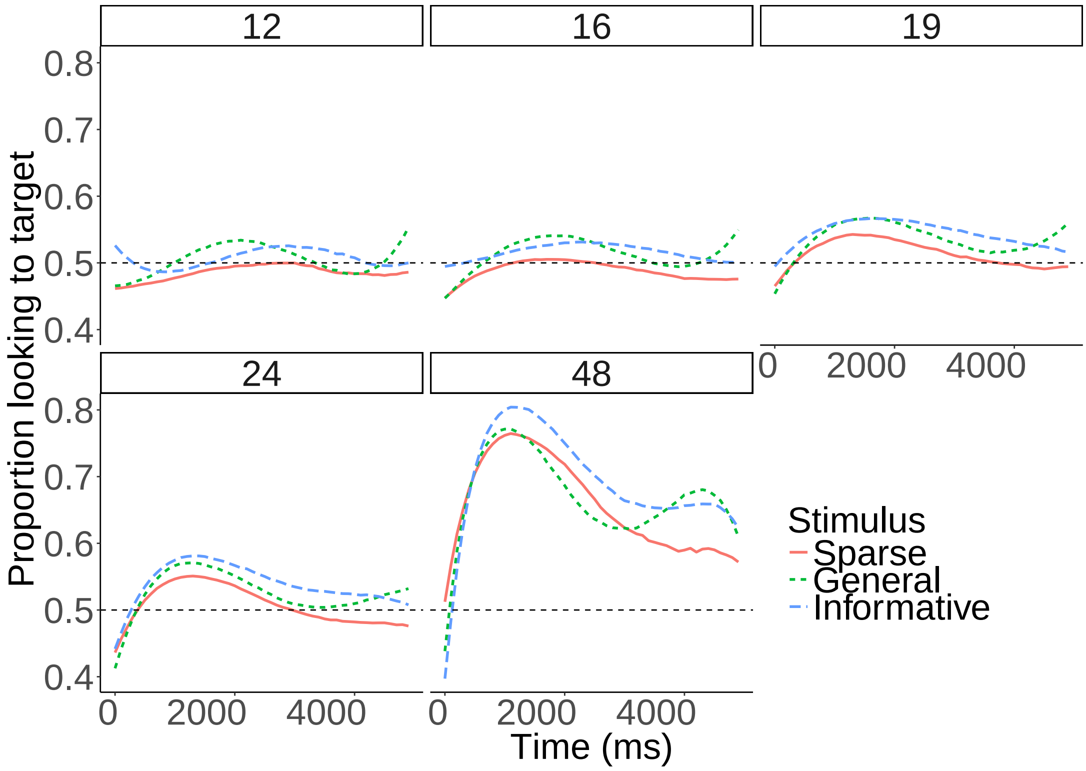

Figure 3.5: Looking proportion to the target for different sentence types.

Figure 3.5 demonstrates the model fit. At the peak of their target looking, around 1000ms after word onset, infants are less likely to look to the target when hearing "*look, red,*" than when hearing a sentence with more information, which provides more detail on what to fixate. When they are provided with the most amount of information ("*look at the red chair!*"), they also maintain attention to the target better than when they are provided with less information. This is demonstrated in the model coefficients as well, with strong ev-

Table 3.5: Model coefficients of model fit comparing the type of sentence. Stimulus Type refers to the sentence type, and are compared to sentence type 1.

	Value	Std.Error	DF	*t*-value	*p*-value
(Intercept)	0.254	0.273	107746	0.929	0.353
ot1	-7.528	3.180	107746	-2.368	0.018
ot2	7.527	2.797	107746	2.691	0.007
ot3	-4.705	1.755	107746	-2.681	0.007
ot4	1.465	0.752	107746	1.948	0.051
Age	-0.372	0.126	144	-2.960	0.004
Stimulus2	0.734	0.374	107746	1.961	0.050
Stimulus3	1.025	0.373	107746	2.747	0.006
ot1:Age	7.480	1.466	107746	5.102	<0.001
ot2:Age	-7.257	1.292	107746	-5.618	<0.001
ot3:Age	4.146	0.811	107746	5.112	<0.001
ot4:Age	-1.142	0.346	107746	-3.304	0.001
ot1:Stimulus2	-7.685	4.514	107746	-1.702	0.089
ot1:Stimulus3	-11.726	4.496	107746	-2.608	0.009
ot2:Stimulus2	6.665	3.970	107746	1.679	0.093
ot2:Stimulus3	9.669	3.956	107746	2.444	0.015
ot3:Stimulus2	-4.733	2.491	107746	-1.900	0.057
ot3:Stimulus3	-5.664	2.482	107746	-2.282	0.023
ot4:Stimulus2	2.720	1.066	107746	2.552	0.011
ot4:Stimulus3	1.723	1.061	107746	1.624	0.104
Age:Stimulus2	-0.481	0.172	107746	-2.790	0.005
Age:Stimulus3	-0.531	0.173	107746	-3.068	0.002
ot1:Age:Stimulus2	5.757	2.079	107746	2.769	0.006
ot1:Age:Stimulus3	6.884	2.089	107746	3.296	0.001
ot2:Age:Stimulus2	-4.700	1.831	107746	-2.567	0.010
ot2:Age:Stimulus3	-5.587	1.840	107746	-3.036	0.002
ot3:Age:Stimulus2	3.284	1.149	107746	2.857	0.004
ot3:Age:Stimulus3	3.225	1.156	107746	2.790	0.005
ot4:Age:Stimulus2	-1.597	0.488	107746	-3.270	0.001
ot4:Age:Stimulus3	-1.028	0.491	107746	-2.095	0.036

idence of a difference between the sentence types, and also in their interaction effects with age and the polynomial time terms. These interaction effects reflect the finding that the impact of sentence type is apparent from 19 months, but not before, as can be seen in Figure 3.5. The structure of the sentence in which an infant hears the colour word has a discernible affect on whether they fixate on the target.

3.4 Discussion

The results of this chapter have demonstrated several important aspects of children's colour word learning. First, the study has found strong evidence for early colour word learning in British infants, supporting the parental report analysis in the previous chapter of this thesis. The results of this chapter find evidence of comprehension of basic colour word knowledge as early as 19 months, much earlier than was found in previous behavioural analyses (e.g. Pitchford & Mullen, 2002; Sandhofer & Smith, 1999). In fact, the parental report study in the previous chapter found that colour words were only comprehended by around 25% of infants; in contrast the present chapter found reliable looking to the target at that age across all 6 colours tested, suggesting that British parents may be very conservative when estimating the comprehension of abstract word categories of their children, a possibility previously suggested by Styles and Plunkett (2009a).

The chapter also demonstrates the validity of parental report in evaluating children's knowledge of abstract categories of words such as colour words. Here, strong evidence was found for an association between how parents perceive their children to understand colour words, and their performance in a behavioural comprehension task. Participants who were judged to have understood the colour word by their parents looked to the target differently to those who did not.

The parental report can be seen to have predictive value even for word classes as abstract as colour words, at least on a large-scale level.

Third, the present chapter has highlighted the role that the sentence structure plays in revealing children's understanding of the meaning of colour words. When infants heard the colour word in a sentence that contained minimal contextual information (*'Look, red'*,) they looked to the target less than when they were presented with additional linguistic context (*'Look at the red one'*). Similarly, when presented with the name of the object at which they were looking (*'Look at the red chair'*), infants maintained their attention to the target for longer, even though both objects were the same. In addition to reinforcing previous findings (Mintz & Gleitman, 2002), this raises important considerations about the role that a colour word plays in the context of a sentence. While Mintz and Gleitman employed a very different paradigm to investigate the nature of adjectives, they found that when the target was named, rather than just using "one," subjects were more likely to extract the property information.

The present findings suggest that infants expect colour words to describe a noun, rather than be the object of a sentence itself. As colour words, and adjectives in general, primarily describe the properties of an object, they appear to be most informative and easiest to decipher when more context is provided about the object which they describe. Mintz and Gleitman (2002) infer from their results that labelling the noun gives parameters to understand the adjective, an explanation that may have some weight in this circumstance given the real-world stimuli employed in the present study. The results are consistent with an account that supposes that even 19-month old infants have developed an expectation that colour words should occur in adjectival position in a sentence and are processed more effectively when this expectation is fulfilled.

Finally, the study demonstrates strong preferences for red objects over other colours, and a preference for non-white objects. While this will likely be affected to an extent by the choice of background colour, it lends further support to previous studies that have examined colour preference in infants and toddlers, which suggest a preference for colours on the red-green axis (Franklin, Bevis, et al., 2010). In line with this, the present study found that blue is preferred only compared to white, and yellow only to blue and white. The results of this study also suggest a possible preference for black, which is preferred by infants to most colours that are not red.

There are some key differences found between the present study and the previous chapter. One such example is that very little difference was found between the six colours, whereas the parental report study found that black and white were learned after the four basic chromatic colour words. An explanation for this may be that the gap between each of the age groups in the current study is sufficiently large that the differences in when the colour words are learned are not apparent. Chapter 2 of this thesis found that the gap between the colours was at most a few months, suggesting it may not be apparent in the present study, where the gap between age groups is 3-4 months at least. In addition, in Chapter 2 we find that there is a possible slight advantage to learning blue over other colours in British English, again not reflected in the present study, due to the size of the age gap between each group.

It is important to note that this chapter only utilises six basic colour words, and only typical examples of each of these terms. While the results suggest that comprehension occurs much earlier than previously thought, the findings may reflect an extremely basic comprehension of typical examples, not an adult-like understanding of the colour word. Early comprehension may reflect only the beginning of a gradual, slow-mapped process (Wagner et al., 2013, 2014), after which the infant's comprehension will subsequently become more adult-like with each interaction, until their understanding extends to the boundaries of each colour word. An area of interest for further studies would be to demonstrate how this process occurs, through longitudinal analysis.

It is possible that while a great deal of attention has been given to the difficulties that infants have with learning colour words, it is simply the case that their errors are more obvious than those involving concrete nouns (Yurovsky et al., 2015). This chapter thus demonstrates that British infants begin the process of comprehending colour words as early as 19 months, and slowly start refining their comprehension with age. While there is little doubt that there are manifold reasons that infants may struggle with the mapping of colour words onto the continuous spectrum of colour (Franklin, 2006; Kowalski & Zimiles, 2006; O'Hanlon & Roberson, 2006), they are still able to learn colour words with great efficiency, in much the same way as they do for other classes of words.

Chapter 4

Infants Display an Early Preference for Dark Colours

4.1 Introduction

In the previous chapter, one topic of discussion was the role of colour preferences and how they shape colour word learning. Early research had concluded that while colour preferences were confusing, and varied greatly between individuals, it was possible that there was a universal preference for "blue" (Granger, 1952; McManus, Jones, & Cottrell, 1981). Since then, systematic analyses have allowed greater insight into colour perception, with debate focussing on the universal aspects of colour preference, versus those that vary by sex or culture (Hurlbert & Ling, 2007), and how that affects colour word mapping and learning.

Early studies came to the conclusions that much of colour preferences was universal (e.g. Al-Rasheed, 2015b; Eysenck, 1941; Hurlbert & Ling, 2007; Witzel, 2015). This lead to universal mechanisms being proposed for colour, where colour preferences were thought to evolve from affective responses to colour-related objects in the environment (Palmer & Schloss, 2010). This theory was thought to be a universal explanation for colour preferences around the world, and was measured against colour preferences from different cultures. Subsequent enquiry into these claims (Taylor & Franklin, 2012), found that there may be constraints on this theory when looking at individual data, and a further study comparing British colour preferences to those of Himba adults (a non-industrialised culture from Namibia), found that not only do colour preferences vary, but the mechanisms for those colour preferences do as well, with strong preferences toward colours on the red-green cone-opponent processes (L-M) over the blue-yellow processes (S-(L+M))[1] (Taylor, Clifford, & Franklin, 2012).

A more recent area of enquiry has shown that rather than being universal, colour preferences may vary with age (Taylor, Schloss, Palmer, & Franklin, 2013). While there was a great deal of existing research into the preferences of hue in infants, as examined in Chapters 2 and 3 (Franklin, Bevis, et al., 2010; Peeples & Teller, 1970; Pitchford & Mullen, 2005; Teller, 1998; Teller, Civan, & Bronson-Castain, 2004), little was known about why these preferences were found to vary. In a looking preference study, Taylor et al. demonstrated that adults have dramatically different colour preferences to those of pre-linguistic infants. The infants, for example, demonstrated a strong preference for dark yellow over light yellow, whereas the adult participants displayed precisely the opposite. This study demonstrated that lightness affects the preference for colours, even in infancy. However a conclusive picture of chromatic preferences in infants is yet to be found, which is one of the aims of the present study.

[1] Retinal cones are sensitive to short (S), medium (M), and long (L) wavelengths, which are known as blue, green and red, respectively, although these are not the exact colours (Jameson & D'Andrade, 1997). Contrasts between them create the cone opponent processes (De Valois & De Valois, 1975).

The previous chapter has been able to demonstrate when colour words are learned, but not when modifiers of colours, such as “dark” and “light” are learned. In Chapter 3 it was demonstrated that colour words are learned early by infants (See also Chapter 2), and used the pre-naming phase of the eye-tracking study to examine the colour preferences, which supported the colour preferences of hue found previously (Franklin, Bevis, et al., 2010). However, this was found only when infants are able to learn basic hues, and not when they can modify those hues – a far more complex task. To modify the hues, a greater knowledge of the boundary of each colour category is required (Wagner et al., 2013, 2014; Yurovsky et al., 2015), as the infant has to comprehend that each colour can appear in dark or light forms, and not simply as a prototypical example. The present study employed the same methodology to examine preferences of dark and light, and also to examine when the terms “dark” and “light” are learned by infants as a modifier for colours, and in doing so, this study will demonstrate when infants can achieve a more adult-like understanding of colour categories.

It was hypothesised in the pre-naming phase (see Chapter 3 for an explanation of the different phases) that infants would attend more to one of the colours if they possessed a chromatic preference. Based on early evidence for some preference for dark colours in Taylor et al. (2013), it was expected that they would display a preference for dark over light colours, but that this preference may change with age, due to the differences between adults and infants found in the literature. In the post-naming phase it was expected that they may not learn the meanings of “dark” and “light” until they were much older, as Chapter 3 found that colour words were known, but the understanding was still developing during the second year of life.

4.2 Methods

4.2.1 Participants

The same 146 participants, across 5 age groups were used as in Chapter 3, for this second eye-tracking task. Participants details can be found in Table 4.1. Participants completed the present experiment along with the previous experiment, with the order randomized. An additional 28 participants were excluded from the task, for either not completing a trial, or for fussiness or parental interference in the task. Participants known to have a family history of colour vision problems that could affect their own vision were not tested for this experiment.

Table 4.1: Descriptive statistics for participants included in the present study.

Age Group	*N*	Mean Age (months)	SD (months)
12	30	11.84	0.70
16	29	15.96	0.70
19	31	19.69	0.73
24	28	24.30	0.36
48	28	53.46	18.78

4.2.2 Materials

For this eye-tracking task, participants were presented with an auditory stimulus and a visual stimulus. Auditory stimuli were recorded by a female native speaker of Southern British English, speaking in an infant-directed manner. The sentences consisted of an object, and a shade (either "dark" or "light"). As in Chapter 3 there were three sentence types: sparse ("*look, light!*"), general ("*look at the light one!*"), or informative ("*look at the light chair!*"). The object was only named in the informative condition, while the target shade was named in all three conditions.

The visual stimuli for the present experiment again consisted of real-world objects, rather than patches of colour, in order to demonstrate infants' ability to comprehend these terms in a real-world context. The objects were familiar objects that the infants should encounter in daily life. Each object chosen was one that lacked a typical colour, and thus could be presented in any hue, such as T-shirts, for example. The hue of each object was either red, blue or green. Yellow was avoided as a light yellow object often looked far too highlighted, and a dark yellow object sometimes appeared orange or brown. The same object, coloured the same basic hue was presented on the left and the right of the screen, against a neutral grey background. The only point of difference between the two was that one was a dark shade and one was a light shade. Objects were recoloured using the GNU Image Manipulation Program (GIMP, www.gimp.org).

Each participant saw 6 trials, 3 dark and 3 light (one of each hue). Trials were counterbalanced within-participant, so that each item appeared as both a target and a distractor. Participants were assigned to one of 6 different lists, which differed only in which objects were used. Trial order was randomised. Participant gaze data was measured using a Tobii TX300 eye-tracker, and trials were run using custom code in MATLAB.

4.2.3 Procedure

Following a brief play session to acclimatise themselves with the surroundings of the lab, where informed consent forms were also signed, participants were seated on a caregivers lap. They were positioned such that they were about 75cm from the eye-tracker screen. A nine-point calibration sequence featuring either a ball or star was run, and the experiment only commenced when at least 7 of the 9 points were successfully calibrated. In each trial, par-

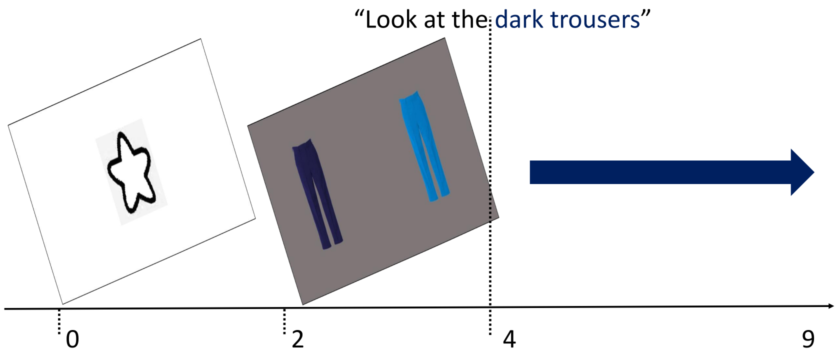

Figure 4.1: Time course of a typical trial.

ticipants saw an attractive attention-getter for 2 seconds, designed to focus their attention to the centre of the screen. Following that, the pictures immediately appeared. A further 2 seconds after picture onset, the target word was heard (dark or light), in the context of one of the auditory stimuli. The images remained on the screen for a further 5 seconds following target word onset.

4.2.4 Analysis

Infant fixations were defined with a custom MATLAB script, where a fixation was determined to be gaze in one location for at least 100ms, allowing for a certain amount of jitter, which is normal in infant gaze. The area around each picture was expanded by 25% to allow for dispersions in gaze. Trials were removed if the infant was not looking to either image for greater than 60% of the trial duration.

Data analysis was completed in R, using the MASS package (Ripley et al., 2017) and eyetrackingR (Dink & Ferguson, 2015). Analysis techniques were carried out in a similar fashion to Chapter 3. Pre-naming data were aggregated to examine looking for each shade for every participant. The post-naming phase was analysed to remove the bias of shade preference, so that rather than comparing dark to light for each participant and each time bin, looking when dark is the target is instead compared to looking when dark is the distractor. This technique allows for the examination of the naming effect without the colour bias.

4.3 Results

4.3.1 Post-naming phase

The following section examines the post-naming phase of the trials, which examines when the infants comprehend the words "dark" and "light." The data were examined in the same way as in Chapter 3, such that for a given participant i and colour j:

$$Proportion_{ij} = \frac{Target_{ij}}{Target_{ij} + Distractor_{ij}}$$

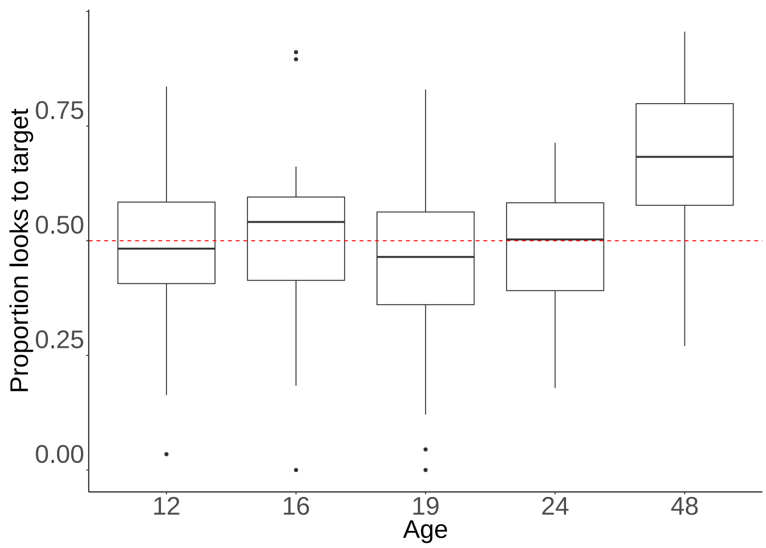

Figure 4.2: Looking proportions to dark and light in the post-naming phase, by age group. The dotted line indicates chance looking.

Overall looking proportions to the target can be seen in Figure 4.2. In each of the age groups, there is little evidence that the looking is mediated by the label, with target looking appearing to be close to chance looking for each of the age groups. These were compared to chance looking with a series of One Sample t-tests. There was no evidence that looking in either the 12 month ($t(59) = -0.775$, $p = 0.442$) or 16 month ($t(57) = 0.225$, $p = 0.823$) age groups differed from chance looking. There was evidence that at 19 months, looking to the target was less than chance ($t(61) = -2.362$, $p = 0.021$), but at chance again at 24 months ($t(55) = -1.083$, $p = 0.284$). At 48 months, however, there was strong evidence for looking to the target differing from chance ($t(55) = 9.159$, $p < 0.001$).

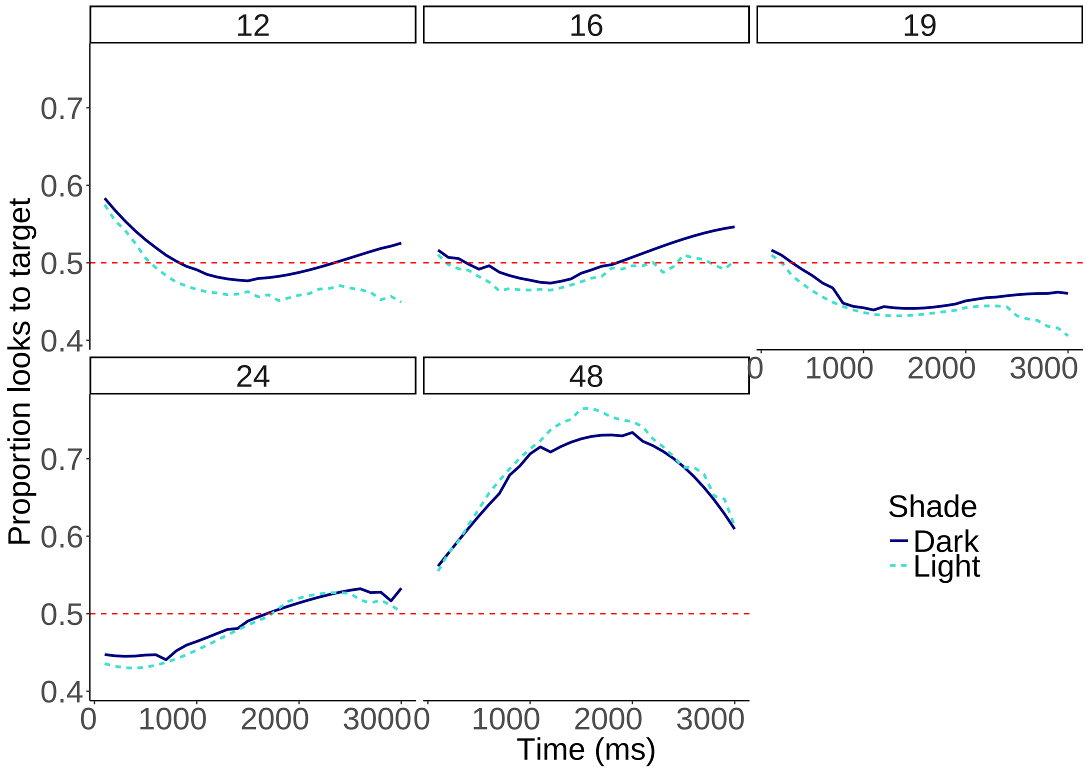

Figure 4.3: Looking proportions to dark and light in the post-naming phase, by age group. The dotted line indicates chance looking.

Looking to the target was then analysed for each 100ms time bin, to see how looking varied with time passed after target word onset. The data were fitted with a binomial hierarchical mixed-effect model, fitted with cubic polynomials of time (Mirman, 2014). This type of model allows for treatment of time as a continuous variable, rather than collapsing across it, which permits examining the overall looking patterns of participants. Fixed effects were the numeric variable of participant age (scaled by a factor of 12), and the treatment-coded variable of shade (dark vs light). The linear orthogonal time term was allowed to vary for each participant, to avoid over-dispersion. The model was only fitted to the first three seconds after target word onset, as any looking beyond that was deemed *a priori* to not be an effect of naming.

Table 4.2: Model coefficients of binomial model on post-naming phase data. "ot" refers in turn to each of the polynomial time terms.

	Value	Std.Error	DF	t-value	p-value
(Intercept)	0.482	0.159	8249	3.027	0.002
ot1	-208.792	80.247	8249	-2.602	0.009
ot2	166.550	17.643	8249	9.440	<0.001
ot3	-88.361	38.459	8249	-2.298	0.022
Age	-0.182	0.072	144	-2.523	0.013
Shade	0.162	0.239	8249	0.677	0.498
ot1:Age	46.115	36.476	8249	1.264	0.206
ot2:Age	-83.043	8.050	8249	-10.316	<0.001
ot3:Age	8.175	17.573	8249	0.465	0.642
ot1:Shade	181.059	148.303	8249	1.221	0.222
ot2:Shade	-1.439	36.078	8249	-0.040	0.968
ot3:Shade	92.551	78.278	8249	1.182	0.237
Age:Shade	0.036	0.108	8249	0.335	0.738
ot1:Age:Shade	-50.730	67.763	8249	-0.749	0.454
ot2:Age:Shade	14.135	16.445	8249	0.860	0.390
ot3:Age:Shade	-26.398	35.828	8249	-0.737	0.461

The model fit (Table 4.2) demonstrated strong evidence for an effect of each of the polynomial time terms, as well as an effect of participant age. There was no demonstrable evidence for an effect of dark versus light. Figure 4.3 shows fitted looking from the model. Over the first three seconds of the trial after the target word onset, there is no demonstrable looking to the target, with the exception of the 48 month-old group, who consistently look to the target, regardless of the lightness. These results indicate that the terms “dark” and “light” are not learned by participants until well after their third or fourth year of life.

4.3.2 Pre-naming phase

The following section analyses the preference for dark colours against light colours, by using the pre-naming phase of the trials. The data were aggregated to include all target and distractor trials for each, as in the pre-naming phase, participants would be unable to distinguish target from distractor. To begin, first the overall effects of shade are examined.

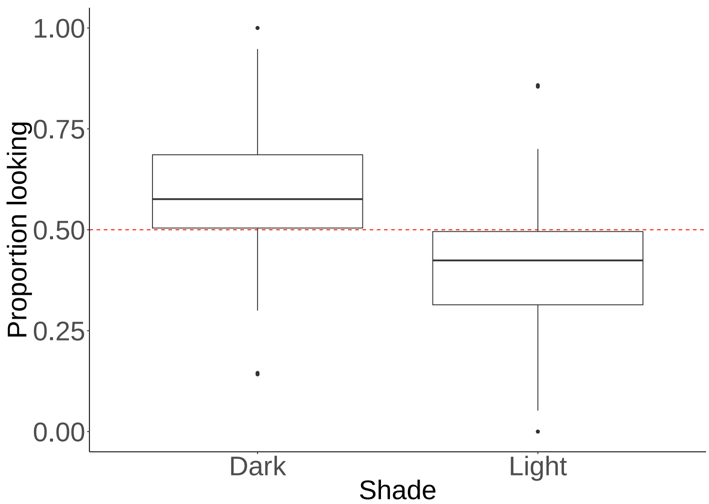

Figure 4.4: Looking proportions to dark and light in the pre-naming phase. The dotted line indicates chance looking.

Proportions of looking to the dark and light objects were compared with a Paired t-test. Participants looked more to the dark object than the light object before they heard a prompt to look at either ($t(145) = 7.3925$, $p < 0.001$). This demonstrates an overall preference for the dark objects over the light objects.

Due to the hypothesis that adults may display different colour preferences to infants, participants were then compared in two groups: the younger participants, who largely showed no evidence of knowing the words (up to and including two years of age), and the older participants (the oldest age group – 48 months), who largely demonstrated that they knew the terms "dark" and "light," and who may behave more like adults, in terms of their colour preferences.

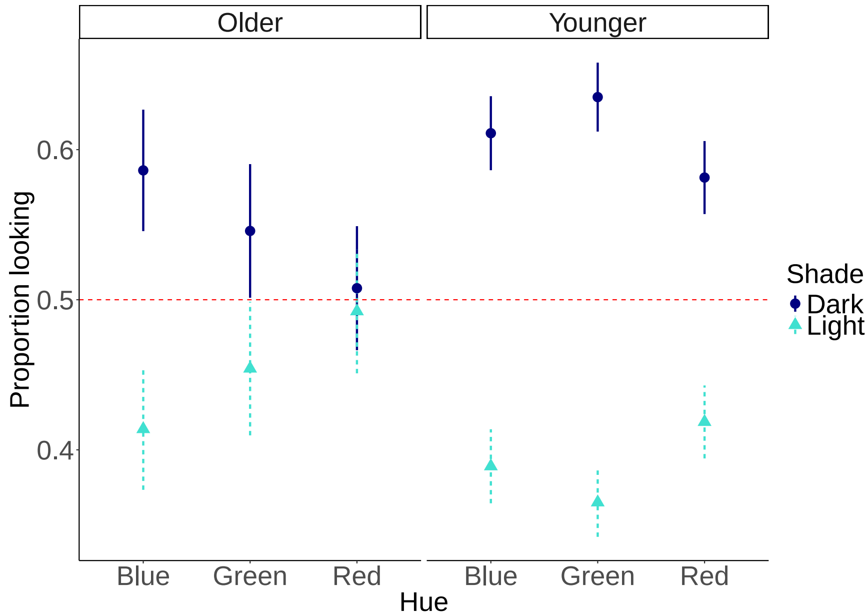

Figure 4.5: Looking proportions to dark and light for each hue and age group in the pre-naming phase. Plotted are means and standard errors. By design, light and dark preferences are necessarily symmetrical.

The proportion of looking to the dark object over the light object was compared for the two groups, with a Welch Two Sample t-test. There was evidence that participants' looking to the dark object decreases in the older age group ($t(142.25) = -2.234$, $p = 0.027$). Similarly, Figure 4.5 demonstrates a clearly changing pattern of colour preference between the 48 month-old age group and the other age groups. The effects of hue (red, green or blue) and age group (older or younger) on the proportion of looking to the dark object were compared with a series of binomial generalised linear models, where the fixed effects were added individually and the change in deviance was assessed with a Chi-squared test.

The addition of group improved model fit (Dev(1)= 45.246, $p < 0.001$), as did the addition of the effect of hue (Dev(2)= 76.309, $p < 0.001$), and the interaction between hue and group (Dev(2)= 30.215, $p < 0.001$). The model coefficients (Table 4.3) also suggest strong evidence for a difference between red and blue, and strong evidence for an interaction between age group, and green. Older toddlers demonstrate different preferences of lightness to those of younger toddlers and infants, most noticeably when the hue is green.

Table 4.3: Model coefficients of GLM examining effects of hue and age group on looking to the dark object. Hues are compared to blue.

	Value	Std.Error	z-value	p-value
(Intercept)	-0.506	0.010	-48.632	<0.001
Group	-0.016	0.021	-0.779	0.436
Green	-0.023	0.015	-1.575	0.115
Red	-0.086	0.015	-5.741	<0.001
Group:Green	-0.156	0.030	-5.288	<0.001
Group:Red	-0.039	0.030	-1.305	0.192

4.4 Discussion

The present chapter has uncovered three main findings. First, that the words “light” and “dark” are learned very late, compared to when the colour words are comprehended; and second, that infants possess a basic colour preference for dark objects; and third, that their chromatic preferences appear to change as they age. This change in preference appears to be most noticeable for green hues.

There is a great deal of evidence that demonstrates that learning “dark” and “light” late is not surprising. Wagner et al. (2013) have demonstrated how an infants' understanding of a colour can change over time. To begin with, they start with very broad, over-extended categories, and these are slowly refined until there is a narrowed, adult-like understanding of the colour. Thus peripheral colours in any category are the last to be properly understood. "Dark" and "light" as modifiers refer exclusively to colours on the periphery of a colour category – it would be erroneous to refer to the centre of the category *blue* as "light blue" or "dark blue." Clearly in this instance, toddlers learn the words for "dark" and "light" only after they have comprehended the colour categories to an adult-like degree, and thus understand that these words modify the hues on the periphery of the colour categories.

The strong preference for dark items over light items is stronger than was expected. Taylor et al. (2013) demonstrated that infants have some preferences for dark over light colours, most obviously in yellow, but also possibly in blue. In the present study, where yellow was not included due to ambiguity, a strong preference is found for dark objects regardless of the hue (although this is diminished somewhat for red). While the present study confirms the results of Taylor et al. for the preference for dark objects when they are blue, they did not find this to be the case for red or green. It is, however, worth noting that in the previous chapter, infants indicated a strong preference for red over other colours, it was also in this colour that the least difference between dark and light preferences were found. This may speak to some overall pattern, where the more a colour is preferred by the individual, the less strong the lightness preference.

The preference for dark over light objects may also be a result of the stimuli choices. Similar to Taylor et al. (2013), a neutral grey colour was used for the background, which will undoubtedly affect infant looking. For the objects, a major point of difference between the present study and that of Taylor et al. is that images of real objects were used, as opposed to flat circles. The decision to use real objects was to make the experiment as close to real life as possible, in particular when dealing with object properties such as colour or lightness, as real objects are much more likely to suffer from the *gavagai* problem (Quine, 1960), where the word in question could refer to any number of the object's properties. By using real objects, the scenario most closely maps to that of the real world, increasing the features to which the word can refer, unlike when using circles. However, by using real objects, it allows the shading of the object to vary throughout, and thus the object can appear a different colour in different parts. In this sense the strength of the present study for examining the word learning may also be the weakness for colour preference, as the colours of objects can not be so closely controlled as in Taylor et al., introducing a great deal of variability. This may be a reason for the differences between that study and the present study, in that the variability in colour and shading of real objects might cause the differences, although these differences are more likely to reflect real-world situations.

The change seen between the age groups that had not learned the words "dark" and "light," and those that had, more closely reflects that of previous results. Taylor et al. (2013) demonstrated different colour preferences between adults and infants; in the present study, a similar difference is shown between younger infants and toddlers around four years of age. In the present study, the older age group demonstrates a less evident propensity for looking to the dark object than the younger age group. Equally, in Taylor et al., the adults demonstrate a preference for the lighter objects, in stark contrast to the infant preferences. The present results, combined with past results, raise the possibility that adult colour preferences may become lighter with age. In addition, the present experiment demonstrated an obvious shift between the two age groups when the colour is green, and much less of a shift when the colour is blue, mirroring the results of Taylor et al.. There may be some biological mechanism of chromatic contrast associated with this, as chromatic contrasts are continuously changing until they are optimised at around 16 years of age, and in decline thereafter (Knoblauch, Vital-Durand, & Barbur, 2001).

It remains to be seen whether the changes in lightness preferences demonstrated in the present study have underpinnings in colour vision, or in associations with objects (Palmer & Schloss, 2010). In addition, further investigations into colour preference with changing background colours will be needed to determine the universality of the claims made in the present study. It is clear that infants have a strong preference for dark colours, but that may change as they learn the terms that define those shades.

Chapter 5

The Role of Colour Labels in Mediating Infant Visual Attention

5.1 Introduction

In the previous chapters, the timing and variability in colour word learning have been analysed. In the present chapter, that research is extended to examine how colour labels mediate infant visual attention. Individuals spontaneously attend to an object when it is named, and will even direct their attention towards objects in the visual scene that share properties with the named object (Cooper, 1974). More recently, it has been shown that both adults and infants attend to objects based on semantic (Huettig & Altmann, 2005; Huettig & McQueen, 2007), phonological (Allopenna et al., 1998; Desroches, Newman, & Joanisse, 2009; Chow et al., 2017), and shape information (Huettig & Altmann, 2004) evoked by auditory words.

One recent field of enquiry into language-mediated attention has focused on the domain of colour. Huettig and Altmann (2011) demonstrated that when adults hear a target word with a typical diagnostic colour (e.g. *"frog"*), they attend to a green object. The finding was extended to 3 year-old participants with similar results (Johnson & Huettig, 2011). These results suggest that colour knowledge is a part of the meaning of some words, for both adults and children.

While it is clear that knowledge of the typical colour of objects has the potential to mediate attention, even in young children, it is unclear how this colour information is stored and accessed in the lexical-semantic system. One possibility is *direct activation*, where the label of a typically-coloured object evokes a mental representation which directly activates a representation of the associated colour feature. Another possibility is *label-mediated activation*, where the auditory label evokes a colour *label*, which subsequently activates an abstract representation of the colour.

In adults and children, the contrast between direct and label-mediated colour representations is difficult to operationalise behaviourally because both possibilities are potentially available to them: they know colour names *and* may represent the typical colours of objects independently of those names. However, very young children either learn colour names extremely late (Franklin, 2006; Mervis et al., 1995; Pitchford & Mullen, 2002; Sandhofer & Smith, 1999; Soja, 1994), or the process of gaining comprehension of colour terms is slow (demonstrated in the previous chapters, as well as Wagner et al., 2013, 2014) and therefore offer a natural opportunity to explore whether knowledge of colour labels is necessary to direct their attention towards objects in the visual scene that share colour properties with a named object.

Johnson et al. (2011) tested 24 month-old toddlers, who did *not* know colour labels according to a naming task, and discovered they also looked at an object that shared the same colour as the referent of an auditory word, e.g., the toddlers looked at the red plate upon hearing "*strawberry*." This result lends strong evidential weight to the direct access model, whereby the auditory label evokes a mental representation that encodes the associated colour feature of the target object, and activation of the associated colour feature can mediate attention towards other objects that share the same colour.

In the Johnson et al. (2011) study, participants were shown two pictures side-by-side and heard a target word on each trial. Participants were shown four target-present trials, four related distractor trials, four unrelated distractor trials, seven colour trials (used to test participants' colour word knowledge) and seven filler trials. The focus of the study was the four related distractor trials, which used labels for animals and food with typical colours. In these trials, participants saw an object matching either the colour of the object label (e.g. a red plate when hearing "*strawberry*)" or an object matching the semantic category of the object label (e.g. a sandwich when hearing "*strawberry*"), paired with an unrelated object. An analysis of these trials showed that in the second second after target word onset, participants look to both types of related distractor above chance. The semantic effect was, however, found to be stronger than the colour effect (see also Mani et al., 2013).

Johnson et al. (2011) also tested the colour word knowledge of participants with a preferential looking trial that used coloured smiley faces. The results suggested that only two of the colours (green and grey) were known to participants. Parents were also asked to report on children's production of colour terms and how well they were used. While 30 of the 48 participants were reported to produce some colour terms, only five were identified as being able to use the colour terms relating to the experiment. When the relevant trials from these five participants were excluded, participants still demonstrated systematic looking at the colour-related distractor. Thus, Johnson et al. (2011) concluded that colour labels did not appear to be mediating attention towards the colour-related distractors in these 24-month old toddlers, offering support for the direct activation account.

5.1.1 Comprehending Colour Words

The Johnson et al. (2011) study makes some critical assumptions about the nature of infant colour word knowledge. First, the authors assert that 24 month-old toddlers are unlikely to comprehend colour words. This claim concurs with previous reports about the timing of colour word learning (Franklin, 2006; Mervis et al., 1995; Pitchford & Mullen, 2002; Sandhofer & Smith, 1999; Soja, 1994). However, in Chapter 2, if was shown that the majority of infants learn colour words well before their second birthday, according to parental reports [1]. In Chapter 3, that research was verified with a behavioural study, lending weight to the suggestion that most toddlers know many colour words before their second birthday.

Second, the Johnson et al. (2011) study discounts the fact that nearly two-thirds of their participants were reported to have some knowledge of the relevant colour words for the task, since they were reported to often use these colour words incorrectly. However, colour words are probably learned in much the same way as other categories of word, in that initial production is preceded by a partial comprehension, after which comprehension and production continues to develop to a more adult-like form (Wagner et al., 2013, 2014). Partial comprehension of the relevant colour terms by the toddlers in the Johnson et al. (2011) study may have been sufficient to support label-mediated attentional processes. Furthermore, in Chapter 3 of this thesis, parental reports of comprehension of colour terms have also been found to

[1] Though it should be noted that no data was reported for Dutch—the language spoken by the participants in the Johnson et al. (2011) study.

predict behavioural data for infants as young as 19-months. While children do often apply colour words incorrectly, these errors are not random, and are mostly due to over-generalisation of the term. These misuses of colour words mask the fact that the focal point of the colour word is often understood correctly (Wagner et al., 2013).

5.1.2 The Present Study

In light of the findings regarding the earlier timeline of colour word learning by toddlers reported so far in this thesis, the present study aimed to replicate and extend the Johnson et al. (2011) study. In particular, I aimed to evaluate toddlers' performance on a similar task to that of Johnson et al. (2011), comparing toddlers who did and did not understand the colour labels in question, using the methods reported in Chapters 2 and 3. Since these previous chapters have shown that most 24-month olds have at least a partial understanding of several basic colour terms, this chapter also includes a second, younger age group. Colour word learning in English has been shown to begin around 19 months in Chapter 3, making them an ideal comparison group to the original study cohort. Thus, in this chapter, participants from two age groups were recruited; 19-month olds as well as a 24-month old age group as in the original study.

The types of trial were kept identical to the original Johnson et al. (2011) study, save for two exceptions. First, the colour trials used to test participants' colour word knowledge were removed, as parental reports are a strong predictor of tightly controlled experimental findings on colour word comprehension, as seen in Chapter 3. Second, four additional *colour discrimination* trials were added for each participant. These trials were similar to the colour related distractor trials in the original study, but the two objects were kept identical, except for colour, to remove any bias stemming from a preference for one object over another. These extra trials also allow for greater power when analysing the colour related trials.

In this chapter, it was predicted that only those infants who understand the relevant colour labels will systematically direct their attention to colour related distractors, if attention to an object's colour feature is mediated through the colour label. In contrast, if colour labels are not a prerequisite for mediating attention to colour related distractors, then both infants who know the colour label and those who do not should show similar levels of systematic attention to colour related distractors.

5.2 Methods

5.2.1 Participants

$N = 39$ nineteen month-old toddlers (17 female, mean age = 19.44 months, s.d. age = 0.49 months) and $N = 31$ twenty-four month-old toddlers (14 female, mean age = 24.30 months, s.d. age = 0.35 months) were recruited from maternity wards or online in the Oxfordshire region of the United Kingdom. All participants were learning English as their first language. A further 4 participants (all twenty-four months) were excluded from analyses due to fussiness.

5.2.2 Design

The experiment consisted of two parts: parental reports, and an automatic eye-tracking task. The parental report section utilised the Oxford Communicative Development Inventory (Oxford CDI, Hamilton et al., 2000). In addition, participants' caregivers were asked to fill out an additional report as to whether their child understood and produced each of 12 colour terms (*red, blue, green, yellow, black, white, orange, brown, purple, pink, grey,* & *aqua*), as analysed in Study 1 of Chapter 2.

The eye-tracking portion of the experiment was designed based on that of Johnson et al. (2011). In the eye-tracking task, each participant was randomly assigned to one of six stimuli lists, each of which consisted of several trial types. In each trial participants saw two different images, side-by-side, and were prompted to look at one. Trials were as follows:

4 Target trials were used to ensure that participants comprehended the target words. One of the objects on the screen was the object named.

4 Unrelated Distractor trials were used as a baseline for comparisons. Neither of the objects on the screen bore any relationship to the object named.

4 Related Distractor trials were used for the primary analysis. One of the objects on the screen was related either by colour or semantically to the object named. Two of these trials used a food, and two used an animal as auditory stimulus. Two of these trials tested a semantic mediation (e.g. a food as auditory stimulus was matched with a food as visual stimulus), and two tested a colour mediation (e.g. a typically red object as auditory stimulus was matched with a red object as visual stimulus).

4 Colour Discrimination trials operated the same way as the colour related trials above, but instead of hearing *"strawberry"* and seeing a red plate and a yellow t-shirt, on this occasion they saw a red plate and a yellow plate. This manipulation gives greater power to analyse the colour related data, and also eliminated a bias due to the shape of the object.

10 Filler trials were used to maintain participant interest. As with the target trials, the object named was present on the screen. The only difference was that the filler trials did not contain any of the objects of interest of the study.

The full lists of trials and stimuli used in the study is listed in Appendix B. Stimuli were counterbalanced so that in each list, participants heard all 12 target stimuli: four as Target trials, to evaluate knowledge of the target objects, four as Related Distractor trials, and four as Colour Discrimination Trials. The lists differed as to which stimuli the participants heard in each context, and the objects that were matched with them.

5.2.3 Stimuli

Auditory stimuli consisted of sentences encouraging participants to look at objects. In the filler trials, sentences consisted of the format, "where's the xxx?" All remaining trials were recorded in the format, "look at the xxx." In the target-present trials, the related distractor trials, and the colour discrimination trials, there were 12 target words, 6 animal and 6 food, each of which can be characterised by a typical colour (e.g. "*strawberry*"). In the unrelated distractor trials, target words were objects that lacked a typical colour (e.g. "*hat*"). Ten additional words were chosen for use during filler trials. Each target word was chosen to be familiar to participants in both age groups, and which they were likely to know based on previous parental reports. All auditory stimuli were recorded by a native female speaker of Souther British English (SBE).

Visual stimuli consisted of high-resolution images of objects, manipulated so that each took up roughly the same amount of space on the screen. Where necessary, images were recoloured using the GNU Image Manipulation Program (GIMP). All images that were chosen for colour discrimination trials or related distractor trials were manipulated to be very typical examples of the colour category, in order to avoid ambiguity.

5.2.4 Procedure

Participants' caregivers were asked to fill out the Oxford CDI and additional parental report on colour terms as in Study 1 of Chapter 2, prior to arriving at the lab, or on their arrival. Experimental sessions began with a short warm-up and play phase to allow participants to feel comfortable in the laboratory surroundings, during which time written consent was obtained from the caregiver. Once the participant had acclimatised to the surroundings, the eye-tracking experiment would begin. Participants were seated on the caregivers lap, at approximately 75cm from the screen. A nine-point calibration sequence was performed until at least seven of the nine points were calibrated successfully.

In each trial (see Figure 5.1), participants first saw an attractive attention-getter for two seconds, before the main trial began. The beginning of the trial was marked by the appearance of two images on the screen, one on the left and one on the right. Trials each lasted seven seconds, with the target word onset occurring two seconds into the trials. All trials were presented in a randomised order, and counterbalanced so that each participant only heard each target word once.

5.2.5 Analysis

Participants' looking was recorded using a Tobii eye tracker sampling at 120Hz. Looking to the target was determined by whether or not an infant was deemed to have fixated on the target, with a fixation defined as a stable gaze in one location for upwards of 100ms. Trial automation and gaze processing were completed with custom Matlab code. Data was then analysed using the R package

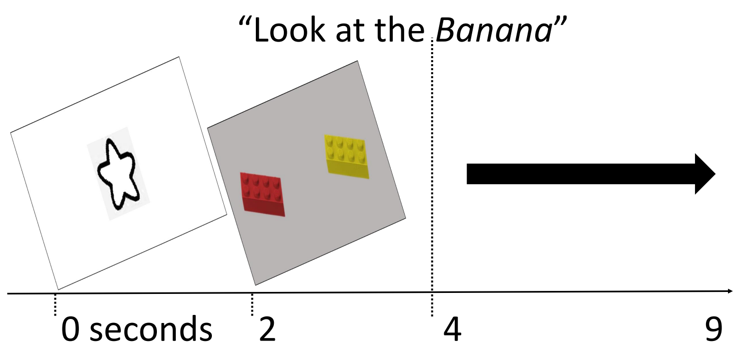

Figure 5.1: Time-line in seconds of eye-tracking trials: this is a typical Colour Discrimination trial, where the intended target would be the yellow brick.

eyetrackingR (Dink & Ferguson, 2015) and lme4 (D. Bates et al., 2017). In this chapter, the decision to use the glmer function in lme4 was made in order to facilitate model comparison, options not included in the MASS package used in Chapter 3. Data was cleaned such that trials wherein a participant did not fixate for more than 60% of the trial duration were removed, in order to reduce noise. Data was only included for a trial if the caregiver had marked that the child comprehended the target word in the auditory stimulus.

5.3 Results

5.3.1 Overall Results

Overall average looking proportions for all participants for the main trial types can be viewed in Figure 5.2. The overall proportions in the Target trials remained consistently above chance (0.5) after the onset of the target word, suggesting that participants on average correctly understood the target word, and looked primarily at the corresponding object.

In the Unrelated trials, looking was at chance.

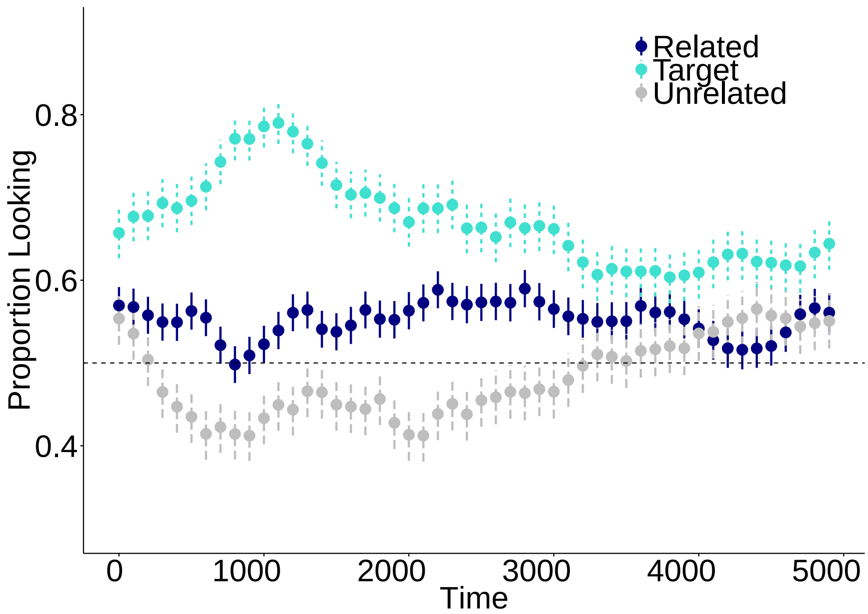

Figure 5.2: Overall results, collapsing over age, from target word onset. Related trials include colour related, semantically related, and colour discrimination trials. The dotted line indicates chance looking at 50%.

The Related trials in Figure 5.2 pool the Semantically Related trials, the Colour Related trials, and the Colour Discrimination trials (they can be viewed separately in Figure 5.3. In these trials the proportion of looking to the related image is only slightly above chance. The following analyses examine the results of these trials in more depth. All proportions of looking to the related image in this study take the proportions of looks to the Related image, divided by the total proportion of looking to either image:

$$Prop = \frac{looks_{target}}{looks_{target} + looks_{distractor}}$$

Looking in semantic trials is consistently above chance in both age groups, in concordance with Johnson et al. (2011). Collapsing over time in each trial, participants looked consistently more to the related distractor in the Semantic condition than in the Colour conditions (Welch two-sample t-test $t = -3.271$, $p = 0.001$, $df = 117.85$). Looking to the related distractor in both the Colour

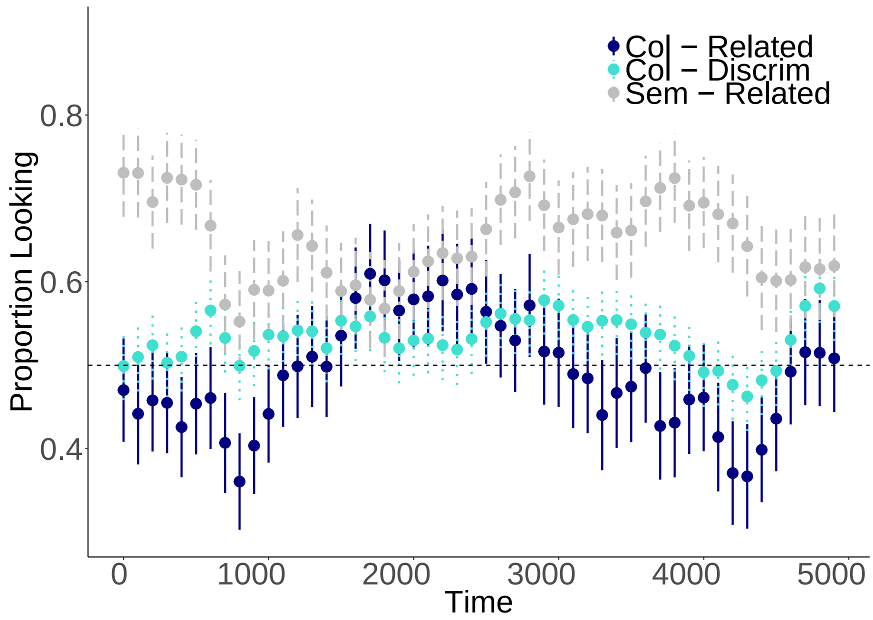

Figure 5.3: Overall results, comparing colour-related, colour discrimination, and semantically-related trials.

(one-sample t-test $t = 2.136$, $p = 0.034$, $df = 211$) and the Semantic (one-sample t-test $t = 4.929$, $p < 0.001$, $df = 71$) conditions were both above chance.

5.3.2 Colour results by age

In these analyses, only the Colour Related, and Colour Discrimination trials are analysed. Colour trials combine Colour Related trials and Colour Discrimination trials, for a total of 6 trials per participant, in order to increase statistical power. In the time window of interest (0 to 3000ms), no strong evidence was found for a difference in proportion of looks to the target in the Colour Related trials and the Colour Discrimination trials (Welch two-sample t-test $t = -0.952$, $p = 0.343$, $df = 122.44$), which allows combining the two.

The proportion of looking to the colour-related image was modelled with a hierarchical multilevel binomial model, which operates by using the binary outcome of fixation in the target area versus fixation in the distractor area. The use of this statistical method, as opposed to collapsing over time, allows a thorough examination of effects as they appear over time, as well as the contributing variables. These results were modelled with orthogonal quartic polynomials of time (elapsed after the target word onset during each trial, Mirman, 2014), and included the main effects of age group (19 or 24 months), as well as all interaction effects between the polynomials and age group. The intercept term and the linear polynomial were allowed to vary by participant, in order to maximize model fit while not over-fitting the data. All fixed effects were difference-coded, and added in to the model with the optimal random effect structure individually.

The model coefficients (Table 5.1) suggest that there was strong evidence for an effect of age, and of an interaction with the linear, quadratic, cubic, and quartic polynomial terms. These results suggest there was a substantial difference between both the overall amount that each age group looked to the colour-related distractor, as well as a difference in the looking patterns of each age group. The model fit to the raw data points can be viewed in the top frame of Figure 5.4.

Table 5.1: Model coefficients of age-based model. “ot” refers to the orthogonal time terms.

	Estimate	Std. Error	z value	Pr > z
(Intercept)	0.90	0.51	1.74	0.081
ot1	4.86	3.74	1.30	0.194
ot2	-0.03	1.94	-0.02	0.987
ot3	0.09	1.05	0.08	0.933
ot4	0.21	0.34	0.60	0.545
Age	-4.43	0.94	-4.73	<0.001
ot1:Age	-46.22	3.76	-12.30	<0.001
ot2:Age	-37.95	2.98	-12.75	<0.001
ot3:Age	-17.56	1.62	-10.83	<0.001
ot4:Age	-3.99	0.54	-7.34	<0.001

19-month old participants look to the related picture only at chance for the first three seconds following target word onset, suggesting that they are not

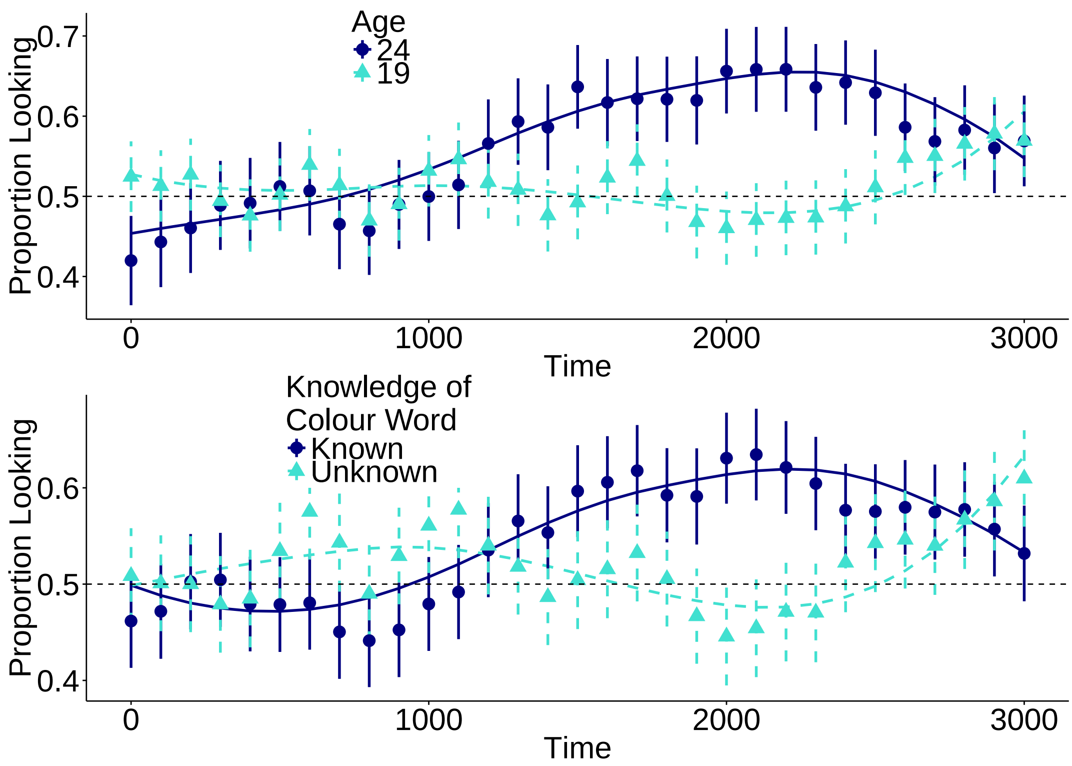

Figure 5.4: Colour trials vs semantic trials, fitted with a model by age in the top frame, and by knowledge of the relevant colour word, in the bottom frame.

influenced by the label to look at the colour-related distractor. By contrast, the 24-month old participants begin shifting their gaze toward the colour-related picture after approximately 1000ms, showing a strong preference for the object of the same colour as the labelled object in the second and third seconds of the trial following target word onset.

The results of the model suggest that while 24-month old toddlers tend to fixate a colour-matched object when hearing a target word, 19-month old toddlers do not. However, based on results of previous studies (Johnson et al., 2011), participants should look to the colour-related distractor regardless of their knowledge of the colour word. Since colour word knowledge will, by necessity and as shown in Chapters 2 and 3, improve with age, examining the difference by age does not completely disentangle colour word knowledge from other age-related factors that may mediate attentional distribution.

5.3.3 Colour results by colour word knowledge

In this analysis, the results were analysed in the same manner as in the model by age above, but with one critical difference: the variable of age was replaced by the variable of whether or not the participant was deemed to comprehend the relevant colour term (i.e. *"red"* in the case of *"strawberry"*, or *"yellow"* in the case of *"banana"*), based on the parental report.

The model coefficients (Table 5.2) show strong evidence for an effect of participants' comprehension of the relevant colour term, as well as strong evidence for that knowledge interacting with each of the orthogonal time terms. There was also evidence of effects of each of the orthogonal time terms. This result highlights that participants look differing amounts based on whether they know the colour word, and they pattern of looking is also different.

This model showed the same patterns of looking in the colour-word knowledge model as in the age model (the bottom panel of Figure 5.4). The participants who did not understand the colour word relevant to the target word show looking to the target largely at chance in fixation choices throughout the first three seconds of the trial, with some late looking at the colour-related distractor around the three-second mark. The participants who understand the relevant colour term look to the target in exactly the same way as the 24-month olds do, with a peak in looking during the 1000-2000ms time window. The match between the two models suggests that it may be colour word knowledge, rather than another, separate factor which is contributing to this effect.

Finally, the knowledge-based model (AIC 85294) was compared to the age-based model (AIC 85375) on the bases of the Akaike Information Criterion (AIC Akaike, 1998), with the lower AIC suggesting the knowledge based model fits the raw gaze fixation data better than the age-based model. This

Table 5.2: Coefficients of knowledge-based model.

	Estimate	Std. Error	z value	Pr> z)
(Intercept)	1.51	0.51	2.93	0.004
ot1	10.88	3.60	3.02	0.003
ot2	4.94	2.03	2.43	0.015
ot3	2.43	1.10	2.20	0.028
ot4	0.77	0.36	2.14	0.032
Knowledge	-4.05	0.38	-10.78	<0.001
ot1:Knowledge	-39.15	3.67	-10.67	<0.001
ot2:Knowledge	-30.61	2.96	-10.35	<0.001
ot3:Knowledge	-13.73	1.61	-8.54	<0.001
ot4:Knowledge	-1.97	0.54	-3.66	<0.001

analysis further suggests that participants' looking to the colour-related object was driven more by their knowledge of the colour term than by another property relating to their age.

5.3.4 Results at 19 Months

In order to prove that looking to the target was an effect of knowing the colour word, rather than age, a subsequent analysis was done only on a subset of participants. Since a large majority of the 24-month olds knew the relevant colour words whereas a substantial number of 19-month olds did not[2], we evaluated the patterns of looking for just the 19-month olds, comparing those reported to know the relevant colour word with those who did not. In this analysis of the 19 month-old participants, and as with the previous analyses, only the Colour trials were included. The model was fit in the same way as in the previous analysis, by knowledge of the relevant colour word. The only difference was that only cubic polynomial time terms were used to fit this model, due to the nature of the data.

The model coefficients (Table 5.3) demonstrate strong evidence toward an

[2]Unsurprisingly, there was very strong evidence that Age (19 vs. 24 months) and Colour Knowledge (Known vs. Unknown) were associated ($\chi^2(1) = 35.279$, $p < 0.001$).

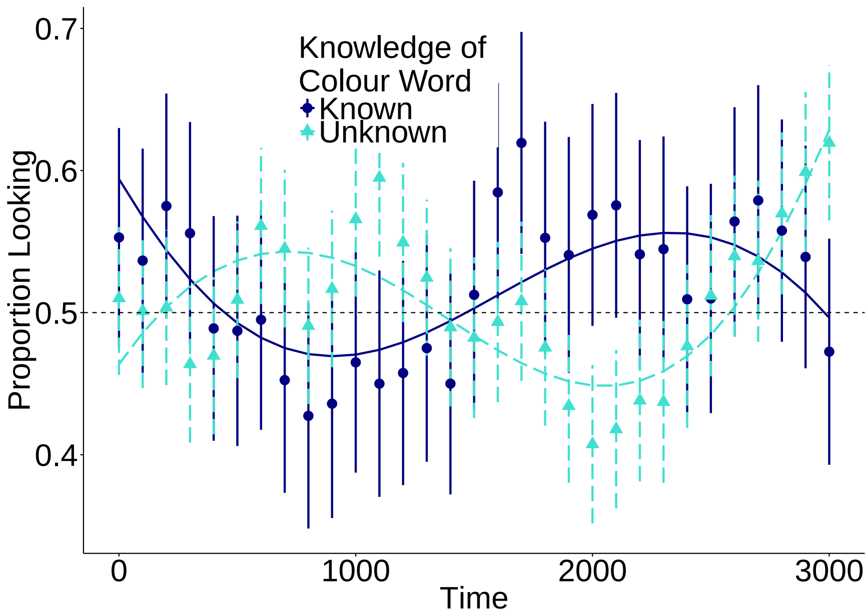

Figure 5.5: Model fit on nineteen month-old participants, by colour word knowledge

effect of the linear and quadratic time terms, as well as an effect of knowing the colour word, and interactions between that knowledge and each of the time terms. Nineteen month old participants show a similar pattern of looking as the overall group (Figure 5.5), if they know the relevant colour word: 19-month olds look systematically at the colour related distractor, in a similar time window as the overall group. If the participants do not know the colour word at nineteen months, they show no pattern of systematic looking during the same time window as that of the overall data.

5.3.5 The Role of Total Vocabulary

Infants learn words rapidly as they age (Schafer & Plunkett, 1998; Werker, Cohen, Lloyd, Casasola, & Stager, 1998). Not only is there a strong association between the age of the participant and their knowledge of the relevant colour word, but also between their knowledge of the colour word and their knowledge of all other words. To confirm that their behaviour in this task is

Table 5.3: Model coefficients of knowledge-based model run only with 19 month-olds

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.93	0.33	2.80	0.005
ot1	7.54	2.21	3.42	<0.001
ot2	3.20	0.77	4.15	<0.001
ot3	0.62	0.33	1.89	0.059
Knowledge	-2.88	0.22	-13.16	<0.001
ot1:Knowledge	-28.62	2.04	-14.03	<0.001
ot2:Knowledge	-21.71	1.46	-14.86	<0.001
ot3:Knowledge	-9.51	0.63	-15.18	<0.001

driven primarily by their knowledge of the relevant colour word and not their overall vocabulary size,[3] two models were fit to the 19 month-olds data: the first, identical to the 19 month-old knowledge-based model run previously; the second, also the same, but replacing knowledge of the relevant colour term with overall receptive vocabulary as a proportion of terms in the Oxford CDI. The lower AIC of the knowledge-based model (AIC 48709) compared to the CDI-based model (AIC 48880), suggests that the knowledge-based model fits the data better than the CDI-based model. Thus infants' behaviour in this task appears to be driven by their knowledge of the relevant colour term, rather than their overall receptive vocabulary.

Finally, since knowing any one term increases the chance of understanding any other term, a final test was conducted to ensure that the results above were the result of knowing the relevant colour word, and not knowing just another, random word. Only fixations during the 1500 - 2500ms time window were examined. First, each of the 416 terms in the Oxford CDI were taken, then proportions of looks to the related distractor in that time window were calculated for when each of those terms were comprehended by the participants. Those proportions were compared to the proportions of looks to the

[3]Participants who did not complete the Oxford CDI (10 participants) were removed from the subsequent analyses.

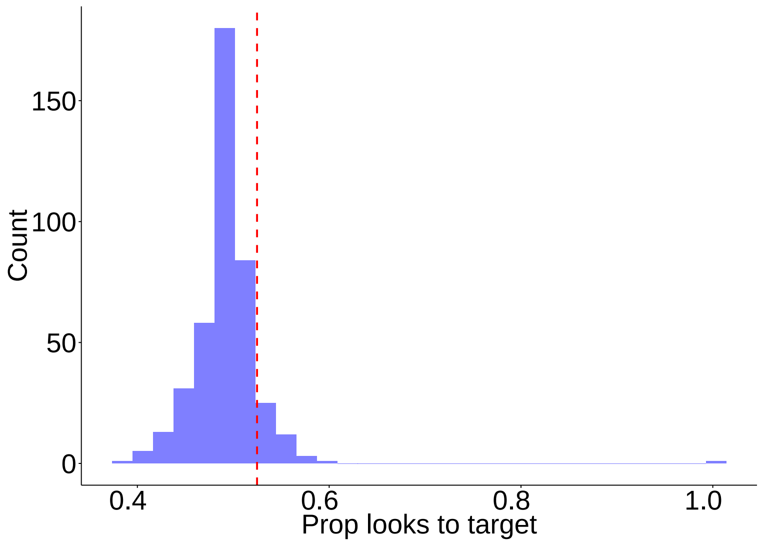

Figure 5.6: Proportions of looking to related distractor in the 1500 - 2500ms time bin when each of the words in the Oxford CDI is known. The dotted line indicates the proportion when the relevant colour word is known.

related distractor when the relevant colour word was known, using a one sample t-test. In the target time window, looking to the related distractor given knowledge of any given word in the Oxford CDI (Figure 5.6) was less than the proportion when they knew the relevant colour word ($t = -17.289$, $p < 0.001$).

In a colour word mediated looking task, shifts in attention toward a colour related object are driven by whether the colour word is known to participants, not by the participants' age, nor by their total vocabulary size, nor by their knowledge of another, random word.

5.4 Discussion

In accordance with previous findings (Johnson & Huettig, 2011; Johnson et al., 2011), the results of this chapter have found that even young toddlers systematically attend to both semantic and colour related distractors when hearing an associated object label. Contrary to these findings, however, the present study has found that in a language-mediated attention task, the ability to systematically fixate on the colour related distractor when hearing the name of a typically coloured object is determined by whether the toddler understands the relevant colour word. This finding was also verified by examining a group of 19 month-olds, confirming that the effect was one of colour word knowledge, not just of age, nor of total vocabulary size. These findings support two possibilities regarding the role of colour words in mediating attention in a task such as the one described:

1. An associated colour word is activated upon hearing the name of an object that typically possesses that colour. Lexical activation of the associated colour word directs the listener's attention to objects that possess this colour. This explanation is an example of *label-mediated activation* described in the introduction.

2. Learning colour words enhances the perceptual or cognitive salience of those colours, leading to heightened attention to objects sharing those colours. This explanation is an example of *direct activation*.

Both of these explanations, which need not be mutually exclusive, offer plausible accounts of the pattern of results reported in this experiment. For example, semantically-related words in the toddler's mental lexicon rapidly priming each other, e.g., the word *dog* primes the word *cat* (Arias-Trejo & Plunkett, 2009, 2013; Styles & Plunkett, 2009a). Although these studies report priming effects for words that are taxonomically and/or thematically related, it is not unreasonable to suppose that *property* words might also be activated by a related object word. Thus, hearing the word 'strawberry' may prime the word 'red' in the toddler's mental lexicon and direct attention to another red object. Picture priming tasks with toddlers (Mani & Plunkett, 2010; Mani, Durrant, & Floccia, 2012) also point to the efficacy of internally-generated labels directing attention to related objects. These lexical-semantic networks are already in place by 18-months of age (Delle Luche et al., 2014).

Research into how colour words are learned has shown that colour categories change as a colour word is learned (Wagner et al., 2013). Initially, colour terms are over-extended. Toddlers divide the spectrum with the terms that are available to them, resulting in boundaries for each colour category that are excessively broad. As additional colour terms are learned, the existence of additional colour categories is highlighted, promoting language-specific categorisation of the colour spectrum. This is part of a general process by which infants move from pre-linguistic, biological colour categories (Skelton et al., 2017), to an adult-like understanding of colour words. This process varies greatly by language and by frequency of exposure to the colour words as seen in Chapter 2 of this thesis. Again, it is not unreasonable to suppose that the fine-tuning of colour categories through exposure to language-specific labelling events highlights the salience of those colour categories, and facilitates the recognition of colour properties that are shared amongst objects. In the present study, toddlers need not necessarily activate the appropriate colour word in order to activate the colour concept, but their ability to match, from memory and abstraction, the colours, may be affected by their ongoing language-driven formation of these colour categories. As discussed in previous chapters, nineteen months marks the earliest stages of colour word learning in British English, and as such the shift that occurs in colour categories during this period may affect their performance in this task.

Further studies into colour abstraction and memory at different ages may serve to determine further the plausibility of the direct activation and label-mediated accounts. For example, if attention is mediated via a colour label then we expect attention to be directed by a prototype or some abstract representation for that colour. In contrast, if attention is mediated directly by the mental representation of the object itself, then object-specific colour characteristics may be more influential in directing attention, e.g., frogs may be considered 'green' but not necessarily a prototypical green.

What are the implications of this study for other types of object property labels, such as those referring to texture, size and shape? It has been established that toddlers are able to generalise word meanings on the basis of such properties (Landau, Smith, & Jones, 1988; Jones & Smith, 2002). It is not clear that knowledge of labels for these properties is required for such generalisation to occur. However, a clear implication of the current study is that knowledge of property labels is likely to highlight attention to such properties in a language-mediated attention task, either through direct activation or indirect label-mediated activation.

For the case of colour properties, this chapter demonstrates that there is an undeniable role for colour words in a language-mediated attention task. These words may achieve their effect either through implicit activation via word priming, or through highlighting the importance of colour in evaluating the similarity between objects. In either case, even the *unheard* word impacts attention in a visual search task.

Chapter 6

Preferential Looking in Categorical Perception of Colour

6.1 Introduction

In the previous chapters, I have examined colour word learning in depth, demonstrating various aspects of colour word comprehension and processing. However, colour word comprehension is just one aspect of colour knowledge, the other is a colour *category* knowledge, as discussed in Chapter 1. In both this, and the following chapter, the arguments made in the previous chapters are expanded to examine categorical perception of colour in infants.

Categorical perception (CP) in colour was first demonstrated in the 1980s (Bornstein & Korda, 1984; Kay & Kempton, 1984, see Chapter 1), but possible effects of language were first brought to light by Gilbert et al. (2006), using a reaction time measure in an oddball task. CP is often defined as a faster or more accurate discrimination of stimuli that cross a category boundary (Regier & Kay, 2009). Gilbert et al. not only successfully demonstrated that CP of colour was found, but that it was found only in the right visual field (RVF) rather than the left in adult participants. Participants in the study oriented faster to oddball stimuli that crossed the green-blue colour boundary than those that remained within it, despite equally discriminable differences. This finding, supported by similar claims (Al-Rasheed et al., 2014; Al-Rasheed, 2015a, 2016; Clifford, Franklin, Davies, & Holmes, 2009; Drivonikou et al., 2007; Franklin, Catherwood, et al., 2010; Gilbert et al., 2008; Goldstone & Hendrickson, 2009; Zhou et al., 2010), gave rise to ideas of a left-hemispheric specialisation for CP (due the crossover between neural hemisphere and visual field, described in Chapter 1), relating to language.

These findings were extended with a series of experiments that tested CP in pre-linguistic infants (Franklin, Clifford, et al., 2005; Franklin, Drivonikou, Clifford, et al., 2008; Franklin, Drivonikou, Bevis, et al., 2008). Franklin, Drivonikou, Clifford, et al. used latency to first look in an eye-tracking paradigm based on the Gilbert et al. results, showing no CP in the RVF, but instead, showing CP of colour in the left visual field (LVF). The authors took this to mean that the right hemisphere of the brain encodes biological, pre-linguistic colour categories, while the left hemisphere, due to its connection with language, encodes the linguistic categories (for discussion of hemodynamic responses see Yang, Kanazawa, Yamaguchi, & Kuriki, 2016). This finding was conceptually replicated with two groups of toddlers, one which knew colour names, and one which did not (Franklin, Drivonikou, Bevis, et al., 2008).

However, CP of colour has since come under considerable scrutiny. Witzel and Gegenfurtner (2011) noted that the stimuli used by Gilbert et al. (2006) were not controlled for chromacity, and in a careful replication controlling for all factors, found that the effect did not replicate (for other criticisms and replication failures see also A. M. Brown et al., 2009; Jraissati, 2012; Lindsey & Brown, 2009; Ocelak, 2016; Witzel & Gegenfurtner, 2013, 2016). The methodology used in these experiments was also criticised; A. M. Brown et al. (2011) noted that reaction times could be predicted from perceived differences and sensory differences, suggesting that there may be a lower-level explanation for the CP observed. An aim of both this, and the next chapter of this thesis is to examine these claims, carefully controlling for stimuli and avoiding reaction time measurements.

The previous research into CP of colour in infants presents similar debate. As noted by Hanley and Roberson (2008), data from only half of the 26 prelinguistic infants examined were included in the Franklin, Drivonikou, Clifford, et al. (2008) study, and the infants were presented with different stimuli to the adults. There is also the possibility that the colour vision of the infants was still developing at around 26 months, the age tested in the study (for more discussion see Roberson et al., 2009; Roberson & Hanley, 2009). The replication with toddlers (Franklin, Drivonikou, Bevis, et al., 2008) avoids these criticisms by testing two groups of toddlers around 40 months. In that study, they were split into two groups based on their production of the colour terms (learners and namers). However, as discussed in Chapters 2 and 3 of this thesis, by 40 months, toddlers would be expected to have a strong understanding of at least typical examples of the basic colour categories, suggesting that this distinction may not be appropriate. As the measurement was between groups, rather than within groups, the N of 37 (around 18 in each group) is again not very large. That infants can perceive colour categorically is thus a claim worthy of revisiting.

Recent research suggests the presence of infant colour categories (e.g. Franklin, Clifford, et al., 2005; Skelton et al., 2017), with findings suggesting that infants possess biological categories of colour for red, yellow, green, blue and purple.[1] Skelton et al. (2017) discovered that infant colour categories have biological origins, based on retinogeniculate pathways (see also Franklin, Skelton, & Catchpole, 2014). The finding suggests that despite criticisms of the methodology used to measure CP of colour in infants, there are grounds to believe it exists. As discussed in Chapter 1, the Skelton et al. finding is based on a novelty-preference task, which measures only 10 infants per between-category group. The results are extremely promising for the existence of infant colour categories, but further investigation is doubtless required to verify those claims.

The differing theoretical accounts of the origins of CP hinge on colour words, and by extension, participant age. On one hand, claims that colour categories are innate, and biological, and then change with the acquisition of colour terms, suggest that infants should display CP of colour prior to their acquisition of colour words (Franklin, Clifford, et al., 2005; Franklin, Drivonikou, Clifford, et al., 2008; Franklin, Drivonikou, Bevis, et al., 2008; Skelton et al., 2017). On the other hand, consistent results from studies with adults in different languages demonstrate that the presence of a category boundary is dependent upon the size and boundaries of the colour *word* category in the native language of the speaker (e.g. Roberson et al., 2009, 2008). While these two accounts need not be mutually exclusive, a hard interpretation of the latter would suggest colour categories only being learned when colour terms are acquired, while the former would indicate that they should be a native, perceptual property. One of the aims of this chapter is to examine this aspect of infant CP.

In this chapter, I revisit the claims made about categorical perception of

[1] Although it should be noted that yellow is not a typical "adult-like" understanding of yellow, and that green and blue do not fall into a typical distinction based on the World Color Survey (Kay et al., 2011).

colour in infants, with a carefully-controlled eye-tracking study that avoids the pitfalls mentioned above. First, a forced-choice preferential looking task is used, rather than an oddball task, avoiding the criticisms of reaction times as a measure of CP. Second, stimuli were carefully chosen based on past experiments. Third, a larger sample size (and lower drop-out rate due to the nature of the task) was tested to reinforce the claims made. In addition, to examine the role that colour words play in CP of colour, two age groups are tested: 12 month-olds, who are unlikely to know colour words, and 19 month-olds, for whom the colour word learning process is beginning.

6.2 Methods

6.2.1 Participants

A total of $N = 64$ participants took part in this study. Participants were divided into two groups: 12 month-olds ($N = 33$, mean age 12.23 months, S.D. age 0.39 months), and 19 month-olds ($N = 31$, mean age 19.29 months, S.D. age 0.47 months). Of those, 9 12 month-old participants were excluded due to failure to complete a trial successfully with less than 60% trackloss, and 1 19 month-old participant was excluded due to failure to calibrate. The final analysed sample of participants was $N = 54$. No participant in this study recorded a family history of colour vision problems. Participants were all recruited in the manner highlighted in the previous chapters. Most participants in this study took part in an additional study (Chapter 7) with the order counterbalanced.

6.2.2 Materials

Stimuli for this experiment were carefully chosen, based on the analyses of Witzel and Gegenfurtner (2011, Study 1). In their re-analysis of Gilbert et al. (2006), the authors cite the need to keep both *Munsell value* and *chroma* constant for the stimuli, settling on a *Munsell value* of 5 and a *chroma* of 6 (Fairchild, 1998; Munsell Color Services, 2007). The present study uses the same values, which are separated by 5 steps, described in Table 6.1. 4 simulated *Munsell chips* were used (7.5G5/6, 2.5BG5/6, 7.5BG5/6, 2.5B56), using the CIE 1931 Yxy values given in Witzel and Gegenfurtner (2011, supplementary material). These were converted into a calibrated RGB for display on the monitor. These were the stimuli used in the *blue* and *green* conditions of this study.

Due to concerns by Franklin, Drivonikou, Clifford, et al. (2008) that younger infants may not be responsive to fine differences in stimuli, an additional condition was included in this study. Stimuli for the *far* condition were instead separated by 10 steps, and designed based on the stimuli used in Franklin, Drivonikou, Clifford, et al. (2008). On this occasion, a constant *Munsell value* of 6 and a *chroma* of 8 was used, as in the aforementioned study. Illuminant C was created using the x and y values described by Fairchild (1998).

6.2.3 Design

Each condition was designed so that participants saw a background stimulus that was close to the category boundary between green and blue. Each background colour had two squares on it, each an equal number of steps from the background colour. One of those squares would be within the same colour category (within-category), and the other would cross the category boundary (between-category):

Table 6.1: Chromaticity coordinates (CIE 1931) for the three conditions in the present study. In the *blue* condition, the first 3 are used; in the *green* condition, the second to fourth stimuli are used. The three denoted with “far” are used in the *far* condition.

Stimulus	Type	Y	x	y
2.5B5/6	Blue	22.0	0.218	0.276
7.5BG5/6	Blueish	23.8	0.229	0.305
2.5BG5/6	Greenish	23.6	0.241	0.343
7.5G5/6	Green	23.1	0.260	0.368
2B6/8	Blue (far)	19.47	0.209	0.282
2BG6/8	Greenish (far)	19.47	0.235	0.355
2G6/8	Green (far)	19.47	0.286	0.431
Illuminant C		23.8	0.310	0.316

1. In the *blue* condition, participants saw a blueish (Table 6.1) background, with a greenish square on one side and a blue square on the other.

2. In the *green* condition, participants saw a greenish background, with a green square on one side and a blueish square on the other.

3. In the *far* condition, participants saw a greenish background (denoted with *far* in Table 6.1, with a green square on one side and a blue square on the other.

Each condition had four trials, with each colour appearing on the left twice and the right twice. This gave a total of 12 trials, which were randomised in order. As *Munsell chips* are only standardised in the presence of Illuminant C (Fairchild, 1998), and to avoid participants habituating to any one colour, the screen was rendered Illuminant C between each trial for a minimum of 2 seconds. The monitor (1920 x 1080 pixels) used for presentation was calibrated using a datacolor Spyder 5 Elite calibration device, ensuring that the colours presented matched the intended hues. Each square was presented as a 531 x 531 pixel square.

If infants demonstrate categorical perception of colour, as suggested in the literature (Franklin, Clifford, et al., 2005; Franklin, Drivonikou, Clifford, et al., 2008; Franklin, Drivonikou, Bevis, et al., 2008), participants will systemically orient to either the within-category colour or the between-category colour. For the purposes of this study, the hypothesis was that they would fixate the between-category colour, but strong systematic looking to either across all conditions could interpreted as evidence for CP. Additionally it was thought that both participant age, but more specifically, colour word knowledge would predict behaviour in the task.

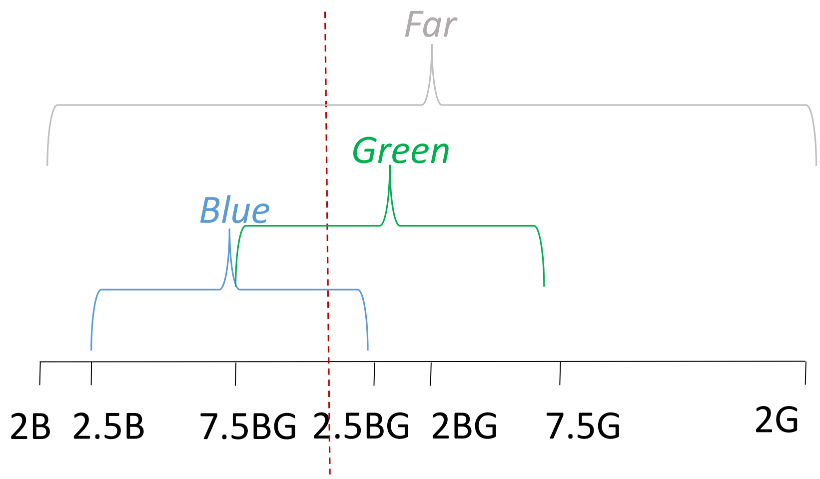

Figure 6.1: Stimuli selections for each condition. The dashed line indicates an approximate category boundary.

6.2.4 Procedure

After filling out consent forms, as well as the colour word supplement to the Oxford CDI (from Study 1 of Chapter 2), and a short play session, participants were seated on the caregivers lap, approximately 75cm from the screen. The standard 9-point calibration sequence was adapted to avoid priming the participant, using a solid black star moving across all 9 points, on a background approximating Illuminant C. Calibration was run until the eye-tracker successfully calibrated at least 7 of the 9 points.

At the beginning of each trial, a black and white flapping star oriented participant attention to the centre of the screen. During that time, a recorded voice said "*Look!*." After 2 seconds the trial began, and the background colour and both squares appeared immediately on the screen. The trial continued for another 3 seconds. Infant gaze was measured with a Tobii TX300 remote eye-tracker, recording at 120Hz. Trials were automated using a custom Matlab script.

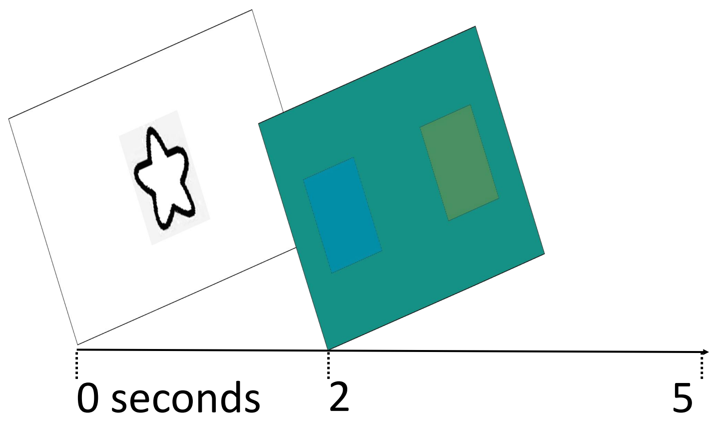

Figure 6.2: A sample trial. Borders around each square are for ease of viewing in this figure, and were not included in the experiment.

6.2.5 Analysis

Analysis was performed in the same manner as in many of the previous chapters, extracting looking for each side (expanding regions of interest by 25% in the same manner as Chapter 3). Looking was calculated in each 100ms time bin for each participant using eyetrackingR (Dink & Ferguson, 2015). Data was then modelled using a hierarchical binomial growth curve, using glmm-

PQL from the package MASS in R (Ripley et al., 2017).

6.3 Results

6.3.1 Overall analysis

Results were calculated to examine the proportion of time in each 100ms time bin that participants looked to the between-category colour. The proportion was taken to be:

$$Prop = \frac{Looking_{\text{Between}}}{Looking_{\text{Between}} + Looking_{\text{Within}}}$$

Thus the null hypothesis in this test would be that looking remained around chance (50%) for the 3000ms of the trial.

Based on an examination of the data, the data was modelled with quartic orthogonal polynomials of the time term (Mirman, 2014), along with a dummy-coded fixed effect of condition (*green*, *blue* or *far*), and a difference-coded fixed effect of age (12 or 19 months). Interaction effects between condition and each polynomial time term were also included as fixed effects. The intercept was allowed to vary by participant.

The model output can be examined in Table 6.2. Importantly, there was strong evidence that looking in both *far* and *green* (both with a green background) differed greatly from *blue*. In addition, their interactions with the polynomial time terms (with the exception of the linear time term) suggest that there is strong evidence for different looking patterns in each condition. Also of note, however, is that there was no evidence for looking varying by age. A graph of the model fit, varying by condition, can be seen in Figure 6.3.

In both the *blue* and the *far* conditions, participants demonstrate system-

Table 6.2: Coefficients of model on participant looking behaviour. Condition was compared to *blue*.

	Value	Std.Error	DF	*t*-value	*p*-value
(Intercept)	-1.236	0.154	2790	-8.017	<0.001
ot1	1.031	0.421	2790	2.453	0.014
ot2	1.992	0.395	2790	5.049	<0.001
ot3	-1.810	0.350	2790	-5.169	<0.001
ot4	2.128	0.318	2790	6.691	<0.001
Far	0.371	0.097	2790	3.835	<0.001
Green	0.884	0.107	2790	8.258	<0.001
Age	0.037	0.288	52	0.128	0.899
ot1:Far	0.348	0.677	2790	0.514	0.607
ot1:Green	-0.598	0.694	2790	-0.862	0.389
ot2:Far	-1.795	0.646	2790	-2.779	0.006
ot2:Green	-1.454	0.657	2790	-2.214	0.027
ot3:Far	1.821	0.538	2790	3.383	<0.001
ot3:Green	2.160	0.569	2790	3.798	<0.001
ot4:Far	-1.064	0.446	2790	-2.384	0.017
ot4:Green	-2.272	0.511	2790	-4.448	<0.001

atic looking toward the within-category object. This is most apparent in the *blue* condition. In the *green* condition there is still some evidence of looking toward the within-category object, but looking is much closer to chance throughout the trial. This indicates a pattern of systematic looking, albeit with some variation between conditions, toward the within-category colour. The same fitted model, instead separated by participant age, can be seen in Figure 6.4.

As suggested by the model output, there is very little variation in participant looking by age – the overall systematic pattern of looking to the within-category colour is maintained regardless of the age of the participant. This result indicates that the behaviour does not change as participants edge closer to learning colour words. In order to test this finding, and to disentangle participant age from their knowledge of the colours being tested, a further analysis was run using only the 19 month-old age group.

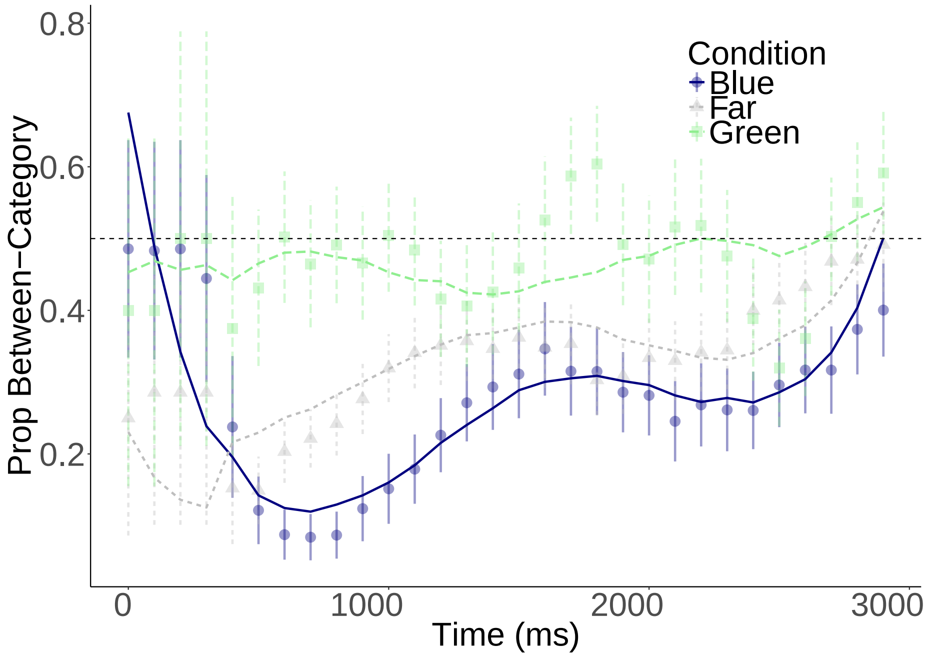

Figure 6.3: Model fit of participant looking to the between-category colour, by condition. In the Green and Far conditions, squares are presented on a green background, in the Blue condition, squares are presented on a blue background.

6.3.2 Colour word knowledge analysis

For this analysis, only the 30 19 month-old participants were included. Participants were coded as knowing colour words if their parents indicated that they comprehend at least one of the words "green" or "blue," in the colour words supplement to the Oxford CDI. Thus there were by this distinction 11 participants who knew the colour words and 19 to whom the colour words were unknown.

The looking data was modelled again as above, using a binomial hierarchical model, with quartic polynomials of the time term. The variable of knowledge was included in the model, as was its interaction with each of the polynomial time terms. Knowledge was specified as being dichotomous (known or unknown), and the resulting categorical variable was difference-coded for ease of interpretation. Again, the intercept was allowed to vary by participant

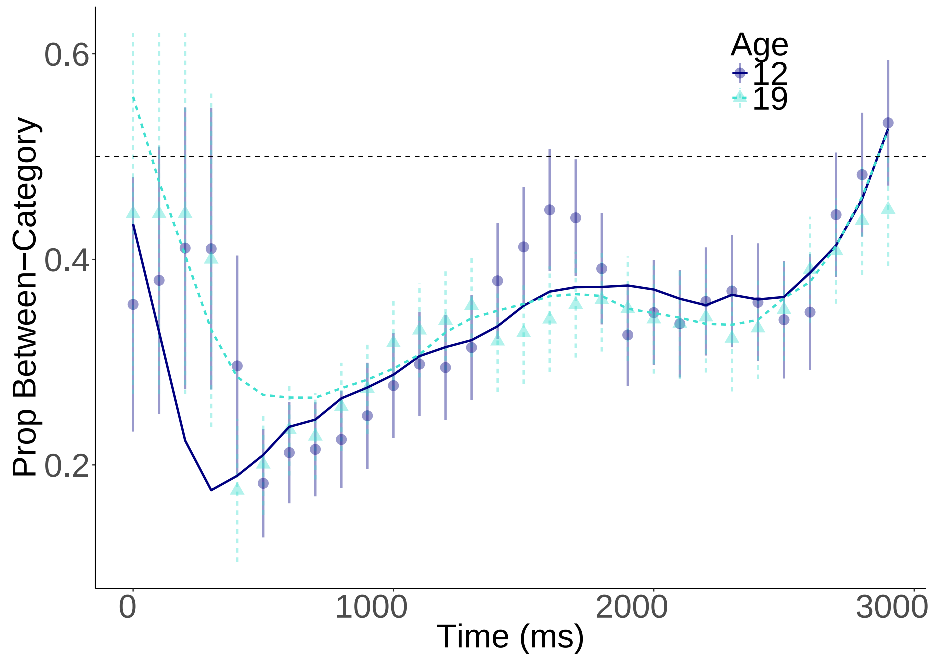

Figure 6.4: Model fit of participant looking to the between-category colour, by participant age.

to allow for individual differences.

Table 6.3 shows the coefficients of the fitted model. Of interest is the colour word knowledge variable, which the model output suggests there is no overwhelming evidence that it has an effect on looking. There is strong evidence of an effect of the interaction between the cubic time term and colour word knowledge, suggesting the possibility of slightly different looking patterns, if not in the amounts of looking.

Both groups of participants systematically look to the within-category colour, regardless of their knowledge of colour words (Figure 6.5). The group who did not comprehend either of the colour words maintained more attention throughout the trial to the within-category colour, however, indicating a possible mediation of looking patterns by colour word knowledge. Toddlers demonstrate a systematic preference to the within-category colour, regardless of age or colour word knowledge acquisition.

Table 6.3: Output of the model on looking mediated by colour word knowledge in 19 month-olds.

	Value	Std.Error	DF	*t*-value	*p*-value
(Intercept)	-0.847	0.186	722	-4.540	<0.001
ot1	1.463	0.379	722	3.862	<0.001
ot2	0.522	0.367	722	1.421	0.156
ot3	-0.124	0.299	722	-0.415	0.678
ot4	0.798	0.240	722	3.325	0.001
Knowledge	0.738	0.373	28	1.980	0.058
ot1:Knowledge	-0.274	0.757	722	-0.362	0.718
ot2:Knowledge	1.208	0.735	722	1.645	0.101
ot3:Knowledge	-1.857	0.599	722	-3.101	0.002
ot4:Knowledge	0.772	0.480	722	1.608	0.108

6.4 Discussion

In this chapter, infant categorical perception of colour was tested with a forced-choice Preferential Looking test, the first to use this paradigm to test colour perception, measuring whether participants demonstrate a systematic preference to either in within-category or the between-category colour. The results indicate a systematic preference to the within-category objects, demonstrating some level of category sensitivity, but running counter to expectations about which they would fixate. This result is consistent with those shown in previous studies (Franklin, Clifford, et al., 2005; Franklin, Drivonikou, Clifford, et al., 2008; Franklin, Drivonikou, Bevis, et al., 2008), demonstrating a category awareness of colours in infants and toddlers.

The results from this chapter can, in some sense, be interpreted in a manner similar to an infant categorisation task (e.g. Plunkett et al., 2008; Younger, 1985). In such a task, a strong novelty preference or a strong familiarity preference could both be considered to be evidence of category knowledge, while a preference for neither (i.e. looking at chance) would suggest no comprehension of the categories in question. It is unquestionably clear from the results of this chapter that the participants demonstrate an ability to distinguish be-

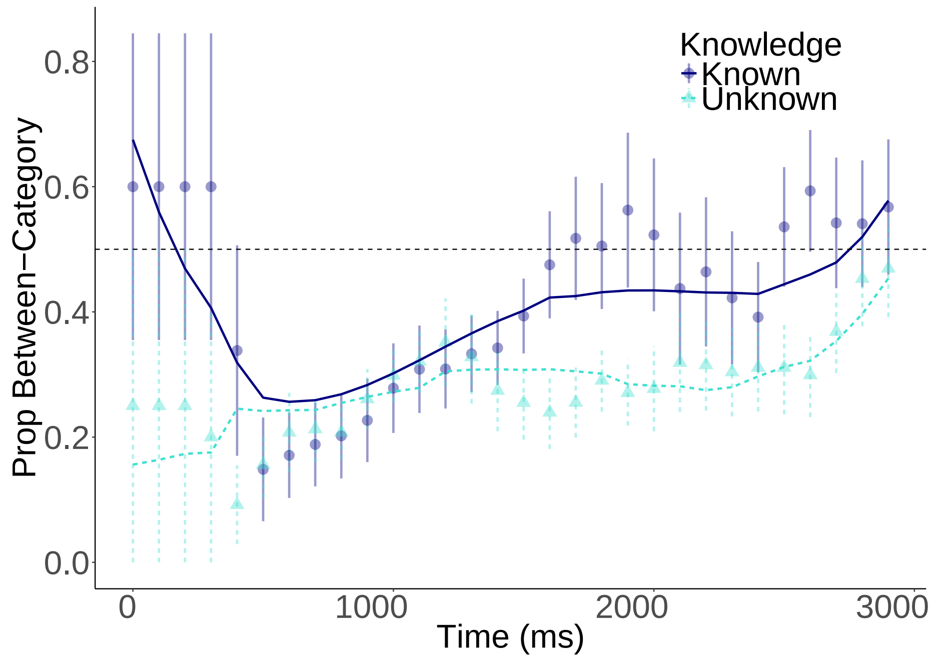

Figure 6.5: Model fit of 19 month-olds looking to the between-category stimulus, by colour word knowledge.

tween the two colour categories, and thus a knowledge of the category boundary, despite the equally perceptible differences between the hues.

One possible alternative explanation to this result would be that infant looking was mediated by colour preferences (Roberson & Hanley, 2009, make similar arguments about the Franklin et al. studies), rather than by a knowledge of the colour category. Upon examination of the results indicated in Figure 6.3, however, this appears unlikely. In the *blue* condition, where the pattern of looking to the within-category colour is most dramatic, the within-category colour is blue, on a blueish background. As demonstrated in Chapter 3 of this thesis, green is a preferred colour over blue in most cases (as well as Taylor et al., 2013), thus an account based purely on colour preferences would indicate a systematic pattern of looking in the *other direction* in the *blue* condition. Similarly, a colour preference-based account would suggest the highest probability of looking to the within-category object in the *green* and *far* conditions, not the case demonstrated in the present study.

In this chapter neither age nor colour word knowledge substantially change the behaviour of categorical looking observed here. Some evidence is demonstrated for an interaction effect between the third-order polynomial and knowledge, but no main effect of knowledge, nor does it interact with any other polynomial term. This may demonstrate some evidence that colour word knowledge mediates small shifts in attention, but the evidence is insufficient to make such a claim. This finding that neither age nor colour word knowledge substantially modify looking behaviour, suggests that rather than being acquired when colour words are learned, infant colour categories pre-date their comprehension of colour words, as discussed in the introduction to this chapter. However, some differences in looking patterns were observed between participants who knew colour words and those who did not, but the overall preference remained the same. A possible explanation of this phenomenon could be that lateralisation of CP of colour reverses when colour words are learned (Franklin, Drivonikou, Clifford, et al., 2008; Franklin, Drivonikou, Bevis, et al., 2008). It is possible that the minor differences observed here are indicative of the shift from biological categorisation of colour to linguistic categorisation of colour, but there is no evidence for this claim. The present study does not test lateralised effects, making it impossible to determine. Answering this pertinent question is an area worthy of further investigation.

Since neither age nor colour word acquisition mediated infant preferences to the within-category colour in the present study, a strong interpretation of a top-down role for language in the processing of colour can be rejected (Lucy & Shweder, 1979). What to make, then, of findings that colour category boundaries change dependent on the language of the speaker (Roberson et al., 2008, 2009)? Infant biological colour categories, and linguistically-determined colour categories in adults, need not be mutually exclusive, rather the results of research to date suggest an account wherein infants possess biologically determined colour categories (Skelton et al., 2017), but then these adapt to the languages learned by the speaker. Indeed the evidence presented in Chapter 2 of this thesis demonstrate that the rates and order of colour word acquisition varies greatly between languages, perhaps indicating the beginning of a process of establishing linguistically-determined colour categories.

Theories regarding the presence of infant colour categories are well-established (e.g. Franklin, Clifford, et al., 2005; Skelton et al., 2017), but it was hitherto debated whether they can use colour categories to demonstrate CP of colour outside of the heavily-criticised paradigms used previously (Franklin, Drivonikou, Clifford, et al., 2008). The paradigms used by Skelton et al. (2017) were novelty preference measures, suggesting that pre-linguistic infants can familiarise to colours and show a novelty preference across colours. Equally, the paradigm used by studies such as Franklin, Drivonikou, Clifford, et al. (2008) was a reaction time measure. The findings of this chapter suggest that infants have knowledge of the blue-green colour category boundary independent of the type of task used, strengthening the overall claims made previously in the literature.

The present study constitutes strong evidence that infants display categorical perception of colour across the blue-green pairing, confirming past reports. In this chapter, infants systematically look to the within-category colour, demonstrating their implicit knowledge of, and perception of, a category boundary between the two colours. The findings also suggest that infants' knowledge of the category boundary extends to their preferential looking, consistently choosing a within-category colour over a between category colour when they are both presented on a coloured background. Infants possess a blue-green category boundary, and respond to it categorically, independent of their age and colour word knowledge.

Chapter 7

Visual Attention and Categorical Perception of Colour

7.1 Introduction

The previous chapter provides an analysis of infant categorical perception (CP) of colour, and describes behavioural evidence for categorical responses to colour. In the present chapter, I expand on that body of work, with another novel paradigm to examine categorical responses to colour in infants. On this occasion, the aim is to test how infant categorical responses to colour affect their ability to maintain attention.

As noted in the previous chapter, categorical perception entails responding differently to a stimulus when it straddles a category boundary. The results of the previous chapter, combined with past studies (Franklin, Clifford, et al., 2005; Franklin, Drivonikou, Clifford, et al., 2008; Franklin, Drivonikou, Bevis, et al., 2008) indicates that there is strong evidence that infants form colour categories before they acquire colour words (Davies & Franklin, 2002; Franklin, Clifford, et al., 2005), and that these categories have a biological root

(Skelton et al., 2017). Infant colour categories seem to be consistent across the green-blue boundary.

Despite this, there has been considerable debate about the "categorical" nature of categorical perception (CP) when it comes to colour in infancy (Hanley & Roberson, 2011; Roberson & Hanley, 2009). While it appears that categorical differences in colour are encoded differently in the brain to within-category differences in colour (Bird, Berens, Horner, & Franklin, 2014; Yang et al., 2016), it is apparent that there exists a distinct possibility of CP of colour in pre-linguistic infants being task-dependent. This could explain the variety of interpretations that exist about CP of colour in infants. The aim of the present task is to assess many of the same claims made in the previous chapter, with a different paradigm.

In the previous chapter, no effect of participant age or of colour word knowledge on the CP of colour was observed. This is consistent with reports of infants displaying CP of colour prior to their acquisition of colour words (Franklin, Drivonikou, Clifford, et al., 2008). However, there is a possibility with any finding of interest, that it is an artefact of the task used, or due to the nature of the data. Thus the findings of the previous chapter are still worthy of re-examination, especially with replication becoming increasingly important to the field of developmental psychology (Frank et al., 2017), a further task was employed to test these findings.

In the current study, a novel infant looking time procedure is utilised. This is a common paradigm used as learning trials in many categorisation studies, where infants are presented with one or more objects on a screen, and the total fixation time is used as a measure (e.g. Althaus & Plunkett, 2015a, 2015b; Plunkett et al., 2008; Younger, 1985). Similar paradigms are also used as habituation trials in infant colour category studies (Skelton et al., 2017).

Other similar measures (overall attention) are also used in examining infant speech perception (the Headturn Preference Procedure, Fernald, 1985; Johnson & Zamuner, 2010; Jusczyk & Aslin, 1995). To date, the Headturn Preference Procedure has been an effective manner of distinguishing infant CP of speech sounds. In this chapter, the paradigm used is similar to the looking time mechanism used in both categorisation trials and in the Headturn Preference Procedure, but a single dynamic visual stimulus is presented, and the measure is the length of time infants attend to the visual stimulus as a result of the dynamic changes. As opposed to the Headturn Preference Procedure, where the image is kept constant but the sound changes, here the sounds are kept constant but the image changes. In this way the experiment presented in this chapter operates almost as a headturn procedure in reverse – the outcome measurement is the same as the Headturn Preference Procedure, but of interest is the visual, rather than the auditory stimulus. The dynamic stimulus changes between two colours – either within one colour category, or across two different colour categories.

Therefore in this chapter, manipulations are tested against the infants ability to maintain attention to the stimulus throughout the trial. The primary manipulation in this experiment is whether the visual stimulus dynamically changes between-category or within-category, with a null hypothesis that participants look for the same duration of time for both stimulus types. As with the previous chapter, participant age and whether or not they comprehend colour words are of interest. For the purposes of this experiment, as with the last chapter, the hypothesis was that participant age and their knowledge of colour words will have an effect on the attention they pay to the different stimuli. Additionally, an attempt was made in this chapter to verify the results from the last chapter, namely that infants categorically prefer stimuli that remains within the category boundaries, and for that purpose the first trial for each participant was examined, the hypothesis being that infants would look more to the within-category stimulus.

7.2 Methods

7.2.1 Participants

Participants in this experiment were $N = 62$ infants, 32 aged 12 months and 30 aged 19 months. All participants took part in the experiment listed in Chapter 6, with the order counterbalanced. There were 2 participants who completed the experiment in Chapter 6 but were unwilling to participate in this experiment. Of those participants, $N = 7$ were removed due to lack of attention in the task (failure to complete one successful trial, see analysis section). Thus the final analysable sample was $N = 29$ 19 month-olds (mean age 19.27 months, S.D. age 0.47 months), and $N = 26$ 12 month-olds (mean age 12.24 months, S.D. age = 0.39 months).

Recruitment of all participants in this study took place in the same manner as in previous chapters, either through online sign-up, or through recruitment at the maternity ward of a local hospital.

7.2.2 Materials

In the present study, the materials remain the same as four of the stimuli used in Chapter 6 of this thesis. 4 *Munsell tiles* (Munsell Color Services, 2007) were simulated using CIE information provided in Witzel and Gegenfurtner (2011, supplementary material). These stimuli were created with a *Munsell value* of 5 and a *chroma* of 6 (Fairchild, 1998), and each stimulus was separated by 5 steps. There were two blue squares, one of which (Blue-green) was close to the category boundary, and two green squares, with one (Green-blue) close to the category boundary. Additionally, Illuminant C was created using the x and y values described by Fairchild (1998). The stimuli are described in Table 7.1.

Table 7.1: Chromaticity coordinates (CIE 1931) for the three conditions in the present study.

Stimulus	Type	Y	x	y
2.5B5/6	Blue	22.0	0.218	0.276
7.5BG5/6	Blue-green	23.8	0.229	0.305
2.5BG5/6	Green-blue	23.6	0.241	0.343
7.5G5/6	Green	23.1	0.260	0.368
Illuminant C		23.8	0.310	0.316

7.2.3 Design

The 4 stimuli listed above were combined to make three conditions.

- The *within-blue* condition alternated between presenting the Blue stimulus and the Blue-green stimulus.
- The *within-green* condition alternated between presenting the Green stimulus and the Green-blue stimulus.
- The *between* condition alternated between presenting the Blue-green stimulus and the Green-blue stimulus.

Each stimulus pair was shown alternating for 8 seconds, accompanied by an auditory stimulus consisting of three tones presented repeatedly in ascending order on loop throughout the trial, designed to maintain the infants' attention to the task. The same auditory stimuli were presented for all conditions, with the ascending tones not designed to be synchronous to the changing colours.

7.2.4 Procedure

After signing consent forms, and filling out the colour word supplement to the Oxford CDI (as in previous chapters) participants were sat in front of the monitor and eye-tracker, approximately 75cm from the screen. The same modified calibration procedure described in Chapter 6 was used for this study, with a solid black star moving across 9 different points, on a background approximating Illuminant C. At the beginning of each trial, an attention-getter was presented in the middle of the screen for 2000ms, to bring fixation back to the middle of the screen. The trial began immediately after, with the trial images appearing on the screen immediately. Stimuli were presented as a square (531 x 531 pixels) on the screen, which alternated between the two colours listed for each every 500ms. In order to avoid infants looking purely because it was in their gaze, stimuli were presented on one side of the screen, rather than in the middle. The background of the trial was kept to Illuminant C in order to standardize and limit habituation to any of the colours in the trial (Fairchild, 1998), and following the 8 seconds of trial, Illuminant C was shown for a minimum of 2 seconds before the start of the next trial. The monitor (1920 x 1080 pixels) used for presentation was calibrated using a datacolor Spyder 5 Elite calibration device.

Each participant saw 12 trials. In order to maintain maximal attention throughout the experiment, 6 trials were presented on one side of the screen, and then 6 trials were presented on the other side of the screen. Each block of 6 trials contained 2 trials from each condition, counterbalanced by starting stimulus. Trials within each block were presented in a random order. Infant looking data was collected by a Tobii TX300 eye-tracker operating at 120Hz, and trials were automated with a custom MATLAB script.

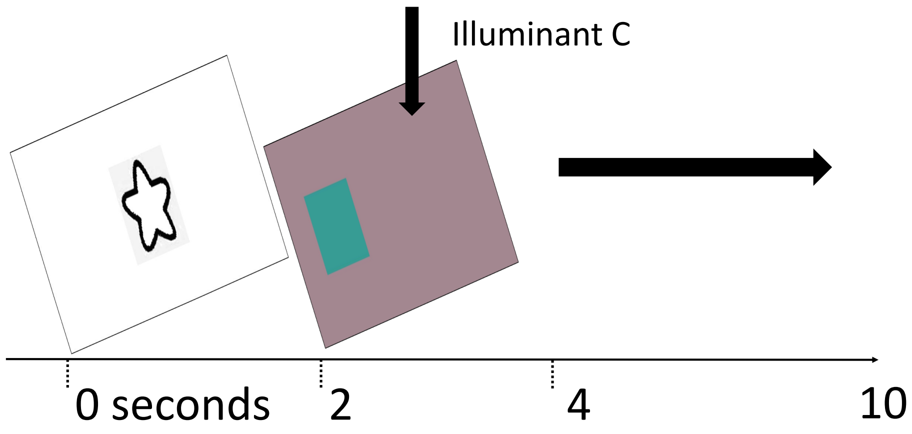

Figure 7.1: A typical trial. The patch of colour would alternate between 2 shades until the trial ends, every 500ms.

7.2.5 Analysis

Methods of analysis for infant looking time experiments differs greatly across experiments and design, and thus great care was taken in analysing this experiment. Data was first extracted with custom scripts in MATLAB, which passed the infant looking data through a fixation filter, where a fixation was defined as looking in the same location (with a certain amount of dispersion to allow for the instability of infant gaze) for a minimum of 100ms (for a discussion of fixation filters, and why they are debated in looking time analyses, see Wass, Smith, & Johnson, 2013).

As total looking time in each trial is the primary measurement of this experiment, as opposed to an indicator of category learning found in habituation trials of categorisation experiments, outliers were not removed as an *a priori* decision[1]. However, since it was still important to remove trials where partic-

[1]Treatment of outliers vary greatly across this type of analysis, but if participants who look less than a certain amount are excluded, then there can be no variability in looking time. Removing outliers also appears to have no major impact on the results. I am indebted to Dr K. Twomey for discussion of filtering and treatment of outliers.

ipants had no interest at all in the task, trials with greater than 90% trackloss were removed from the analysis. The 7 participants who were removed for failure to complete a trial, did not complete a trial in which trackloss was less than 90%. Data was analysed using the eyetrackingR package (Dink & Ferguson, 2015) and lme4 (D. Bates et al., 2017) in R.

7.3 Results

First, participant numbers for each trial were analysed. As the design of the experiment measures total attention paid to a stimulus, it can be expected that participants lose interest over the course of the experiment, such that at some point they stop paying attention to the stimulus altogether. In the first and second trials, 51 participants are paying attention; by the last trial only 20 participants attended to the trial. Despite the stimulus swapping sides of the screen in the second block to maintain participant interest for longer, most trials in the second block were completed by fewer than 50% of participants (Figure 7.2). As such, the following analyses contain data only from the first block of 6 trials.

7.3.1 Overall analyses

The proportion of time spent looking to the target for each participant and each trial was calculated, and then fit with a hierarchical multilevel linear model, using the lmer function from lme4 (D. Bates et al., 2017). The trial number (scaled) and the categorical variable of condition (*within-green*, *within-blue*, or *between*) was added to the model. An interaction effect was also included between the condition and the trial number. There was a random intercept of participant, to allow for each participant having a different measure

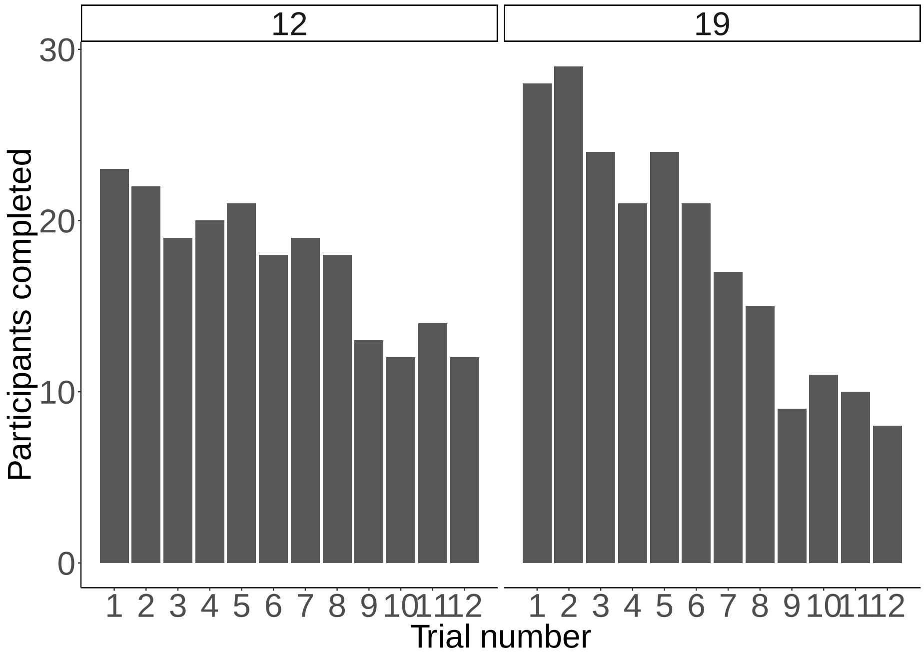

Figure 7.2: Numbers of participants who completed each trial, separated by age group.

of baseline attention.

Table 7.2: Model output of multilevel model on first 6 trials. Conditions are compared to the *between* condition.

	Estimate	Std.Error	DF	t-value	p-value
(Intercept)	0.523	0.045	253.59	11.55	<0.001
Trial	-0.352	0.128	234.73	-2.741	0.007
Within_Blue	-0.002	0.059	236.23	-0.04	0.968
Within_Green	0.083	0.061	235.47	1.363	0.174
Trial:Within_Blue	0.016	0.193	239.73	0.083	0.934
Trial:Within_Green	-0.341	0.193	235.77	-1.767	0.079

The model output (Table 7.2) demonstrates that there is strong evidence for an effect of trial, suggesting that participant attention diminished as the trials went on. There is very limited evidence for an interaction effect with trial numbers and condition, suggesting that the attention may have reduced at different rates depending on the condition that was being shown, but the evidence for that claim is not strong. There is no evidence for an overall difference between the looking proportions in each of the conditions based on this model.

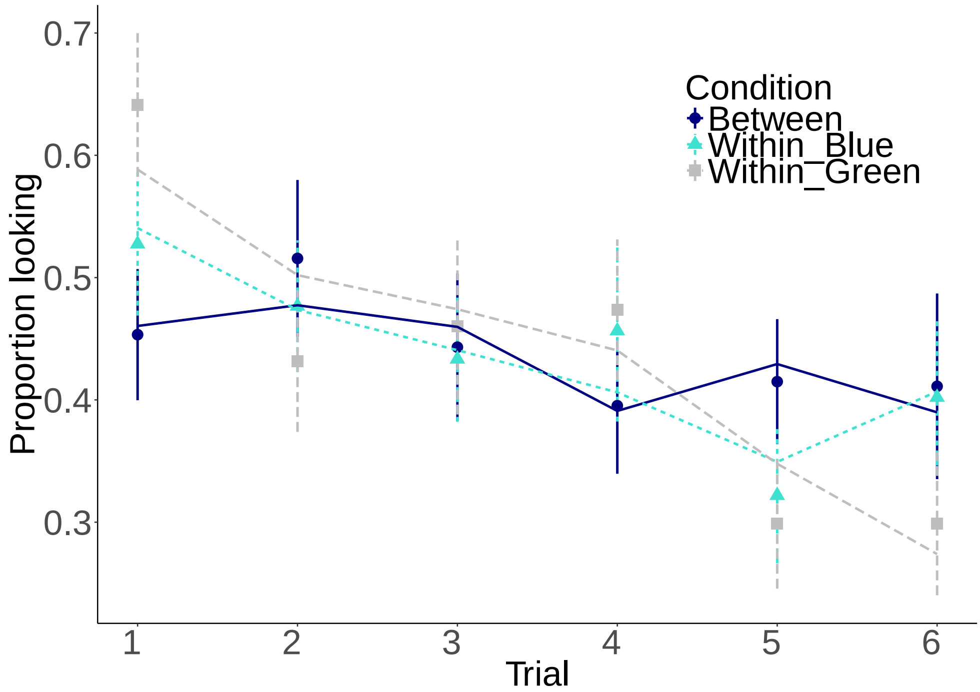

Figure 7.3: Model fit to overall looking proportion data in the first 6 trials. The bars indicate standard errors of the mean.

The model fit can be viewed in Figure 7.3. An examination of the figure lends some weight to the claims of differences in how the trial number interacts with the condition. In the *between* condition, the proportion of looking to the target is comparatively low in the first trial, but continues largely at that level throughout the first 6 trials. By contrast, the two *within* conditions start with much higher proportions of fixating the target, but decrease dramatically throughout the course of the trials. The high standard errors, however, reinforce the point that there is not enough evidence to be assured of that claim.

7.3.2 Individual effects

The data was aggregated for window analysis, by condition and participant ID. In addition, as the overall domain of interest for this study is within- versus between-category attention, the two conditions *within-blue* and *within-green* were aggregated.

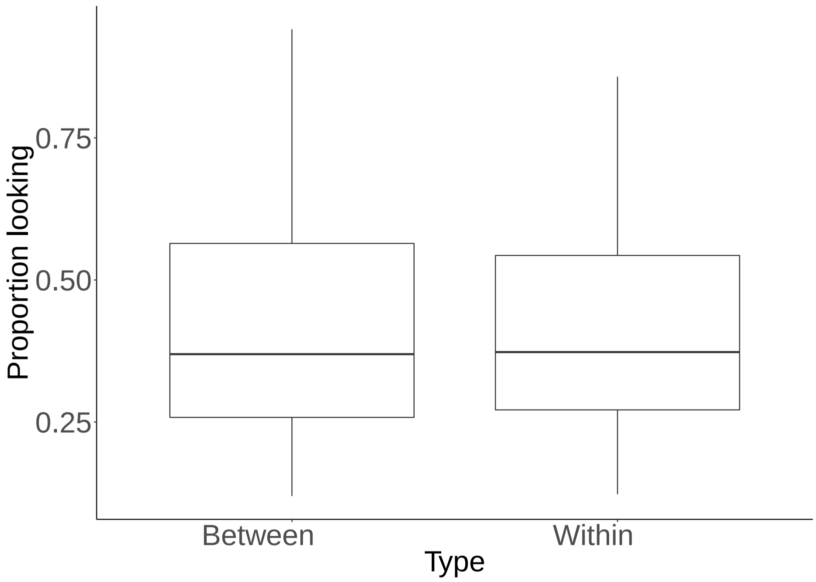

Figure 7.4: Looking proportions in each condition.

The overall proportion of time spent looking to the target in the within-participant conditions of *within* and *between* category colours (Figure 7.4) were compared by way of a paired t-test. The results indicate no evidence for an effect of condition on participant looking proportions ($t(51) = 0.167$, $p = 0.868$). Overall, as suggested by the multilevel model run previously, as well as Figure 7.4, there is no discriminable difference in looking between the within-category trials and the between-category trials.

The effect of participant age, as well as colour word knowledge (both between-participant effects), on looking times was also examined. Age was taken to be the participants' age group as tested in the study, either 19 months or 12 months. Colour word knowledge was determined in the same way as in Chapter 6, for consistency; that is participants who comprehended any one of either "blue" or "green" as determined by their caregivers in the colour word supplement to the Oxford CDI were classed as knowing the relevant colour terms, those who knew neither were classed as not knowing the colour terms.

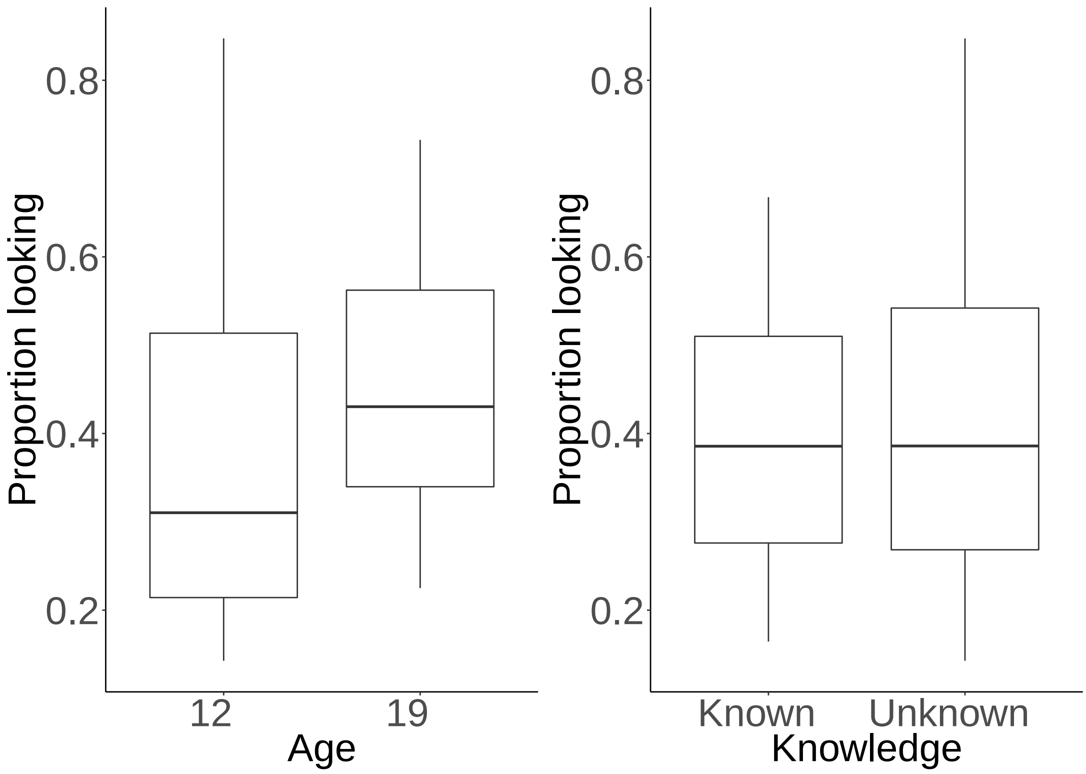

Figure 7.5: Looking proportions to the target in each age group, on the left panel, and separated by colour word knowledge, on the right panel.

The looking proportions of participants separated both by age and by colour word knowledge (Figure 7.5) were compared separately with Welch two-sample t-tests. There was no strong evidence for an effect of participant age ($t(43.969) = -1.465$, $p = 0.150$), nor was there any evidence for an effect of colour word knowledge on participant looking ($t(24.028) = -0.489$, $p = 0.629$). Neither the age of the participant nor their colour word knowledge had any effect on their overall proportions of looking to the target.

7.3.3 Looking patterns

Participant looking to the target was aggregated for each condition type (*within* or *between*) and modelled together with the participants' colour word knowledge to determine whether either their knowledge of the colour words or the trial type mediated when they looked at the target within each trial. Time in the trial was aggregated into 100ms time bins, as in previous chapters. As this study measures looking patterns within trials rather than overall measures, and is thus more sensitive, all valid trials were included, rather than just the first block. Colour word knowledge and condition type were both contrast-coded.

A binomial generalised linear model with quartic orthogonal polynomials on the time term was run on fixations in the target region of interest. Colour word knowledge and condition type were added as fixed effects, and all interaction effects were included. The intercept, as well as the slope of the condition type were allowed to vary by participant. The key difference between this model and those of previous chapters is that there is only one target, so the proportion instead measures looks in the target area over looks anywhere else. Because of that, in order to allow time for participants to launch a successful saccade, the model was run beginning at 1000ms and ending at 5000ms (much longer stretches the effectiveness of the model). The model was fit with the glmmPQL function from the package MASS (Ripley et al., 2017).

The output of the model can be examined in Table 7.3. There is no evidence for looking patterns differing with the effect of colour word knowledge, and there is only weak evidence for looking patterns differing by condition type. There is, however, very strong evidence for an interaction between colour word knowledge and condition type, as well as the interaction between that and each of the orthogonal time terms, suggesting that whether participants comprehend colour words may affect their looking patterns differently for each condition type.

The model fit in Figure 7.6 lends some weight to the evidence indicated in

Table 7.3: Output of model examining looking patterns within trials.

	Value	Std.Error	DF	t-value	p-value
(Intercept)	2.526	1.182	4100	2.137	0.033
ot1	24.005	13.382	4100	1.794	0.073
ot2	36.529	12.713	4100	2.873	0.004
ot3	14.167	6.243	4100	2.269	0.023
ot4	9.735	3.211	4100	3.032	0.002
Knowledge	2.822	2.364	53	1.194	0.238
Type	4.105	2.350	4100	1.747	0.081
ot1:Knowledge	36.556	26.764	4100	1.366	0.172
ot2:Knowledge	32.034	25.426	4100	1.260	0.208
ot3:Knowledge	17.625	12.486	4100	1.412	0.158
ot4:Knowledge	4.962	6.422	4100	0.773	0.440
ot1:Type	46.056	26.764	4100	1.721	0.085
ot2:Type	44.088	25.426	4100	1.734	0.083
ot3:Type	22.165	12.486	4100	1.775	0.076
ot4:Type	11.208	6.422	4100	1.745	0.081
Knowledge:Type	15.552	4.700	4100	3.309	0.001
ot1:Knowledge:Type	176.687	53.528	4100	3.301	0.001
ot2:Knowledge:Type	174.490	50.853	4100	3.431	0.001
ot3:Knowledge:Type	87.555	24.972	4100	3.506	0.001
ot4:Knowledge:Type	45.742	12.844	4100	3.561	<0.001

the table. The latter part of the trial is of particular interest, as it suggests that when participants know the colour words, their interest seems to diminish possibly faster than when they do not, particularly for the *between* condition type. Attention does diminish throughout the trial, as expected. There may be differences in looking patterns between the *between* and *within* condition types, but they are subtle, and possibly only mediated by whether participants understand colour words.

7.3.4 First trial analysis

Another aim of this chapter was to verify the claims of the previous chapter, with respect to the preference infants have for a within-category colour over a between-category colour. In order to see if that preference was observed in

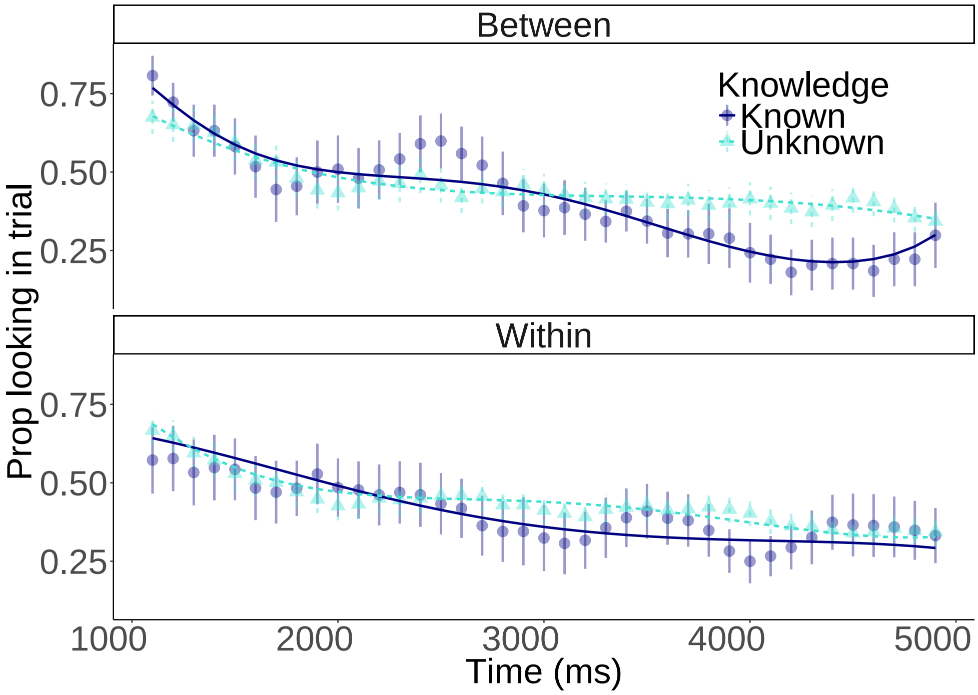

Figure 7.6: Looking proportions to the target over time within each trial type.

the present study, participant responses to only the first trial were examined. As seen from the previous model, looking in later trials may be mediated by a number of factors, thus the first trial is the only one that measures a pure preference.

Overall looking proportions were compared for each participant by way of a Wilcoxon rank sum test with continuity correction, due to the skew in proportions. As each participant only experienced one first trial, the condition type (*within* or *between*) was a between-subjects factor. There was some weak evidence that participants displayed a preference for the within-category colour pair over the between-category colour pair ($w(n = 51) = 219$, $p = 0.052$). The medians are different, as can be seen in Figure 7.7, but the variation in the data and the lack of power in the test means there is not strong enough evidence for this to be considered a meaningful difference.

The higher proportion of looks when the first trial was a within-category

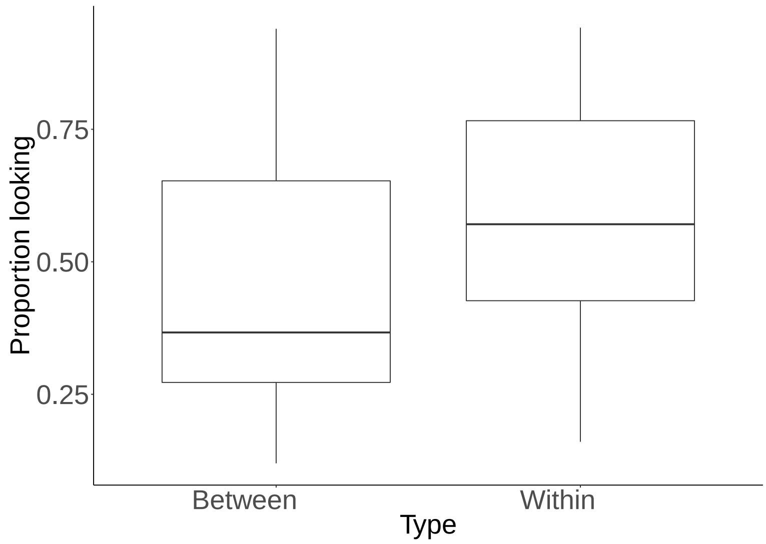

Figure 7.7: Looking proportions to the between-category pair or the within-category pair in the first trial.

pair lends some weight to the findings of the previous chapter, but the evidence in favour of this is not significant. Participants may display a preference to the within-category colour pair, but there is no substantial evidence for that from the present study.

7.4 Discussion

In the present chapter, infants were tested for their looking times and looking patterns when presented with a carefully-controlled set of stimuli that flashed between two colours. While the differences between both colours were the same, one set crossed a category boundary, while the other remained within the category boundary. The result of this study suggests that there was no evidence for a difference in overall looking proportions in these conditions, nor was there strong evidence for differences in looking proportions based on the participants' age or on their colour word knowledge, in this sense confirming the results of the previous chapter. Patterns of looking within trials do reveal the possibility of interaction effects between condition type and participants colour word knowledge, suggesting that these factors may combine for some mediation of looking proportions. Finally, the first trial was analysed for a raw preference for the within-category pair over the between-category pair, revealing a possible difference, but not a detectable one.

The lack of an overall difference between the two conditions is of interest, partly due to the lack of agreement with the previous chapter. In the present design, participants pay no more attention to a stimulus when it flashes across a category boundary than between one, suggesting no observation of the category boundary (Franklin, Clifford, et al., 2005; Franklin, Pilling, & Davies, 2005). The experimental conditions and set up remained identical to that of the previous chapter, where there was strong evidence of CP of colour in infants the same age. There is thus a strong possibility that infant CP of colour is task-dependent (Hanley & Roberson, 2008). Infants may show CP when measuring a reaction time (Franklin, Drivonikou, Clifford, et al., 2008; Franklin, Drivonikou, Bevis, et al., 2008), or when asked to choose between a between-category stimulus and a within-category stimulus on a coloured background, as in the previous chapter, but that CP may not extend to maintaining their attention for longer in a task where the stimulus changes between two colours.

Of course there are alternative possibilities. Franklin, Drivonikou, Clifford, et al. (2008) employed stimuli that were much farther apart (10 steps) perceptually than the ones used in the present experiment (the same were used in the *far* condition of Chapter 6). The same stimuli were not used for the present experiment, due to the lack of a balanced design, as there are only 3 hues that can be used when measuring that way, rather than the 4 required for the present study. Franklin, Drivonikou, Bevis, et al. (2008) used stimuli differing by 5 steps, such as those used here, but with an older age group. It is possible that due to the nature of the design, with the colours switching back and forth, the difference between the hues was not wide enough for participants to observe the category boundary, whereas when they were placed side by side, as in the previous chapter, the boundary is more noticeable. Wider ranging stimuli may have allowed to capture the effect of CP in participants of this age with the current design, but the current result demonstrates that with the same stimuli, this task does not capture CP of colour, suggesting some modicum of task-dependency, as has been previously alluded to (Hanley & Roberson, 2008; Roberson & Hanley, 2009).

The results of this chapter verify, in part, the results of the previous chapter. Neither participant age or colour word knowledge mediated participant looking proportions to the target, demonstrating again a finding that ran counter to expectations from previous studies (Franklin, Drivonikou, Bevis, et al., 2008). Based on the results of that study, it could be expected that colour word knowledge would change participant CP, but that was not the case in either this or the previous chapter.

In addition, an analysis of the first trial of each participant highlighted the possibility that there is some preference for the within-category colour pairing, that manifests itself early and disappears as participants experience more trials. The design of the study was such that there was not enough evidence to fully verify this claim. While the participant attention throughout the course of the experiment was not mediated by condition type, as was the original hypothesis, there may be a preference for the within-category pairing.

Overall looking time experiments are generally not as powerful as designs that test the looking proportions in each time bin, so that may be a factor in some of the marginal results in this study. Despite this, it has been proven to be a reliable design for infant CP of speech perception, and as such could be expected to be for the current experiment as well. Often total looking time experiments are not used with participants as old as 19 months, such as in the present chapter, but the comparably high proportion of participant looking in that age group (as well as the small drop out rate) demonstrates that, at least with dynamic stimuli, this is a viable paradigm for that age group. This finding, that dynamic stimuli can maintain interest even at 19 months, may be of interest to fields that regularly use looking time as a measure, such as the field of infant categorisation.

In conclusion, CP of colour in infancy appears to a degree to be task-dependent, and only affects certain aspects of development. In Chapter 6, the colour boundary along the blue-green line clearly has a strong effect on infant colour preferences, given a choice between the within-colour and the between-colour groups. In this chapter, by contrast, no difference is observed between the within-category colours and the between-category colours, suggesting that the colours that cross a category boundary are not inherently more, or less, interesting than those that remain within the colour boundary.

Chapter 8

The development of visual closure in infancy: Implications for colour vision testing

8.1 Introduction

In the previous experimental chapters, participants have been tested with eye-tracking paradigms to examine the impact of their colour word knowledge upon various aspects of their cognition. In doing so, participants were frequently asked to look at objects of varying colours, and respond appropriately. However, this method is contingent on participants' ability to perceive colours in a standard way. In previous chapters, as with other studies (e.g. Franklin, Drivonikou, Clifford, et al., 2008; Franklin, Drivonikou, Bevis, et al., 2008; Maule et al., 2017; Skelton et al., 2017), participants were excluded based on a high probability of colour vision problems owing to a family history of these problems. This method of excluding participants is standard, due to the lack of infant- and toddler-appropriate colour vision tests. One aim of this chapter is to examine the feasibility of a type of colour vision test in young toddlers.

A lack of certainty surrounding infant colour vision has led to a variety of techniques being used to overcome it. Booth and Waxman (2009), for example, remove participants with low rates of participation in the experiment. This can be problematic, as it might remove participants who could see the stimuli, but are uninterested in the task at hand. It also has the downside of inadvertently removing participants who support the null hypothesis, making it a problematic addition to researcher degrees of freedom (Simmons, Nelson, & Simonsohn, 2011; Wicherts et al., 2016). Other authors (e.g. Clifford et al., 2009, 2012; Franklin & Davies, 2004; Franklin, Drivonikou, Clifford, et al., 2008; Franklin, Sowden, et al., 2010) apply a different technique wherein infants are excluded based on a family history of colour vision deficiencies. This method is more robust, but still is reliant on accurate family histories being given to the researcher, and on other family members being aware that they may have a colour vision deficiency. However, there are no appropriate tests at that age, and toddlers find colour vision tests extremely difficult to complete (Franklin, Clifford, et al., 2005).

One potential method for testing colour vision would be in a modified Ishihara Test (Ishihara, 1917), using pseudoisochromatic plates such as those in Figure 8.1. In an Ishihara test, a shape or number is hidden in an image containing circles of varying sizes, and the image is only discriminable by hue. The plates vary in their colours in order to test different colour deficiencies (although the original plates do not test for tritan deficiencies). However, there are two main problems involved with the implementation of such a test for infants. First, many of these tasks require a response, asking the participant to say what they saw, or to at least trace a figure of what they see; an

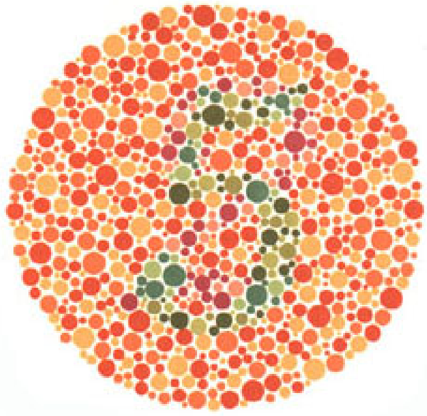

Figure 8.1: A sample of a standard plate used in an Ishihara plate test. From http://www.color-blindness.com/

infant is highly unlikely to be able to complete this task before three years of age. Second, the implementation of such a test is contingent on infants being able to recognise images presented in this format. It is this feature of infant perception that the current chapter examines.

While infant colour vision is largely in place by around 4 months of age (Bornstein, 1985; Franklin & Davies, 2004; Peeples & Teller, 1970, 1978; Teller, 1998; Teller et al., 2004), the rest of infant perception develops at varying rates (Atkinson, Braddick, & Moar, 1977; Atkinson, Braddick, & French, 1979; Atkinson & Braddick, 1998; Atkinson, 2002; Braddick & Atkinson, 2011; Harris, Atkinson, & Braddick, 1976). In order to complete a task like the Ishihara Plate test, infants need to be able to fill in the outline of the image, a perceptual property known as *visual closure*: a key skill in the development of reading abilities, among others (Kavale, 1982). Visual closure develops throughout early childhood as children learn to read (Mee Bell, McCallum, & Cox, 2003), but it is unknown when infant visual closure will be developed sufficiently for such a task.

In order to address these questions, in the current chapter a modified Ishihara Plate test is employed, but rather than requiring a response from participants, two plates are presented in a forced-choice preferential looking task (Teller, 1979), encouraging infants only *look* to the correct plate. In order to examine the stability of the development of visual closure in the second year of life, infants at two different ages, 16 and 19 months were recruited for the task. The aim of the present study was to explore the feasibility for such a task to test colour vision at these ages, with a two-fold hypothesis: First, if stimuli such as those proposed below are suitable, participants should systematically look to the target; Second, if participants' visual closure is not yet developed to the point that it is sufficient to complete this task, then they should fixate the images later at 16 months than they do at 19 months if it is only slightly behind that of 19 month-olds, or not at all if it is insufficient for the task. Given visual closure is still developing during the ages of learning to read, as discussed above, it is expected that the younger group will fixate later than the older group, as it will take them longer to identify the target.

8.2 Experiment 1

8.2.1 Methods

8.2.1.1 Participants

A total of 115 participants participated in this task, N = 58 16 month-olds (mean age 16.22 months, S.D. 0.46 months), and N = 57 19 month-olds (mean age 19.47 months, S.D. 0.66 months). All participants were recruited either online or at the local maternity ward. None of the participants demonstrated any signs of a serious vision deficiency, either to parents or to the researcher. An additional 5 (2 16 month-old) participants were removed due to failure to calibrate, or due to no successful trials being completed.

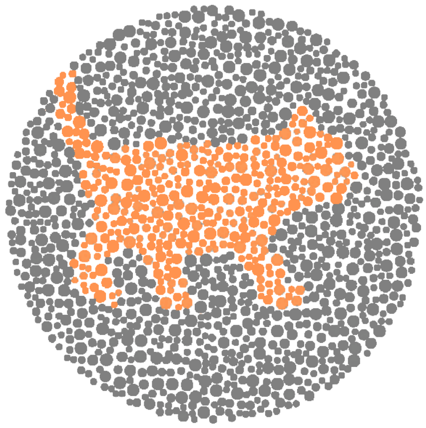

Figure 8.2: An example plate, depicting a cat, used in the study.

8.2.1.2 Visual stimuli design

Visual stimuli were created based on a set of stimuli created by Aisch (2011). Stimuli consisted of common everyday images that would be familiar to young toddlers from everyday life (and were commonly known based on early Oxford CDI estimates; Hamilton et al., 2000). These stimuli were presented in a format similar to the Ishihara plates. Unlike those created by Aisch, the current stimuli were instead coloured with hues as close as possible to the first Ishihara plate, that should be visible to anyone with better than monochromatic vision, in grey (CIE L*ab = 53.585, 0.000, 0.000) and orange (CIE L*ab = 61.349, 35.270, 52.523). Recolouring was performed in the Gnu Image Manipulation Program (www.gimp.org). Stimuli were designed so that the image would appear in orange, and the background would appear in grey, thus being visible to any individual with colour vision. 7 plates were created for this sample, containing an apple, a bottle, a butterfly, a cat, a dog, a duck and a tree. An example plate is included as Figure 8.2.

8.2.1.3 Plate selection

All seven plates were rated by 36 (20 female) adults online. None of the adults reported any vision deficits. Those selecting the stimuli were asked to input as text what they thought the image represented in each plate was, and how clear that image was (on a scale of 0 - 100). 1 adult participant of those was excluded for misunderstanding the task, and only rating clarity up to 20. It was decided that any plate that was misidentified at least 25% of the time, or any plate with a clarity rating below 0.75 would be excluded. This 75-75 rule provided a simple heuristic for plate selection, and ensured that adult participants were satisfied that the plates were as designed, as well as being sufficiently clear. For each stimulus, over 80% of participants correctly identified the stimulus. For 4 of the plates (butterfly, dog, cat, and duck), all participants correctly identified the intended stimulus. However, the mean clarity rating of bottle was 0.68, meaning only the remaining 6 stimuli were included in the experiment. The ratings of all remaining stimuli are included in Table 8.1.

Table 8.1: Stimulus ratings for all stimuli included in the experiment. Identification is the proportion of adults who correctly identified the stimulus, while the clarity score is the rating out of 100, divided by 100.

Stimulus	Identification	Clarity
Apple	0.89	0.79
Butterfly	1	0.87
Cat	1	0.85
Dog	1	0.90
Duck	1	0.82
Tree	0.83	0.83

8.2.1.4 Auditory stimuli

All auditory stimuli were recorded by a female native speaker of Southern British English, speaking slowly and clearly, in an infant-directed manner.

Stimuli consisted of sentences of the format: *"look at the xxx,"* where xxx was the name of each of the stimuli contained in the experiment. Recordings were cleaned and edited in audacity (www.audacityteam.com).

8.2.1.5 Procedure

Before arriving at the lab, participants' caregivers were asked to fill out the Oxford CDI (Hamilton et al., 2000) online. Participants arrived at the lab, and then had a short play session to acclimatise to surroundings, when consent forms were filled out by the caregiver. Participants were then seated on their caregivers laps, roughly 75cm from the eye-tracker and associated video screen. A nine-point calibration sequence was performed until at least 7 of the calibration points were calibrated correctly, after which the experimental trials commenced. In each trial, participants saw two stimuli, presented one on either side of the screen, on a white background, and were prompted to look at one by the auditory stimulus. In each trial, participants saw an attractive attention-getter for 2 seconds, after which point the stimuli immediately appeared on the screen. Following that, the auditory stimulus began, such that the target word appeared exactly 2 seconds after the visual stimuli onset. The trial continued for another 5 seconds after target word onset. Each participant saw a total of 6 experimental trials; each participant was randomly assigned to one of 5 experimental lists, that differed in their pairings of each stimulus. Trials were counterbalanced within subjects, such that each pair appeared twice, with each object in the pairing appearing as target and distractor. The timecourse of a sample trial can be seen in Figure 8.3. Trials were presented using a custom script in MATLAB, and infant gaze data was recorded using a Tobii TX300 eye-tracker, recording at 120 Hz.

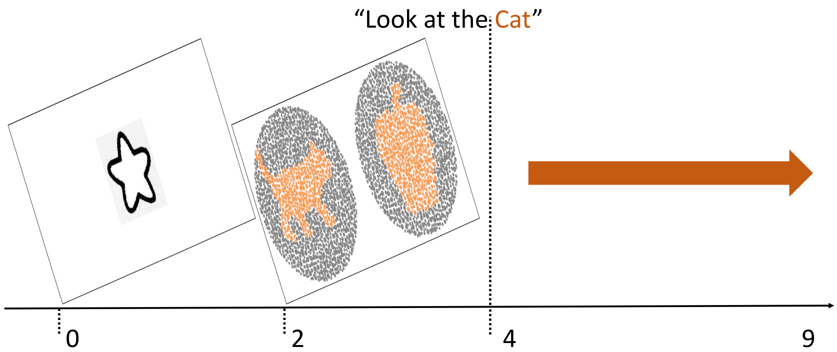

Figure 8.3: The timecourse of a typical trial.

8.2.1.6 Analysis

Fixations for each participant were extracted using a custom MATLAB script, where a fixation was defined as stable gaze in one area of the screen for a minimum of 100ms, allowing for a certain amount of instability in the infant gaze. Data was cleaned so that only trials where participants looked to either of the images for 50% of the total trial time were included. Only trials where parents indicated that the participant comprehended the label of the target object according to the Oxford CDI were included in the analysis. This exclusion, as well as a number of participants who did not fill out the Oxford CDI at all, meant that the final analysis included data from $N = 87$ participants (44 16 month-olds).

Data was then analysed in R, using the glmmPQL function from the package MASS (Ripley et al., 2017), and eyetrackingR (Dink & Ferguson, 2015). Fixations were aggregated for each participant and each 100ms time bin, and then modelled using a generalised linear model, as in the previous chapters.

8.2.2 Results

As with previous chapters, the first three seconds of the trial following target word onset were fitted with a binomial generalised linear model, which took cubic orthogonal polynomial time (elapsed after target word onset) terms and the difference-coded measure of age (16 or 19 months) as independent variables. On this occasion, because of the interest in the effect of timing, only fixations from the 300ms mark onwards were used, as anything before then is unlikely to be a reaction to the target word (Bergelson & Aslin, 2017). The model also took the interaction effects between age and each of the three time terms as independent variables. Dependent variables were target looking versus distractor looking in each time bin. The model was hierarchical, allowing the intercept to vary by participant.

The model output (Table 8.2) indicates no main effect of age, or of any of the polynomial time terms. However, there was relatively strong evidence for an effect of the interaction between age and the linear time term, and age and the cubic time term. This indicates the possibility that the patterns of looking between the two age groups followed a different time course.

Table 8.2: Output of generalised linear model.

	Value	Std.Error	DF	t-value	p-value
(Intercept)	-0.099	0.196	2145	-0.506	0.613
ot1	-2.472	6.028	2145	-0.410	0.682
ot2	-2.335	1.488	2145	-1.569	0.117
ot3	-1.955	2.984	2145	-0.655	0.513
Age	-0.648	0.393	85	-1.649	0.103
ot1:Age	26.731	12.057	2145	2.217	0.027
ot2:Age	-4.657	2.975	2145	-1.565	0.118
ot3:Age	14.577	5.968	2145	2.443	0.015

From examining the top panel of Figure 8.4, it can be seen that while the 19 month-old participants tend to fixate the target objects around the 1000ms mark, the peak in target looking for the 16 month-olds is much later, around

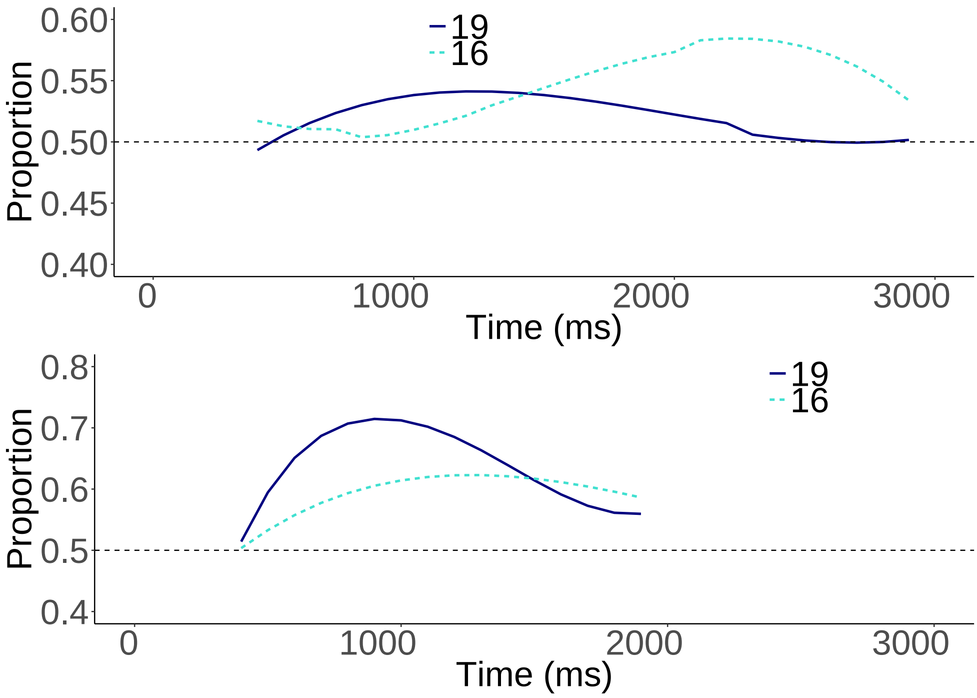

Figure 8.4: The fitted models from the two experiments, showing proportions of looking to the target, with Experiment 1 in the top panel, and Experiment 2 in the bottom panel.

the 2000ms mark. Both groups of participants do systematically fixate the target, however, demonstrating some level of efficacy for the task at hand. The difference in looking patterns may indicate an additional level of difficulty for the younger participants.

It appears that both age groups can successfully locate the target image, but it is unknown whether the difference is symptomatic of a general difference in processing and fixation speed between the two age groups (Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998), or specific to the stimuli utilised here. Experiment 2 examines these questions under similar conditions to those employed in Experiment 1.

8.3 Experiment 2

8.3.1 Methods

8.3.1.1 Participants

A subset of the participants who completed Experiment 1 (N = 31 19 month-olds) and (N = 30 16 month-olds) also completed a control condition. These participants were chosen as the subset by being the last participants to partake in this experiment.

8.3.1.2 Stimuli

The plates used in Experiment 1 were replaced with photos of each of the objects used in the experiment.

8.3.1.3 Procedure

The procedure for Experiment 2 was almost identical to that of Experiment 1, but the participants who partook in the experiments were given a short break before taking part in Experiment 2. Trials for this experiment were cut short to allow for more trials, leaving only 2 seconds after target word onset, as it was expected that participants would have less difficulty in locating the target image when presented with real photographs. Participants saw 30 trials in a randomised order in this experiment for improved statistical power (due to the smaller N); due to the increased number of trials, all participants saw the same experimental list, which paired each object with each other object.

8.3.1.4 Analysis

Data for this experiment was analysed in the same way as Experiment 1. After excluding participants who did not know the target words according to the

Oxford CDI, a total of 47 participants (27 16 month-olds) remained.

8.3.2 Results

An identical model was fit to the one specified in the previous experiment. From Table 8.3, it is apparent that there is strong evidence for all effects included in the model, suggesting overall differences between the two age groups in the amount of looking to the target, as well as the time course for doing so. The model fit can be seen in the bottom panel of Figure 8.4. The peak looking period for 19 months is slightly earlier than in the previous experiment, representative of the decreased difficulty participants in this age group have when visually searching for the target. Proportions of looking to the target are also much stronger in Experiment 2 than in Experiment 1 for this age group.

Table 8.3: Model output for Experiment 2.

	Value	Std.Error	DF	t-value	p-value
(Intercept)	-7.573	1.248	699	-6.068	<0.001
ot1	59.658	9.479	699	6.294	<0.001
ot2	-32.412	5.532	699	-5.859	<0.001
ot3	7.167	1.551	699	4.622	<0.001
Age	-10.058	2.496	45	-4.029	<0.001
ot1:Age	77.573	18.958	699	4.092	<0.001
ot2:Age	-45.316	11.064	699	-4.096	<0.001
ot3:Age	11.358	3.102	699	3.662	<0.001

At 16 months there is a considerable difference in looking pattern between the first and second experiments. In Experiment 2, the peak in looking comes over 1000ms earlier than the peak in Experiment 1, with also much less of a difference between the two age groups, suggesting that for 16 month-olds, locating the target was much more difficult in Experiment 1 than in Experiment 2. It is also possible that infants respond faster in this Experiment due to the faster trial times.

8.4 Discussion

In this chapter, two experiments were run. In the first, infants were tested with modified Ishihara-style plates, depicting outlines of familiar images, and were prompted to look at one of them. In the second, the plates were replaced with real images of the same objects. While both age groups looked to the target at around the 1000ms mark in Experiment 2, in Experiment 1 only the 19 month-old age group performed in the same manner. In Experiment 1, the 16 month-old age group did show a pattern of looking to the target, but much later, after the 2000ms mark.

These results raise several important considerations. First, the results present strong evidence for the viability of a test such as the one presented for examining colour vision in infants. In Experiment 1, both groups of infants successfully locate the target, suggesting that images displayed in that format are visible to infants at these ages. The ages tested in the present study are younger than the lower limits of existing colour vision tests, displaying the potential of a test such as the one described in this chapter to be utilised as a colour vision test for younger toddlers. Fixating an outline of an image being presented in pseudoisochromatic plates is not an insurmountable challenge to toddlers as young as 16 months. Additionally, presenting the task as an Intermodal Preferential Looking task allows reliance on the participants ability to fixate the image, rather than describe the image, provided the object label is known to the participant.

The second consideration raised by the results of this chapter is that while it appears that a test such as this would work for toddlers as young as 16 months, it is unlikely to work with age groups *younger* than 16 months. While the 16 month-old age group were able to locate the target image, they did so much slower than the 19 month-old age group, and much slower than their own performance in Experiment 2. This suggests that while the younger participants were able to fixate the target, they did so with a considerable amount of difficulty. In addition, it is probable that many of the object labels will not be known by most participants younger than 16 months, severely limiting the effectiveness of the task.

The difficulty experienced by the 16 month-olds in Experiment 1 is likely attributable to the stage of development of their visual closure, as expected by the experimental hypothesis. Based on these findings, it is likely that visual closure is a perceptual property that develops slowly, much slower that, for example, colour perception (Bornstein et al., 1976; Bornstein, 1985; Atkinson, 2002). A partially-occluded image or an image with an incomplete outline is likely to present varying levels of difficulty for the individual, depending on the developmental stage of the individual in question.

An alternative conclusion would be that infants at 16 months have a slower processing speed than 19 month-olds, which may interact with the additional difficulty of the occluded image to cause the slower processing speed in Experiment 1. The results of Experiment 2 suggest that the processing speed for 16 month-olds is not dramatically slower than for 19 month-olds, but the possibility of the interaction, rather than a less developed visual closure, which causes the differences between Experiment 1 and Experiment 2 in 16 month-olds.

In conclusion, the present chapter has examined the visual closure of infants at both 16 and 19 months with a view to a prototype of a colour vision test for these age groups. The results demonstrate that it may be an effective task as young as 16 months, younger than any currently available colour vision task. However, visual closure is still under development at these ages, providing further insight into infant perceptual development.

Chapter 9

General Discussion and Conclusions

9.1 Introduction

In the previous chapters, there has been systematic analysis into two areas of children's learning about colours: colour *word* learning, and colour *category* learning. The former, assessed through experiments in Chapters 2 - 5, have examined how infants learn colour words, and how the colour words affect their ability to process colour-related concepts. The latter, assessed broadly in Chapters 6 - 8, examines how infants develop colour categories, and assesses perception more widely in order to ascertain the feasibility of colour vision testing in infants prior to their second birthday. In the present chapter, the main findings from the previous chapters are discussed, and the impact on the field is assessed.

In Chapter 1, several overarching fields of enquiry were postulated, namely: the timing of colour word learning; constraints on colour word learning; the role of colour words in mediating visual attention; categorical perception of colour; and colour vision testing in infants. These overarching fields of enquiry have formed the basis of the investigation which has been presented thus far in this thesis, and shall be addressed in this chapter.

9.2 Timing of colour word learning

In Chapter 2, Bayesian models on the Oxford CDI (Hamilton et al., 2000) and MB-CDIs (Frank et al., 2016) reveal that, although languages differ greatly in the timing and the order of colour word learning, colour word learning in many cases is well under way by 2 years of age. Additionally, Chapter 2 demonstrates that in many languages, the majority of infants can be expected to comprehend colour words by 30 months, and that frequency of input is a big factor in predicting onset of colour word production (Goodman et al., 2008; Yurovsky et al., 2015). This finding suggests that there are no universal timings for colour word learning, but that it occurs much earlier than reported previously (Andrick & Tager-Flusberg, 1986; Mervis et al., 1995; Pitchford & Mullen, 2003; Sandhofer & Smith, 1999; Soja, 1994).

In Chapter 3, an eye-tracking task with nearly 150 British infants and toddlers demonstrated that, at least in the case of British English, colour word comprehension occurs on a basic level as early as 19 months. While this only includes examples from the centre of each colour category, the results from Chapter 3 taken in conjunction with those of Chapter 2, demonstrate that the process of colour word learning begins much earlier than had been hypothesised, and that there are certainly no cortical obstructions to learning colour names at that age (Bornstein, 1985). Chapter 3 additionally demonstrates that infants' expectations for colour words are for them to appear in the adjectival position, suggesting that at some level, infants expect colour words to modify nouns (Mintz & Gleitman, 2002).

The literature to date suggests that colour word learning in English in many cases occurs in the third year of life, due to the difficulties involved with mapping colour words to colours (Andrick & Tager-Flusberg, 1986; Franklin, 2006). The results of Chapters 2 and 3 are in direct contrast to this suggestion. These chapters strongly suggest that the timeline posited is in line with that of other classes of words. Fenson et al. (1994) demonstrates that many infants can be expected to comprehend around 150 - 200 terms by 16 months of age (similar but slightly lower numbers are demonstrated in British English, Hamilton et al., 2000; Mayor & Plunkett, 2014). This suggests that, rather than being learned late, as often expected (Pitchford & Mullen, 2003), colour words are learned more or less in line with other words, with the possibility of some delay due to the relative abstractness of the category.

There are two possibilities that arise from this investigation: one is that the age of colour word learning has been incrementally dropping over time (Franklin, 2006), perhaps due to some external influence such as greater exposure to the terms due to an increase in coloured plastic toys; the second is that the measures used in this thesis allow for a more sensitive measurement of colour word learning, capturing an earlier stage of comprehension than past studies. There is some merit to the first suggestion, as with an increase in coloured toys and household goods, parents are more likely to differentiate similar items by colour. As demonstrated in Chapter 2 of this thesis, the increased exposure in that case would likely lead to earlier colour word learning. It is also important to note in this case that the languages examined in Chapter 2 are all from industrialised countries. However, in English and in many other of the languages examined, industrialisation occurred a considerable distance in the past, and it is post-industrialisation that the proliferation of coloured plastic goods could be expected. Thus if this theory were correct, the previous investigations into this topic (e.g. Pitchford & Mullen, 2001, 2002; Sandhofer & Smith, 1999; Soja, 1994) performed in the last 20 years would also be expected to show early colour word learning. However, a comparative study on colour word learning in industrialised countries versus pre-industrialised countries would be more conclusive in examining the role industrialisation plays in diminishing the age of colour word learning.

It is more likely that the results of Chapters 2 and 3 of this thesis constitute a more sensitive measurement of colour word comprehension than previous studies. Chapter 2 employs parental report as a primary measure, which is likely to capture a very basic comprehension, before it would necessarily be born out in laboratory-based tests (Styles & Plunkett, 2009b). Chapter 3 tests comprehension of only very early examples of basic colour categories, with no testing of colours peripheral to the category, and no later-learned terms, such as *"brown."*

That these two chapters find early colour word learning with more sensitive measures, demonstrates considerable support for the slow-mapping hypothesis of Wagner et al. (2013). The very earliest stages of colour word learning are captures in these studies, but that comprehension continues to be refined over time, even after production of the term begins. Colour word learning thus begins much earlier than previous studies have reported – as early as 19 months, but develops slowly over time.

In Chapter 2, the internal timing of colour word acquisition in multiple languages is also discussed. Contrary to the Berlin and Kay (1969) expectations that the infant colour word learning order would follow the same order as the order in which colour words develop in a language, due to neural and perceptual mechanisms; the results of Chapter 2 suggest that there is no universal order for colour word learning. While pre-linguistic infant colour categories may exist on a foundation that is biological and based on retino-geniculate pathways, colour words are learned in different orders in different languages, due largely to the frequency with which they are heard by developing infants.

9.3 Constraints on colour word learning

Colour word learning begins as early as 19 months in English, and marginally earlier or later in some other languages, as established in Chapters 2 and 3 of this thesis. Research into infant colour word learning to date, has often focussed on why infants find colour word learning so difficult (Franklin, 2006; Pitchford & Mullen, 2002; Sandhofer & Smith, 1999). In light of the findings in this thesis, the constraints, and the difficulty, are worthy of reassessment.

As discussed in Chapter 1, past research has debated two types of mechanisms causing colour words to be learned late. On one hand, varying perspectives on children's cortical and cognitive development have indicated that the difficulties could stem from either cognitive or conceptual difficulties in learning colour words (Andrick & Tager-Flusberg, 1986; Bornstein, 1985; Kowalski & Zimiles, 2006; Shatz et al., 1996). On the other hand, attempts to compare colour word learning with other classes of words has raised the possibility of linguistic constraints on colour word learning, with the primary difficulty being mapping words onto categories, rather than a conceptual difficulty (O'Hanlon & Roberson, 2006; Pitchford & Mullen, 2001; Sandhofer & Smith, 1999; Soja, 1994). There is considerable evidence for both positions, and it is likely that both conceptual and linguistic factors effect the onset of colour word learning.

The results demonstrated in Chapters 2 and 3 suggest that the real difficulty may not be in learning the term, or mapping the term onto basic exemplars of the colour category. The findings in Chapter 2 suggest that colour words are learned much earlier than previously thought, albeit with great variability in different languages. Chapter 3 further demonstrates that colour words are comprehended as early as 19 months, in British infants learning English as their first language. How, then, to reconcile these claims with those suggesting that young toddlers have great difficulty learning colour words?

It is likely that there is a constraint; however the constraint is not so much on their ability to map colour words correctly, but rather on developing an *adult-like understanding* of colour words. It has been shown previously that children who still produce colour words incorrectly at times rarely do so for the category centroids, but that the problem is overextending the term to include adjacent colours (Wagner et al., 2013, 2014). Similarly, in Chapter 4 of this thesis, it is demonstrated that toddlers only learn "*dark*" and "*light*" in the context of colour words comparatively late, suggesting that comprehension of atypical examples of colour words develops later than typical examples. That toddlers produce colour words in response to "what colour is this?" but do so haphazardly (Pitchford & Mullen, 2003), suggests not that they have failed to comprehend colour words, but rather that they are yet to identify correctly the boundaries of the colour terms that they have already learned.

Thus the problem for infants attempting to learn colour words is more subtle than has been previously alluded to; colour words still have to be abstracted from context (the *gavagai* problem, Quine, 1960), they still have to take note of colour as a relevant domain of interest (Franklin, 2006), and they still have to perform a series of mappings to comprehend colour words (Sandhofer & Smith, 1999, 2001). That infants can do so by around 19 months is a testament to their formidable language-learning abilities. However, the greatest constraint on toddlers developing an adult-like comprehension of colour words appears to be in correctly identifying the boundaries of each colour term, and in understanding its limits.

9.4 Mediating visual attention with colour words

In Chapter 5, toddlers at 19 months and 24 months were tested to see whether hearing a label for an object with a typical colour (such as "*strawberry*") would elicit attention to an object of the same colour (such as a red plate). This is a task that has been shown to be effective with adults (Huettig & Altmann, 2011). The question addressed in Chapter 5, is whether a comprehension of colour words is necessary to attend to the colour-matched distractor in that instance. In Chapter 5, two possible explanations are put forward: the *direct activation* account; and the *label-mediated activation* account.

Previous studies with 36 month-olds and 24 month-olds (Johnson et al., 2011; Johnson & Huettig, 2011), have shown that toddlers at these ages do systematically attend to the colour-matched distractor. At 24 months, it was previously shown that those toddlers had no knowledge of colour words, and as such, the study constituted strong support for *direct activation*.

The results of Chapter 5 challenge that account. Based on the findings of Chapters 2 and 3, it was found that there was a strong chance that 24 month-olds would know basic colour words, at least for typical exemplars of each colour word. Thus the main point of difference between Chapter 5 of this thesis and the previous results by Johnson et al. (2011) is the inclusion of a 19 month-old participant group in addition to the 24 month-olds. The younger age group would be less likely to comprehend colour words, making an interesting comparison with the older age group. The findings showed that the 19 month-olds failed to systematically attend to the colour-matched distractor, but in addition, that only participants who were believed to comprehend colour words showed a pattern of systematic attention to the colour-matched distractor.

Based on these findings, I conclude that both the *direct* and *label-mediated* explanations are plausible, and may work in concert to allow attention to the colour-matched distractor. It is possible that just as semantic or taxonomic information can be rapidly primed (Arias-Trejo & Plunkett, 2009, 2013; Styles & Plunkett, 2009a), it is possible that object properties, such as colour, may equally be able to be primed. Thus in the example given above, the word "*red*" may be primed by the word "*strawberry*." Previous work suggests that there is strong evidence that colour concepts are primed by typically-coloured objects, albeit slower than semantically-related words (Mani et al., 2013), but further work is required to determine whether a task such as that of Mani et al. requires a knowledge of colour labels for successful colour-related priming.

9.5 Categorical perception of colour

Chapters 6 and 7 of this thesis employ contrasting methodology to examine the phenomenon of Categorical Perception (CP) of colour in infants. In Chapter 6, participants are presented with a blue and a green stimulus on either side of a screen with either a blue or a green background, testing participants' CP of colour with respect to systematic choices of either a within- or between-category colour, as compared to the background. In Chapter 7, participants are presented with a square that flashes between two colours, either within- or between-category, in order to test CP of colour with respect to their attention to the stimulus.

The results of the two chapters are contrasting in some respects, and in agreement in others. In Chapter 6, a consistent effect is found, where participants constantly look to the within-category colour over the between-category colour. By contrast, Chapter 7 finds no difference in attention to the within- and between-category stimuli. Both studies use a measure of gaze, rather than reaction times, as a measurement (see A. M. Brown et al., 2011, for criticisms of RTs in this type of experiment), and a consistent theory of CP in infants would expect the same pattern in both cases. Why do these two studies find such a disparity in results?

One possible explanation is that there are differing mechanisms at play in the two types of task. Previous reaction time experiments have demonstrated that pre-linguistic infants look to a between-category stimulus faster than a within-category stimulus (Franklin, Drivonikou, Clifford, et al., 2008). Based on that pattern, one could assume that there would be a systematic pattern by which infants look to one colour first, and then the other in the task in Chapter 6, but one that would not affect performance in the task in Chapter 7. However, if that were the sole mechanism at play, the systematic pattern of looking in Chapter 6 should be first to the between-category colour, not the within-category colour as demonstrated in that chapter.

Another possible explanation for the disparity between the results of the two tasks, is that the measurements employed in Chapter 7 lack the statistical power of the measurements in Chapter 6. It is undoubtedly true that looking-time paradigms, where aggregates of time spent looking to the target are taken across a whole trial, are lacking in sensitivity compared to growth curve analysis (Barr, 2008, 2013; Barr, Levy, Scheepers, & Tily, 2013; Mirman, 2014). However, it should be recalled that in Chapter 7, a growth curve across trials was also performed, and still demonstrated no difference between the within- and between-category conditions.

A further possible reason for the disparity in results between the two chapters could be related to habituating to the task. It is possible that while there may initially be a difference in the attention to within- and between-category conditions in Chapter 7, that element of surprise or arousal that is required to maintain interest over the duration of the trials is lost at an equal rate, due to participants simply not finding the task interesting enough to demonstrate the subtle differences being employed. Looking time experiments are usually run with much younger infants than 19 months, and it may be that their relative lack of interest in the task is the cause of the disparity in results. It is particularly notable that the analysis of the first trial in Chapter 7 indicates possible, albeit weak, support for greater attention being paid to the within-category pairing, in concordance with the results of Chapter 6. It may be the case that there is an initial preference for the within-category pairing, but that the participants lose interest in the task rapidly enough that this difference is overshadowed. More powerful techniques, such as pupil dilation responses (Jackson & Sirois, 2009; Sirois & Jackson, 2012) may demonstrate this case more effectively, but are unlikely to be reliable with stimuli that change colour.

Finally, it is also possible that the reason for the disparity in the effect of CP between Chapters 6 and 7 is due to the task. It may simply be the case that the effects of CP of colour are demonstrable when infants can choose a within-category versus a between-category colour (i.e. when they are side by side), but that the differences are too subtle to allow backwards comparison of a current trial to previous trials. If this were to be the case, then it is possible that with a much larger variation in colour changes (e.g. 10 steps instead of

5), the effect could be seen in the type of task used in Chapter 7, as well as in Chapter 6.

The two chapters do have some points of concordance. Contrary to the findings of Franklin, Drivonikou, Bevis, et al. (2008), neither chapter finds any support for CP of colour changing with colour word acquisition. This is not necessarily unusual. Franklin, Drivonikou, Bevis, et al. found that hemispheric lateralisation of CP changes with colour term acquisition; while it might be expected that colour word acquisition might also affect other aspects of CP, this does not necessarily run counter to the previous findings. It may be that while there is a mechanistic change between biological infant colour categories (Davies & Franklin, 2002; Franklin, Clifford, et al., 2005), and linguistically-mediated adult colour categories, which is demonstrated through the change in hemispheric lateralisation, the extent of their categorisation of colour does not change throughout this process.

However, Franklin, Drivonikou, Bevis, et al. also had a much older group of participants than the studies in Chapters 6 and 7, and a different definition of colour word knowledge. In that study, the authors split participants into two groups, based on their mastery of the colour terms: "namers" and "learners." In that study, the namers were those which possessed an adult-like comprehension of a colour term, whereby they both comprehended the term and did not regularly over-extend the term, while the learners were those with an imperfect comprehension of each colour term. By contrast, in both of the studies in this thesis, participant comprehension of colour terms is assessed with CDIs. While in Chapter 3, CDIs are found to be reasonable indicators of colour word learning, and to predict performance in colour-mediated tasks in Chapter 5, the level of comprehension is a different one to that assessed by Franklin, Drivonikou, Bevis, et al..

It should also be noted that neither study found any effect of age on participant CP. This is not entirely unexpected – the age groups included in the two studies in this thesis were chosen as they were on either side of the colour word learning process beginning for British infants (as demonstrated in Chapter 3). By contrast, Franklin, Drivonikou, Bevis, et al. (2008) assessed participants between 2 and 5 years of age, and Franklin, Drivonikou, Clifford, et al. (2008) assessed participants around 5 months old. If the sort of change in CP that has been reported occurs, it is possibly at a substantially older age than 19 months. It is interesting to note that in Franklin, Drivonikou, Bevis, et al., the learners were around 32 months old, while the namers were around 46 months old, and a difference between the two groups was found in CP. If these results are to be believed, then the change in CP, possibly moderated by colour term knowledge, occurs substantially later than 19 months.

9.6 Colour vision testing and visual closure

In Chapter 8, a test experiment and a control experiment were presented. In the test experiment, infants saw outlines of familiar objects presented as a series of small circles, similar to the design commonly used in pseudoisochromatic plates (Ishihara, 1917). In the control experiment, those images were replaced with pictures of the objects themselves. The results indicated that infants were much slower to look to the target at 16 months than they were at 19 months, a differential that was not demonstrated in the control experiment. The findings may demonstrate that infant visual closure is still under development at around that age, causing them difficulty in finding the image named.

The ability to infer an outline from stimuli such as those used in Chapter

8 is a crucial component of many variants of colour vision tests. Participants in colour vision tests such as the Ishihara test (1917) are often required to be able to make the outline, and verbally or physically demonstrate what pattern they see, where the correct answer is visible only with normal colour vision. While the prototype demonstrated in Chapter 8 removes the need for participants to respond, thus making it more accessible for toddlers and infants who are not yet speaking consistently, the ability to make out the outline is still an assumption of the test. The prototype presented in Chapter 8 worked successfully for infants as young as 16 months old, but the much slower fixation time suggests that it may not be successful with infants much younger than 16 months. That, as well as the the fact that many infants under 16 months will not comprehend most of the target words, means that the prototype may have a lower limit of 16 months.

Are there any plausible ways of testing colour vision below 16 months of age? Dynamic colour vision tests such as the City University Colour Vision Test (Barbur et al., 1994) have had considerable success with adults[1], but still require the participant to tell the experimenter when they lose sight of the target. There may be promise for colour vision testing below the benchmark reported in Chapter 8 of this thesis if dynamic tests can be altered to make them usable with infants, but such changes would need to ensure that infant disinterest and an inability to see the target could somehow be differentiated. Until such designs can be implemented, the prototype investigated in Chapter 8 is the first successful prototype of a colour vision test for toddlers under 2 years of age, to my knowledge.

Visual closure is often monitored as an important skill in reading (Kavale, 1982), but the developmental trajectory of visual closure in infants is other-

[1] A sample of the test can be seen here: https://www.city.ac.uk/health/research/centre-for-applied-vision-research/a-new-web-based-colour-vision-test

wise not well documented. The results of Chapter 8 raise the possibility that the second year of life is a crucial time period for the development of visual closure. The visual closure related skills required in reading are obviously different to those required for the task in Chapter 8 of this thesis, but the results suggest that development may accelerate during the second year of life.

9.7 Concluding Remarks

The results of the experiments presented in this thesis demonstrate that infants learn colour words much earlier than previously demonstrated, and that they do with great variety, dependent on the languages they speak. Colour words are learned as a function of the frequency of input, and as such the order of learning colour words changes with the language, rather than following a universal, perceptual order. However, the results of this thesis also show that this only describes the very beginning of the process; words such as *"dark"* and *"light"* to finer shades of colour are learned much later than the basic colour terms.

The results presented here also demonstrate that colour word learning is a crucial part of infant processing of colour-related concepts. The findings in this thesis show that knowledge of a colour word is crucial for toddlers to attend to a colour-matched objects when they hear the name of a typically-coloured object, raising new questions about the role of feature labels in object processing.

The results of this thesis also demonstrate that there is a certain element of task dependency when assessing categorical perception of colour in infants. I present in this thesis, one study that shows strong effects of a category preference, and one that shows no such effect. The results of two different studies, however, show no effect of colour word knowledge on categorical perception of colour, suggesting that if there are mechanistic changes in how infants process colour, it may be at an older age than those tested in this thesis.

Finally, a preliminary prospective paradigm for a colour vision test that would be applicable to infants is presented, showing that the test is viable for infants as young as 16 months. However, the findings demonstrate that the second year of life may be a critical period for the development of visual closure, and as such, similar paradigms are unlikely to be successful with age groups below 16 months.

The findings presented in this thesis demonstrate the importance of further investigation into this field. Much of how infants process visual features, such as colour, is still uncategorised; and it is unclear how infants combine features to make a full representation of an object. In addition, the present research highlights that there is cause for further research into infant perception, particularly with regards to colour vision testing and visual closure.

References

Abbott, J. T., Griffiths, T. L., & Regier, T. (2016). Focal colors across languages are representative members of color categories. *Proceedings of the National Academy of Sciences*, 11178–11183.

Aisch, G. (2011). *Generating Color Blindness Test Images with Processing.* Retrieved from `www.vis4.net`

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected papers of hirotugu akaike* (pp. 199–213). New York, NY: Springer New York.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439.

Al-Rasheed, A. S. (2015a). Categorical perception of color: Evidence from secondary category boundary. *Psychology Research and Behavior Management*, *8*, 273–285.

Al-Rasheed, A. S. (2015b). An experimental study of gender and cultural differences in hue preference. *Frontiers in Psychology*, *6*, 1–5.

Al-Rasheed, A. S. (2016). Categorical perception of colour is lateralised to the left hemisphere: Evidence from present-absent and pop-out tasks.

International Journal of Social Science and Humanity, *6*(2), 113–118.

Al-Rasheed, A. S., Franklin, A., Drivonikou, G., & Davies, I. (2014). Left hemisphere lateralization of categorical color perception among Roman and Arabic script readers. *Psychology*, *5*(March), 255–270.

Althaus, N., & Plunkett, K. (2015a). Categorization in infancy: Labeling induces a persisting focus on commonalities. *Developmental Science*, 1–11.

Althaus, N., & Plunkett, K. (2015b). Timing matters: The impact of label synchrony on infant categorisation. *Cognition*, *139*, 1–9.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the Environment in Memory. *Psychological Science*, *2*(6), 396–408.

Andrick, G. R., & Tager-Flusberg, H. (1986). The acquisition of colour terms. *Journal of Child Language*, *13*(1), 119–134.

Arias-Trejo, N., & Plunkett, K. (2009). Lexical-semantic priming effects during infancy. *Philosophical transactions of the Royal Society of London*, *364*(1536), 3633–3647.

Arias-Trejo, N., & Plunkett, K. (2013). What's in a link: Associative and taxonomic priming effects in the infant lexicon. *Cognition*, *128*(2), 214–227.

Atkinson, J. (2002). *The Developing Visual Brain*. Oxford University Press.

Atkinson, J., & Braddick, O. (1998). Research methods in infant vision. In *Vision research: A practical approach* (pp. 2–8). Oxford University Press.

Atkinson, J., Braddick, O., & French, J. (1979). Contrast sensitivity of the human neonate measured by the visual evoked potential. *Investigative ophthalmology & visual science*, *18*, 210–213.

Atkinson, J., Braddick, O., & Moar, K. (1977). Infants' detection of image defocus. *Vision research*, *17*, 1125–1126.

Barbur, J. L., Harlow, J., & Plant, G. T. (1994, 12). Insights into the different exploits of colour in the visual cortex. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *258*(1353), 327–334.

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474.

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in psychology*, *4*(December), 328.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 1–43.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... Eigen, C. (2017). Fitting linear mixed-effects models using lme4. *Journal of statistical software*, *67*(1), 1 - 113.

Bates, E., Bretherton, I., & Snyder, L. (1988). *From first words to grammar: individual differences and dissociable mechanisms*. New York: CUP.

Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J. S., ... Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language*, *21*(1), 85–123.

Bergelson, E., & Aslin, R. N. (2017). Nature and origins of the lexicon in 6-mo-olds. *Proceedings of the National Academy of Sciences*(Early Edition), 1–6.

Berlin, B., & Kay, P. (1969). *Basic Color Terms; Their Universality and Evolution*. Berkeley: University of California Press.

Bird, C. M., Berens, S. C., Horner, a. J., & Franklin, A. (2014). Categorical encoding of color in the brain. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(12), 4590–4595.

Bloom, P. (2000). *How Children Learn the Meaning of Words*. Cambridge, Mass:

MIT Press.

Booth, A., & Waxman, S. R. (2003). Mapping words to the world in infancy: Infants' expectations for count nouns and adjectives. *Journal of Cognition and Development*, *4*, 37–41.

Booth, A., & Waxman, S. R. (2009). A horse of a different color: Specifying with precision infants' mappings of novel nouns and adjectives. *Child Development*, *80*(1), 15–22.

Bornstein, M. H. (1985). On the development of color naming in young children: Data and theory. *Brain and Language*, *26*(1), 72–93.

Bornstein, M. H., Kessen, W., & Weiskopf, S. (1976). The categories of hue in infancy. *Science (New York, N.Y.)*, *191*(4223), 201–202.

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*(3), 207–222.

Braddick, O., & Atkinson, J. (2011). Development of human visual function. *Vision Research*, *51*(13), 1588–1609.

Brown, A. M., Lindsey, D. T., & Guckes, K. M. (2011). Color names, color categories, and color-cued visual search: Sometimes, color perception is not categorical. *Journal of Vision*, *11*(12), 2–2.

Brown, A. M., Lindsey, D. T., Rambeau, R. S., & Shamp, H. A. (2009, 8). Visual search for colors as a test of the Sapir-Whorf hypothesis. *Journal of Vision*, *9*(8), 366.

Brown, R. W., & Lenneberg, E. H. (1954). A study in language and cognition. *J Abnorm Psychol*, *49*(3), 454–462.

Bürkner, P.-C. (2015). *brms : An R Package for Bayesian Generalized Linear Mixed Models using Stan.*

Bürkner, P.-C. (2017). Bayesian Distributional Non-Linear Multilevel Modeling with the R Package brms.

Childers, J. B., & Tomasello, M. (2006, 4). Are Nouns Easier to Learn Than Verbs? Three Experimental Studies. In *Action meets word* (Vol. 16, pp. 311–335). Oxford University Press.

Chow, J., Aimola Davies, A., & Plunkett, K. (2017). Spoken-word recognition in 2-year-olds: The tug of war between phonological and semantic activation. *Journal of Memory and Language*, *93*, 104–134.

Clifford, A., Franklin, A., Davies, I. R. L., & Holmes, A. (2009). Electrophysiological markers of categorical perception of color in 7-month old infants. *Brain and Cognition*, *71*(2), 165–172.

Clifford, A., Franklin, A., Holmes, A., Drivonikou, V. G., Özgen, E., & Davies, I. R. L. (2012). Neural correlates of acquired color category effects. *Brain and Cognition*, *80*(1), 126–143.

Cohen, L. B., Chaput, H. H., & Cashon, C. H. (2002). A constructivist model of infant cognition. *Cognitive Development*, *17*(3-4), 1323–1343.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language. *Cognitive Psychology*, *107*(1), 84–107.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(1), 7–29.

Dale, P. S. (1991). The validity of a parent report measure of vocabulary and syntax at 24 months. *Journal of speech and hearing research*, *34*(3), 565–571.

Dale, P. S., Bates, E., Reznick, J. S., & Morisset, C. (1989). The validity of a parent report instrument of child language at twenty months. *Journal of Child Language*, *16*(2), 239–249.

Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers*, *28*(1), 125–127.

Daoutis, C. A., Franklin, A., Riddett, A., Clifford, A., & Davies, I. R. L. (2006). Categorical effects in children's colour search: a cross-linguistic comparison. , 373–400.

Davies, I., & Franklin, A. (2002). Categorical similarity may affect colour pop-out in infants after all. *British Journal of Developmental Psychology*, *20*(2), 185–203.

Delle Luche, C., Durrant, S., Floccia, C., & Plunkett, K. (2014). Implicit meaning in 18-month-old toddlers. *Developmental Science*, *6*, 948–955.

Desroches, A. S., Newman, R. L., & Joanisse, M. F. (2009). Investigating the Time Course of Spoken Word Recognition: Electrophysiological Evidence for the Influences of Phonological Similarity. *Journal of Cognitive Neuroscience*, *21*(10), 1893–1906.

De Valois, K., & De Valois, R. L. (1975). Neural coding of color. In E. Cartarette & M. Friedman (Eds.), *Handbook of perception* (pp. 117–166). New York, NY: Academic Press.

Dink, J. W., & Ferguson, B. (2015). *eyetrackingR: An R Library for Eye-tracking Data Analysis.*

Drivonikou, G. V., Kay, P., Regier, T., Ivry, R. B., Gilbert, a. L., Franklin, a., & Davies, I. R. L. (2007). Further evidence that Whorfian effects are stronger in the right visual field than the left. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(3), 1097–1102.

Eysenck, H. J. (1941). A Critical and Experimental Study of Colour Preferences. *The American Journal of Psychology*, *54*(3), 385–394.

Fairchild, M. D. (1998). *Color Appearance Models*. Reading, MA: Addison-Wesley.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., ... Stiles, J. (1994). Variability in Early Communicative Development. *Monographs of the Society for Research in Child Development*, *59*(5), 174–185.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories: User's Guide and Technical Manual* (2nd Editio ed.). Baltimore, MD: Brookes Publishing Co.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, *8*(2), 181–195.

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid Gains in Speed of Verbal Processing by Infants in the 2nd Year. *Psychological Science*, *9*(3), 228–231.

Frank, M. C., Bergelson, E., Bergmann, C., Christia, A., Floccia, C., Gervain, J., ... Yurovsky, D. (2017). A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. *Psyarxiv*, *27b43*, 1–26.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2016). Wordbank: an open repository for developmental vocabulary data. *Journal of Child Language*, 1–18.

Franklin, A. (2006). Constraints on children's color term acquisition. *Journal of Experimental Child Psychology*, *94*(4), 322–327.

Franklin, A., Bevis, L., Ling, Y., & Hurlbert, A. (2010). Biological components of colour preference in infancy. *Developmental Science*, *13*(2), 346–354.

Franklin, A., Catherwood, D., Alvarez, J., & Axelsson, E. (2010). Hemispheric asymmetries in categorical perception of orientation in infants and adults. *Neuropsychologia*, *48*(9), 2648–2657.

Franklin, A., Clifford, A., Williamson, E., & Davies, I. (2005). Color term knowledge does not affect categorical perception of color in toddlers. *Journal of Experimental Child Psychology*, *90*(2), 114–141.

Franklin, A., & Davies, I. R. (2004). New evidence for infant colour categories. *British Journal of Developmental Psychology*, *22*(3), 349–377.

Franklin, A., Drivonikou, G. V., Bevis, L., Davies, I. R. L., Kay, P., & Regier, T. (2008). Categorical perception of color is lateralized to the right hemisphere in infants, but to the left hemisphere in adults. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(9), 3221–3225.

Franklin, A., Drivonikou, G. V., Clifford, A., Kay, P., Regier, T., & Davies, I. R. L. (2008). Lateralization of categorical perception of color changes with color term acquisition. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(47), 18221–18225.

Franklin, A., Gibbons, E., Chittenden, K., Alvarez, J., & Taylor, C. (2012). Infant color preference for red is not selectively context specific. *Emotion*, *12*(5), 1155–1160.

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*(3), 227–248.

Franklin, A., Skelton, A. E., & Catchpole, G. (2014). The case for infant colour categories. In W. Anderson, C. P. Biggam, C. Hough, & C. J. Kay (Eds.), *Colour studies: A broad spectrum* (pp. 169–180). John Benjamins Publishing Company.

Franklin, A., Sowden, P., Notman, L., Gonzalez-Dixon, M., West, D., Alexander, I., . . . White, A. (2010). Reduced chromatic discrimination in children with autism spectrum disorders. *Developmental Science*, *13*(1), 188–

200.

Franklin, A., Wright, O., & Davies, I. R. L. (2009). What can we learn from toddlers about categorical perception of color? Comments on Goldstein, Davidoff, and Roberson. *Journal of Experimental Child Psychology*, *102*(2), 239–245.

Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., ... Conway, B. R. (2017). Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences*, *114*(40), 10785–10790.

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 Sciences of the United States of America*, *103*(2), 489–494.

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2008). Support for lateralization of the Whorf effect beyond the realm of color discrimination. *Brain and Language*, *105*(2), 91–98.

Goldstein, J., Davidoff, J., & Roberson, D. (2009). Knowing color terms enhances recognition: Further evidence from English and Himba. *Journal of Experimental Child Psychology*, *102*(2), 219–238.

Goldstone, R. L., & Hendrickson, A. T. (2009). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(1), 69–78.

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, *14*(01), 23.

Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: The case for a developmental lexical principles framework. *Journal of Child Language*, *21*, 125–155.

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, *35*(3), 515–531.

Granger, G. W. (1952). Objectivity of Colour Preferences. *Nature*, *170*(4332), 778–780.

Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a British communicative development inventory. *Journal of Child Language*, *27*(3), 689–705.

Hanley, J. R., & Roberson, D. (2008, 5). Do infants see colour differently? *Scientific American - Mind and Brain*.

Hanley, J. R., & Roberson, D. (2011). Categorical perception effects reflect differences in typicality on within-category trials. *Psychonomic bulletin & review*, *18*(2), 355–363.

Harris, L., Atkinson, J., & Braddick, O. (1976). Visual contrast sensitivity of a 6-month-old infant measured by the evoked potential. *Nature, 264*, 570–571.

Heibeck, T. H., & Markman, E. M. (1987, 8). Word learning in children: an examination of fast mapping. *Child development*, *58*(4), 1021–1034.

Heider, E. R. (1971). *"Focal" color areas and the development of color names.* (Vol. 4) (No. 3). US: American Psychological Association.

Heider, E. R. (1972). Universals in color naming and memory. *Journal of Experimental Psychology*, *93*(1).

Hidaka, S., & Smith, L. B. (2010). A Single Word in a Population of Words. *Language Learning and Development*, *6*(3), 206–222.

Holmes, K. J., & Wolff, P. (2012). Does categorical perception in the left hemisphere depend on language? *Journal of Experimental Psychology: General*, *141*(3), 439–443.

Houston-Price, C., Mather, E., & Sakkalou, E. (2007). Discrepancy between parental reports of infants' receptive vocabulary and infants' behaviour in a preferential looking task. *Journal of Child Language*, *34*(4), 701–724.

Houston-Price, C., Plunkett, K., & Harris, P. (2005). 'Word-learning wizardry' at 1;6. *Journal of Child Language*, *32*(1), 175–189.

Huettig, F., & Altmann, G. T. M. (2004). The On-Line Processing of Ambiguous and Unambiguous Words in Context: Evidence from Head-Mounted Eyetracking. *The On-Line Study of Sentence Comprehension: Eyetracking, ERPs and Beyond*, 187–208.

Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*(1), 23–32.

Huettig, F., & Altmann, G. T. M. (2011). Looking at anything that is green when hearing "frog": how object surface colour and stored object colour knowledge influence language-mediated overt attention. *Quarterly Journal of Experimental Psychology*, *64*(1), 122–145.

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, *57*(4), 460–482.

Hurlbert, A. C., & Ling, Y. (2007). Biological components of sex differences in color preference. *Current Biology*, *17*(16), 623–625.

Ishihara, S. (1917). *Tests for Color-Blindness*. Tokyo: Hongo Harukicho.

Jackson, I., & Sirois, S. (2009). Infant cognition: Going full factorial with pupil dilation. *Developmental Science*, *12*(4), 670–679.

Jameson, K., & D'Andrade, R. G. (1997). It's not really red, green, yellow, blue: an inquiry into perceptual color space. In C. L. Hardin & L. Maffi (Eds.), *Color categories in thought and language* (pp. 295–319). Cambridge: Cambridge University Press.

Johnson, E. K., & Huettig, F. (2011). Eye movements during language-mediated visual search reveal a strong link between overt visual attention and lexical processing in 36-month-olds. *Psychological Research, 75*(1), 35–42.

Johnson, E. K., McQueen, J. M., & Huettig, F. (2011). Toddlers' language-mediated visual search: They need not have the words for it. *Quarterly Journal of Experimental Psychology, 64*, 1672–1682.

Johnson, E. K., & Zamuner, T. (2010). Using infant and toddler testing methods in language acquisition research. In E. Blom & S. Unsworth (Eds.), *Experimental methods in language acquisition research* (pp. 73–93). Amsterdam: John Benjamins Publishing Company.

Jones, S. S., & Smith, L. B. (2002). How children know the relevant properties for generalizing object names. *Developmental Science, 5*(2), 219–232.

Jraissati, Y. (2012). Categorical Perception of Color: Assessing the Role of Language. *Croatian Journal of Philosophy, XII*(36), 439–463.

Jusczyk, P. W., & Aslin, R. N. (1995). *Infants' Detection of the Sound Patterns of Words in Fluent Speech* (Vol. 29) (No. 1).

Kavale, K. (1982). Meta-Analysis of the Relationship Between Visual Perceptual Skills and Reading Achievement. *Journal of Learning Disabilities, 15*(1), 42–51.

Kay, P. (2003). Resolving the question of color naming universals. *Proceedings of the National Academy of Sciences, 100*(15), 9085–9089.

Kay, P., Berlin, B., Maffi, L., Merrifield, W., & Cook, R. (2011). *The World Colour Survey.* University of Chicago Press.

Kay, P., & Kempton, W. (1984). What Is the Sapir-Whorf Hypothesis? *American*

Anthropologist, *86*(1), 65–79.

Kay, P., & Regier, T. (2007). Color naming universals: The case of Berinmo. *Cognition*, *102*(2), 289–298.

Knoblauch, K., Vital-Durand, F., & Barbur, J. L. (2001). Variation of chromatic sensitivity across the life span. *Vision Research*, *41*(1), 23–36.

Kowalski, K., & Zimiles, H. (2006). The relation between children's conceptual functioning with color and color term acquisition. *Journal of Experimental Child Psychology*, *94*(4), 301–321.

Kruschke, J. K. (2013). Bayesian Estimation Supersedes the t Test. *Journal of Experimental Psychology: General*, *142*(2), 573–603.

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, *3*(3), 299–321.

Lindsey, D., & Brown, A. M. (2009). Color difference scaling at the blue-green color category boundary as a test of the Sapir-Whorf Hypothesis. *Journal of Vision*, *9*(8), 340.

Lucy, J. A. (1997). The linguistics of "color". In C. Hardin & L. Maffi (Eds.), *Color categories in thought and language* (pp. 320–346). Cambridge: Cambridge University Press.

Lucy, J. A., & Shweder, R. A. (1979). Whorf and his critics: Linguistic and nonlinguistic influences on color memory. *American Anthropologist*, *81*(3), 581–615.

Macario, J. F. (1991). Young children's use of color in classification: Foods and canonically colored objects. *Cognitive Development*, *6*(1), 17–46.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk* (3rd ed.). Malwah, NJ: Lawrence Erlbaum Associates.

Mani, N., Durrant, S., & Floccia, C. (2012). Activation of phonological and semantic codes in toddlers. *Journal of Memory and Language*, *66*(4), 612–

622.

Mani, N., Johnson, E. K., McQueen, J. M., & Huettig, F. (2013). How Yellow Is Your Banana? Toddlers' Language-Mediated Visual Search in Referent-Present Tasks. *Developmental Psychology*, *49*(6), 1036–1044.

Mani, N., & Plunkett, K. (2010). In the infant's mind's ear: evidence for implicit naming in 18-month-olds. *Psychological science : a journal of the American Psychological Society / APS*, *21*(7), 908–913.

Maule, J., Stanworth, K., Pellicano, E., & Franklin, A. (2017). Ensemble perception of color in autistic adults. *Autism Research*, *10*(5), 839–851.

Maule, J., Witzel, C., & Franklin, A. (2014). Getting the gist of multiple hues: metric and categorical effects on ensemble perception of hue. *Journal of the Optical Society of America. A, Optics, image science, and vision*, *31*(4), 93–102.

Mayor, J., & Plunkett, K. (2009). A mathematical insight into the size of infant vocabularies. *Proceedings of the 31st Annual Cognitive Science Society*, 437–442.

Mayor, J., & Plunkett, K. (2011). A statistical estimate of infant and toddler vocabulary size from CDI analysis. *Developmental Science*, *14*(4), 769–785.

Mayor, J., & Plunkett, K. (2014). Shared understanding and idiosyncratic expression in early vocabularies. *Developmental Science*, *17*(3), 412–423.

McManus, I. C., Jones, A. L., & Cottrell, J. (1981). The aesthetics of colour. *Perception*, *10*(6), 651–666.

Mee Bell, S., McCallum, R. S., & Cox, E. A. (2003). Toward a Research-Based Assessment of Dyslexia. *Journal of Learning Disabilities*, *36*(6), 505–516.

Mervis, C. B., Bertrand, J., & Pani, J. R. (1995). Transaction of cognitive-linguistic abilities and adult input: A case study of the acquisition of colour terms and colour-based subordinate object categories. *British Journal of Developmental Psychology*, *13*(3), 285–302.

Mills, D. L., Coffey-Corina, S. A., & Neville, H. J. (1993). Language Acquisition and Cerebral Specialization in 20-Month-Old Infants. *Journal of Cognitive Neuroscience*, *5*(3), 317–334.

Mills, D. L., Coffey-Corina, S. A., & Neville, H. J. (1997). Language comprehension and cerebral specialization from 13 to 20 months. *Developmental Neuropsychology*, *13*(3), 397–445.

Mintz, T. H., & Gleitman, L. R. (2002). Adjectives really do modify nouns: The incremental and restricted nature of early adjective acquisition. *Cognition*, *84*(3), 267–293.

Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R Analysis and Visualization Using R*. Boca Raton, FL: Chapman & Hall / CRC Press.

Morrone, M. C., Burr, D. C., & Fiorentini, A. (1993). Development of infant contrast sensitivity to chromatic stimuli. *Vision Research*, *33*(17), 2535–2552.

Munsell Color Services. (2007). *The Munsell Book of Color: Glossy Collection*. Grandville, MI: X-rite.

Ocelak, R. (2016). "Categorical Perception" and Linguistic Categorization of Color. *Review of Philosophy and Psychology*, *7*(1), 55–70.

O'Hanlon, C. G., & Roberson, D. (2006). Learning in context: Linguistic and attentional constraints on children's color term learning. *Journal of Experimental Child Psychology*, *94*(4), 275–300.

Palmer, S. E., & Schloss, K. B. (2010). An ecological valence theory of human color preference. *Proceedings of the National Academy of Sciences*, *107*(19), 8877–8882.

Peeples, D. R., & Teller, D. Y. (1970). Color Vision and Brightness Discrimination in Two-Month-Old Human Infants. *Science*, *189*.

Peeples, D. R., & Teller, D. Y. (1978). White-Adapted Photopic Spectral Sensitivity in Human Infants. *Vision Research*, *18*, 49–53.

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

Pitchford, N., Davis, E. E., & Scerif, G. (2009). Does colour preference have a role in colour term acquisition? *British Journal of Developmental Psychology*, *27*(4), 993–1012. Retrieved from http://doi.wiley.com/10.1348/026151008X399916

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

Pitchford, N., & Mullen, K. (2002). Is the acquisition of basic-colour terms in young children constrained? *Perception*, *31*(11), 1349–1370.

Pitchford, N., & Mullen, K. (2003). The development of conceptual colour categories in pre-school children: Influence of perceptual categorization. *Visual Cognition*, *10*(1), 51–77.

Pitchford, N., & Mullen, K. (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.

Plunkett, K., Hu, J. F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition*, *106*, 665–681.

Quine, W. V. O. (1960). *Word & Object*. MIT Press.

Ramscar, M., Thorpe, K., & Denny, K. (2007). Surprise in the Learning of

Color Words. *Proceedings of the 29th Annual Cognitive Science Society*, 575–580.

Regier, T., & Kay, P. (2009). Language, thought, and color: Whorf was half right. *Trends in Cognitive Sciences*, *13*(10), 439–446.

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

Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(4), 1436–41.

Regier, T., & Xu, Y. (2017). The Sapir-Whorf hypothesis and inference under uncertainty. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*(6).

Rice, M. (1980). *Cognition to language categories: Categories, word meaning, and training*. Baltimore MD: University Park Press.

Ripley, B. D., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., & Firth, D. (2017). *Package ' MASS '*.

Roberson, D., & Davidoff, J. (2000). The categorical perception of colors and facial expressions: The effect of verbal interference. *Memory and Cognition*, *28*(6), 977–986.

Roberson, D., Davidoff, J., Davies, I. R. L., & Shapiro, L. R. (2004). The development of color categories in two languages: a longitudinal study. *Journal of experimental psychology. General*, *133*(4), 554–71.

Roberson, D., Davidoff, J., Davies, I. R. L., & Shapiro, L. R. (2005). Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology*, *50*(4), 378–411.

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

Roberson, D., & Hanley, J. R. (2009). Only half right: Comment on Regier and Kay. *Trends in Cognitive Sciences*, *13*(12), 497–500.

Roberson, D., Hanley, J. R., & Pak, H. (2009). Thresholds for color discrimination in English and Korean speakers. *Cognition*, *112*(3), 482–487.

Roberson, D., Pak, H., & Hanley, J. R. (2008). Categorical perception of colour in the left and right visual field is verbally mediated: Evidence from Korean. *Cognition*, *107*(2), 752–762.

Robinson, C. W., & Sloutsky, V. M. (2004). Auditory dominance and its change in the course of development. *Child Development*, *75*(5), 1387–1401.

Sandhofer, C. M., & Smith, L. B. (1999). Learning color words involves learning a system of mappings. *Developmental Psychology*, *35*(3), 668–679.

Sandhofer, C. M., & Smith, L. B. (2001). Why children learn color and size words so differently: evidence from adults' learning of artificial terms. *Journal of Experimental Psychology. General*, *130*(4), 600–620.

Schafer, G., & Plunkett, K. (1998). Rapid word learning by fifteen-month-olds under tightly controlled conditions. *Child Development*, *69*(2), 309–320.

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*(1), 177–199.

Shatz, M., Tare, M., Nguyen, S. P., & Young, T. (2010). Acquiring non-object terms: The case for time words. *Journal of Cognition and Development*, *11*(1), 16–36.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology. *Psychological Science*, *22*(11), 1359–1366.

Sirois, S., & Jackson, I. R. (2012). Pupil dilation and object permanence in infants. *Infancy*, *17*(1), 61–78.

Skelton, A., Catchpole, G., Abbott, J., & Franklin, A. (2017). Biological origins of color categorization. *Proceedings of the National Academy of Sciences*, *114*(21), 5545–5550.

Smith, L. B. (1984). Young children's understanding of attributes and dimensions: A comparison of conceptual and linguistic measures. *Child Development*, *2*, 363–380.

Soja, N. N. (1994). Young children's concept of color and its relation to the acquisition of color words. *Child Development*, *65*(3), 918–937.

Stan Development Team. (2016). *RStan: the R interface to Stan.*

Styles, S., & Plunkett, K. (2009a). How do infants build a semantic system? *Language and Cognition*, *1*(1), 1–24.

Styles, S., & Plunkett, K. (2009b). What is 'word understanding' for the parent of a one-year-old? Matching the difficulty of a lexical comprehension task to parental CDI report. *Journal of Child Language*, *36*(4), 895–908.

Styles, S., Plunkett, K., & Duta, M. D. (2015). Infant VEPs reveal neural correlates of implicit naming: Lateralized differences between lexicalized versus name-unknown pictures. *Neuropsychologia*, *77*, 177–184.

Taylor, C., Clifford, A., & Franklin, A. (2012). Color preferences are not universal. *Journal of Experimental Psychology: General*, *142*(4), 1015–27.

Taylor, C., & Franklin, A. (2012). The relationship between color–object associations and color preference: Further investigation of ecological valence theory. *Psychonomic Bulletin & Review*, *19*(2), 190–197.

Taylor, C., Schloss, K., Palmer, S. E., & Franklin, A. (2013). Color preferences in infants and adults are different. *Psychonomic bulletin & review*, *20*(5), 916–22.

Teller, D. Y. (1979). The forced-choice preferential looking procedure: A psychophysical technique for use with human infants. *Infant Behavior and Development*, *2*(1), 135–153.

Teller, D. Y. (1998). Spatial and temporal aspects of infant color vision. *Vision research*, *38*(21), 3275–3282.

Teller, D. Y., Civan, A., & Bronson-Castain, K. (2004). Infants' spontaneous color preferences are not due to adult-like brightness variations. *Visual neuroscience*, *21*(3), 397–401.

Tomasello, M., & Mervis, C. B. (1994). The instrument is great but measuring comprehension is still a problem. *Monographs of the Society for Research in Child Development*, *59*(5), 174–179.

Tomasello, M., Strosberg, R., & Akhtar, N. (1996). Eighteen-month-old children learn words in non-ostensive contexts. *Journal of Child Language*, *23*(1), 157–176.

Wagner, K., Dobkins, K., & Barner, D. (2013). Slow mapping: Color word learning as a gradual inductive process. *Cognition*, *127*(3), 307–317.

Wagner, K., Jergens, J., & Barner, D. (2014). Partial Color Word Comprehension Precedes Production. *Proceedings CogSci*, 1724–1729.

Wass, S. V., Smith, T. J., & Johnson, M. H. (2013). Parsing eye-tracking data of variable quality to provide accurate fixation duration estimates in infants and adults. *Behavior Research Methods*, *45*(1), 229–250.

Waxman, S. R. (1999). Specifying the scope of 13-month-olds' expectations for novel words. *Cognition*, *70*(3), 35–50.

Waxman, S. R., & Booth, A. (2003). The origins and evolution of links between word learning and conceptual organization: new evidence from 11-month-olds. *Developmental Science*, *62*, 128–135.

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: evidence from 12- to 13-month-old infants. *Cognitive psychology*, *29*(3), 257–302.

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word-object associations by 14-month-old infants. *Developmental Psychology*, *34*(6), 1289–1309.

Wicherts, J. M., Veldkamp, C. L., Augusteijn, H. E., Bakker, M., van Aert, R. C., & van Assen, M. A. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid P-hacking. *Frontiers in Psychology*, *7*(1832), 1–12.

Witzel, C. (2015). Commentary: An experimental study of gender and cultural differences in hue preference. *Frontiers in Psychology*, *6*, 1–3.

Witzel, C., & Gegenfurtner, K. R. (2011). Is there a lateralized category effect for color? *Journal of Vision*, *11*(12), 1–25.

Witzel, C., & Gegenfurtner, K. R. (2013). Categorical sensitivity to color differences. *Journal of Vision*, *13*(2013), 1–33.

Witzel, C., & Gegenfurtner, K. R. (2016). Categorical perception for red and brown. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(4), 540–570.

Yang, J., Kanazawa, S., Yamaguchi, M. K., & Kuriki, I. (2016). Cortical response to categorical color perception in infants investigated by near-infrared spectroscopy. *PNAS*, *113*(9), 2370–2375.

Younger, B. A. (1985). The Segregation of Items into Categories by Ten-Month-Old Infants. *Child Development*, *56*(6), 1574–1583.

Yurovsky, D., Wagner, K., Barner, D., & Frank, M. C. (2015). Signatures of Domain-General Categorization Mechanisms in Color Word Learning. *Proceedings CogSci*.

Zhou, K., Mo, L., Kay, P., Kwok, V. P. Y., Ip, T. N. M., & Tan, L. H. (2010). Newly trained lexical categories produce lateralized categorical perception of color. *Proceedings of the National Academy of Sciences, 107*(22), 9974–9978.

Appendix A

Supplementary Information for Chapter 2

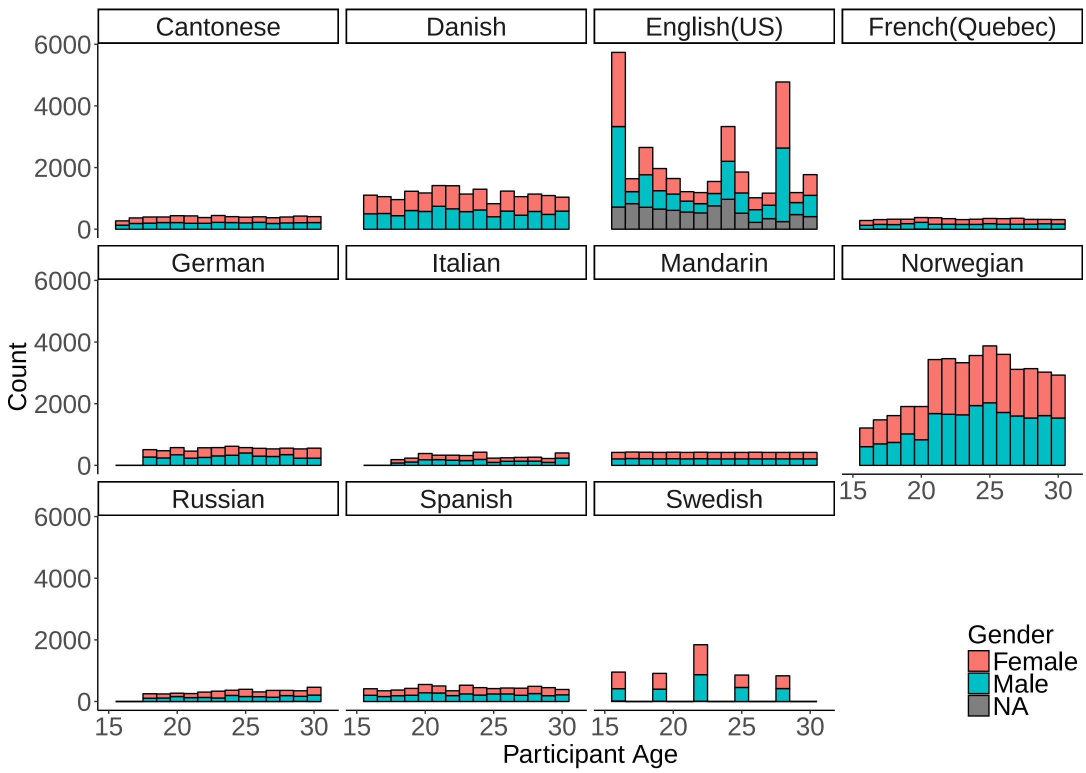

Figure A.1: Participant information for all data used from the Wordbank Repository, separated by age and gender. Grey indicates that there was no gender information available.

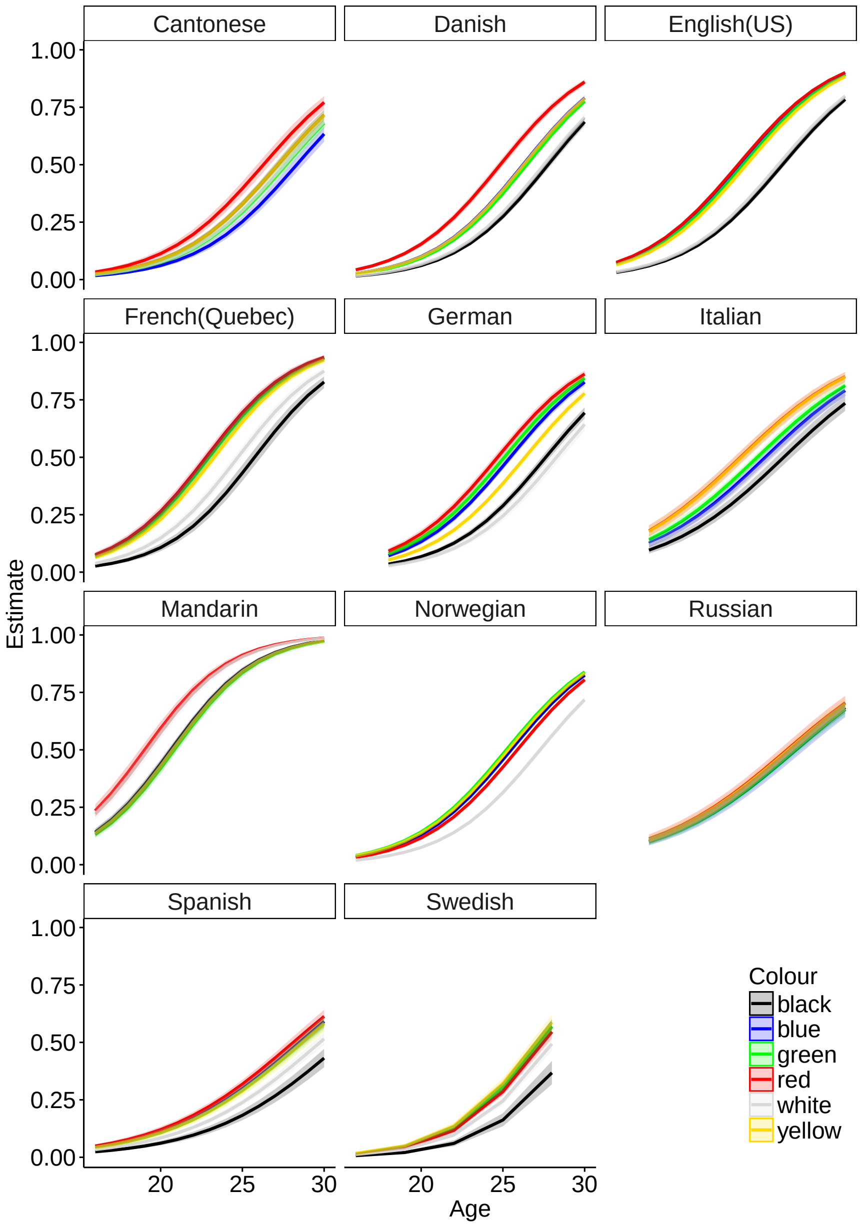

Figure A.2: Fitted final model from Study 3, for comparison to Study 2. The results show *blue* further behind than in Study 2, indicating other possible factors.

Table A.1: Corpora used in analysing frequency of occurrence for each colour word.

Language	Corpora Used
Cantonese	Lee/Wong/Leung
Danish	Plunkett
French	Champaud, Hunkeler, Lyon, Pauline
English	Bates, Bernstein, Bloom70, Bloom73, Bohannon, Brent, Higginson, Kuczaj, Nicholas-NH, Post, Rollins, Soderstrom
German	Caroline, Szagun
Italian	D'Odorico, Klammler, Tonelli
Mandarin	Beijing, Context, Tong
Norwegian	Garmann, Ringstad
Russian	Protassova, Tanja
Spanish	Aguirre, JacksonThal, LlinasOjea, Nieva, OreaPine
Swedish	Lacerda

Table A.2: Full frequency calculations for each Childes corpus used in Study 3. *Bernstein-Ratner corpus did not include the "interview" section as there was little parent data. **Szagun corpus only included the Normal-Hearing group.

Corpus	Language	Colour	Token	Unique	Total
Bates	English	black	1	1617	36518
Bates	English	blue	10	1617	36518
Bates	English	green	27	1617	36518
Bates	English	red	42	1617	36518
Bates	English	white	18	1617	36518
Bates	English	yellow	32	1617	36518
Bernstein-Ratner*	English	black	8	1632	35715
Bernstein-Ratner	English	blue	9	1632	35715
Bernstein-Ratner	English	green	10	1632	35715
Bernstein-Ratner	English	red	15	1632	35715
Bernstein-Ratner	English	white	1	1632	35715
Bernstein-Ratner	English	yellow	10	1632	35715
Bloom70	English	black	0	1493	17283
Bloom70	English	blue	1	1493	17283
Bloom70	English	green	1	1493	17283
Bloom70	English	red	9	1493	17283
Bloom70	English	white	0	1493	17283
Bloom70	English	yellow	1	1493	17283
Bloom73	English	black	1	702	9333

Bloom73	English	blue	0	702	9333
Bloom73	English	green	2	702	9333
Bloom73	English	red	1	702	9333
Bloom73	English	white	1	702	9333
Bloom73	English	yellow	2	702	9333
Bohannon	English	black	1	640	2871
Bohannon	English	blue	3	640	2871
Bohannon	English	green	1	640	2871
Bohannon	English	red	1	640	2871
Bohannon	English	white	0	640	2871
Bohannon	English	yellow	4	640	2871
Brent	English	black	91	8250	493113
Brent	English	blue	261	8250	493113
Brent	English	green	284	8250	493113
Brent	English	red	270	8250	493113
Brent	English	white	93	8250	493113
Brent	English	yellow	267	8250	493113
Higginson	English	black	6	2114	39197
Higginson	English	blue	31	2114	39197
Higginson	English	green	21	2114	39197
Higginson	English	red	23	2114	39197
Higginson	English	white	12	2114	39197
Higginson	English	yellow	14	2114	39197
Kuczaj	English	black	5	3367	52763
Kuczaj	English	blue	6	3367	52763
Kuczaj	English	green	11	3367	52763
Kuczaj	English	red	24	3367	52763

Kuczaj	English	white	13	3367	52763
Kuczaj	English	yellow	14	3367	52763
Nicholas-NH	English	black	28	3582	161705
Nicholas-NH	English	blue	117	3582	161705
Nicholas-NH	English	green	101	3582	161705
Nicholas-NH	English	red	109	3582	161705
Nicholas-NH	English	white	18	3582	161705
Nicholas-NH	English	yellow	89	3582	161705
Post	English	black	7	2897	86698
Post	English	blue	86	2897	86698
Post	English	green	65	2897	86698
Post	English	red	94	2897	86698
Post	English	white	20	2897	86698
Post	English	yellow	70	2897	86698
Rollins	English	black	15	1746	57421
Rollins	English	blue	132	1746	57421
Rollins	English	green	116	1746	57421
Rollins	English	red	136	1746	57421
Rollins	English	white	16	1746	57421
Rollins	English	yellow	121	1746	57421
Soderstrom	English	black	12	4205	89285
Soderstrom	English	blue	25	4205	89285
Soderstrom	English	green	16	4205	89285
Soderstrom	English	red	45	4205	89285
Soderstrom	English	white	9	4205	89285
Soderstrom	English	yellow	14	4205	89285
LeeWongLeung	Cantonese	black	20	2725	59222

LeeWongLeung	Cantonese	blue	9	2725	59222
LeeWongLeung	Cantonese	green	14	2725	59222
LeeWongLeung	Cantonese	red	37	2725	59222
LeeWongLeung	Cantonese	white	15	2725	59222
LeeWongLeung	Cantonese	yellow	20	2725	59222
Plunkett	Danish	black	34	4187	128419
Plunkett	Danish	blue	66	4187	128419
Plunkett	Danish	green	60	4187	128419
Plunkett	Danish	red	121	4187	128419
Plunkett	Danish	white	38	4187	128419
Plunkett	Danish	yellow	65	4187	128419
Champaud	French	black	1	2329	18056
Champaud	French	blue	10	2329	18056
Champaud	French	green	4	2329	18056
Champaud	French	red	11	2329	18056
Champaud	French	white	5	2329	18056
Champaud	French	yellow	7	2329	18056
Hunkeler	French	black	0	1139	14323
Hunkeler	French	blue	2	1139	14323
Hunkeler	French	green	0	1139	14323
Hunkeler	French	red	7	1139	14323
Hunkeler	French	white	9	1139	14323
Hunkeler	French	yellow	24	1139	14323
Lyon	French	black	103	10733	502145
Lyon	French	blue	449	10733	502145
Lyon	French	green	436	10733	502145
Lyon	French	red	481	10733	502145

Lyon	French	white	164	10733	502145
Lyon	French	yellow	358	10733	502145
Pauline	French	black	6	2489	32031
Pauline	French	blue	16	2489	32031
Pauline	French	green	21	2489	32031
Pauline	French	red	11	2489	32031
Pauline	French	white	8	2489	32031
Pauline	French	yellow	6	2489	32031
Caroline	German	black	86	8253	177076
Caroline	German	blue	135	8253	177076
Caroline	German	green	119	8253	177076
Caroline	German	red	157	8253	177076
Caroline	German	white	62.91513	8253	177076
Caroline	German	yellow	86	8253	177076
Szagun**	German	black	31	9207	320448
Szagun	German	blue	98	9207	320448
Szagun	German	green	143	9207	320448
Szagun	German	red	141	9207	320448
Szagun	German	white	32.08102	9207	320448
Szagun	German	yellow	89	9207	320448
D’Odorico	Italian	black	7	1601	23722
D’Odorico	Italian	blue	11	1601	23722
D’Odorico	Italian	green	6	1601	23722
D’Odorico	Italian	red	16	1601	23722
D’Odorico	Italian	white	1	1601	23722
D’Odorico	Italian	yellow	13	1601	23722
Klammler	Italian	black	0	525	2202

Klammler	Italian	blue	0	525	2202
Klammler	Italian	green	1	525	2202
Klammler	Italian	red	1	525	2202
Klammler	Italian	white	0	525	2202
Klammler	Italian	yellow	1	525	2202
Tonelli	Italian	black	56	5221	82348
Tonelli	Italian	blue	84	5221	82348
Tonelli	Italian	green	123	5221	82348
Tonelli	Italian	red	188	5221	82348
Tonelli	Italian	white	89	5221	82348
Tonelli	Italian	yellow	187	5221	82348
Beijing	Mandarin	black	64	5048	106553
Beijing	Mandarin	blue	31	5048	106553
Beijing	Mandarin	green	6	5048	106553
Beijing	Mandarin	red	109	5048	106553
Beijing	Mandarin	white	124	5048	106553
Beijing	Mandarin	yellow	39	5048	106553
Context	Mandarin	black	10	2948	56273
Context	Mandarin	blue	41	2948	56273
Context	Mandarin	green	40	2948	56273
Context	Mandarin	red	77	2948	56273
Context	Mandarin	white	61	2948	56273
Context	Mandarin	yellow	24	2948	56273
Tong	Mandarin	black	24	2935	82141
Tong	Mandarin	blue	20	2935	82141
Tong	Mandarin	green	40	2935	82141
Tong	Mandarin	red	81	2935	82141

Tong	Mandarin	white	51	2935	82141
Tong	Mandarin	yellow	29	2935	82141
Garmann	Norwegian	black	0	2455	18848
Garmann	Norwegian	blue	7	2455	18848
Garmann	Norwegian	green	6	2455	18848
Garmann	Norwegian	red	7	2455	18848
Garmann	Norwegian	white	1	2455	18848
Garmann	Norwegian	yellow	4	2455	18848
Ringstad	Norwegian	black	45	6313	126890
Ringstad	Norwegian	blue	38	6313	126890
Ringstad	Norwegian	green	69	6313	126890
Ringstad	Norwegian	red	16	6313	126890
Ringstad	Norwegian	white	1	6313	126890
Ringstad	Norwegian	yellow	57	6313	126890
Protassova	Russian	black	2	3356	17586
Protassova	Russian	blue	1	3356	17586
Protassova	Russian	green	2	3356	17586
Protassova	Russian	red	7	3356	17586
Protassova	Russian	white	2	3356	17586
Protassova	Russian	yellow	7	3356	17586
Tanja	Russian	black	0	668	1667
Tanja	Russian	blue	0	668	1667
Tanja	Russian	green	1	668	1667
Tanja	Russian	red	3	668	1667
Tanja	Russian	white	1	668	1667
Tanja	Russian	yellow	0	668	1667
Aguirre	Spanish	black	4	2600	37880

Aguirre	Spanish	blue	14	2600	37880
Aguirre	Spanish	green	8	2600	37880
Aguirre	Spanish	red	8	2600	37880
Aguirre	Spanish	white	4	2600	37880
Aguirre	Spanish	yellow	59	2600	37880
JacksonThal	Spanish	black	0	2515	34053
JacksonThal	Spanish	blue	0	2515	34053
JacksonThal	Spanish	green	4	2515	34053
JacksonThal	Spanish	red	8	2515	34053
JacksonThal	Spanish	white	0	2515	34053
JacksonThal	Spanish	yellow	11	2515	34053
LlinasOjea	Spanish	black	10	5221	87337
LlinasOjea	Spanish	blue	10	5221	87337
LlinasOjea	Spanish	green	17	5221	87337
LlinasOjea	Spanish	red	35	5221	87337
LlinasOjea	Spanish	white	28	5221	87337
LlinasOjea	Spanish	yellow	21	5221	87337
Nieva	Spanish	black	3	2777	51756
Nieva	Spanish	blue	109	2777	51756
Nieva	Spanish	green	69	2777	51756
Nieva	Spanish	red	121	2777	51756
Nieva	Spanish	white	22	2777	51756
Nieva	Spanish	yellow	73	2777	51756
OreaPine	Spanish	black	14	3692	47430
OreaPine	Spanish	blue	31	3692	47430
OreaPine	Spanish	green	52	3692	47430
OreaPine	Spanish	red	35	3692	47430

OreaPine	Spanish	white	20	3692	47430
OreaPine	Spanish	yellow	32	3692	47430
Lacerda	Swedish	black	1	1481	19399
Lacerda	Swedish	blue	9	1481	19399
Lacerda	Swedish	green	9	1481	19399
Lacerda	Swedish	red	7	1481	19399
Lacerda	Swedish	white	4	1481	19399
Lacerda	Swedish	yellow	11	1481	19399

Appendix B

Supplementary Information for Chapter 5

Table B.1: Complete list of stimuli used during eye-tracking experiment.

List	Trial Condition	Pictured Objects	Named Target
1	Target	Crocodile, Blue Ball	Crocodile
1	Target	Red Trousers, Elephant	Elephant
1	Target	Red Toothbrush, Cheese	Cheese
1	Target	Carrot, Green Bike	Carrot
1	Unrelated Distractor	Brown Dog, Black Shoe	House
1	Unrelated Distractor	Green Car, Red Apple	Table
1	Unrelated Distractor	Blue Helicopter, White Truck	Hat
1	Unrelated Distractor	Black Watch, Yellow Plane	Book
1	Related Distractor (SA)	White Plate, Fish	Pig
1	Related Distractor (CA)	Green Bowl, White T-Shirt	Sheep

1	Related Distractor (SF)	Sandwich, Green Sock	Chocolate
1	Related Distractor (CF)	Yellow Nappy, Purple Dress	Banana
1	Colour Discrimination (CDA)	Black Button, Brown Button	Monkey
1	Colour Discrimination (CDA)	Green Bib, Blue Bib	Frog
1	Colour Discrimination (CDF)	Red Zip, White Zip	Milk
1	Colour Discrimination (CDF)	Red Block, Yellow Block	Strawberry
1	Filler	Telephone, Hand	Telephone
1	Filler	Balloon, Doll	Doll
1	Filler	Tree, Fork	Tree
1	Filler	Train, Spoon	Train
1	Filler	Door, Cat	Door
1	Filler	Cat, Door	Cat
1	Filler	Train, Spoon	Spoon
1	Filler	Fork, Tree	Fork
1	Filler	Doll, Balloon	Balloon
1	Filler	Telephone, Hand	Hand
2	Target	Blue Ball, Crocodile	Crocodile
2	Target	Elephant, Red Trousers	Elephant
2	Target	Cheese, Red Toothbrush	Cheese
2	Target	Green Bike, Carrot	Carrot
2	Unrelated Distractor	White Truck, Blue Helicopter	Table
2	Unrelated Distractor	Yellow Plane, Red Apple	Hat
2	Unrelated Distractor	Black Shoe, White T-Shirt	House
2	Unrelated Distractor	Black Watch, Brown Dog	Book
2	Related Distractor (SA)	Purple Dress, Lion	Sheep
2	Related Distractor (CA)	Pink Mitten, Green Sock	Pig
2	Related Distractor (SF)	Blue Nappy, Green Apple	Banana

2	Related Distractor (CF)	Flower, Brown Bag	Chocolate
2	Colour Discrimination (CDA)	Green Zip, Black Zip	Frog
2	Colour Discrimination (CDA)	White Block, Brown Block	Monkey
2	Colour Discrimination (CDF)	Yellow Button, Red Button	Strawberry
2	Colour Discrimination (CDF)	White Bib, Blue Bib	Milk
2	Filler	Telephone, Hand	Telephone
2	Filler	Balloon, Doll	Doll
2	Filler	Tree, Fork	Tree
2	Filler	Train, Spoon	Train
2	Filler	Door, Cat	Door
2	Filler	Cat, Door	Cat
2	Filler	Train, Spoon	Spoon
2	Filler	Fork, Tree	Fork
2	Filler	Doll, Balloon	Balloon
2	Filler	Telephone, Hand	Hand
3	Target	Black Shoe, Pig	Pig
3	Target	Pink Mitten, Sheep	Sheep
3	Target	Green Sock, Chocolate	Chocolate
3	Target	Banana, White Truck	Banana
3	Unrelated Distractor	Green Bike, Blue Ball	Table
3	Unrelated Distractor	Yellow Plane, Red Toothbrush	Hat
3	Unrelated Distractor	Black Watch, Yellow Nappy	House
3	Unrelated Distractor	Red Apple, Purple Dress	Book
3	Related Distractor (SA)	Black Dog, Blue Cup	Monkey
3	Related Distractor (CA)	Brown Table, Green Bowl	Frog
3	Related Distractor (SF)	Sandwich, Pink Mitten	Milk
3	Related Distractor (CF)	Blue Trousers, Red Plate	Strawberry

3	Colour Discrimination (CDA)	Green Zip, Yellow Zip	Crocodile
3	Colour Discrimination (CDA)	Green Bib, Grey Bib	Elephant
3	Colour Discrimination (CDF)	Yellow Button, Blue Button	Cheese
3	Colour Discrimination (CDF)	Red Block, Orange Block	Carrot
3	Filler	Telephone, Hand	Telephone
3	Filler	Balloon, Doll	Doll
3	Filler	Tree, Fork	Tree
3	Filler	Train, Spoon	Train
3	Filler	Door, Cat	Door
3	Filler	Cat, Door	Cat
3	Filler	Train, Spoon	Spoon
3	Filler	Fork, Tree	Fork
3	Filler	Doll, Balloon	Balloon
3	Filler	Telephone, Hand	Hand
4	Target	Pig, Black Shoe	Pig
4	Target	Sheep, Pink Mitten	Sheep
4	Target	Chocolate, Green Sock	Chocolate
4	Target	White Truck, Banana	Banana
4	Unrelated Distractor	Black Watch, Red Toothbrush	Table
4	Unrelated Distractor	Yellow Nappy, Purple Dress	House
4	Unrelated Distractor	Blue Ball, Yellow Plane	Hat
4	Unrelated Distractor	Red Apple, Green Bike	Book
4	Related Distractor (SA)	Fish, Sandwich	Frog
4	Related Distractor (CA)	Green Bowl, Brown T-Shirt	Monkey
4	Related Distractor (SF)	Brown Table, Green Grapes	Strawberry
4	Related Distractor (CF)	Pink Mitten, White Chair	Milk
4	Colour Discrimination (CDA)	Grey Button, Black Button	Elephant

4	Colour Discrimination (CDA)	Green Block, Yellow Block	Crocodile
4	Colour Discrimination (CDF)	Orange Zip, White Zip	Carrot
4	Colour Discrimination (CDF)	Yellow Bib, Blue Bib	Cheese
4	Filler	Telephone, Hand	Telephone
4	Filler	Balloon, Doll	Doll
4	Filler	Tree, Fork	Tree
4	Filler	Train, Spoon	Train
4	Filler	Door, Cat	Door
4	Filler	Cat, Door	Cat
4	Filler	Train, Spoon	Spoon
4	Filler	Fork, Tree	Fork
4	Filler	Doll, Balloon	Balloon
4	Filler	Telephone, Hand	Hand
5	Target	Black Watch, Frog	Frog
5	Target	Blue Helicopter, Monkey	Monkey
5	Target	Strawberry, Brown Table	Strawberry
5	Target	Black T-Shirt, Milk	Milk
5	Unrelated Distractor	Yellow Plane, Red Toothbrush	Hat
5	Unrelated Distractor	Black Shoe, Green Sock	Table
5	Unrelated Distractor	Red Plate, Blue Trousers	House
5	Unrelated Distractor	Green Apple, Purple Dress	Book
5	Related Distractor (SA)	Lion, Blue Cup	Crocodile
5	Related Distractor (CA)	Green Bowl, White Truck	Elephant
5	Related Distractor (SF)	Red Chair, Cookie	Cheese
5	Related Distractor (CF)	Orange Dress, Brown Dog	Carrot
5	Colour Discrimination (CDA)	White Button, Red Button	Sheep
5	Colour Discrimination (CDA)	Green Bib, Pink Bib	Pig

5	Colour Discrimination (CDF)	Black Zip, Yellow Zip	Banana
5	Colour Discrimination (CDF)	Brown Block, Blue Block	Chocolate
5	Filler	Telephone, Hand	Telephone
5	Filler	Balloon, Doll	Doll
5	Filler	Tree, Fork	Tree
5	Filler	Train, Spoon	Train
5	Filler	Door, Cat	Door
5	Filler	Cat, Door	Cat
5	Filler	Train, Spoon	Spoon
5	Filler	Fork, Tree	Fork
5	Filler	Doll, Balloon	Balloon
5	Filler	Telephone, Hand	Hand
6	Target	Monkey, Blue Helicopter	Monkey
6	Target	Frog, Black Watch	Frog
6	Target	Milk, Black T-Shirt	Milk
6	Target	Brown Table, Strawberry	Strawberry
6	Unrelated Distractor	Yellow Plate, Blue Cup	Hat
6	Unrelated Distractor	Orange Dress, Green Sock	Table
6	Unrelated Distractor	Red Plate, Blue Ball	Book
6	Unrelated Distractor	Yellow Plane, Black Shoe	House
6	Related Distractor (SA)	Red Chair, Brown Dog	Elephant
6	Related Distractor (CA)	Pink Mitten, Green Bowl	Crocodile
6	Related Distractor (SF)	White Truck, Green Grapes	Carrot
6	Related Distractor (CF)	Yellow Nappy, Red Toothbrush	Cheese
6	Colour Discrimination (CDA)	Pink Zip, Blue Zip	Pig
6	Colour Discrimination (CDA)	White Bib, Green Bib	Sheep
6	Colour Discrimination (CDF)	White Button, Brown Button	Chocolate

6	Colour Discrimination (CDF)	Black Block, Yellow Block	Banana
6	Filler	Telephone, Hand	Telephone
6	Filler	Balloon, Doll	Doll
6	Filler	Tree, Fork	Tree
6	Filler	Train, Spoon	Train
6	Filler	Door, Cat	Door
6	Filler	Cat, Door	Cat
6	Filler	Train, Spoon	Spoon
6	Filler	Fork, Tree	Fork
6	Filler	Doll, Balloon	Balloon
6	Filler	Telephone, Hand	Hand