Category Learning in Infants and Adults

Jelena Sučević

St Hugh's College

Department of Experimental Psychology
University of Oxford



A Thesis presented for the degree of

Doctor of Philosophy

Trinity term 2018

"Truth is beautiful and divine, no matter how humble its origin; it is the same in the musty boiler-room as it is in the glorious stars of heaven."

Mihajlo Pupin

Category Learning in Infants and Adults

Jelena Sučević

St Hugh's College, University of Oxford

Submitted for the degree of Doctor of Philosophy

Trinity term 2018

Abstract

The human brain has a remarkable ability to organise knowledge into a structured system. Categorisation represents one of the main mechanisms from which this system of structured knowledge is derived. The aim of this thesis is to provide a better understanding of novel category learning in infants and in adults.

In Part I of the thesis, I explore how various parameters of the learning situation shape category learning in infants. In Experiments 1 – 3, I investigate the role of working memory demands and perceptual load in category learning. The number of items simultaneously available during learning was systematically varied in these experiments. Using a standard familiarisation-novelty preference procedure, infants were first presented with a set of training items, followed by the test phase in which category formation was assessed with the presentation of two items. Preference for one of the two objects was used as an index of category formation. Familiarisation was interleaved with

test blocks in order to explore category formation at different levels of familiarity and to evaluate the dynamics of the learning process. To determine how an opportunity to compare items affects category learning, infants were familiarised with one item at a time (Experiments 1 and 1a), with a pair of items (Experiment 2), and with four items simultaneously (Experiment 3). The results revealed that paired presentation provides an optimal learning context, as infants showed evidence of category learning already after one familiarisation block. In Experiments 4–7, I explore the role of labels and motion in category learning with a gaze-contingent adaptation of the standard familiarisation-novelty preference procedure. The results indicate that adding motion as a category feature hinders visual category learning. In contrast, labels facilitate category learning, but only when both categories are accompanied by a category-specific label. When only one category is labelled, no learning can be observed.

In Part II of the thesis, I investigate whether the same kind of implicit unsupervised category learning seen in infants is also present in adults, and how category learning shapes the visual processing of objects at the neural level.

In Experiment 8, I develop a novel paradigm to explore implicit, unsupervised category learning in adults. There was no explicit mention of categories, and participants were not aware of the main aim of the task. The results revealed that adults, in the same way as infants, use statistical regularities to spontaneously organise a stream of visually presented items into categories.

In Experiment 9, I investigated neural signatures of novel category learning in adults. By comparing neural signatures of visual object processing at the very beginning of the category learning process and after an extensive categorisation practice, I tested how categorisation experience alters the way in which exemplars are processed. The results revealed that categorisation experience is linked to a modulation of several event-related potentials components suggesting that categorisation may affect visual object processing at an early stage.

In summary, this thesis demonstrates that category learning is a dynamic process shaped by different factors, which determine outcomes of the learning situation. In addition, the findings suggest that similar kinds of statistical category learning are present in infancy and adulthood.

Acknowledgements

First and foremost, I would like to thank my supervisors Kim Plunkett and Nadja Althaus for their knowledge, guidance and patience. I am very appreciative of their insightful and tireless suggestions. I would also like to thank Mihaela Duta for her help and continuous support.

I am grateful to St Hugh's College for its valuable support, inspiring academic atmosphere, and making me feel part of the family. I would particularly like to thank my college advisor Martin O'Neil for his support.

I am also deeply grateful to Aditi Lahiri and Colin Brooks, and all members of the Language & Brain Lab, for generously providing alternative lab facilities during the times when my own department was closed and for their support and incredible kindness.

Special thanks to my labmates Irina Lepadatu, Janette Chow, Klara Horvath, Samuel Forbes, Marlene Spangenberg and Amanda Griffin for making the BabyLab such a nice and friendly place to work.

I owe great thanks to Vanja Ković for encouraging me to explore this wonderful and exciting world of science and guiding my first steps into research. Suzy Styles and Olivera Savić showed me that research can be a creative and

fun process and I thank them for that.

I would like to thank my wonderful friends Jovana Mitrović, Purnima Kumar and Agota Marton. I feel so lucky for having them close and sharing this journey with them. Thank you, Nora Rouast for all the awesome tea times. Thanks also to my friends Melanie Scheriner, Bojan Banjac, Ivana Kuzmanović, Marija Paunović and Marija Minić for staying in touch and creating the world out-there for me.

I owe my deepest gratitude to my family, my parents Nada and Jovan Sučević, my brother, Nikola Sučević and my sister-in-law Ljiljana Sučević for their unconditional love, support and faith in me. This thesis is dedicated to them.

Contents

Abstract **iii**

Acknowledgements **vi**

1 General Introduction **1**

1.1 Defining categories and concepts 3

1.2 Conceptual development during the first year of life 4

1.2.1 Paradigms investigating infants' learning 5

1.2.2 Categorisation development 8

1.2.3 Learning correlation-based categories 10

1.2.4 Visual categorisation in a multimodal world 14

1.3 What enables infants to form categories? 16

1.4 Nature of category representations 18

1.5 Statistical learning underlying category formation 19

1.6 Perceptual constraints on infants' statistical learning 20

1.7 Task-dependent category learning 22

1.8 Implicit category learning in adults 23

1.9 Outline of the thesis 27

2 Methodology **31**

2.1 Eye-tracking 31

2.2 Event-related potentials technique 34

I Category Learning in Infants **38**

3 Memory and perceptual load in infants' category learning **39**

3.1 Overview of the experiments 41

3.2 Methods 44

3.2.1 Stimuli 44

3.2.2 Experimental design 49

3.2.3 Participants 52

3.2.4 Statistical analysis 52

3.3 Experiment 1: single-item familiarisation 56

3.3.1 Introduction 56

3.3.2 Method 56

3.3.3 Results 58

3.3.4 Experiment 1a: control experiment 60

3.3.5 Discussion 65

3.4 Experiment 2: paired familiarisation 69

3.4.1 Introduction 69

3.4.2 Method 72

3.4.3 Results 74

3.4.4 Discussion 86

3.5 Experiment 3: massed familiarisation 91
3.5.1 Introduction 91
3.5.2 Method 92
3.5.3 Results 94
3.5.4 Discussion 107
3.6 Habituation rates across experiments 110
3.7 Chapter 3: Discussion 112

4 Motion and labels in infants' category learning 117

4.1 Overview of the studies 121
4.2 Methods 124
4.2.1 Stimuli 124
4.2.2 Experimental design 124
4.2.3 Participants 129
4.2.4 Statistical analysis 129
4.3 Experiment 4: Contrasting dynamic and static categories . . . 132
4.3.1 Introduction 132
4.3.2 Method 133
4.3.3 Results 134
4.3.4 Discussion 150
4.4 Experiment 5: Dynamic categories 152
4.4.1 Introduction 152
4.4.2 Method 153
4.4.3 Results 155

4.4.4 Infants' exploration strategies in training 159
4.4.5 Exploration strategies in training and performance in test 161
4.4.6 Moving objects test 162
4.4.7 Discussion 162
4.5 Experiment 6: Contrasting labelled and silent categories . . . 165
4.5.1 Method 166
4.5.2 Results 167
4.5.3 Exploration strategies in training and performance in test 176
4.5.4 Discussion 179
4.6 Experiment 7: Labelled categories 182
4.6.1 Introduction 182
4.6.2 Method 183
4.6.3 Results 184
4.6.4 Exploration strategies in training and performance in test 190
4.6.5 Discussion 191
4.7 Chapter 4: Discussion 192

II Category Learning in Adults 197

5 Implicit categorisation in adults 198

5.1 Introduction 198

5.2 Method . . . 202
5.2.1 Participants . . . 202
5.2.2 Stimuli . . . 203
5.2.3 Experimental design . . . 203
5.2.4 Equipment and experimental procedure . . . 208
5.2.5 Analysis plan . . . 208
5.3 Results . . . 209
5.3.1 Training . . . 210
5.3.2 Test . . . 210
5.4 Discussion . . . 214

6 Neurophysiological signatures of category learning 217

6.1 Introduction . . . 217
6.2 Method . . . 220
6.2.1 Participants . . . 220
6.2.2 Experimental design . . . 221
6.2.3 Stimuli . . . 224
6.2.4 Equipment . . . 225
6.2.5 Experimental procedure . . . 226
6.3 Experimental measures . . . 227
6.3.1 Behavioural measures . . . 227
6.3.2 EEG measures . . . 227
6.4 Analysis plan . . . 228
6.4.1 Behavioural analysis plan . . . 228

6.4.2 EEG analysis plan 230
6.5 Results: Behavioural data 233
6.5.1 Training 233
6.5.2 Test 235
6.6 Results: Event-related potentials 238
6.6.1 Training 238
6.6.2 Test 243
6.7 Discussion 253

7 General Discussion 258

7.1 Overview of main findings 258
7.1.1 Main findings part I - category learning in infants . . . 258
7.1.2 Main findings part II - category learning in adults . . . 262
7.2 Implications 265
7.2.1 Optimising the category learning context 265
7.2.2 Labels and motion in category learning 269
7.2.3 Implicit category learning in adults 272
7.2.4 Neural signatures of category learning 272
7.3 Methodological limitations 273
7.4 Theoretical limitations 275
7.5 Future directions 276
7.6 Conclusion 277

Appendix 279

A Experiment 1 **279**

B Experiment 1a **281**

C Experiment 2 **282**

D Experiment 9 **284**

List of Figures

1.1 Dot pattern stimuli used by Bomba and Siqueland (1983) to study infants' ability to learn artificial categories: triangle category prototype and training exemplars. 9

1.2 Stimuli used by Younger (1985) and Plunkett et al. (2008) to study infants ability to learn correlation-based categories. . . . 12

1.3 Stimuli used by Younger (1992) to study infants ability to learn categories of faces based on correlations among facial features: three training and two test items. 14

1.4 An example of a category structure used in a study by Clapper and Bower (1994). Columns in grey depict less prognostic features. 25

2.1 Setup used in eye-tracking studies. 34

2.2 Main ERP components (reproduced from Daltrozzo & Conway 2014). 36

2.3 Participant preparation for the ERP study (photo reproduced with permission of the Language & Brain Laboratory, University of Oxford). 37

3.1 Outline of the chapter: The experiments presented use an identical experimental design with interleaved training and test trials (grey boxes refer to training blocks, red boxes depict test blocks). Across Experiments 1, 2 and 3 the amount of information immediately available was manipulated by varying the number of items simultaneously presented. All experiments employed an identical test procedure based on novelty preference (illustrated in the bottom part of the figure). 43

3.2 Schematic depiction of the category structure and stimuli examples in Plunkett et al. (2008) and in the present study: each stimulus can be described in 4-dimensional perceptual space (each dimension resembles one feature, i.e. arms, hands, body and legs). 45

3.3 The Suki set: feature range. 46

3.4 The Suki set with correlated feature values. 47

3.5 Novel, out-of-category test item. 48

3.6 The Suki set with uncorrelated feature values. 48

3.7 Schematic depiction of the experimental design: three familiarisation blocks interleaved with three test blocks. 49

3.8 Structure of the test blocks: two category formation trials followed by a novelty preference test. 51

3.9 Outline of the test design: inter-category item (4444) and a category prototype (either 2222 or 6666) are presented in test (a); If infants formed two categories, the inter-category item will look more interesting and novel as it lies between the two categories (b); if infants formed one category, the inter-category item will be less interesting as it is closer to the average of this category (c). 55

3.10 Schematic depiction of the experimental procedure: three familiarisation blocks were interleaved with three test blocks. . . 57

3.11 Experiment 1: Structure of familiarisation trials. 58

3.12 Experiment 1: Mean looking time across infants in each training block (dashed line depicts mean looking times in each trial, solid line represents model values, vertical dotted lines represent time points when test blocks were presented, grey envelope depicts standard error). 59

3.13 Experiment 1: Looking preferences in the test blocks (error bars depict standard error, red dotted line depicts expected preference if performance was at chance level, higher preference value indicates preference for the inter-category item). . . 60

3.14 Experiment 1a: Mean looking time across infants in each training block (dashed line depicts mean looking times in each trial, solid line represents model values, vertical dotted line represents time points when test blocks were presented, grey envelope depicts standard error). 63

3.15 Experiment 1a: Looking preferences in the test blocks (error bars depict standard error, red dotted line depicts expected preference if performance was at chance level, higher preference value indicates preference for the category prototype). . . 64

3.16 Experiment 1a: Looking preferences in novelty preference tests (error bars depict standard error, red dotted line depicts expected preference if performance was at chance level, higher preference value indicates preference for the novel item). . . . 65

3.17 Experiment 2: Structure of the familiarisation trials. 73

3.18 Experiment 2: Mean looking time across infants in each training block (dotted line depicts mean looking times in each trial, solid line represents model values, vertical dotted line represents time points when test blocks were presented; grey envelope depicts standard error). 74

3.19 Experiment 2: Looking preferences in 3 test blocks (higher preference value means that they prefer the inter-category item, whereas lower preference value indicates preference for the category prototype). 75

3.20 Mean number of shifts in each trial over the course of three familiarisation blocks (grey squares depict mean number of shifts for within category comparisons, red triangles depict between categories comparisons; error bars depict standard errors; grey and red lines depict model predictions for within and between categories comparisons, respectively). 77

3.21 Looking preferences of low and high shifters in the first test block (x-axis depicts trial time course in seconds; y-axis depicts elog-transformed novelty preference). 80

3.22 Mean looking time of low- and high-shifters in each training block in Experiment 2 (grey squares depict low-shifters, red triangles high-shifters; error bars depict standard error; grey and red lines depict model predictions for low- and high-shifters, respectively). 82

3.23 Fixation characteristics of low- and high- shifters (values on y-axis are given in milliseconds; boxplot depicts median and interquartile range). 83

3.24 The proportion of looking time outside of the areas of interest in infants with low and high shifting rates (error bars depict standard error). 84

3.25 Total looking time accumulated during familiarisation blocks in Experiment 1 (successive familiarisation, depicted in light grey) and Experiment 2 (paired familiarisation, depicted in dark grey; error bars depict standard error). 86

3.26 Experiment 3: Structure of familiarisation trials. 93

3.27 Comparison of corresponding sequences of one familiarisation blocks across Experiment 1, 2 and 3: (a) items presented one at a time in Experiment 1; (b) pairs of items presented in Experiment 2; (c) quadruplets presented in Experiment 3. . . 94

3.28 Experiment 3: Mean looking time across infants in each training block (dotted line depicts mean looking times in each trial, solid line represents model values; vertical dotted lines represent time points when test blocks were presented; grey envelope depicts standard error). 95

3.29 Experiment 3: Looking preferences in the test blocks (error bars depict standard error, red dotted line depicts expected preference if performance was at chance level, higher preference value indicates preference for the inter-category item). . . 96

3.30 Experiment 3: Mean looking preference for the novel item across infants in the test blocks (error bars depict standard error, red dotted line depicts expected preference if performance was at chance level, higher preference value indicates preference for the novel item). 97

3.31 Experiment 3: Mean number of shifts across infants in each trial of the three familiarisation blocks (dashed line depicts mean number of shifts in each trial, solid line represents model values, vertical dotted lines represent time points when test blocks were presented, grey envelope depicts standard error). . 99

3.32 Markov chain with two states and four potential transitions between the states. 101

3.33 Mean transitional probabilities for *between-* and *within-category* shifts in the three familiarisation blocks (red dotted line depicts expected transitional probability if performance was at chance level). 103

3.34 Relationship between mean number of shifts and between-category transitional probabilities. 104

3.35 Time-course of the novelty preference in the low- and high-shifting groups (a) and in the low- and high-transitional probability groups (b), (error bars depict standard error). 107

3.36 Mean looking time in three experiments (squares, triangles and circles depict mean looking time in single, paired and massed experiment, respectively; solid, dashed and dotted lines represent model values for each experiment; vertical dotted line represents time points when test blocks were presented).111

3.37 Perceptual load hypothesis: Novelty preference as a function of perceptual load. 114

4.1 Outline of the chapter: The experiments presented use an identical experimental design with interleaved training and test trials (grey boxes refer to training blocks, red boxes depict test blocks). Experiment 4 and 5 explore the role of motion in category learning, whereas Experiment 6 and 7 investigate the role of labels in categorisation. All experiments used a gaze-contingent familiarisation paradigm and an identical test procedure based on novelty preference (illustrated in the bottom part of the figure). 123

4.2 Schematic depiction of the experimental design: three familiarisation blocks interleaved with three test blocks. 125

4.3 Schematic depiction of the test trials: category formation (a and b) and category preference test trial (c), (novelty preference test not depicted). 128

4.4 Experiment 4: Structure of familiarisation trials. 134

4.5 Experiment 4: The mean amount of looking time to dynamic and static category exemplars in each training block (error bars depict standard error). 135

4.6 Experiment 4: Mean looking time for static and dynamic category exemplars over the course of familiarisation (squares and triangles depict mean looking times in each trial for static and dynamic category, respectively; solid and dashed line represents model values for static and dynamic category; error bars depict standard error). 136

4.7 Experiment 4: Looking preferences in the dynamic category formation test (a), and when presented as the first test trial (b), (values higher than .5, depicted by a horizontal dashed red line, suggest a preference for the inter-category item, error bars depict standard error, error bars depict standard error). . 138

4.8 Experiment 4: Looking preferences in the static category formation test (a) and when presented as the first test trial (b), (values higher than .5, depicted by a horizontal dashed red line, suggest a preference for the inter-category item error bars depict standard error). 139

4.9 Experiment 4: Looking preferences in the between categories preference test (a) and when presented as the first trial (b), (values higher than .5, depicted by a horizontal dashed red line, suggest a preference for the static category prototype, error bars depict standard error). 141

4.10 Experiment 4: Looking preferences in the novelty preference test (values higher than .5, depicted by a horizontal dashed red line, suggest a preference for the inter-category item, error bars depict standard error). 142

4.11 Experiment 4: Mean shifting rates (a); Proportions of the between-category and within-category shifts (b); proportion of shifts within the dynamic category and within the static category (c), (dotted line depicts shifting rates at chance). . . 144

4.12 Experiment 4: Transitional probabilities in each training block for within-category shifts (a) and for the between-category shifts (b), (dotted line depicts performance at chance, error bars depict standard error). 146

4.13 Experiment 4: Performance of infants engaged in similarity-focused and contrast-focused exploration strategies (error bars depict standard error). 149

4.14 Experiment 4: Structure of familiarisation trials. 154

4.15 Experiment 5: Mean looking time in each training block (category 1 was a set of stimuli defined by low values on feature range, whereas category 2 was defined by high values on the feature range, error bars depict standard error). 155

4.16 Experiment 5: Mean looking time for exemplars from the two categories over the course of familiarisation (squares and triangles depict the two categories, solid and dashed line represent model values, vertical dotted lines represent time points when test blocks were presented, error bars depict standard error). . 156

4.17 Experiment 5: Mean looking preferences in category formation test (higher preference value means that they prefer the inter-category item, whereas lower preference value indicates preference for the category prototype). 157

4.18 Experiment 5: Mean looking preferences in the between category formation test 158

4.19 Experiment 5: Mean looking preferences in the novelty preference test (higher preference value higher than .5 indicates preference for the novel item; value of 0 indicates familiarity preference). 158

4.20 Experiment 5: Mean shifting during familiarisation blocks (a) and proportion of within-category and between-category shifts (dotted line depicts performance at chance). 160

4.21 Experiment 5: Mean transitional probabilities during familiarisation (circles and triangles depict mean transitional probabilities for within- and between shifts, respectively; error bars depict standard error). 161

4.22 Experiment 6: Structure of familiarisation trials. 166

4.23 Experiment 6: Mean looking time in the three familiarisation blocks (error bars depict standard error). 167

4.24 Experiment 6: Mean looking time across infants over the course of familiarisation (red triangles and grey squares depict mean looking times for labelled and silent categories, respectively; dashed red line and grey solid line represent model values for labelled and silent category, respectively; vertical dotted lines represent time points when test blocks were presented, error bars depict standard error). 168

4.25 Experiment 6: Mean looking preferences in the labelled category formation test (a) and in the first test trial only (b), (values higher than .5 (horizontal dashed red line) suggest a preference for the inter-category item). 169

4.26 Experiment 6: Looking preferences in the silent category formation test (a) and in the subset of trials presented first in each test (b), (values higher than .5 (horizontal dashed red line) suggest a preference for the inter-category item). 170

4.27 Experiment 6: Looking preferences in the between categories preference test (a) and in the subset of trials presented first in each test (b) (values higher than .5 (horizontal dashed red line) suggest a preference for the silent category). 171

4.28 Experiment 6: Looking preferences in the novelty preference test (values higher than .5 (horizontal dashed red line) suggest a preference for the novel object) 172

4.29 Experiment 6: Mean shifting rates in the three familiarisation blocks (a); Proportions of the between-category and within-category shifts (b); proportion of shifts within the labelled category and within the silent category (c), (dotted line depicts shifting rates at chance). 174

4.30 Experiment 6: Transitional probabilities in each training block for within-category shifts (a) and for the between-category shifts (b), (dotted line depicts performance at chance, error bars depict standard error, however, not visible due to small values). 176

4.31 Experiment 6: Performance of low- and high-shifters in test. . 177

4.32 Experiment 6: Test performance of infants engaging contrast-focused (high probability of label-to-silent transition) and similarity-focused strategies (low label-to-silent transitional probability). 178

4.33 Experiment 6: Test performance of infants with low probability of silent-to-label transition and infants with high silent-to-label transitional probability. 179

4.34 Experiment 7: Structure of familiarisation trials. 183

4.35 Experiment 7: Mean looking time in three familiarisation blocks (red and grey bars represent looking time accumulated for the two categories separately, error bars depict standard error). . . 184

4.36 Experiment 7: Mean looking time across infants in each training block (red triangles and grey squares depict mean looking times for the two categories; dashed red line and grey solid line represent model predictions; error bars depict standard error). 185

4.37 Experiment 7: Looking preferences in the category formation test (higher preference value indicates preference for the inter-category item). 186

4.38 Experiment 7: Looking preferences in the between-categories preference test. 186

4.39 Experiment 7: Looking preferences in the novelty preference test (higher preference value indicates novelty preference). . . 187

4.40 Experiment 7: Mean shifting rates in the three familiarisation blocks (a) and mean shifting proportions for between and within-category shifts (b). 189

4.41 Experiment 7: Transitional probabilities for the within- and between-category shifts in the three training blocks (dashed grey line depicts chance level performance for *within-category* and dotted grey line for *between-category* transitions). 190

5.1 Visual similarity for *within-* and *between-category* comparisons was variable in the training blocks (a), but kept constant in the test blocks (b). 206

5.2 Experiment structure in training (a) and test blocks (b). . . . 207

5.3 Example of the trial sequence. 207

5.4 Mean reaction time over the course of the seven blocks for different comparisons in training blocks (a) and in test blocks (b and c), (error bars depict standard error). 212

6.1 Training phase: Trial timeline. 222

6.2 Test phase: Trial timeline. 223

6.3 Stimuli used in the training phase (a) and in the test phase (b), (numbers below each item represent the value of each for the four features). 225

6.4 BioSemi ActiveTwo 64-channel position outline used for data acquisition (retrieved from *BioSemi* (n.d.)) and response box used to record behavioural responses (retrieved from *Cedrus* (n.d.)). 226

6.5 Electrode sites included in the analyses: three bands with coronal orientation subdivided by hemisphere: anterior left (Fp1, AF3, AF7, F1, F3, F5, F7), anterior right (Fp2, AF4, AF8, F2, F4, F6, F8), central left (FC1, FC3, FC5, FT7, C1, C3, C5, T7, CP1, CP3, CP5, TP7), central right (FC2, FC4, FC6, FT8, C2, C4, C6, T8, CP2, CP4, CP6, TP8), posterior left (P1, P3, P5, P7, P9, PO3, PO7, O1), and posterior right (P2, P4, P6, P8, P10, PO4, O2). 230

6.6 Mean amplitude was analysed for three time windows: N200, P300 and late positivity time window. 232

6.7 Mean accuracy (a) and mean reaction time (b) across participants over the course of training blocks (the dashed line depicts the mean accuracy in each trial; the solid line represents the model prediction; grey envelope depicts standard error). 234

6.8 Mean reaction time for trained-item test trials (a) and for novel-item test trials (b), (solid line represents model prediction, error bars depict standard error). 237

6.9 N200 amplitudes in training blocks. (*) $p < .07$; * $p < .05$; ** $p < .01$. 240

6.10 P300 amplitudes in training blocks. (*) $p < .07$; * $p < .05$; ** $p < .01$. 241

6.11 late positivitu: Mean amplitudes in the training blocks. (*) $p < .07$; * $p < .05$; ** $p < .01$. 242

6.12 N200 amplitudes in within-category priming trials in test blocks. (*) $p < .07$; * $p < .05$; ** $p < .01$. 246

6.13 P300 amplitudes for within-category priming trials in test blocks. (*) $p < .07$; * $p < .05$; ** $p < .01$. 247

6.14 Late positivity amplitude differences for within-category priming trials in test block. (*) $p < .07$; * $p < .05$; ** $p < .01$. . . . 248

6.15 N200 amplitudes for between-category priming in test blocks. (*) $p < .07$; * $p < .05$; ** $p < .01$. 250

6.16 P300 amplitudes for between-category priming in test blocks. (*) $p < .07$; * $p < .05$; ** $p < .01$. 251

6.17 Late positivity effects for between-category priming in test blocks. (*) $p < .07$; * $p < .05$; ** $p < .01$. 252

6.18 Summary of the ERP analyses: (significance levels: (*) $p <$ .07; * $p < .05$; ** $p < .01$). 253

7.1 Thesis Part I: Chapter 3 demonstrated that paired familiarisation represents optimal learning context for category learning in 10-month-old infants. 260

7.2 Thesis Part I: Chapter 4 demonstrated that motion hinders and labels facilitate the process of visual category learning in 10-month-old infants. 262

7.3 Thesis Part II: Chapter 5 demonstrated that adults are able to extract category structure from a stream of visually presented objects. 263

7.4 Thesis Part II: Chapter 6 demonstrated that categorisation experience was linked to a modulation of several ERP components. 264

7.5 The role of memory and perceptual load in category learning: paired presentation represents the optimal learning context (area depicted in grey). 267

List of Tables

3.1 Experiment 1: Growth Curve Analysis. Parameter estimates. . 58

3.2 Experiment 1a: Growth Curve Analysis. Parameter estimates. 63

3.3 Experiment 2: Growth Curve Analysis. Parameter estimates. . 74

3.4 Experiment 2: Mean looking preferences on test (SDs provided in brackets. 76

3.5 Experiment 2: Growth Curve Analysis. Parameter Estimates. 81

3.6 Experiment 3: Growth Curve Analysis. Parameter estimates. . 95

3.7 Habituation in Experiment 1, 2 and 3: Growth Curve Analysis (parameter estimates). 111

4.1 Experiment 4: Growth Curve Analysis. Parameter estimates. 136

4.2 Experiment 5: Growth Curve Analysis. Parameter estimates . 156

4.3 Experiment 6: Growth Curve Analysis. Parameter estimates . 168

4.4 Experiment 7: Growth Curve Analysis. Parameter estimates. 185

5.1 Trial statistics (trials were excluded from analysis if the reaction time was above two standard deviations from the mean). 210

5.2 Mean accuracy (proportion of correct response) and standard deviation in seven blocks of training (upper section) and test (lower section of the table). 210

5.3 Descriptive statistics for reaction time in training and test (in milliseconds). 211

5.4 Growth Curve Analysis: Parameter estimates. 213

6.1 The growth curve analysis on the mean reaction time in training. 234

6.2 Mean accuracy in test blocks. 235

6.3 Growth curve analysis on mean reaction time across test blocks. 237

A.1 Experiment 1: Stimuli structure used in three familiarisation blocks (each number represents a value on each of the four features). 279

B.1 Experiment 1a: Stimuli structure used in three familiarisation blocks (each number represents a value on each of the four features). 281

C.1 Experiment 2: Growth curve analysis for the decrease in the amount of looking time over the course of the experiment. Parameter Estimates. 282

C.2 Experiment 2: Parameter Estimation for the Growth Curve Analysis - modelling novelty preference in low- and high-shifters). 283

D.1 Structure of the stimulus set used in the ERP experiment. . . 284

Chapter 1

General Introduction

Imagine visiting a new library for the first time. While admiring the hundreds of books lying on the shelves, you decide to search for your favourite novel. What is your method? You assume the books must be organised in some way. You consider several criteria to base your search on: perhaps book size, front cover colour or publishing year. Quickly you discard these strategies and assume that the books must be organised in a more systematic way. A quest for your favourite novel proves to be very efficient thanks to the fact that the library is indeed organised in a meaningful way.

Human knowledge can be seen as a library, with the mind as a librarian who organises our ideas in the form of mental books. But how does the human mind achieve this? And how does this *mindbrarian* come up with a meaningful and sophisticated principle for sorting thousands of mental books, instead of just focusing on the surface characteristics? In other words, how does the *mindbrarian* formulate abstract, general principles of organisation (for instance, a literature epoch), instead of employing merely

visual characteristics, such as book cover colour?

One of the most fascinating facts about the human mind is that, despite the infinite nature of variations of objects, knowledge is a well-structured system where objects are organised into categories and mentally represented as concepts (Murphy, 2002). Concepts take the "blooming, buzzing confusion" of sensory experiences and establish order through imposing categories (Goldstone et al., 1978). In addition, concepts represent a "mental glue", binding our past experiences to our present interactions with the world (Murphy, 2002). Hence, the categorisation ability, i.e. the ability to recognise perceptually distinct entities as the same kinds of things, is considered to be critical for the organisation and stability of cognition (Mareschal & Quinn, 2001) and fundamental to intelligent behaviour (Younger, 2010).

The main aim of the research presented in this thesis is to explore the key aspects underlying the development of such categorisation abilities, as well as the mechanisms involved in the formation of concepts. In particular, the focus is on learning novel statistical categories and determining how different factors may shape category formation in infants and adults, as well as how categorisation shapes object perception.

In the remainder of this chapter, I introduce the main theoretical concepts, provide an overview of the key studies of conceptual development during the first year of life, outline current views on the nature of these representations, and explore implicit category learning in adults. Lastly, I provide an overview of the work presented in this thesis.

1.1 Defining categories and concepts

If the human mind had no categorisation ability, each object would be unique and no generalisation of knowledge would be possible (Machery, 2009). In everyday life, that would be like living every day as it was the first day of our lives, since our inability to use our existing knowledge and memories flexibly would render them useless. In such a situation, the mind would constantly be overwhelmed, even in the context of a simple everyday activity.

Despite the fact that the terms "concepts" and "categories" are often used interchangeably, Murphy (2002) makes a distinction between the two. He sees concepts as mental representations of classes of things and categories the classes themselves. Similarly, Machery (2009) suggests that they are two sides of the same psychological phenomenon - our disposition to put individual objects into equivalence classes.

Concepts and categories are formed in a process called *categorisation*. Categorisation is often defined as the capacity to produce judgements as to which category an item belongs to (Machery, 2009). It involves treating two or more distinct entities as in some way equivalent in the service of accessing knowledge and making predictions (Medin, 1989). For many years, the main focus of categorisation research was on exploring the nature of mental representations of categories (e.g. Rosch et al, 1976), yielding several influential theories: for example, prototype theory and exemplar theory. The former assumes that a category is represented as an average, idealised representation of the properties contained within a category (Lin & Murphy, 1997),

whereas the latter proposes that more than one exemplar of the category is memorised and then used as a category representative (Reisberg, 2006).

Against this background, the goal of the present thesis is to further explore categorisation; in particular, the focus is on processes at work during novel category formation. Since understanding categorisation in infancy provides answers to some of the key questions surrounding categorisation, the main focus of the thesis is on the process of novel category learning in infants. In addition, this thesis explores whether a similar kind of category learning is present in adults. In the following section, I provide an overview of the main methodology used in categorisation studies with infants and review key discoveries that contributed to our understanding of the early stages of categorisation development. Finally, I give a brief outline of studies exploring implicit unsupervised category learning in adults.

1.2 Conceptual development during the first year of life

Concept learning is the capacity to acquire concepts from encountering some members of their extension (Machery, 2009). There are two aspects of conceptual development that need to be distinguished: infants' actual concepts and their concept-learning skills (Murphy, 2002). The present work focuses on concept-learning skills by examining the ability to acquire novel categories.

1.2.1 Paradigms investigating infants' learning

Exploring infants' looking behaviour has been one of the main tools used to investigate cognitive development for over five decades. In the late fifties, Robert Fantz (1961) established a novel method to investigate cognitive processes in infants - *the preferential looking paradigm*. Minimal response demands and a wide range of potential applications made preferential looking the most common method in infant research and an important methodological breakthrough (Houston-Price & Nakai, 2004). Fantz (1961) designed a special "looking chamber" to explore infants' perception of form. In this chamber, a baby would be lying on her back while two pictures were presented above her. An experimenter recorded how much time she spent looking at each of the pictures. Preference for one of the stimuli was interpreted as an index of various perceptual, attentional or memory processes (Houston-Price & Nakai, 2004).

Since Fantz' initial studies, various adaptations of the preferential looking task have been developed. One instance of the paradigm is *the intermodal looking preference procedure*. This task has become one of the key tools used to study word learning in young children. Here, infants are typically presented with a pair (or several) pictures while the name of one of the depicted objects is played. If, upon naming, infants look more at the named picture than a distractor image, this is interpreted as an indicator of word comprehension.

Another variant of the preferential looking task is the *familiarisation - novelty preference task*. An initial familiarisation phase in which infants are presented with a series of pictures is followed by a test trial in which a pair of pictures is presented. If infants express greater interest in one of the pictures, this preference is interpreted as an index of recognition. For instance, after being familiarised with a set of cats, when presented with a new cat and a dog, a preference for the dog is interpreted as evidence that infants recognise that the novel cat is similar to previously presented ones, i.e. they all belong to the category of cats. Therefore, categorisation is inferred if infants generalise responsiveness to a novel instance from the familiar category and display differential responsiveness to the novel instance from a novel category (Younger, 2010).

Several studies demonstrated that the level of familiarity with to-be-learned categories is one of the factors driving infants looking preferences in categorisation experiments (Hurley & Oakes, 2012; Kovack-Lesh et al., 2012). Most categorisation studies, however, use only one familiarisation block followed by one test block. Therefore, it remains unclear what is the nature of the relationship between the level of familiarity and performance. This is particularly important since category representations are dynamic in nature and constantly updated (Jones & Smith, 1993). To test whether increased exposure leads to better categorisation, in this thesis I introduce an experimental design with interleaved training and test blocks. This enables probing category formation at various familiarisation levels, each defined by

increasing levels of familiarisation.

What drives infants' looking preferences?

Common to all these tasks is that looking preferences are used as indices of various cognitive processes. A key underlying assumption in most of these tasks is that infants' looking behaviour is driven by a novelty preference. The novelty preference is a tendency to find novel objects more interesting than familiar ones and so more attention is directed towards the novel object. As pointed out by Houston-Price and Nakai (2004), this preference implies that infants discriminate between the stimuli, and that preferential attention reflects the relative salience of the stimuli.

The novelty preference represents one of the main mechanisms driving infants' behaviour. Sokolov's model of the orienting reflex is one of the widely accepted models for understanding novelty preference in infants (Sokolov, 1963). According to this model, an internal representation of the stimulus is being formed during a familiarisation phase. In the subsequent test phase, two presented stimuli are compared to this internal model. Infants tend to direct more attention to the novel stimulus because it does not match with the internal model. As pointed out by Pascalis and De Haan (2003), this model provides a valuable explanation of novelty preference, but offers no such explanation for cases in which infants express a preference for a familiar object.

In addition to novelty and familiarity preference, infants can also express no preference for either of the test items. Some authors propose that this random looking behaviour should not be seen as a failure to discriminate, but instead as an initial stage of learning. Hunter and Ames' (1988) model of infants' attentional preferences proposes that, initially, infants will show an equal interest in both familiar and novel objects. With more familiarisation experience, they will shift to a familiarity preference and finally, after a sufficient amount of familiarisation, infants will demonstrate a novelty preference.

Sophian (1980) proposes an alternative model in which the null preference is a transition phase between an initial familiarity and a final novelty preference phase (familiarity reflects incomplete encoding, whereas novelty preference results from a complete encoding of the stimulus). However, these models do not account for any effect of forgetting the stimuli and how it would be expressed in looking preferences (Pascalis & de Haan, 2003). In addition, some studies find no null preference stage in infants' looking behaviour (Roder et al., 2000). Furthermore, they propose that "the random preference" observed following the preference for familiarity in prior research might just be an artifact of data averaging (Roder et al., 2000).

1.2.2 Categorisation development

One of the first experimental studies exploring category-learning skills in young infants was carried out by Bomba and Siqueland (1983). Inspired by

the seminal work of Posner and Keele (1970), Bomba and Siqueland investigated 3- to 4-month-old infants' ability to form categories of dot patterns. Using a familiarisation-novelty preference procedure, they familiarised infants with a set of dot patterns that represented distortions of geometrical shapes (square, triangle or a circle, see Figure 1.1). When tested by presenting a prototype of the familiarised shape along with another novel shape, infants expressed a preference for the novel shape (Bomba & Siqueland, 1983). For instance, after being familiarised with distortions of triangles, infants expressed a preference for the prototype of a square as opposed to the prototype of a triangle. According to these findings, the authors concluded that the infants had formed the category "triangles".

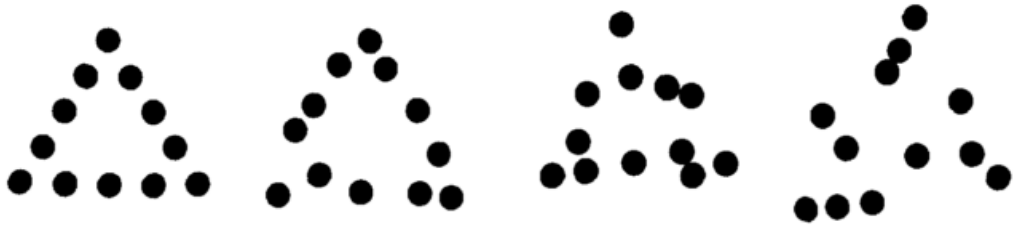


Figure 1.1. Dot pattern stimuli used by Bomba and Siqueland (1983) to study infants' ability to learn artificial categories: triangle category prototype and training exemplars.

Using the same experimental paradigm, Quinn (1987) explored whether the presence of other stimuli affect infants' ability to form a category. Whereas in Bomba and Siqueland's study (1983) infants were exposed to exemplars from only one category, Quinn tested a learning situation in which, in addition to exemplars of one category, infants are presented with other items as well. In one condition, the additional exemplars all came from the same category, whereas in another condition, items come from several different

categories. The study showed that infants are able to form the target category even when exemplars from another category were presented during familiarisation, but failed to do so when additionally presented items lacked category structure, suggesting that adding systematic variations does not hinder the category learning process, but adding random variability during learning inhibits learning in infants at this age.

An important implication of these results is that the number of items presented during familiarisation can impact category formation. In addition to the overall amount of information presented during learning (number of categories presented) and nature of the presented information (structured or random), another relevant factor is the number of items simultaneously presented in each trial. Several studies have shown that presenting items in pairs facilitates category learning (Ankowski et al., 2013; Oakes & Ribar, 2005), even if the two presented items were identical (Oakes et al., 2009). Despite the fact that multiple studies have confirmed the beneficial effect of paired presentation on learning, further research is needed to better understand what is driving these effects. This motivated a set of experiments presented in Chapter 3 in which I explore how an opportunity to compare and contrast items helps infants to uncover category structure.

1.2.3 Learning correlation-based categories

The present thesis focuses on one specific class of categories – correlation-based categories. This class of categories is particularly important because

the world is a structured system in which objects have a high correlational structure and features do not occur independently of each other (Rosch et al., 1976). For instance, animals with feathers usually have wings, whereas furry animals rarely do. Ability to detect this relational information, i.e. that certain features co-occur, is considered to play a central role in category formation (Younger & Cohen, 1986); it is assumed that it underlies categorisation in adults (Rosch et al., 1976). Therefore, mastering correlation-based categories seems to be one of the key steps in the conceptual development.

In one of the early studies, Younger & Cohen (1986) investigated the development of infants' ability to form categories based on relational information. They presented 4-, 7- and 10-month-old infants with a set of drawings of animal-looking creatures that varied in three features: body, tail and feet. For instance, an item could have a giraffe, cow or elephant body combined with a feathered, fluffy or horse tail and a specific shape of feet. Using a standard familiarisation-novelty preference procedure, infants were first familiarised with a set of training items and then tested for category learning. Younger & Cohen (1986) discovered a progression from detecting isolated features at 4 months, followed by the perception of feature correlations of single items at 7 months, and at 10 months infants formed categories based on feature correlations (Younger & Cohen, 1986).

In the aforementioned study stimulus features varied qualitatively (eg., an animal's body could take one of three different shapes). In another study, Younger (1985) tested infants' correlation-based category formation with a

set of drawings of novel animals that consisted of identical features with varying dimensions (for instance, items from two categories were identically shaped; one category had short legs, the other long legs). In one condition, values of features were correlated (Narrow Condition), i.e. values of one feature were predictive for values of another feature (for instance, animals with long legs always had short necks, as shown in Figure 1.2). In the second condition, feature values were distributed randomly (Broad Condition), e.g. animals with long legs could appear with either short or elongated necks. After being familiarised with either a broad or a narrow stimulus set, infants were tested with two novel items - one containing the average values of all features (3333) and one consisting of the extreme values (1111 or 5555).

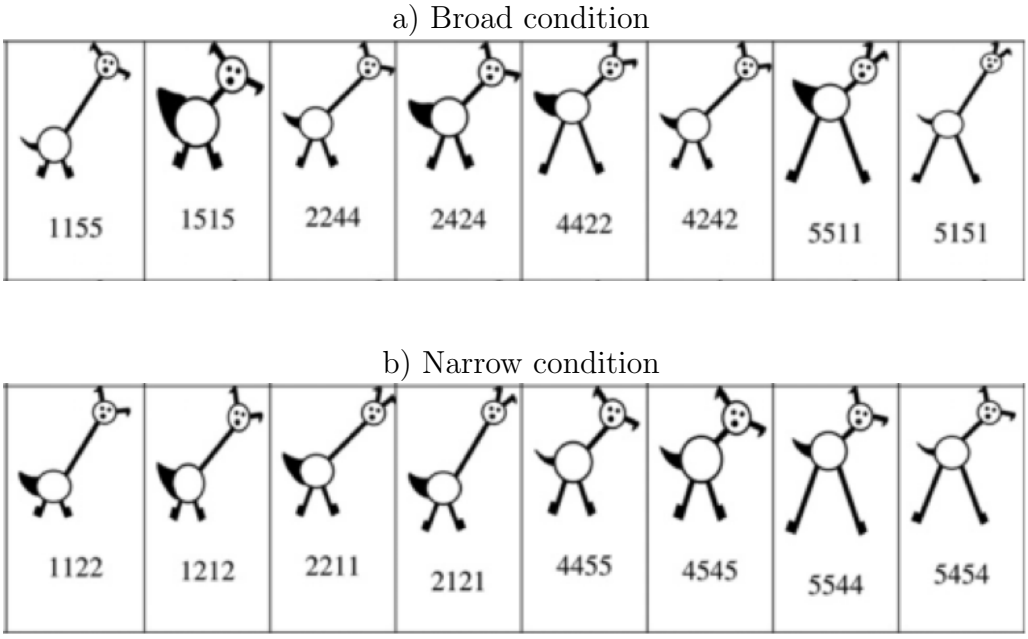


Figure 1.2. Stimuli used by Younger (1985) and Plunkett et al. (2008) to study infants ability to learn correlation-based categories.

If infants formed two categories, the average item 3333 might look more novel as it lies on the boundary of the two categories. If infants grouped all

items into one category, item 3333 would be of less interest as it is more similar to category representation. The extreme items, meanwhile (1111/5555), would look more novel as they lie on the edges of the category and further away from the category representation (detailed explanation of this experimental design is given in Chapter 3). Younger and Cohen (1986) found that infants familiarised with the narrow condition expressed a preference for the average item in the test phase (3333). By contrast, infants familiarised with the broad condition set spent more time looking at the extreme items (1111/5555), therefore suggesting that infants in the narrow condition formed two categories, while infants in the broad condition formed just one. This finding was confirmed and further extended in several studies that have shown that around 10 months of age, infants do indeed become adept at forming correlation-based categories (Mather & Plunkett, 2011; Plunkett et al., 2008; Younger, 1992; Younger & Fearing, 1999). In another study, Younger (1992) showed that the emergence of the ability to detect correlations and use them as a basis for categorisation is tied not only to artificial stimuli, but also to more ecologically valid categories, for example human faces. She familiarised 7- and 10-month-old infants with a set of drawings of faces testing their ability to categorise these faces based on correlations among facial features (for instance, spread of the nose was always paired with wide apart eyes, see Figure 1.3). As in the previous study, at 10, but not 7 months of age, infants were able to categorise faces based on feature correlations.



Figure 1.3. Stimuli used by Younger (1992) to study infants ability to learn categories of faces based on correlations among facial features: three training and two test items.

These studies provided key insights into infants' ability to acquire correlation-based categories. However, there are several limitations of these studies that bring into question their ecological validity. Stimulus sets used in the above mentioned studies were simple line drawings. Therefore, the question is whether these results translate into learning about the natural categories, where items are visually more complex. To address this question, a new stimulus set was designed for the purposes of this thesis to more closely resemble natural categories (3D appearance and texture). Furthermore, I increased the variability of visual features, which enabled the design of test items which are completely novel. None of the feature values seen at test were presented during familiarisation, which was a shortcoming of the previous research.

1.2.4 Visual categorisation in a multimodal world

The discussed studies focused on the process of perceptual categorisation based on purely visual information. Infants' typical learning environment is abundant in various kinds of both relevant and irrelevant information. To fully understand how visual categories are acquired, it seems necessary

to explore whether other types of input, such as auditory input, affect the process of category learning. The focus of my research was on understanding the role of two particular types of information in category learning – labels and motion. In the following section, I present a brief outline of the key findings that inspired experiments presented in Chapter 4.

Plunkett et al. (2008) explored how labels affect the process of forming perceptual categories. Using the same stimulus set as in Younger's (1985) study, infants were presented with a narrow condition (thus inviting infants to form two categories) accompanied either by category-specific labels or by one general label. In a condition where two labels were presented, infants expressed evidence of having formed two categories; by contrast, when one label accompanied all items, infants showed a preference for the extreme items in the test, suggesting they had formed one category. In a condition where two labels were randomly assigned, infants expressed no preference for any of the test items, suggesting that no categories were formed. Several studies demonstrated that labels can impact the process of forming visual perceptual categories (Althaus & Mareschal, 2014; Althaus & Plunkett, 2015; Plunkett, 2010).

In addition to linguistic input, infants learn in a highly dynamic world. Therefore, sensitivity to correlations between dynamic and static features seems crucial for infants' developing knowledge about the world (Rakison, 2004). Despite this fact, there is contrasting evidence about the role of motion in visual category learning: some studies find that it facilitates learning

(Sloutsky & Robinson, 2008), while others find that infants at this age have a limited ability to bind static and dynamic information (Burnham et al., 1988; Rakison & Poulin-Dubois, 2002).

1.3 What enables infants to form categories?

What are the mechanisms underlying infants' ability to form correlation-based categories? Kemler Nelson (1984) proposes that categories with family-resemblance structure, such as correlation-based categories, are more easily learned using holistic processing. Holistic processing is guided by global similarity relations and the overall similarity serves as the fundamental relation among stimuli and their representations, not the abstracted stimulus features (Kemler Nelson, 1984). Holistic processing is accomplished as exemplar learning and generalisation by overall similarity (Brooks, 1978). According to Kemler Nelson (1984) holistic processing is especially suitable when multiple stimulus features are relevant for categorisation and when overall within-category similarity is high and between-similarity is low.

In contrast to this holistic processing account, the analytic processing perspective assumes a process of hypothesis testing based on selected features which leads to discovering which stimulus properties are relevant for which categories. Therefore, stimuli are compared and contrasted according to their constituent features. While Kemler Nelson (1984) suggests a developmental shift from the holistic to the analytic mode of processing, several studies on early acquisition of correlation-based categories speak against

this claim providing evidence that infants' responses in the test phase were driven by selective attention to correlated attributes, not by broad similarity judgements during learning (Younger, 1985; Younger & Mekos, 1992).

A model proposed by Barbara Younger (2010) assumes two core processes underlying structured category representations in infancy: abstracting prototypes and processing correlations among features. The former refers to the fact that infants from early on express evidence of generating an internal graded category representation based on their experience of varying exemplars of that category. This is expressed through prototype effects observed in a number of studies (Bomba & Siqueland, 1983; Plunkett et al., 2008; Younger, 1985). Additional evidence for the prototype abstraction comes from studies of natural categories. Rosch et al. (1976) claimed that the typicality gradient, i.e., the finding that typical category exemplars are more easily categorised that atypical category exemplars, is an inherent characteristic of everyday concepts. Related to the typicality gradient is the second pillar of categorisation: sensitivity to correlated features (Younger, 2010). In light of Rosch's view, "an organism able to detect and represent clusters of correlated attributes is capable of categorisation" (Younger 2010, p. 250). It is these correlated attributes that underlie the structure of natural object categories. To summarise, the experimental evidence shows that very young infants are able to form a summary representation of a category that is graded in structure.

1.4 Nature of category representations

Machery (2009) claims that categorisation processes take representations of distal stimuli as inputs, the proximal stimuli itself. However, infants' early categories are perceptually grounded (Younger, 2010) and dependent solely on the surface appearance of things (Mandler, 2000, 2004a,b). According to Mandler (2000), these perceptual concepts are schemas of what objects look like. The process of perceptual categorisation is an automatic part of perceptual processing that computes the perceptual similarity of one object to another. These early perceptual categories are viewed as a foundation on which concept formation builds. According to Mandler, conceptual categorisation represents a different kind of categorisation process primarily based on what objects are; what they do and what can be done with them. It consists of the redescription of perceptual information into a conceptual form, where perceptual input is analysed and recoded into a new format. This results in notions of kinds, such as animals, plants and furniture (Mandler, 2000).

In contrast to the two processes hypothesis (Mandler, 2000), some authors propose an alternative account – they claim that early concept development represents a continuous process of representational enrichment (Younger, 2010). This process is grounded in a perceptual system that allows for the extraction of regularities in both static and dynamic properties of objects (Rakison & Lupyan, 2008).

1.5 Statistical learning underlying category formation

The "statistical" approach to visual category formation, established by Younger and Cohen (1986) and further developed in many studies, shows the transition from perception of individual features to perception of correlations among features. Infants are excellent statistical learners (Younger, 2010) as they must be able to extract features, reorganise dimensions along which features vary, and distinguish relevant dimensions from irrelevant ones in order to form categories (Aslin, 2010).

A substantial literature on statistical learning in infants comes from the domain of language research (e.g., Saffran et al. 1996). In the case of language, statistical regularities are temporally ordered; in the case of visual category learning information is both *temporally* and *spatially* arranged. In both cases, an infant has to figure out which elements form clusters based on their co-occurence relations (Aslin, 2010).

Several studies explored statistical learning in the visual domain in infants (Fiser & Aslin, 2002; Kirkham et al., 2002). For instance, Kirkham et al. (2002) presented 2-, 5- and 8-month-old infants with sequences of visual stimuli. Using the habituation-dishabituation technique, they found that infants at all three age groups demonstrated evidence of statistical learning, i.e., the ability to encode transitional probabilities of various events. The authors propose a domain general mechanism responsible for statistical learning, where the same learning mechanism underlies different forms

of statistical learning. Furthermore, they propose that this form of learning may have a critical role in infants learning.

1.6 Perceptual constraints on infants' statistical learning

One of the key questions in category learning research is how infants select which information to focus on, i.e., how they extract the correct statistics from the vast array of potential statistics available. Aslin (2010) proposes that there are five constraints on statistical learning that contribute to resolving the problem of "too many statistics":

1. *Gestalt principles* - aspects of a scene are linked together by automatic mechanisms, such as common motion, good continuation, and similarity.

2. *Intrinsic hierarchy of stimulus salience* - there are preferences for certain types of stimuli, such as high-contrast edges. If more salient stimuli capture the attention first, this reduces the set of potentially relevant statistics.

3. *The first-in bias* - after a certain statistic is learned, the potential to learn another set of statistics from the same inventory is inhibited.

4. *Limitations in working memory* - working memory undergoes a rapid development during the first year of life. Whereas infants can retain

only one item at 6 months (Káldy & Leslie, 2005), at 10 months of age four objects can be retained in their working memory (Ross-Sheehy et al., 2003). Limited capacities of the working memory must pose some sort of sub-sampling strategy in order to be able to process the stimuli, but also inhibit fully effective statistical learning. Aslin (2010) reported that whereas 10-month-old infants were able to track statistical coherence when a visual array consisted of three items (Fiser & Aslin, 2002), they failed to do so when the array size increased to four.

5. *The spatial scale of the stimulus* - there is a trade-off between spatial resolution, i.e., the number of visible features, and feature integration, i.e., how many fixations are needed to access the features. Smaller resolution makes integration easier because fewer eye movements and spatial shifts are needed, but a too small resolution would mean that details are not visible. On the other hand, high resolution makes details more accessible, but increases demands in terms of the eye movements and spatial shifts an infant must make in order to integrate features.

Aslin's summary of the visual statistical learning constraints provides a useful framework for understanding infants' statistical category learning. However, these constraints are mainly focused on the properties of individual items and effects of individual constraints. To further this knowledge, it is necessary to better understand how these constraints interact within the same learning situation. In addition, these constraints cannot be seen as fixed parameters, but are also determined by the characteristics of the wider

learning situation.

1.7 Task-dependent category learning

In addition to maturation-dependent constraints on learning, certain characteristics of the learning situation can have a major impact on infants' ability to succeed in the learning task. This has been confirmed by various studies revealing that different tasks may lead to different conclusions, even in infants of the same age and using the same stimuli. For instance, studies based on a visual habituation task find evidence of categorisation earlier than studies using sequential-touching tasks. These differences might be due to different response demands, since control of eye-movements is less complex than the control and execution of motor responses. However, even studies using similar experimental approaches find different outcomes. For instance, 10-month-old infants showed evidence of categorisation in an object-examining task, but failed to do so in a sequential-touching task (Oakes et al., 1996). It is still possible that these effects are due to different motor demands of the task; therefore, it might be the case that infants learned categories, but failed to show it due to limitations in motor and execution domain (Younger & Furrer, 2003).

Studies of visual category learning also highlight the importance of task characteristics for learning. After being familiarised with a series of pictures of dogs, 4-month-old infants presented with a cat and a new dog tended to look more at the cat, distinguishing between the categories of dogs and

cats (Eimas & Quinn, 1994), but only when familiarised with pairs of objects (Kovack-Lesh & Oakes, 2007; Oakes & Ribar, 2005; Younger & Furrer, 2003).

Further evidence for the importance of task-specifics for learning comes from studies demonstrating that the order in which familiarisation items are presented can strongly affect learning outcomes: that is, infants' performance in the category formation test. Mather & Plunkett (2011) presented 10-month-old infants with a set of items in a condition that maximised or minimised perceptual distance between successive exemplars and found that the order in which different items are sampled has a strong impact on category formation.

These studies clearly show the importance of task-dependencies in category learning. Against this background, this thesis aims to explore how specifics of the task and the stimuli used, shape the dynamics of category learning. Understanding how task-dependencies shape categorisation would allow further exploration into how categorisation interacts with other cognitive processes and allow for a broader understanding of the more general mechanisms underlying category learning.

1.8 Implicit category learning in adults

Whereas most learning in infancy represents some kind of unsupervised learning, in adulthood the focus shifts to supervised learning. In line with this, existing research on categorisation in adults has largely focused on the explicit, supervised type of category learning. In this case, participants are

instructed that their task is to learn categories and are given feedback during the learning phase. However, little attention has been given to implicit, unsupervised learning. Here, no feedback is provided and participants are not told about the categories (thus resembling infants' typical learning context). The task is rather disguised as something other than a learning task (Smith, 2008). Implicit category learning can thus be defined as a sort of self-directed pattern detection (Clapper, 2006).

Clapper and colleagues conducted a series of studies investigating unsupervised category learning in adults (Clapper, 2006; Clapper & Bower, 1994, 2002). These studies are particularly interesting because they used categories defined by correlations among visual features, the class of categories this thesis is focusing on. The categories they used consisted of multiple features that differed in their relevance for the category membership; some were highly diagnostic, while others were irrelevant (an example of a category structure is given in Figure 1.4). These studies used the *attribute listing task*, whereby participants were presented with an exemplar from one of the two categories and were asked to list features that are most informative for distinguishing this item from all previously seen items. Since features varied in their prognostic relevance, the difference in the number of relevant and irrelevant features listed was used as an index of category learning. These studies found that participants tend to list more relevant features with more exposure to category exemplars. Clapper & Bower (1994) investigated the conditions under which participants were able to discover the categories in

unsupervised learning conditions. They discovered that presenting exemplars from the same category in one block followed by a block containing exemplars from another category proved to be a better learning context than presenting the categories interleaved. Another study investigated intentional unsupervised learning using the same task, but informing participants that there are two categories, (Zeithamova & Maddox, 2009). In that case, interleaved presentation was found to provide a better learning context than the blocked sequence.

Category 1

Feature							
1	2	3	4	5	6	7	8
1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	2
1	1	1	1	1	1	2	1
1	1	1	1	1	1	2	2
1	1	1	1	1	2	1	1
1	1	1	1	1	2	1	2
1	1	1	1	1	2	2	1
1	1	1	1	1	2	2	2

Category 2

Feature							
1	2	3	4	5	6	7	8
2	2	2	2	2	1	1	1
2	2	2	2	2	1	1	2
2	2	2	2	2	1	2	1
2	2	2	2	2	1	2	2
2	2	2	2	2	2	1	1
2	2	2	2	2	2	1	2
2	2	2	2	2	2	2	1
2	2	2	2	2	2	2	2

Figure 1.4. An example of a category structure used in a study by Clapper and Bower (1994). Columns in grey depict less prognostic features.

The above mentioned studies clearly show that characteristics of the learning situation, such as an awareness of the categorical nature of the task can shape the learning process. For this reason, I designed a novel paradigm which had no explicit mention of categories. The aim was to create the conditions for a completely implicit unsupervised learning that resemble infants' typical learning context.

Another aspect central to this thesis is whether the same kind of unsupervised learning is present in both infants and adults. Several models offer possible ways of understanding categorisation processes from a developmen-

tal perspective.

According to the Competition between Verbal and Implicit Systems model (COVIS, Ashby et al. 1998), the implicit system learns by recognising covariation within the environment. The neural basis of this system (the frontal striatal structures and the posterior caudate) develops within the first year of life (Ashby et al., 1998; Huang-Pollock et al., 2011). Therefore, it is possible that the same kind of learning takes place in infancy and adulthood.

Similarly, the *Supervised and Unsupervised Stratified Adaptive Incremental Network* (SUSTAIN) assumes a continuous trajectory of conceptual development and describes developmental differences as a result of limitations in memory and stimulus encoding during infancy (Gureckis & Love, 2004; Love et al., 2004).

Even though these models make predictions about categorisation at different stages of development, there is less research addressing these claims experimentally. For a long time it was thought that it is impossible to compare categorisation in early and in later stages of development (for a review of this discussion see Murphy, 2002). Recently, however, several studies have addressed this question. Evidence for shared implicit category learning processes across development comes from a study that compared explicit and implicit category learning in adults and 3-7 year-old children (Minda et al., 2008). Performance on explicit categorisation varied with age, but no differences were present in implicit categorisation. Another study explored the effects of implicit novel category learning on similarity judgements in adults

and in 5 year-old children (Livingston & Andrews, 2005). Results revealed that exemplars from different categories were judged as less similar after category training. This pattern of results was observed in both groups, suggesting that fundamental categorisation processes do not change with age (Livingston & Andrews, 2005).

Taken together, these studies demonstrate that it is possible to address the same kind of category learning in both children and in adults. However, the question that remains is whether such learning can take place even earlier, in infancy. Surprisingly, there hasn't been extensive research conducted on this. If brain structures involved in implicit category learning are sufficiently developed towards the end of the first year of life (as proposed by the COVIS model), it is possible that the same kind of learning is present in a developing and a mature brain. One of the aims of this thesis is to fill this gap and understand whether there is a shared implicit category learning at different stages of the development.

1.9 Outline of the thesis

This thesis addresses four aspects of visual category learning. Nine experimental studies were conducted and grouped into four chapters, grouped around these four topics:

- The role of comparison and contrast in category learning;
- The role of dynamic visual features and labels in category learning;

- Implicit category learning in adults;
- The impact of categorisation on object processing at the neural level.

The thesis is organised into two parts. Part I explores categorisation in infancy, whereas Part II focuses on category learning in adults.

Chapter 2 provides a brief description of the methodology used in the thesis. While the first seven studies employed an eye-tracking technique, one study used behavioural measures such as reaction times and accuracy, the final study used an event-related potential technique to explore the neural signatures of categorisation.

Chapter 3 explores the role of comparison in category learning. It consists of three experiments designed to systematically explore how opportunity to compare and the perceptual load, defined as the amount of visual information available during learning, affect infants' ability to learn categories. Inspired by the seminal work of Younger and Cohen (1986) and the existing body of research on statistical category learning, a standard familiarisation-novelty preference procedure was used. Infants were first presented with a set of training items (familiarisation phase), followed by a test phase in which category formation was assessed with the presentation of two items. Preference for one of the two objects was used as an index of category formation. Furthermore, familiarisation was interleaved with test blocks in order to explore category formation at different levels of familiarity and to evaluate not only the outcomes but also the dynamics of the learning process. In particular, Experiment 1 investigated infants ability to form categories when

familiarised with one item at a time. Using the same experimental design as in existing studies, it aimed to replicate and further extend the existing body of research. Whereas Experiment 1 used items with correlated features (narrow condition), Experiment 1a explored infants' performance when presented with an unconstrained set (uncorrelated features, broad condition). Experiment 2 and 3 explore whether providing an opportunity to contrast and compare during learning by presenting multiple items simultaneously affects category learning.

Chapter 4 describes four studies which tackle the role of dynamic visual features and labels in category formation. The impact of dynamic motion features is assessed in Experiments 4 and 5. Experiment 4 tests the effects of having dynamic motion features associated with one category, whereas in Experiment 5 both categories have a category-specific motion. Further, Experiments 6 and 7 investigate the effects of labels on visual category learning. In Experiment 6, exemplars from only one category are labelled, whereas Experiment 7 explores how two distinct labels affect the process of category formation. All four experiments employ a standard familiarisation-novelty preference procedure. In particular, after being familiarised with a set of training items, infants' category formation was measured as a preference for one of the two items presented in test. Again, as in experiments presented in the previous chapter, three familiarisation blocks were interleaved with test blocks, enabling a direct comparison of infants performance across experiments. One important novelty that the present set of experiments in-

troduced is a gaze-contingent familiarisation, thus allowing for the infants to take a more active role during familiarisation.

Part II of the thesis explored the process of category learning in adults. These experiments used the same stimuli as the studies with infants, thus providing an opportunity to compare category learning in the two groups. This part consists of two chapters.

Chapter 5 presents a novel experimental paradigm developed to investigate the process of implicit, unsupervised category learning in adults. Participants were presented with a fast-paced presentation of category exemplars and asked to judge whether exemplars were identical or different. Response time was used to assess whether there are differences between within-category and between-categories judgements. This paradigm resembles the learning context of infants' experiments as participants were not given any instructions about categories and received no feedback on their performance. Using the identical stimulus set as the one used with infants provided further ground for the comparison of the category learning process in infants and adults.

Chapter 6 explores neural signatures of category learning. By comparing neural signatures of visual object processing at the very beginning of the category learning process and then again after an extensive categorisation practice, we test how categorisation alters the way in which exemplars are processed at the neural level.

The final, Chapter 7 presents a discussion of the main findings, main insights and proposes directions for future research.

Chapter 2

Methodology

This chapter provides a brief overview of eye-tracking and event-related potentials, the two methods employed in this thesis, and explains the rationale for their implementation in the present research.

2.1 Eye-tracking

Experiments 1 to 7 (Chapters 3 and 4) exploit eye-tracking measures to investigate category learning processes in infants.

Monitoring oculomotor activity represents an ideal neuroscientific model to explore links between brain mechanisms and behaviour (Luna et al., 2008). Oculomotor behaviour can be reflexive, voluntary and spontaneous, and the three dominant metrics used in eye-tracking research are fixations, saccades and scan paths (Eckstein et al., 2017). Fixations represent settled gaze, and their position and duration are thought to reflect attention engagement and processing time (Eckstein et al., 2017), whereas saccades are rapid eye-

movements, and their accuracy, latency and number are used as indices of shifts in attention (Luna et al., 2008).

The analysis of infants' looking behaviour has been one of the key methodological pillars of developmental research and provided a tool to study aspects of infant development inaccessible by other behavioural measures (Gredebäck & Johnson, 2009). In fact, looking time represents the most common behavioural technique for studying various aspect of early development (Aslin, 2012; Oakes, 2012).

In a typical eye-tracking experiment, an infant is seated in front the computer screen and presented with various visual stimuli while an eye-tracking camera, positioned below the screen, records the infants' eye-movements. Various aspects of looking behaviour can be explored, for instance total looking time, proportion of total looking time, first look, shifting rates etc. The particular measure (or measures) used depends on the research question and the linking hypothesis that connects underlying cognitive processes of interest and a dependent variable (Aslin, 2012). For instance, in one of the seminal studies Fantz (1961) developed a "visual interest test" in order to investigate infants' perception abilities. Infants were lying on their back inside a "looking chamber" and a pair of shapes were presented on the ceiling in front of them. By comparing the amount of time infants spent looking at each of the shapes, he discovered that infants consistently show preferences for certain shapes. Based on this result, Fantz made proposals about the underlying processes of attention and memory.

Experimental setup

After written consent was obtained from a carer, the infant was seated on a carer's lap approximately 60 cm from a 1920x1080 inch monitor screen in a sound-proof experimental booth. The carer was asked to keep their eyes closed for the duration of the experiment. Gaze data was recorded using a Tobii TX300 Eye Tracker with a 120 Hz sampling frequency. The testing session started with a four point calibration procedure. The study was run with a custom Matlab stimulus presentation software PresentMate based on the Psychophysics Toolbox. In case infants got distracted, the experimenter talked to them via microphone between trials in order to attract infants' attention back at the screen. Infants' behaviour was also monitored via a centrally-located camera above the screen (see Figure 2.1).

Data processing

Fixations were calculated from raw gaze data using custom Matlab routines. A second-order Savitsky-Golay filter with a length of seven samples was used for data filtering (Nyström & Holmqvist, 2010). Minimum fixation duration was 100 ms. Eye-tracking data was considered valid if the eye-tracker validation flag indicated that one of the eyes is detected and the recorded gaze was within the screen area. Blinks were detected as sections of the data with an instantaneous rate of change of pupil diameter greater than 0.1 mm for both eyes and the corresponding samples were flagged as invalid. The invalid data for gaze was replaced with last valid value for maximum 100ms.

Data analysis

In addition to standard analyses of preferential looking based on aggregated measures, analysis of the eye-tracking data (and behavioural data in Chapter 5) included growth curve analysis. I decided to use this type of multilevel regression technique because it is particularly suited for the analysis of continuous time-course data (such as looking time course) while accounting for differences on both group and individual level (Mirman, 2014).

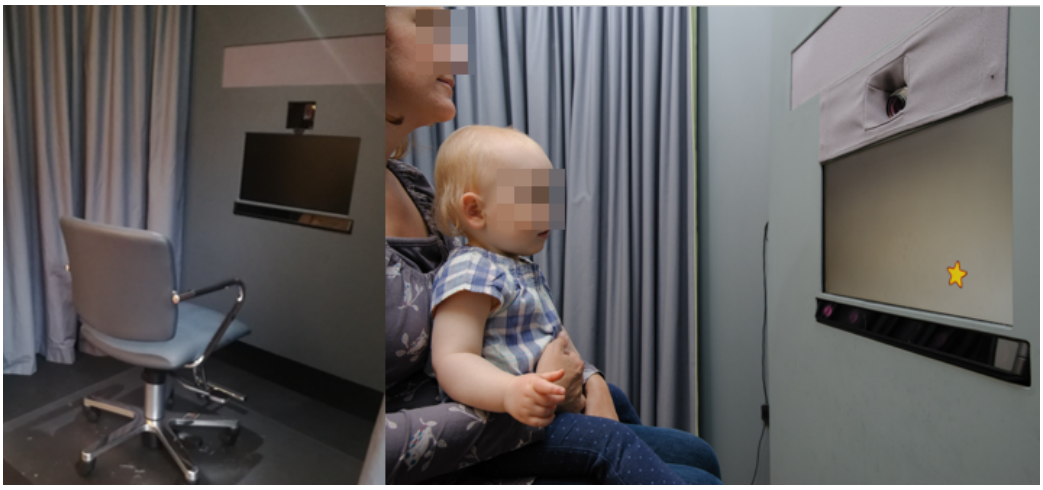


Figure 2.1. Setup used in eye-tracking studies.

2.2 Event-related potentials technique

Experiment 9 presented in Chapter 6 used an event-related potentials technique to investigate neural signatures of category learning.

Hans Berger's studies in the early twentieth century showed that it is possible to record the brain's electrical activity (electroencephalogram) by placing electrodes on the scalp (Luck, 2005). Whereas electroencephalogram (thereafter EEG) represents a very coarse measure of brain activity and re-

flects a mixture of a number of different sources of activity, event-related potentials (thereafter ERP) represent small voltage fluctuations evoked by certain sensory or cognitive events (Luck, 2005). ERP are time-locked to the onset of an event and can be used to investigate how information is processed in the brain over time (Rugg, 2001). ERP reflect electrical currents generated by neural population firing in a synchronous way. More specifically, ERP are thought to reflect summed post-synaptic potentials (Luck, 2005).

To extract ERPs from the EEG, due to a low signal-to-noise ratio, it is necessary to apply signal averaging, i.e. average many EEG segments time-locked to an event of interest in order to obtain reliable measures. Hence, in a typical ERP experiment, multiple trials are presented and various data cleaning steps are undertaken to minimise potential background noise and artifacts.

The main advantage of ERP technique is high temporal precision, therefore enabling researchers to track neural processing on a level of milliseconds. It is important to note that spatial resolution is lower, as mutliple different dipoles of different locations and orientations can result in identical ERP recorded on the scalp (Luck, 2005).

Different sensory events can modulate ERP in different manners and to a different extent. However, depending on the sensory modality, there are modulations in the wave shape that are common. For visually elicited ERP, it is thought that early components (0-200ms post-stimulus) are mainly modulated by characteristics of visual processing, whereas later components reflect

higher cognitive processes such as memory updating and semantic processing (for a summary of ERP components see Figure 2.2).

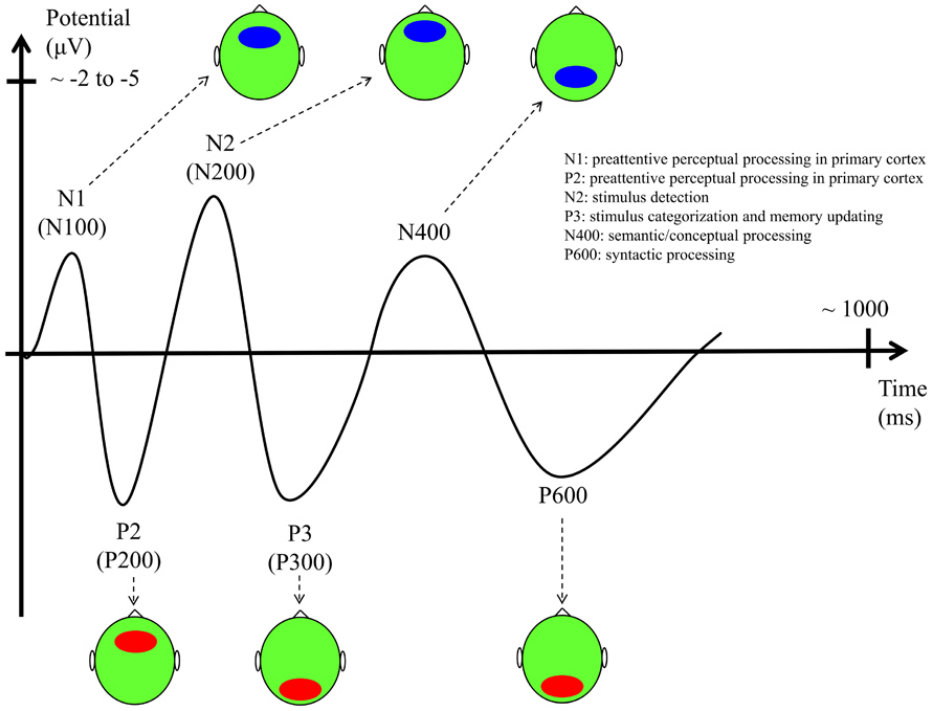


Figure 2.2. Main ERP components (reproduced from Daltrozzo & Conway 2014).

Experimental setup

The EEG signal was recorded using BioSemi ActiveTwo amplifier (BioSemi, Inc., Amsterdam, the Netherlands) from 64 sintered Ag/AgCl pin electrodes arranged in a 10-10 montage online referenced to mastoids (see Figure 2.3). Four additional electrodes were positioned below the eyes and two laterally to the external canthi of the left and right eye, to identify ocular artefacts (in addition, Fp1 and Fp2 were used as upper EOGs). All electrode offsets were kept below 30 μV and the signal was digitised at 2048 Hz.

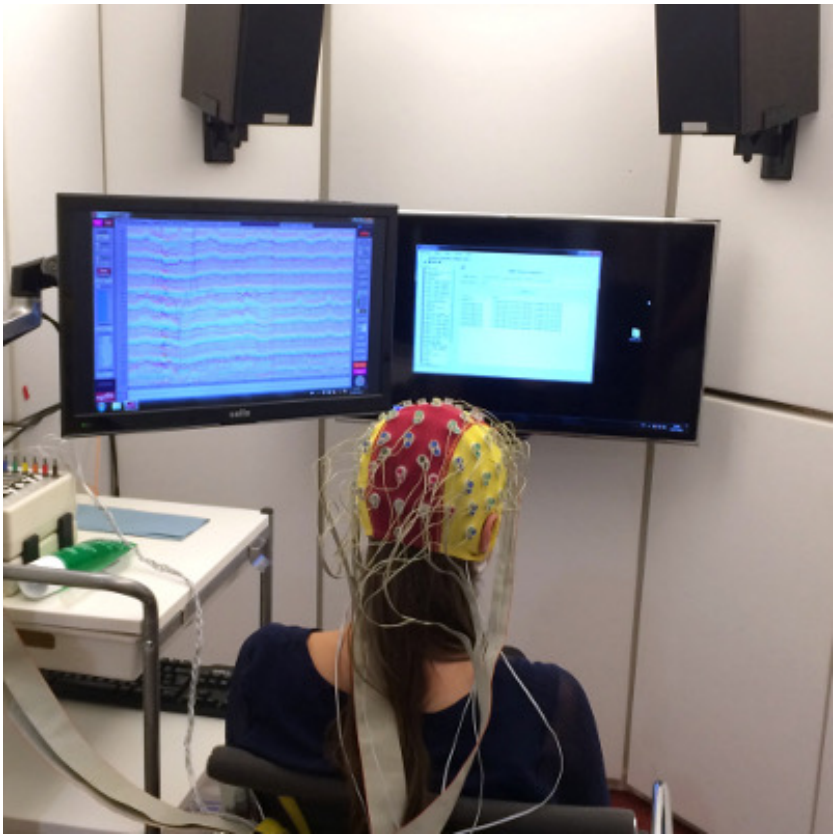


Figure 2.3. Participant preparation for the ERP study (photo reproduced with permission of the Language & Brain Laboratory, University of Oxford).

Part I

Category Learning in Infants

Chapter 3

Memory and perceptual load in infants' category learning

This chapter investigates the role of perceptual complexity and memory load in infants' category learning with a series of three experiments employing a novel set of visual stimuli. Studies show that infants become able to detect feature co-occurrences and use it as a basis for category learning even before their first birthday (Mather & Plunkett, 2011; Plunkett et al., 2008; Younger & Cohen, 1986; Younger & Fearing, 1999). However, there are two key limitations to these studies that restrict their ability to describe infants' category learning in the real world. On the one hand, these studies used simple line drawings as stimuli, thus making it difficult to link these results to formation of real-world categories. On the other hand, the items used to test category formation were not completely novel. Thus, it is possible that infants' looking preferences in test were to some extent influenced by

familiarity with some of the test items.

The approach of this work involved the design of a new set of stimuli to increase ecological validity and allow better comparison to the literature on infants' categorisation of real objects. The new stimulus set was designed as more realistic, textured 3D-looking items. Furthermore, the novel stimulus set had an increased variability of the perceptual space allowing for completely novel test items to be used in test.

In addition, the afore mentioned studies familiarised infants with one item at a time. Some studies suggest that presenting items in pairs facilitates category learning (Oakes et al., 2009; Oakes & Ribar, 2005). Presenting multiple items at a time increases perceptual load during learning, but at the same time it reduces memory demands inherent to sequential familiarisation. To explore how manipulating task demands in terms of perceptual and memory load affect infants' category learning, the number of items infants are presented with during familiarisation was varied in the experiments presented in this chapter.

One of the factors driving infants' looking preferences in categorisation experiments is the level of familiarity with to-be-learned categories (Hurley & Oakes, 2012; Kovack-Lesh et al., 2014, 2012). For instance, Hurley & Oakes (2012) explored how the amount of previous experience affects infants abilities to distinguish two categories. They compared the performance of infants who had pets at home to those who had no pets and found that infants with pet exposure direct more attention towards most category-relevant features

than infants with no pet exposure (Hurley & Oakes, 2012). Several models of categorisation have suggested that the amount of familiarisation has an important impact on categorical representations such that the category representation changes with increasing familiarity with category members (Jones & Smith, 1993; Quinn, 2002). Categorical representations are constantly being updated and adapted according to the incoming input and are therefore dynamic in nature (Jones & Smith, 1993).

To the best of our knowledge, only one study so far explored whether increased exposure leads to better categorisation. Oakes et al. (2009) found no effect of the amount of familiarisation time on infants' categorisation performance. However, this study used existing categories (cats and dogs), with which infants may already have been familiar, which have masked out potential effects of the familiarisation levels manipulated in this experiment. To resolve this, experiments presented in this chapter probed category formation at various familiarisation levels by interleaving training and test blocks. A main hypothesis was that infants' performance on category formation tests will improve with the increase in the amount of familiarisation. Comparison of infants' looking preferences in the three test blocks provides a direct test of this hypothesis.

3.1 Overview of the experiments

To address these questions, three studies were conducted. Experiment 1 and 1a were initial explorations of 10-months-old infants' ability to learn cate-

gories using a novel stimulus set designed for the purposes of this research. Unlike previous studies which mostly used line drawings, this novel set consisted of naturalistic 3D-looking objects. Experiment 2 and 3 explore the role of comparison in category learning and the question of an optimal perceptual load for category formation by varying the number of items presented simultaneously during familiarisation. All experiments are based on a standard familiarisation-novelty preference procedure. Infants are first presented with a set of training items (familiarisation phase), followed by the test phase in which category formation is assessed with the presentation of two items and preferences for one of the two objects is used as an index of category formation. A novelty to this design is that familiarisation and test blocks were interleaved. This manipulation allows testing category formation at different levels of familiarity and evaluates not only outcomes but also the dynamics of the learning process. An outline of the experiments presented in this chapter is given in Figure 3.1.

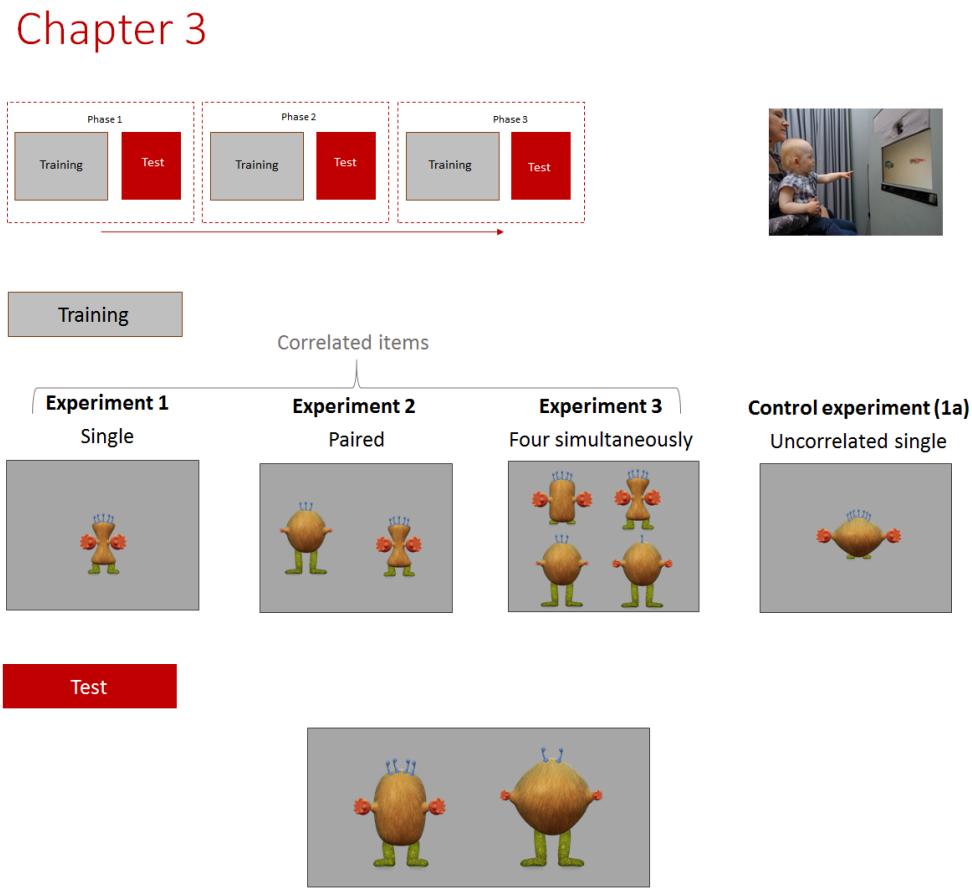


Figure 3.1. Outline of the chapter: The experiments presented use an identical experimental design with interleaved training and test trials (grey boxes refer to training blocks, red boxes depict test blocks). Across Experiments 1, 2 and 3 the amount of information immediately available was manipulated by varying the number of items simultaneously presented. All experiments employed an identical test procedure based on novelty preference (illustrated in the bottom part of the figure).

3.2 Methods

The description that follows is common to all experiments reported in this chapter, with the exception of the familiarisation phase. Specifics of the familiarisation are described separately for each experiment.

3.2.1 Stimuli

A novel set of visual items consisting of coloured and textured 3D-looking anthropomorphic creatures (called Sukis) was designed and created for the purposes of this study. An additional out-of-category visual item was created with similar components, but with features arranged in a non-anthropomorphic fashion. The stimuli were created using Sculptris (Pixologic Inc., Los Angeles, CA, www.pixologic.com) and Photoshop (Adobe Inc., www.adobe.com). The Sukis feature structure resembled stimulus properties used in several categorisation studies (Mather & Plunkett, 2011; Plunkett et al., 2008; Younger & Cohen, 1986), with the exception that test stimuli contained feature levels not seen during familiarisation (see Figure 3.2).

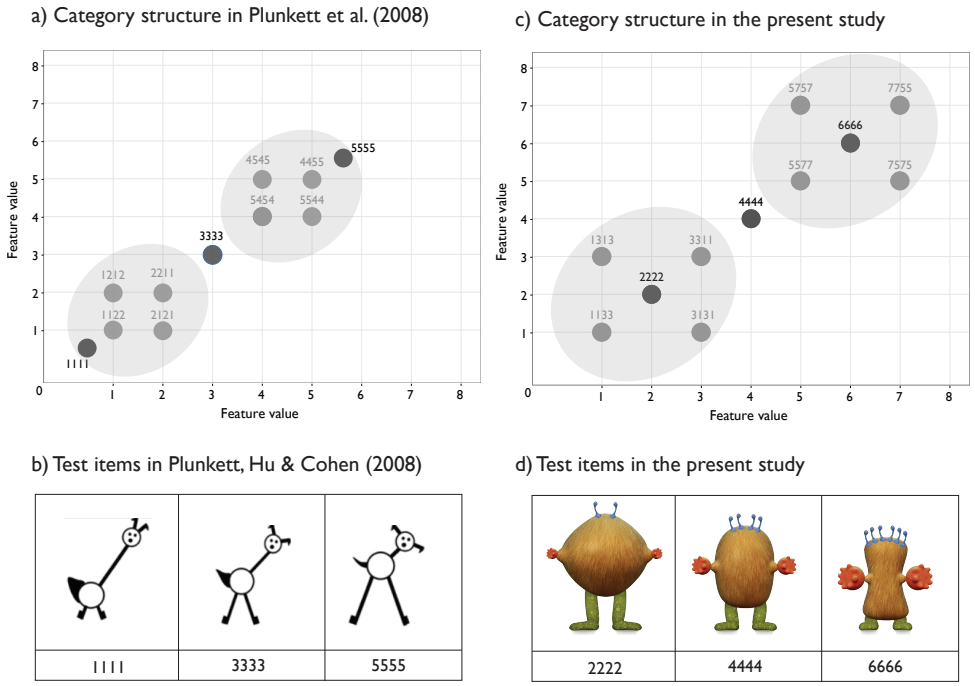


Figure 3.2. Schematic depiction of the category structure and stimuli examples in Plunkett et al. (2008) and in the present study: each stimulus can be described in 4-dimensional perceptual space (each dimension resembles one feature, i.e. arms, hands, body and legs).

Each Suki consisted of four features: body, antennae, hands and legs. Each feature varied systematically on a scale of seven levels: antennae number varied from one to seven (Figure 3.3a), hands shape from small to big (Figures 3.3b), body shape from thin to rounded (Figure 3.3c), legs from short to long (Figure 3.3d).

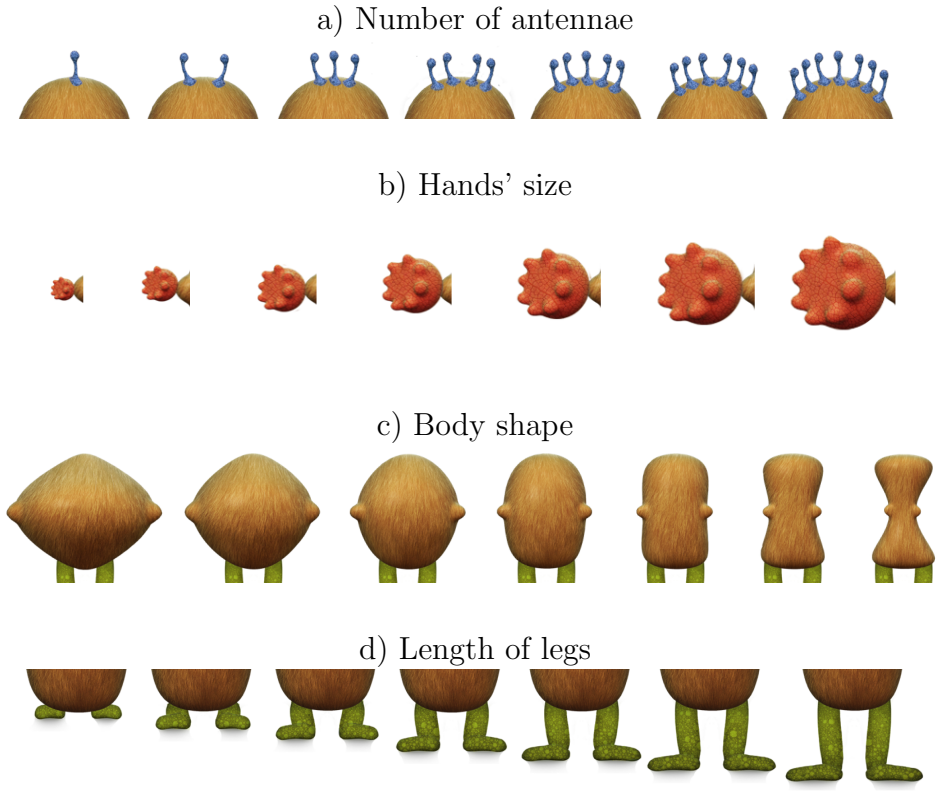


Figure 3.3. The Suki set: feature range.

To explore infants' use of feature correlations to group objects into categories, a set of 24 Sukis was selected with levels of one feature being always associated with certain levels on other features, thus inviting participants to form two categories (defined as the narrow condition in Plunkett et al. (2008), for more details see Appendix A). To illustrate, Sukis with long legs always had small hands, whereas Sukis with a rounded body had fewer antennae (see Figure 3.4). The structure of each Suki can be described with four numbers, each number representing a level of a particular feature (e.g. an item 1133 where the first number stands for number of antennae, the second for the hands size, the third number represents body shape and the fourth number legs length). By contrast to Younger's stimulus set, the range of potential levels each feature can take was extended from five to seven. The variability

of the stimulus set was increased in order to have the possibility to create test items made of completely novel feature levels that have not been presented during the familiarisation phase. Four Sukis were selected to be presented as test items: an inter-category item (consisting of overall mean levels for each feature, i.e. 4444), and two category prototypes (each consisting of average levels for each of the features for its category, i.e. items with values 2222 and 6666 on four relevant features).

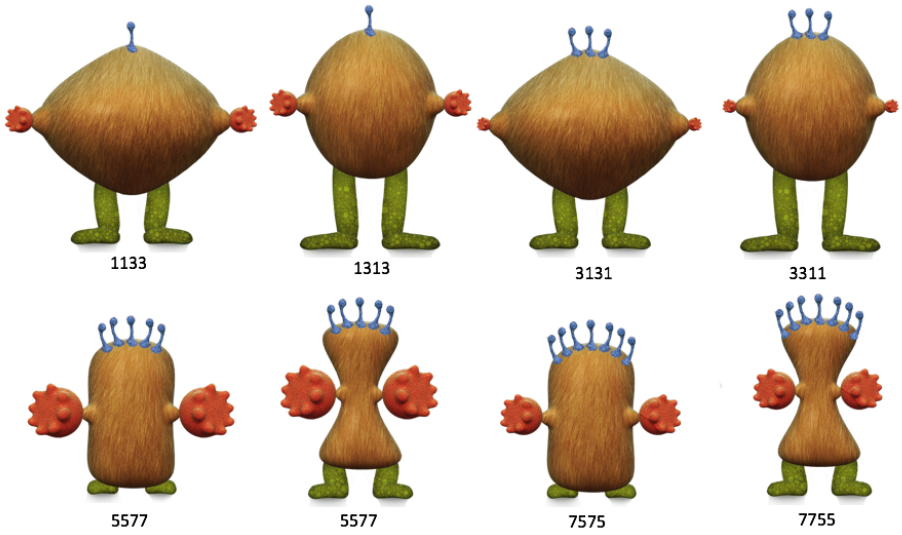


Figure 3.4. The Suki set with correlated feature values.

In addition, a completely novel, out-of-category object which comprised similar features of different shapes organised in a completely different manner were presented in the final trial of the test phase (an example of a novel item is shown in Figure 3.5).

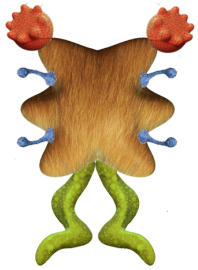


Figure 3.5. Novel, out-of-category test item.

Another set of Sukis was created using identical features, but assembled in a way that values on one feature were not predictive of values on other features (resembling the "broad" condition in Plunkett, Hu & Cohen, 2003). Examples of this set are given in Figure 3.6 and the complete structure of the set is described in the Appendix A.

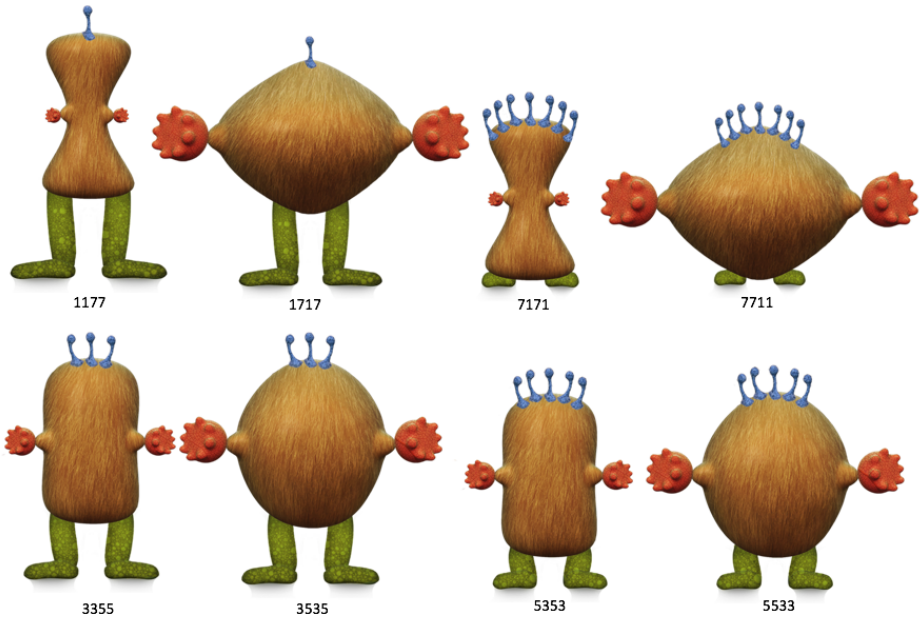


Figure 3.6. The Suki set with uncorrelated feature values.

3.2.2 Experimental design

For a detailed explanation of the experimental setup and signal processing, please see Chapter 2. All stimuli (500 x 500 pixels) were presented against a 5% grey background. Trials were initiated by the experimenter when the infant attended the screen. Experiment consisted of three phases, each comprising of a training and test block (see Figure 3.7).

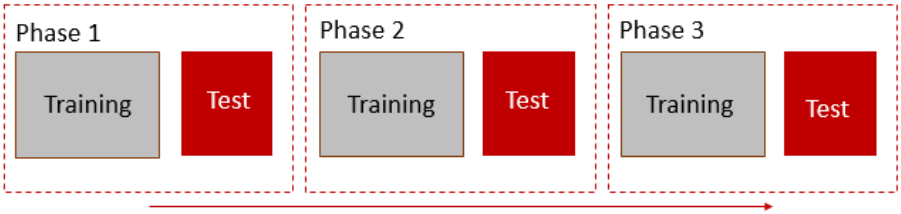


Figure 3.7. Schematic depiction of the experimental design: three familiarisation blocks interleaved with three test blocks.

Training

A total of 24 items were presented during familiarisation (12 from each category). Total familiarisation duration was kept identical across all experiments. More detailed description of trial structure is given in methods section of each experiment.

The order of stimuli presentation was carefully controlled. Mather and Plunkett (2011) showed that the order in which stimuli are presented may affect category formation. Sequences maximising the difference between consecutive stimuli resulted in better performance as compared to sequences that minimised the differences between the consequent stimuli (Mather & Plunkett, 2011). To control for potential effects of the presentation order, we estimated difference between consecutive items for all possible sequences.

To quantify this difference, we used Euclidean distance between consecutive stimuli in a four-dimensional Euclidean space which is a frequently used measure of dissimilarity (Chatfield & Collins, 1981).

To calculate the distance between consecutive Sukis, S_i and S_j in a trial sequence, each Suki was represented by a point in the four dimensional feature space and the distance between the Sukis was calculated as the Euclidean distance:

$$\sqrt{(A_i - A_j)^2 + (H_i - H_j)^2 + (B_i - B_j)^2 + (L_i - L_j)^2}$$

where the $A_{i,j}$, $H_{i,j}$, $B_{i,j}$ and $L_{i,j}$ represent the feature values for the antennae, hands, body size and leg features, respectively. In the next step, the mean Euclidean distance (as an average of distances between eight consecutive objects) for all possible stimuli sequences (40320 sequences) was calculated. Sequences were selected to be used if they fall within the range between the 40^{th} and 60^{th} percentile (8112 sequences). Then, for each participant a particular sequence from this pool was randomly selected.

Test

Test blocks were identical in all experiments presented in this chapter. Each test block consisted of category formation test trials and an additional novelty preference trial. In each test trial, after an attention getter was presented for $2,000$ ms, two test objects were presented simultaneously for $10,000$ ms. The first two trials were category formation test trials in which the inter-category item (object 4444) and category prototype (2222 or 6666) were presented.

The positions of the two objects were counterbalanced across the two trials. The third, novelty preference trial was always a novelty preference test in which one of the training items from the previous familiarisation phase was presented along with a novel, previously unseen out-of-category object (see Figure 3.8). The purpose of this trial was to check whether infants were engaging in the task and expressing the expected novelty preference. The choice of the category prototype (2222 or 6666) presented in a particular test block was balanced across test blocks. The third test block was identical to the first test block for half of the participants, whereas others saw identical items as in the second test block. Which category prototype item was presented first was counterbalanced across participants.

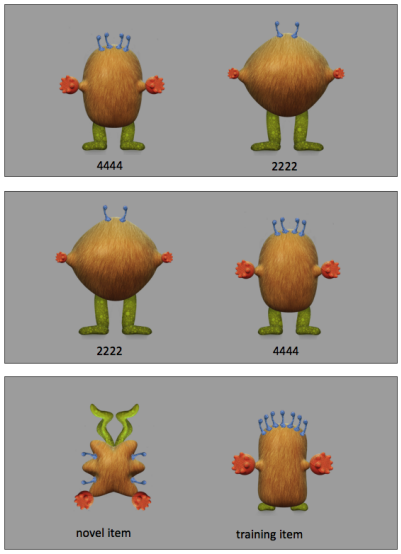


Figure 3.8. Structure of the test blocks: two category formation trials followed by a novelty preference test.

3.2.3 Participants

Experiments presented in this chapter were conducted with 128 10-month-old infants. Participants were recruited at the local maternity ward and all were full-term babies with no known health conditions. Infants who refused to look at the screen and infants whose total looking time in a training block was more than three standard deviations below the mean looking time of infants in the same experiment were excluded from the analysis. All participants came from homes where English was the only language spoken. The study was approved by the University of Oxford Medical Sciences Ethics Committee.

3.2.4 Statistical analysis

Training

To determine whether there is a decrease in the amount of looking time over the course of the three familialiration blocks, growth curve analysis (Mirman, 2014) was performed in R (R Core Team, 2013) using the lme4 package (Bates et al. (2015), version 1.1-12) and lmerTest package (Kuznetsova et al. (2017), version 2.0-33). Mean looking time was modelled using a linear growth model with a fixed effect of Trial and random effect of Participant on the intercept and slope terms. The fixed effect of Trial was added to the base model and the effect was evaluated using model comparisons (Mirman, 2014).

We expect to observe a decrease in looking time across familiarisation blocks. This habituation effect is not unexpected as habituation is one of

the typical indicators of learning and is in line with previous categorisation studies (e.g. Althaus & Plunkett 2015; Mather & Plunkett 2011; Plunkett et al. 2008).

Test

Category formation was tested after every familiarisation block using two trials. The inter-category item (representing the average of all familiarisation items) and one of the category prototypes were presented in each test trial. Looking preference was used as an index of category formation. If infants formed two categories, it is expected they will spend more time looking the inter-category item, whereas if they formed one category, it is expected that infants will find category prototype more novel (see Figure 3.9). Preference scores were calculated by dividing looking at the inter-category item by total looking time to both objects.

A repeated-measures ANOVA with factors Block (1, 2 and 3) and Test (1 and 2) was performed to determine whether there are changes in the preferences over the course of the three test blocks (Greenhouse-Geisser corrections for F-statistics were applied where sphericity assumptions were violated). Planned comparisons against chance were performed for each block to determine whether infants' expressed preference for one of the test items in each test block (t-tests or non-parametric tests in case of non-normally distributed data).

Interpreting a preference in test trials as an index of category formation is based on the assumption that infants' behaviour is driven by novelty prefer-

ence. To validate that infants' looking preferences in test trials were driven by novelty preference, at the end of each test block an additional trial was presented to test whether infants indeed show a preference for novel items. Infants looking to the novel object was divided by the total looking time and a one-way ANOVA with Block as a within-subjects factor was performed. Planned comparisons against chance (.5) were performed for infants' preference for the novel object in each block.

Generalised eta-squared was reported as a measure of the effect size for repeated-measures ANOVAs as it is invariant across different research designs, thus allowing for comparison of effect sizes across different analyses (Olejnik & Algina, 2003) and Cohen's d for t-tests using lsr R package (Navarro (2015), version 0.5).

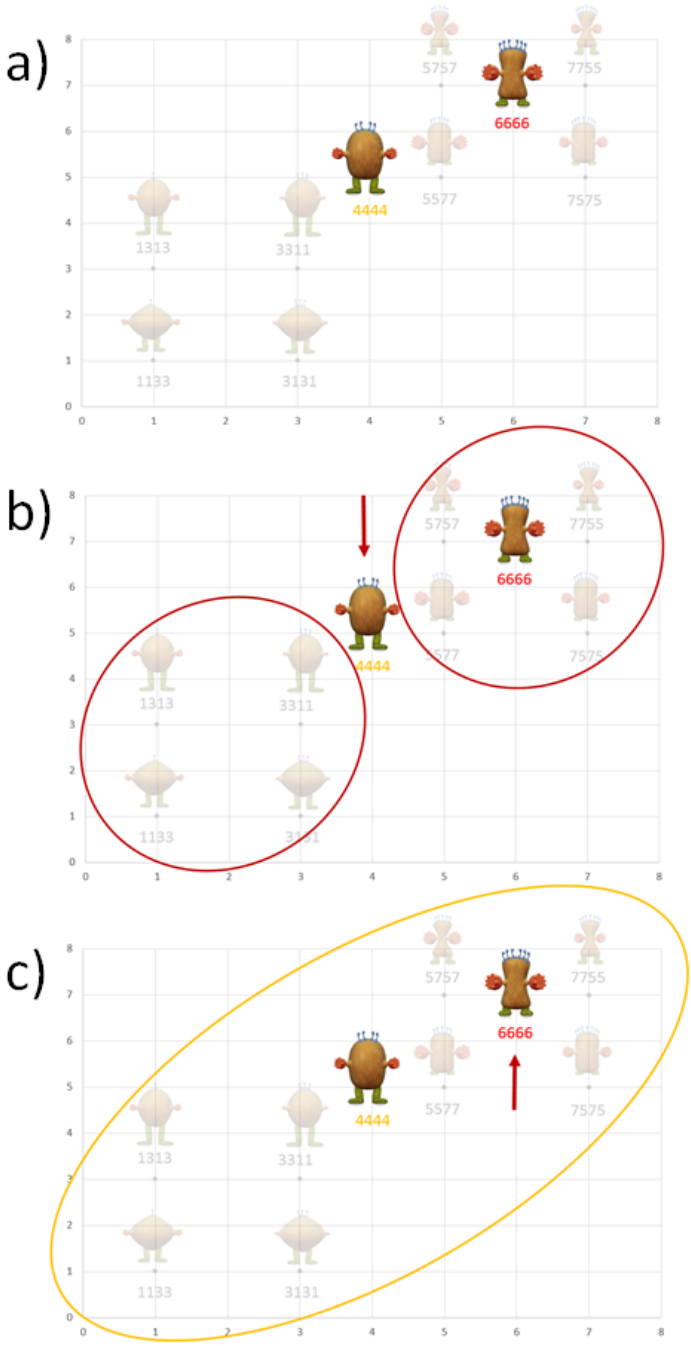


Figure 3.9. Outline of the test design: inter-category item (4444) and a category prototype (either 2222 or 6666) are presented in test (a); If infants formed two categories, the inter-category item will look more interesting and novel as it lies between the two categories (b); if infants formed one category, the inter-category item will be less interesting as it is closer to the average of this category (c).

3.3 Experiment 1: single-item familiarisation

3.3.1 Introduction

Existing research (Mather & Plunkett, 2011; Plunkett et al., 2008; Younger & Cohen, 1986; Younger & Fearing, 1999) has demonstrated that 10-month-old infants are able to learn novel correlation-based categories. However, these studies used line drawings as stimuli, leaving limited possibility to link these results with research on infants' learning of realistic categories. Experiment 1 aims to further extend these results and determine whether infants can succeed in the same task when presented with more realistic and perceptually more complex novel categories.

Another aim of the present experiment is to investigate the role of familiarity in infants' category learning. A main hypothesis was that infants' performance on category formation tests will improve with the increase in the amount of familiarisation. Comparison of infants' looking preferences in the three test blocks provides a direct test of this hypothesis.

3.3.2 Method

Participants

Thirty-two 10-month-old infants took part in this study (14 girls and 16 boys, M_{age} = 10.24, age range = 9.64 - 10.68). Participants were recruited at the local maternity ward and all were full-term babies with no known health conditions. All participants came from homes where English was the

only language spoken. Three participants were excluded due to fussiness and refusal to look at the screen during the experimental session. Of the 29 participants in the final sample, 27 completed the second block and 25 contributed to the data for the final of three blocks.

Stimuli

A set of 24 Sukis were selected for training. All items had correlated features (values of one feature were predictive of the values of other features) as described in Section 3.2.1. Four additional items were used in test blocks: the inter-category item, two category prototypes and a novel out-of-category item.

Experimental design

As described in Section 3.2.2 three test blocks were interleaved with training blocks (see Figure 3.10).

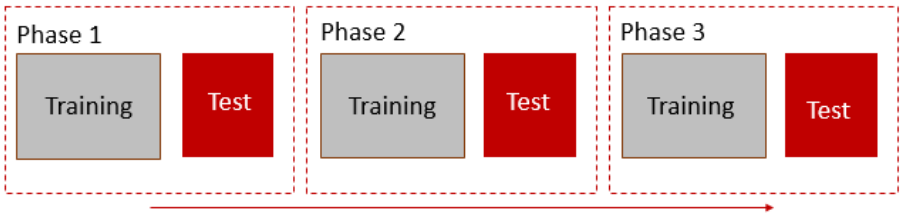


Figure 3.10. Schematic depiction of the experimental procedure: three familiarisation blocks were interleaved with three test blocks.

A total of 24 items was presented during familiarisation (12 from each category). Each familiarisation block consisted of eight trials (thus, no familiarisation stimuli were repeated). Each trial started with a presentation of an animated star in the central location of the screen accompanied by a

chiming sound for a duration of 2,000 ms. Following this, one stimulus (500 x 500 pixels) was presented in the central location for 6,000 ms (see Figure 3.11).

For more details about the design of the test blocks, please see Section 3.2.2.

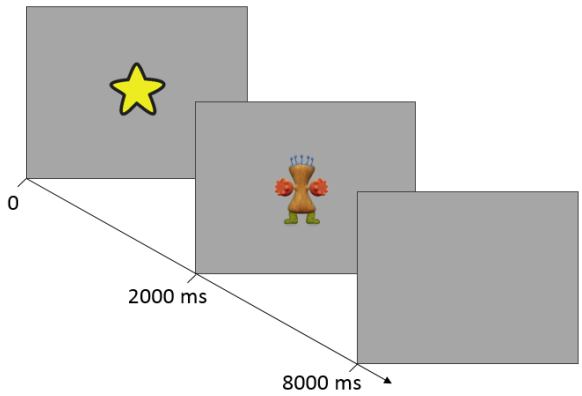


Figure 3.11. Experiment 1: Structure of familiarisation trials.

3.3.3 Results

Training

As depicted in Figure 3.12 there was a gradual decrease in the mean looking time over the course of the three training blocks. This was further confirmed by a growth curve analysis (model specifications are given in Appendix A). ($\chi^2(1) = 8.42,\ p < .01$). There was an average decrease in the amount of looking time of 36.13 ms per trial (see Table 3.1).

Table 3.1. Experiment 1: Growth Curve Analysis. Parameter estimates.

	Estimate	Std. Error	df	t value	p value
Intercept	5131.50	158.59	30.18	32.36	<**.001**
Slope (Trial)	-36.13	11.39	29.03	-3.17	<**.01**

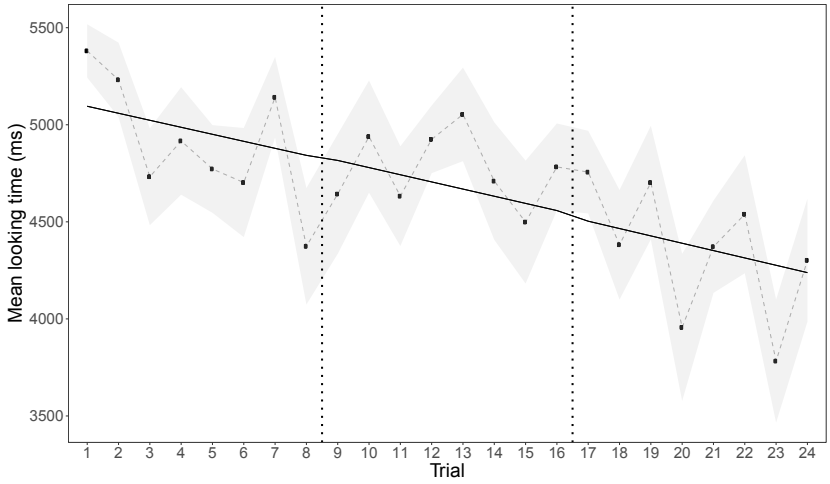


Figure 3.12. Experiment 1: Mean looking time across infants in each training block (dashed line depicts mean looking times in each trial, solid line represents model values, vertical dotted lines represent time points when test blocks were presented, grey envelope depicts standard error).

Test - Category Formation

Preference scores were calculated by dividing looking at the inter-category item by total looking time to both objects. The average score was calculated for the two trials of each test block. An ANOVA on the mean preference score with a within-subject factor *Block* (1, 2 and 3) showed no significant effects ($p > .05$) suggesting that infants' failed to show evidence of category learning (Figure 3.13).[1] Planned comparisons of the mean preference in each test block against chance revealed no significant effects ($all\ \ p_s > 0.5$).

[1] Only participants who contributed to all blocks were included in this analysis (N=25), whereas planned comparisons were performed with all participants who contributed to that particular block.

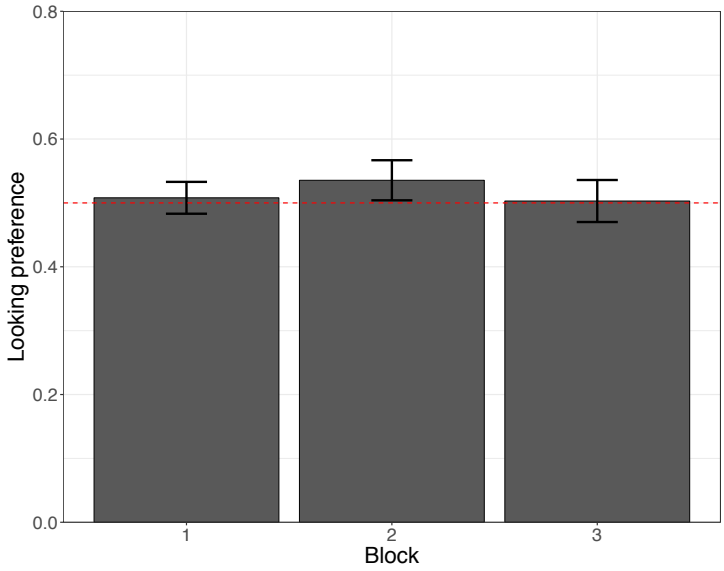


Figure 3.13. Experiment 1: Looking preferences in the test blocks (error bars depict standard error, red dotted line depicts expected preference if performance was at chance level, higher preference value indicates preference for the inter-category item).

Test - Novelty Preference

To validate that infants' looking preferences in test trials were driven by novelty preference, at the end of each test block an additional trial was presented to test whether infants indeed show a preference for novel items. Infants' looking to the novel object was divided by the total looking time and a one-way ANOVA with *Block* (1, 2 and 3) as a within-subjects factor revealed no effects. Planned comparisons showed that infants' preference for the novel object differed from chance only in the second block ($M = .61$, $SD = .23$, $t(26) = 2.58$, $p < .05$, $d = .50$).

3.3.4 Experiment 1a: control experiment

Experiment 1 explored infants' ability to form correlation-based categories using a newly designed stimulus set inspired by the Younger (1985) study.

To rule out a possibility that preference for the inter-category resulted from the interest in this as such, regardless of potential category structure, Younger (1985) conducted an additional experiment in which familiarisation items had uncorrelated features, i.e. values of one feature were not predictive for values on other features, therefore providing no basis for two categories to be formed.

As part of the validation of the Suki stimulus set, Experiment 1a was conducted. Experimental procedure was kept identical as in Experiment 1, except that familiarisation items had uncorrelated features, i.e. values of one feature were not predictive for values on other features, therefore providing no basis for two categories to be formed.

Method

Participants

Forty 10-month-old infants took part in this study (18 girls and 22 boys, M_{age} = 10.12, age range: 9.47 - 10.59). Seven additional participants were tested, but not included in the study due to refusal to look at the screen, fussiness and crying.

Stimuli

A set of 24 Sukis with uncorrelated features (resembling the broad condition in Plunkett et al. 2008, see Figure 3.6) was selected to be presented to infants during familiarisation. Hence, values of one feature were not predictable for values on other features and there was no correlation among features.

Experimental design

The procedure was identical to the one used in Experiment 1. Three familiarisation blocks were presented interleaved with three test blocks. Each familiarisation block consisted of eight trials. Each trial started with a presentation of an animated star in the central location of the screen accompanied by a chiming sound for a duration of $2,000$ ms. Subsequently, one stimulus was presented in the central location for $6,000$ ms. A particular sequence for each participant was randomly selected from a pool of sequences that fall within the range between the 40^{th} and 60^{th} percentile (8920 sequences). Test blocks were kept the same as in Experiment 1. Each test block consisted of two category formation test trials followed by a novelty preference test (see Section 3.2.2 for a detailed description).

Results

Training

Mean looking time was calculated for each of the eight trials of each familiarisation block. There was a gradual decrease in the mean looking time over the course of the three training blocks (Figure 3.14).

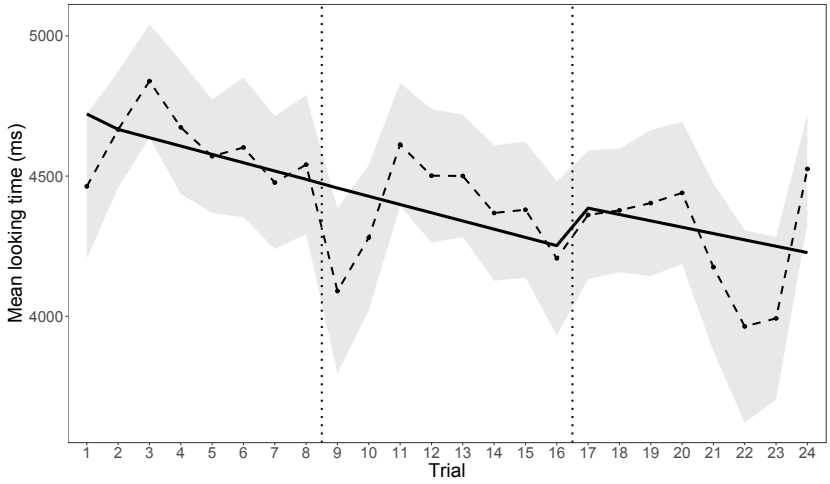


Figure 3.14. Experiment 1a: Mean looking time across infants in each training block (dashed line depicts mean looking times in each trial, solid line represents model values, vertical dotted line represents time points when test blocks were presented, grey envelope depicts standard error).

There was a significant decrease in the amount of looking time of 29.67 ms ($SE = 11.99$) per trial ($\chi^2(1) = 8.42,\ p < .05$, Table 3.2).

Table 3.2. Experiment 1a: Growth Curve Analysis. Parameter estimates.

	Estimate	Std. Error	df	t value	p value
Intercept	4725.68	181.21	41.42	26.08	**<.001**
Slope (Trial)	-29.67	11.59	39.44	-2.56	**0.01**

Test - Category Formation

Category formation was tested after every familiarisation block using two trials. A repeated-measures ANOVA with factors *Block* (1, 2 and 3) and *Test* (1 and 2) revealed no significant effects nor interaction (*all* $p_s > .05$). Planned comparisons against chance were performed for each block. Infants demonstrated significant familiarity preference (i.e. preference for the category prototype) in the third test block suggesting that infants formed one overall category ($M = .44,\ SD = .12,\ t(30) = -2.65,\ p < .05,\ d = .48$),

(see Figure 3.15).

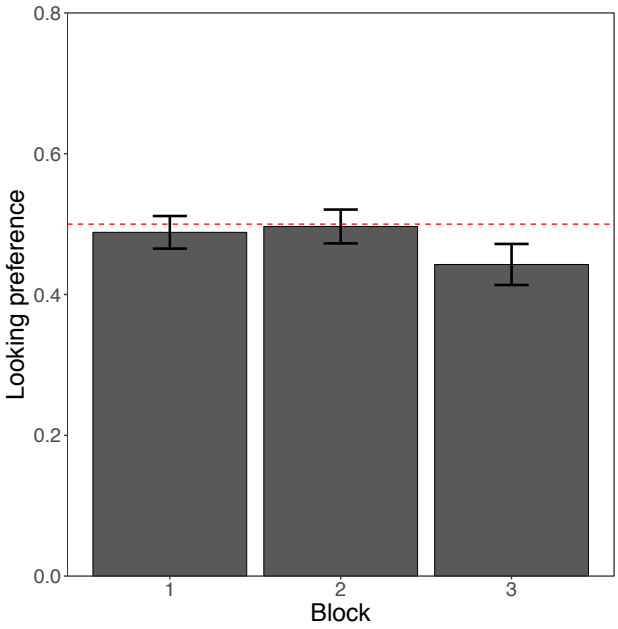


Figure 3.15. Experiment 1a: Looking preferences in the test blocks (error bars depict standard error, red dotted line depicts expected preference if performance was at chance level, higher preference value indicates preference for the category prototype).

Test - Novelty Preference

One-way ANOVA with *Block* as a within-subjects factor revealed no effect of *Block* (Figure 3.16). Planned comparisons showed that infants demonstrated novelty preference in the first two blocks (block 1: $t(38) = 3.16,\ p < .01, d = .51$; block 2: $t(37) = 2.25,\ p < .05,\ d = .36$), whereas in the third block did not reach significance level ($p > .05$). This result supports the claim that infants looking behaviour was driven by a novelty preference in the first two test blocks.

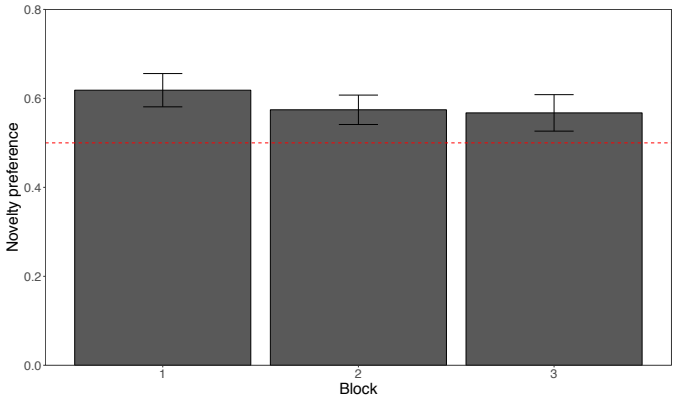


Figure 3.16. Experiment 1a: Looking preferences in novelty preference tests (error bars depict standard error, red dotted line depicts expected preference if performance was at chance level, higher preference value indicates preference for the novel item).

3.3.5 Discussion

In Experiment 1 we explored infants' ability to form novel perceptual categories. In order to learn these categories, infants must be able to detect statistical regularities underlying category structure. Infants were presented with a set of 24 familiarisation items (12 from each category) and category formation was tested after presenting every eight familiarisation trials. Having familiarisation and test blocks interleaved allowed for examination of category formation at different levels of familiarity.

In contrast to the existing body of research that showed that 10-month-old infants can form categories based on feature correlations (Mather & Plunkett, 2011; Plunkett et al., 2008; Younger & Cohen, 1986), we found little evidence of category learning in the present study. There are several factors that could explain this result. As the experimental procedure used in this study was identical to the one reported in Plunkett et al. (2008), the observed differences between these two studies may be due to the increased stimulus

complexity of the stimulus set used. In particular, the increased perceptual variability of the stimulus features could be one of the factors driving the observed results. Instead of a five-point range (used in previous studies), in the current study each feature could take one of seven potential levels (see Figure 3.2). Increasing perceptual variability of stimulus features provided an opportunity to create test items that were entirely new as none of the feature levels that test items consisted of was previously shown in training. In addition, whereas stimuli used in the previous studies were simple line drawings, the Sukis were designed as more realistic, 3D-looking, textured items which further increased complexity of categories. Taken together, this suggests that increasing feature variations and visual complexity hindered infants' ability to learn these categories. A classic model of infants' habituation (Hunter & Ames, 1988) postulates that increased familiarisation leads to novelty preference and that the three key factors driving infants' visual preferences are age, familiarity and complexity (Houston-Price & Nakai, 2004; Hunter & Ames, 1988). Hence, in light of this model results of the present study suggest that the increased complexity of the categories might require extended familiarisation for the infant to succeed in forming categories.

In the present study, category formation was tested using looking preferences in trials in which a category average item (2222 or 6666) and an inter-category item (4444) were presented. It was assumed that if infants form two categories, they will find the inter-category item more interesting as it lies on the boundary between the two categories. Thus, any looking

preference for the inter-category item is interpreted as an indicator of category learning. An alternative explanation would be that the inter-category item had a special status and was more interesting regardless of category formation. To examine whether the special status of the inter-category item is a result of a category learning process or not, in Experiment 1a infants were familiarised with a set of stimuli where features were uncorrelated. No preferences for the inter-category item were expected as there was no relational information to be used as a basis for category formation. The results of Experiment 1a confirmed this as infants showed no preference for the inter-category item. Moreover, in the final test block, infants preferred category prototypes, suggesting that they grouped all items into one category.

One of the aims of this study was to examine the role of familiarity in infants' category formation. To achieve this, test blocks were interleaved with familiarisation blocks. However, testing at various points of familiarisation may also affect category learning itself. Presenting test items and novel, out-of-category items provides infants with additional information that might aid infants to define category boundaries. To address this possibility, instead of using interleaved design future studies should implement a between-subjects design having categorisation tested at the end of the entire familiarisation sequence, for instance after two or three blocks of familiarisation.

To summarise, this study added to our understanding of infants' ability to form categories based on relational information. Whereas previous studies showed that infants are able to learn these categories when presented with

line drawings of novel animals, Experiment 1 revealed that they fail in this task when presented with more complex realistic novel animals. It is possible that this is due to increased complexity of the to-be-learned categories. Since infants in the present study were presented with one item at a time, this process of comparing exemplars inevitably posed high demands in terms of memory capacities. A question we address in the following experiment is how changing familiarisation parameters in terms of perceptual and memory load might affect the process of category formation.

3.4 Experiment 2: paired familiarisation

3.4.1 Introduction

In addition to perceptual properties of to-be-learned categories, other characteristics of a learning situation may have a major impact on the processes at work during categorisation. In a recent study, Carvalho and Goldstone (2014) have shown that the structure of a category may influence how efficiently the category representations will be formed. Nevertheless, they discovered that this effect is closely tied to the way in which category members are presented. They exposed adult participants to a set of category exemplars, either one at a time or in pairs and discovered that these different learning conditions led to different category-specific allocation of attention: whereas simultaneous presentation promotes attention to commonalities among objects, sequential presentation promotes attention to differences among objects. Similarly, another study has shown that, the opportunity to compare or contrast simultaneously presented items can have a different effect on category learning (Ankowski et al., 2013). In a match-to-sample task, four-year-old children were familiarised with a set of items from the same or different categories. To test category extension, children were presented with a new set of items and asked to pick one that also belongs to the familiarised category. They found that the opportunity to compare multiple members of the same category promotes information about category membership while contrasting category members to non-category members highlights information about category boundaries. In addition, the authors found the contrasting process

more useful when members of the to-be-learned category were perceptually highly similar (Ankowski et al., 2013).

These studies suggest that categorisation should not be seen only as an extraction of abstract rules or computation of feature statistics, but emphasise the importance of learning conditions in shaping category formation processes. To further illustrate, after being familiarised with a series of pictures of dogs, when presented with a new dog and a cat, 4-month-old infants tend to look more at the cat. This suggests that infants at this age can already distinguish between the categories of dogs and cats (Eimas & Quinn, 1994). However, this was the case only when infants were familiarised with pairs of objects. In contrast, when presented with the same set of exemplars one-by-one, infants failed to show any evidence of categorisation (Oakes & Ribar, 2005).

In fact, many of the studies that report evidence of categorisation in young infants utilised a paired presentation regime during the familiarisation phase (Bomba & Siqueland, 1983; Eimas & Quinn, 1994; Quinn, 1987). Paired presentation was beneficial even in the case when two identical items were presented in pairs and similar enhancing effects of paired presentation are noted even at the level of mere object recognition (Fagan, 1978; Oakes et al., 2009).

The facilitative effects of paired presentation on category learning in young infants may be driven by the different memory and attention demands posed by exploring items one-by-one or in pairs. One of the factors making

category acquisition difficult in a successive learning regime is the limited memory ability in young infants. Difficulties in keeping and updating information in short-term memory may lead to a failure in extracting category relevant information. In addition to posing different memory demands, successive and paired learning regimes may also elicit different attention strategies. While reducing memory demands, paired presentation increases demands on infants attention because infants must select which stimulus to attend and shift between multiple items (Oakes & Ribar, 2005).

The present study aims to further explore the extent to which paired presentation facilitates category learning in comparison to sequential learning. Presenting items in pairs might simply lead to a more immediate motivation to contrast and compare items. This may result in comparisons at different representational levels in the two learning conditions. In the paired condition, two items may be compared on the same level of representation, whereas in the successive condition this might not be the case. In successive learning, items may be compared on a more abstract level, as a result of the absence of detailed memories of previously presented objects. Thus, the two learning situations may involve comparisons at different representational levels. The present experiment investigates to what extent the opportunity to compare items, while having both visually available, facilitates category learning. With an exception of the familiarisation regime (paired instead of successive presentation), the experimental design was identical to the one used in Experiment 1. This allows for a direct comparison of the effect of

different learning conditions on infants' category formation and exploration of the role of memory and perceptual demands in extracting the similarities and differences between objects that enables category learning.

3.4.2 Method

The design of Experiment 2 was identical to Experiment 1 except for the presentation regime during training. Instead of successive presentation, training items were presented in pairs.

Participants

Twenty-eight 10-month-old infants took part in this experiment (15 girls and 13 boys, $M_{age} = 10.15$, age range = 9.61 - 10.69). Two infants were excluded from the later analyses due to failing to reach a minimum amount of looking time during familiarisation. Therefore, of the 26 participants in the final sample, 25 contributed to the data of the second block and 21 to the final third block.

Stimuli and Procedure

The stimulus set used in this experiment was identical to the one used in Experiment 1. As opposed to the sequential presentation in the first experiment, familiarisation items were presented in pairs. A total of four trials were presented in each learning block. Following the presentation of an attention getter, two objects were presented simultaneously for 12,000 ms. Total familiarisation time was thereby identical to Experiment 1 (see Figure 3.17).

Pairs of objects were selected based on the sequences used in Experiment 1, i.e. each sequence used in Experiment 1 had a corresponding paired sequence in Experiment 2 (for example, trial 1 and 2 in the successive regime corresponded to trial 1 in the paired presentation). It is important to note that despite the difference in the number of trials in each learning block, the total duration of familiarisation was identical across the two experiments. Therefore, any potential differences in the effects of the two familiarisation regimes cannot be due to a difference in the amount of time infants were given to explore the objects.

Test trials were identical as in Experiment 1 (described in Section 3.2.2).

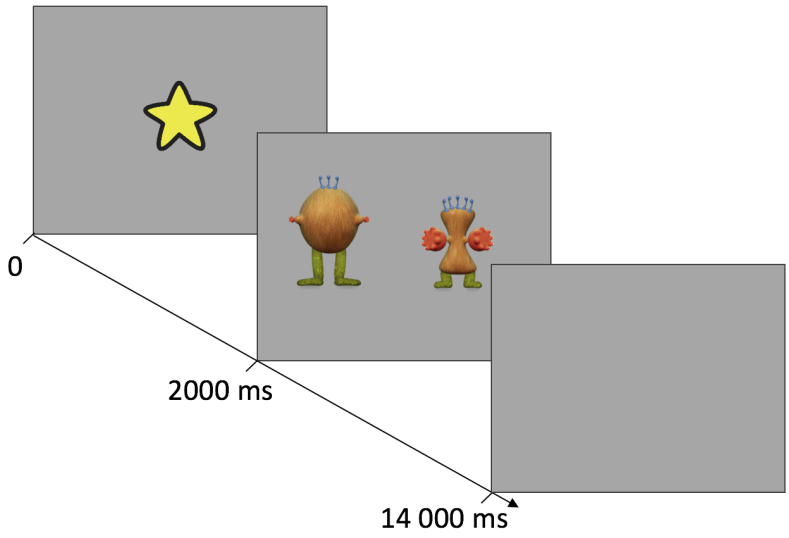


Figure 3.17. Experiment 2: Structure of the familiarisation trials.

3.4.3 Results

Training

Visual inspection of mean looking times in familiarisation trials suggested that there was a decrease in the amount of looking time over the course of the three blocks (Figure 3.18).

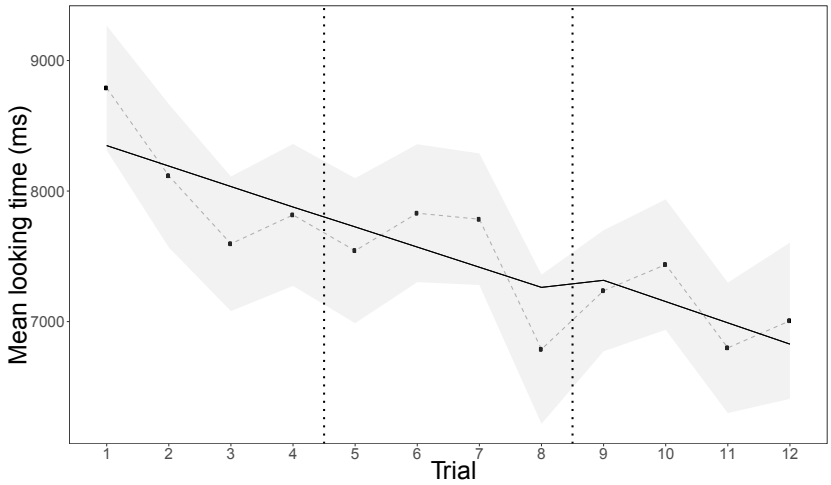


Figure 3.18. Experiment 2: Mean looking time across infants in each training block (dotted line depicts mean looking times in each trial, solid line represents model values, vertical dotted line represents time points when test blocks were presented; grey envelope depicts standard error).

There was a significant decrease in looking time over the course of the experiment ($\chi^2(1) = 7.90,\ p < .01$). Mean looking time decreased approximately 156.89 ms per trial ($SE = 50.85$, table 3.3).

Table 3.3. Experiment 2: Growth Curve Analysis. Parameter estimates.

	Estimate	Std. Error	df	t value	p value
Intercept	8505.30	441.93	25.67	19.25	**<.001**
Slope (Trial)	-156.89	50.85	24.49	-3.09	**<.01**

Test - Category Formation

For each test trial, a preference score was calculated by dividing looking at the inter-category item by total looking time to both the inter-category item and the category prototype, and averaged for two trials of each test block. An ANOVA on the mean preference with the factor *Block* (1, 2 and 3) revealed a near significant main effect of *Block* ($F(2, 40) = 3.07,\ p = .057,\ \eta_g^2 = .1$)[2]. Planned comparisons of the performance in each test block against chance (0.5) revealed that infants showed evidence of category formation already after one block of training, as they fixated the inter-category item significantly above chance ($t(25) = 2.91,\ < .01,\ d = .57$). Mean preference values and dispersion are depicted in Figure 3.19 and Table 3.4.

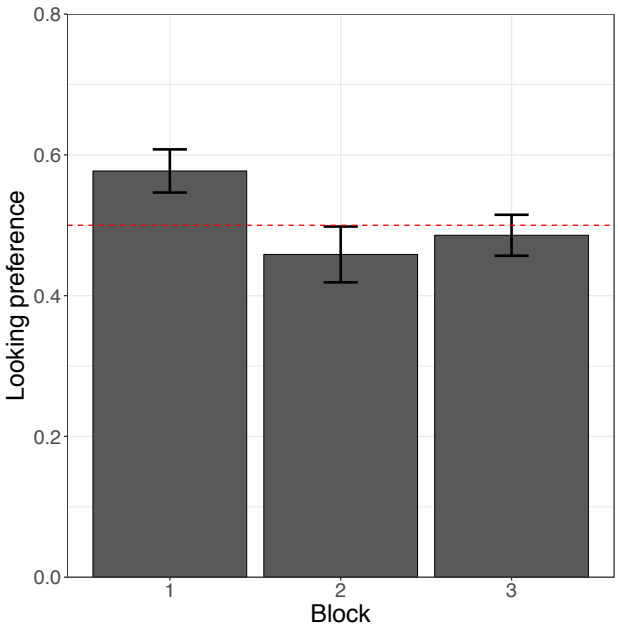


Figure 3.19. Experiment 2: Looking preferences in 3 test blocks (higher preference value means that they prefer the inter-category item, whereas lower preference value indicates preference for the category prototype).

[2]Only participants who contributed to all trials were included in this analysis (N=21)

Test - Novelty Preference

A one-way ANOVA with *Block* as a within-subjects factor revealed a significant effect of *Block* ($F_g(2, 40) = 3.46,\ p < .05,\ \eta_g^2 = .09$). Planned comparisons revealed that infants' preference for the novel object differed from chance only in the first block ($Mdn = .71,\ V = 268,\ p < .05,\ d = .48$, Table 3.4.). This result supports the assumption that infants looking behaviour in the first test was driven by novelty preference.

Table 3.4. Experiment 2: Mean looking preferences on test (SDs provided in brackets.

Block	**Test**	**Novelty**
1	**.57** (.12)*	**.63** (0.26)*
2	.46 (.18)	.48 (0.29)
3	.49 (.14)	.60 (0.27)
		$*p<.05$

Exploring looking patterns in training

In the current experiment infants were familiarised with two items simultaneously which provided them with an opportunity to directly compare the two presented items. In this section we explored to which extent infants exploited the opportunity to compare presented pairs. We tested whether there is a change in the number of shifts between the two presented objects over the course of familiarisation trials. In addition, we aimed to investigate if the number of shifts between the two items may be dependent on whether these two items stem from the same or from two different categories. To test this,

we conducted a growth curve analysis using a similar approach as described in Section 3.2.4. Mean number of shifts across familiarisation was modelled with fixed effects of *Trial* (1-12) and *Comparison Type* (items coming from the same or different categories) and random effect of *Participants* on the intercept. Analyses revealed that there was a significant effect of *Trial* on the slope, indicating a decrease in the number of shifts over the course of the experiment ($\chi^2(11) = 36.89,\ p < .001$). No effect of *Comparison Type* on the number of shifts was present indicating that there was an equal number of shifts in trials consisting of items from the same and from two different categories (see Figure 3.20 and model parameters in Appendix C.1).

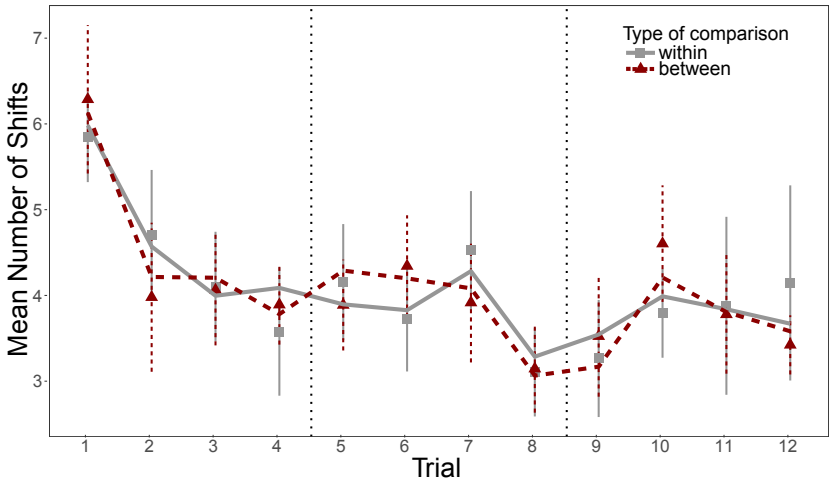


Figure 3.20. Mean number of shifts in each trial over the course of three familiarisation blocks (grey squares depict mean number of shifts for within category comparisons, red triangles depict between categories comparisons; error bars depict standard errors; grey and red lines depict model predictions for within and between categories comparisons, respectively).

The relationship between looking patterns in training and test performance

To explore whether performance in test trials was driven by different attention strategies in training, the relationship between looking patterns expressed during the first familiarisation block and subsequent test was analysed. In Experiment 2, two items were presented simultaneously for 12,000 ms during familiarisation. During this period infants could shift attention from one item to another, providing them with the opportunity for immediate comparison under a reduced memory load. If the opportunity to compare the different items directly is the key factor that makes paired familiarisation beneficial, we assume that the infants who take this opportunity might perform better than those who do not. To explore whether infants who make more shifts differ in test performance as compared to those who shift less, mean and total number of shifts during familiarisation were calculated for each infant.

It was hypothesised that high rates of shifting lead to a more robust category representation and also to quicker emergence of a novelty preference at test. Based on a median split (4.12), the group of low shifters reveal an average of 3.00 shifts ($SD = 0.96$) whereas high shifters 6.25 shifts ($SD = 1.20$) in each trial[3]. As shown in Figure 3.21, high-shifters demonstrate stronger novelty preference for the inter-category item. This is particularly evident in the first half of the trial, rising faster and remaining steady for

[3]High and low-shifting groups were the same when split based on shifting in the first block and based on mean shifting rates across the three familiarisation blocks

the remainder of the trial. In contrast, low-shifters express a slower rise in novelty preference which peaks approximately half way through the trial.

To model these differences in looking preferences among low and high shifters we used an empirical-logit growth curve analysis (Mirman, 2014). Time was modelled using linear, quadratic and cubic polynomials. Low-shifting was coded as a reference level to model how the probability of fixating the novel object changed over time for the two groups of infants. Models included *Participant* as a random effect because we expected that there might be individual differences within the groups. There was a significant main effect of *Shifting Group* ($\chi^2(1) = 1.96,\ p = 0.05$), but the interaction term did not reach significance. These results indicate that high shifters demonstrated stronger novelty preference than low shifters and maintained that novelty preference throughout the duration of the trial (model parameters are given in Appendix C.2).

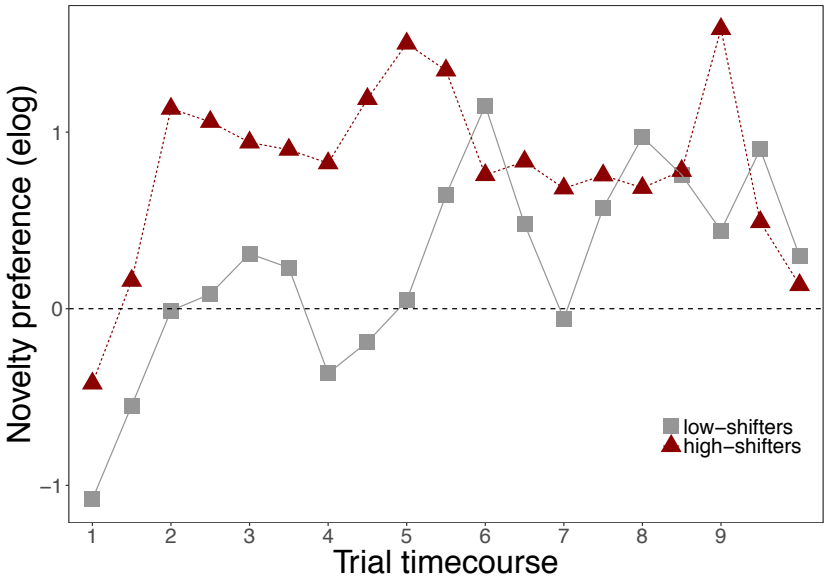


Figure 3.21. Looking preferences of low and high shifters in the first test block (x-axis depicts trial time course in seconds; y-axis depicts elog-transformed novelty preference).

Exploration of low- and high-shifters' looking patterns during training: habituation rates

To explore whether infants who make more shifts also express different habituation trends as compared to low-shifters, mean looking time over the course of familiarisation was modeled using a growth curve analysis (Mirman, 2014). The data were fit using a linear growth model with a fixed effect of *Group* (low- vs. high-shifters) on the intercept and linear terms (decline in the looking time over time) and a random effect of *Participants* on the intercept and slope to model individual differences in initial amount of looking time and rate of habituation. The fixed effects of group were added individually and their effects on model fit were evaluated using model comparisons. Improvements in model fit were modelled using 2-times the change in log-likelihood,

which is distributed as χ^2 with degrees of freedom equal to the number of parameters added (as in Mirman, 2014). The data and model are shown in Figure 3.22.

There was a significant main effect of *Group* on the intercept ($\chi^2(7) = 5.58$, $p < 0.05$), indicating that high-shifters initially accumulated more looking than low shifters. They had overall higher looking times across the entire experiment, whereas there were no differences in the rate of change (adding an interaction term didn't improve the model fit, model estimates are given in Table 3.5).

Table 3.5. Experiment 2: Growth Curve Analysis. Parameter Estimates.

	Estimate	Std. Error	df	t value	p value
Intercept	7827.13	493.29	34.13	15.87	**<.001**
Trial	-161.46	50.89	24.35	-3.17	**<.001**
Group	1380.27	553.47	26.15	2.49	**.02**

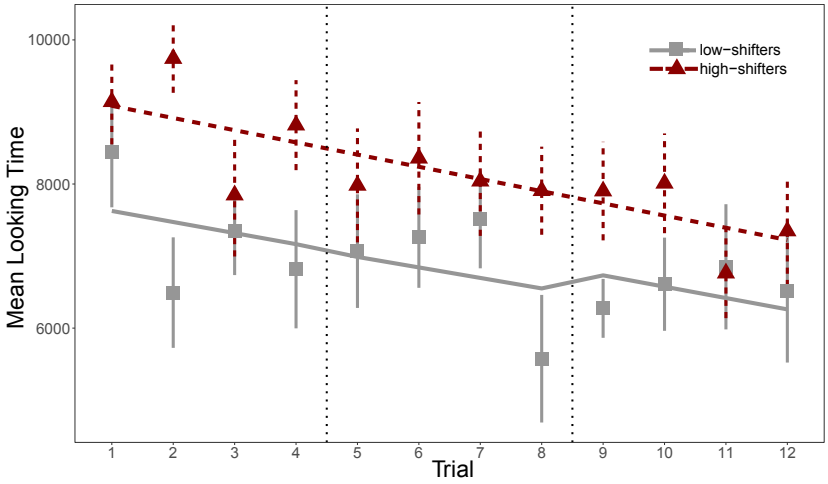


Figure 3.22. Mean looking time of low- and high-shifters in each training block in Experiment 2 (grey squares depict low-shifters, red triangles high-shifters; error bars depict standard error; grey and red lines depict model predictions for low- and high-shifters, respectively).

Exploration of low- and high-shifters' looking patterns during training: fixation characteristics

To further explore whether there are systematic differences in the looking behaviours of infants who exhibit low and high shifting rates, basic parameters of looking were analysed. Several fixation characteristics were examined: minimum and maximum fixation duration, mean fixation duration and time to first look (time needed to land the first fixation). The analysis revealed that there are differences in the mean and maximum fixation duration. Infants in the low-shifting group have longer mean fixations (low shifters: $M = 1559.86$, $SD = 593.36$; high-shifters: $M = 1103.11$, $SD = 244.27$; $W = 152$, $Z = -2.50$, $p < .05$, $d = 0.67$) and longer maximum fixation duration (low-shifters: $M = 4615.41$, $SD = 2494.75$; high-shifters: $M = 2965.48$, $SD = 929.86$; $W = 147$, $Z = -2.25$, $p < .05$, $d = 0.6$), as shown in the bottom row of Figure 3.23). No differences in time to first look and

minimum fixation duration were present (*all ps* > .05). In addition, by definition, high-shifters accumulated more fixations ($M = 23.14$, $SD = 3.84$) than low-shifters ($M = 15.79$, $SD = 6.08$, $t(22) = -3.89$, $p < .001$, $d = 1.45$).

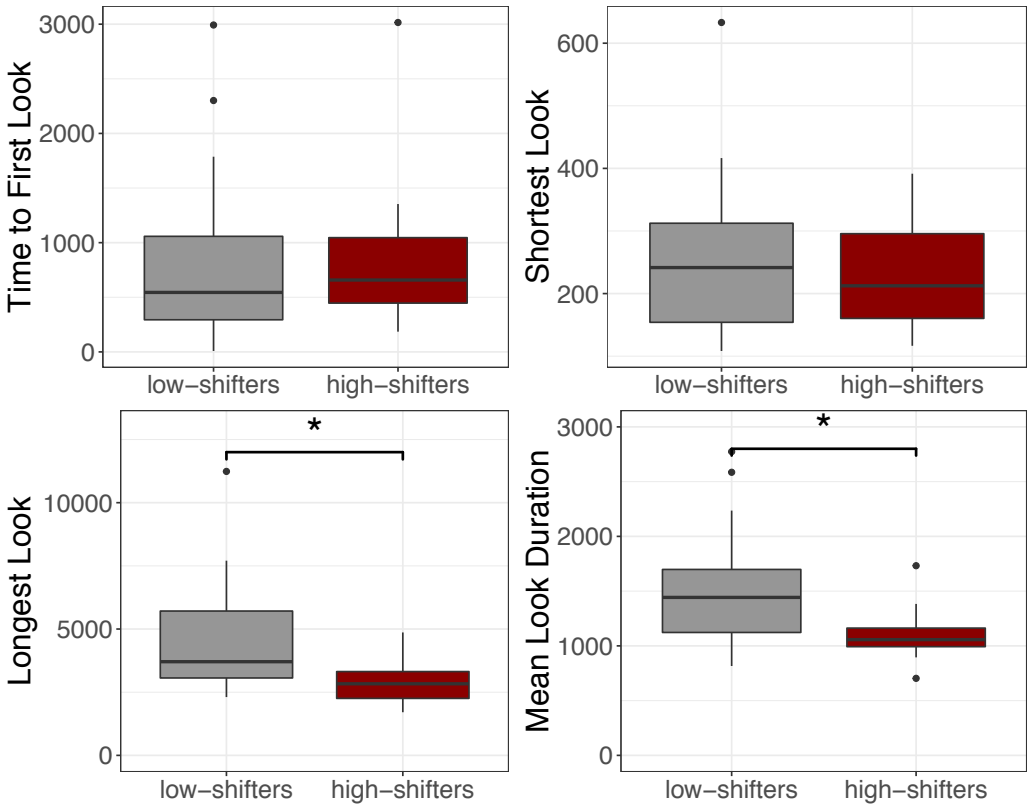


Figure 3.23. Fixation characteristics of low- and high- shifters (values on y-axis are given in milliseconds; boxplot depicts median and interquartile range).

Exploration of low- and high-shifters' looking patterns: Looking outside of areas of interest

The analysis presented in the previous section revealed that high-shifters accumulated more looking time as compared to low-shifters. However, this analysis focused only on looking time to the areas of interest (AOIs) while

fixations outside of AOIs were not taken into account. In the current section, we further explore whether there are also differences between the groups in the amount of looking to areas other than the two familiarisation items. The proportion of the total looking time fixations falling outside of the AOIs was calculated (see Figure 3.24). A mixed measures ANOVA with between-subjects factor Group (low- vs. high-shifters) and a within-subjects factor *Familiarisation Block* (1,2 and 3) revealed a significant main effect of *Familiarisation Block* ($F(2, 40) = 4.88$, $p < .05$, $\eta_g^2 = .08$) suggesting that there is a increase in the amount of time infants spent looking at areas outside of regions of interest. There was no effect of group nor an interaction, suggesting that the observed decline does not differ between the two groups ($all\ ps > .05$).

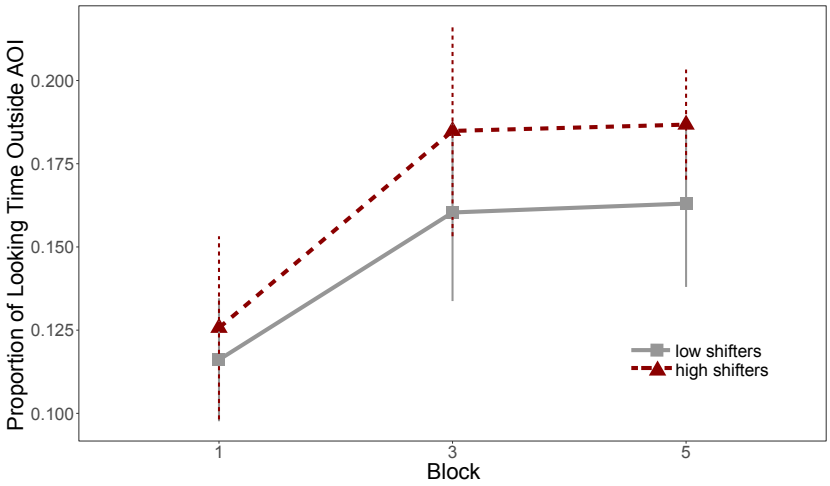


Figure 3.24. The proportion of looking time outside of the areas of interest in infants with low and high shifting rates (error bars depict standard error).

Familiarisation time in Experiments 1 and 2

Familiarisation time was equated across Experiments 1 and 2 in order to rule out the possibility that performance might be due to differences in the levels of familiarisation in the two experiments. Even though infants were given the same opportunity to explore the objects, it is important to explore whether infants in the two experiments accumulated different looking times during familiarisation. A mixed ANOVA with a *Block* as a repeated factor (familiarisation block 1, 2 and 3) and *Familiarisation Regime* (successive and paired) as a between-subjects factor was conducted. The analysis revealed a significant main effects of *Block* ($F(2, 88) = 11.84,\ p < .001,\ \eta_g^2 = .09$) and *Familiarisation Regime* ($F(1, 44) = 16.41,\ p < .001,\ \eta_g^2 = .19$) suggesting that infants in Experiment 1 accumulated more looking time during familiarisation than infants in the Experiment 2 (see Figure 3.25). These findings indicate that the rapidity of categorisation in the paired presentation in Experiment 2 cannot be a result of greater familiarity with the training stimuli.

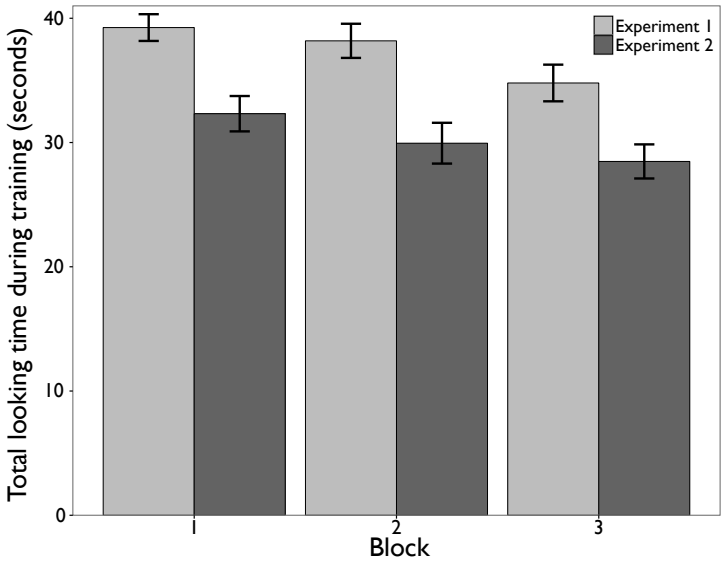


Figure 3.25. Total looking time accumulated during familiarisation blocks in Experiment 1 (successive familiarisation, depicted in light grey) and Experiment 2 (paired familiarisation, depicted in dark grey; error bars depict standard error).

3.4.4 Discussion

In contrast to results of Experiment 1, where infants showed no evidence of category formation after three blocks of familiarisation, Experiment 2 revealed that infants form categories already after one block of familiarisation. These results clearly demonstrate how learning context can affect the dynamics of category learning and that paired presentation leads to more efficient learning. This result is in line with previous studies which revealed that paired presentation may have positive effects on learning and memory in younger infants (Fagan, 1978; Oakes et al., 2009; Oakes & Ribar, 2005). Our results support and further extend these findings by demonstrating that even in the case of older infants, when memory abilities are more developed, paired presentation still proves to be beneficial for learning. It is important to emphasise that these differences were not driven by differences in the amount

of familiarisation preceding the test phase. Infants in both experiments were given the same opportunity to visually explore items. In fact, infants in the successive learning regime accumulated more looking time during familiarisation then did infants in the paired learning regime. Therefore, the observed differences cannot be due to differences in familiarisation experience. In addition, the length of experiments was kept constant across the two experiments, that is, time from start of familiarisation to test was identical, overcoming limitations of some previous studies.

What makes presenting items in pairs so beneficial for learning? As pointed out by Markman and Ross (2003) properties of category representations are dependent on the way they are learnt. Presenting items in pairs provides more opportunity to compare and contrast available information (Markman & Ross, 2003). Since visual information is constantly available in paired presentation, infants can immediately examine similarities and differences between the two items. Furthermore, infants are at liberty to decide for themselves how to allocate attention to the paired visual stimuli. In contrast, in the successive training regime, direct comparisons are not possible. Infants can only compare a visually presented item with a working memory representation of the previously presented item. Inevitably, this leads to comparisons performed across different representational levels. The results of my study suggest that more direct lower-level comparisons are more useful for the process of category formation in infants.

The facilitation effects of paired presentation on category learning are in

line with research on categorisation in children and adults. Moreover, comparison has been emphasised as one of the central factors of category learning in several models of categorisation (Gentner & Namy, 1999; Kotovsky & Gentner, 1996; Kurtz et al., 2013). For instance, Gentner and Namy (1999) suggested that comparing perceptually and conceptually similar objects can provide insights into the structure of a category and that exposure to multiple items facilitates category abstraction by a process of structural alignment. Comparison and alignment of two representations highlights common structures and relational knowledge (Gentner & Namy, 1999). This process is claimed to serve as a bridge from an initial perceptually-based category to a later more sophisticated category representation in children (Kotovsky & Gentner, 1996) and in adults (Kurtz et al., 2013).

Further support for the claim that the opportunity to directly compare items is more beneficial for learning is provided by the more precocious performance of infants who shifted more frequently between stimuli during training. High shifting in training correlated with a better performance at test. Several previous studies have shown that looking behaviour during training predicts test performance (Jankowski et al., 2001; Johnson et al., 2004; Kovack-Lesh et al., 2008; Rose et al., 2003). More shifts and briefer looks were associated with a stronger novelty preference in a recognition task (Rose et al., 2003). A similar pattern of results was found in the present study. We discovered that high-shifters had shorter fixation durations as compared to the low-shifting group. No differences in the habituation rates between infants who

exhibit low- and high shifting rates were registered, although the group of high-shifters accumulated higher amounts of looking time. In addition, there was no difference in the amounts of time the two groups spent looking outside of relevant areas of interest. These results speak in favour of an account which proposes that shifting reflects a propensity for active comparison and individual differences in speed of processing (Kovack-Lesh et al., 2008) and leads to more efficient category formation.

To summarise, in this study we show that paired presentation is beneficial for category learning. Infants fail to learn categories in a successive familiarisation regime, but are successful in a paired familiarisation regime. Thus, presenting items in pairs facilitates learning. Nevertheless, the current study does not provide answers as to what makes comparative information beneficial for learning. Does presenting items in pairs highlight similarities and category membership or does it point to differences and category boundaries?

To answer this question, we explored whether infants engaged more in comparison when presented with two items coming from the same category and with pairs of items from different categories by comparing the shifting rates in the two types of trials. No difference in the shifting rates for pairs of items from the same and from different categories was found suggesting that there is no difference in the comparison strategies. However, the design of the present study was not optimal for addressing this question as trial pairs were selected to resemble the order of presentation used in Experiment 1 (to enable a direct comparison of results). As a consequence, number of

pairs coming from the same and pairs of items from two different categories in every trial was not balanced. In order to investigate to what extent the contrasting information is used, future studies need to manipulate the stimuli pairs in a way that they either emphasise differences (with pairs of items always coming from different categories) or similarities (by presenting pairs of items from the same category). In addition, it would be interesting to explore whether category complexity might modulate the effects of learning condition on the category formation process and determine whether comparative information remains beneficial even in the case of acquisition of less complex visual categories.

The fact that presenting multiple items simultaneously during training facilitates category learning also opens a question of whether there is a task load that is optimal for learning. In other words, would an additional increase in the amount of immediately available information further facilitate learning or, on the contrary, additional processing demands might hinder the learning process? To address this question, we conducted an additional experiment in which we increased the task load by presenting infants with four items simultaneously. The results of this study are presented in the following section.

3.5 Experiment 3: massed familiarisation

3.5.1 Introduction

The previous study highlighted the importance of comparison in category formation. Infants failed to learn categories when familiarised with one item at a time, but succeeded when familiarised with two items at the same time. Hence, presenting items in pairs facilitates category learning. In the present chapter I investigated whether an additional increase in the amount of immediately available information further facilitates category learning.

Presenting multiple items simultaneously might further reduce memory demands since four items can be immediately explored and compared, resulting in a further facilitation of category learning. However, increasing the amount of information immediately available by presenting multiple items simultaneously poses higher demands in terms of attention, particularly in terms of selective attention, as infants must be able to focus on one item while ignoring other, potentially distracting items and also be able to efficiently disengage from one item in order to explore other available items. Hence, it is possible that the increase in available information would not facilitate but on the contrary hinder the category formation process. To address this question, infants in the present study were familiarised with four items simultaneously. Keeping other aspects of the experimental procedure identical to those in previous experiments provides an opportunity to compare performance of infants exposed to different familiarisation regimes and explore how characteristics of familiarisation shape category formation.

3.5.2 Method

The experimental design of Experiment 3 was identical to Experiment 1 and 2 with an exception of the familiarisation regime. In the present experiment, infants were familiarised with four items simultaneously.

Participants

Thirty 10-month-old infants took part in this experiment (16 girls and 14 boys, M_{age} = 10.17, age range = 9.60 - 10.69). Two infants were excluded due to a failure to establish tracking of their eye-movements and an additional one due to fussiness. Of the 27 participants in the final sample, 24 completed the second block and 22 contributed to the data for the final of three blocks.

Stimuli and experimental design

The stimulus set used in this experiment was identical to Experiment 1 and 2. Three familiarisation blocks were presented interleaved with three test blocks.

In each familiarisation block, infants were presented with a set of eight items (thus, total of 24 items were presented over the course of the experiment). In each trial, four items were presented at the same time. The two trials were presented twice. This manipulation was introduced in order to keep the number of trials identical to Experiment 2. Location of the items on the screen was changed from first to second repetition (i.e. first to third and second to fourth trial), so that each items' location was not in any way predictable on the basis of category membership.

Following presentation of an attention getter, four objects were presented simultaneously for 12,000 ms (see Figure 3.26). Total familiarisation time was thereby identical to Experiments 1 and 2.

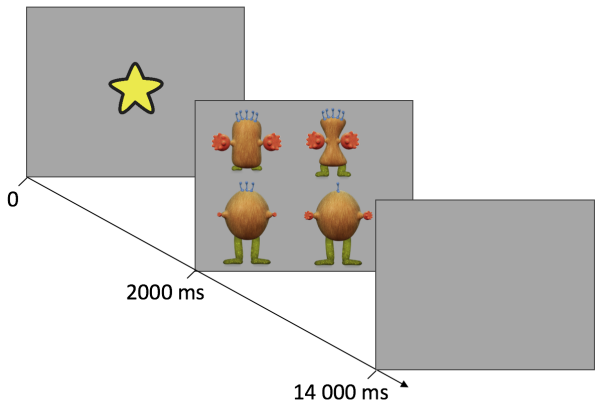


Figure 3.26. Experiment 3: Structure of familiarisation trials.

The choice of sets of four objects presented in a familiarisation trial was based on the sequence used in Experiment 2. The pairs of items presented in two trials in Experiment 2 were presented in one trial in the current experiment. This way, presentation sequences corresponded to sequences used in Experiment 1 and 2. For example, as illustrated in Figure 3.27 four items presented in the first familiarisation trial of Experiment 3 corresponded to the first two trials of Experiment 2 and first four familiarisation trials of Experiment 1. This enabled keeping presentations sequences as similar as possible across the three experiments.

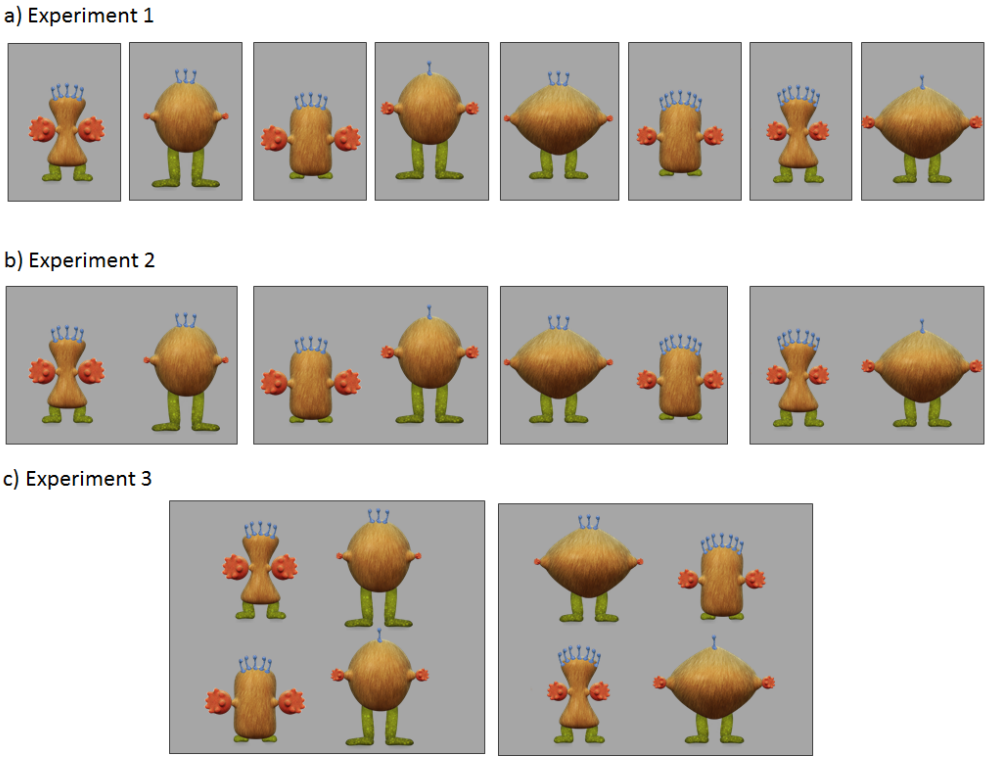


Figure 3.27. Comparison of corresponding sequences of one familiarisation blocks across Experiment 1, 2 and 3: (a) items presented one at a time in Experiment 1; (b) pairs of items presented in Experiment 2; (c) quadruplets presented in Experiment 3.

The test was identical as in Experiment 1 and 2 and details are given in Section 3.2.2.

3.5.3 Results

Training

Mean looking time was calculated for each trial of the three familiarisation blocks. Visual inspection suggested that there is a decrease in the amount of looking time over the course of the experiment (see Figure 3.28). Mean looking time decreased on average 257.21 ms per trial (SE = 59.68, parameter estimates are given in Table 3.6). This was confirmed by a growth

curve analysis ($\chi^2(6) = 14.3,\ p < .001$). In addition, infants spent an equal amount of time fixating items from the two categories ($p > .05$).

Table 3.6. Experiment 3: Growth Curve Analysis. Parameter estimates.

	Estimate	Std. Error	df	t value	p value
Intercept	8692.17	422.69	28.29	20.56	**<.001**
Slope (Trial)	-257.21	59.68	21.54	-4.31	**<.001**

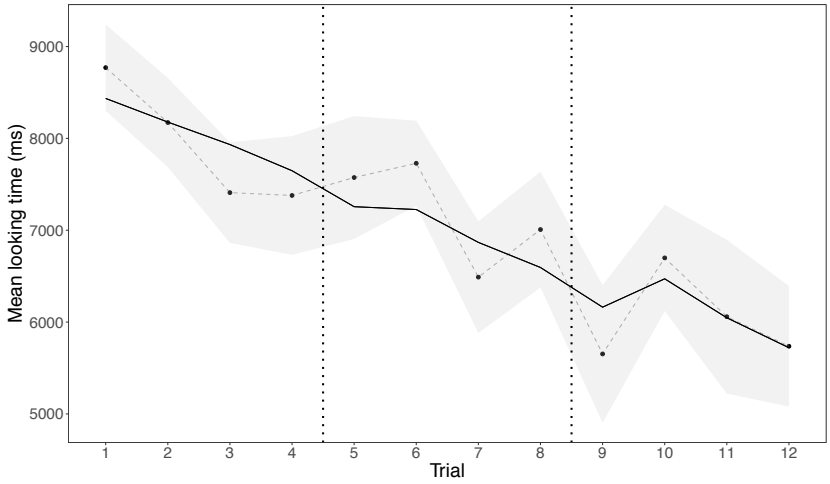


Figure 3.28. Experiment 3: Mean looking time across infants in each training block (dotted line depicts mean looking times in each trial, solid line represents model values; vertical dotted lines represent time points when test blocks were presented; grey envelope depicts standard error).

Test - Category Formation

An ANOVA with a within-subject factor *Block* (1, 2 and 3) on the mean looking preference revealed a significant main effect of Block ($F(2, 42) = 4.04,\ p < .05,\ \eta_g^2 = .12$). To perform planned comparisons, the average looking preference in each test block was compared against chance (.5). There was a significant preference for the category prototype in the first block ($Mdn = .44,\ V = 106,\ p = .05,\ d = .45$), followed by no preference

for any of the test items in the second block ($p > .05$) and lastly, there was a significant preference for the inter-category item in the last test block ($M = .57,\ SD = .16,\ t(21) = 2.21,\ p < .05,\ d = .47$). These results suggest that infants initially grouped items into one category and after three familiarisation blocks infants formed two categories (see Figure 3.29).

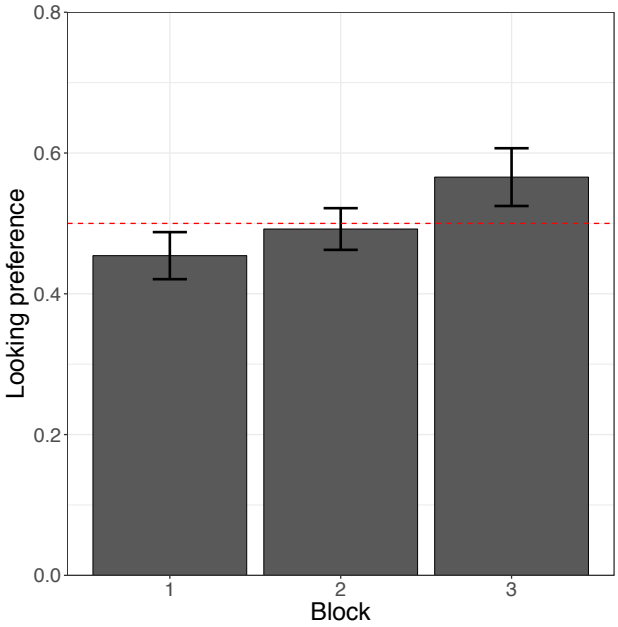


Figure 3.29. Experiment 3: Looking preferences in the test blocks (error bars depict standard error, red dotted line depicts expected preference if performance was at chance level, higher preference value indicates preference for the inter-category item).

Test - Novelty Preference

A one-way ANOVA with *Block* as a within-subjects factor revealed no effect of *Block* ($p > .5$). Planned comparisons revealed that infants expressed a novelty preference in the first block ($M = .63,\ SD = .22,\ t(26) = 3.003,\ p < .05,\ d = .58$), no systematic preference in the second block and a novelty preference in the final test block ($M = .57,\ SD = .16,\ t(21) = 2.08,\ p = .05,\ d = .44$, see Figure 3.30).

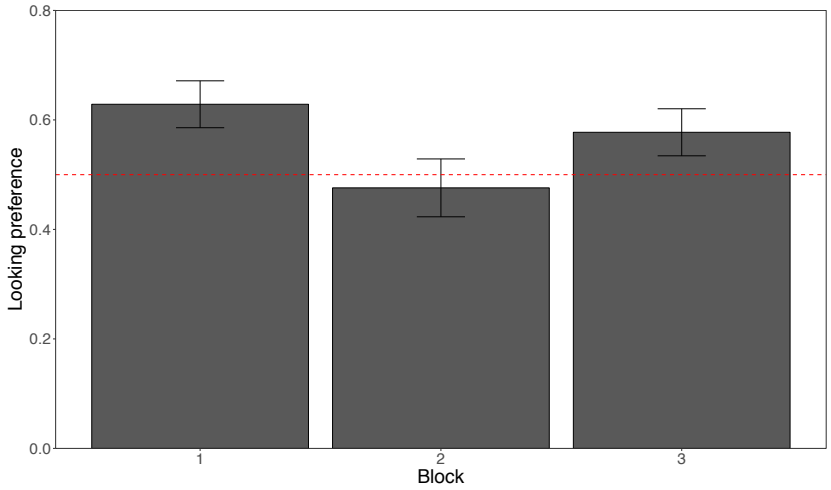


Figure 3.30. Experiment 3: Mean looking preference for the novel item across infants in the test blocks (error bars depict standard error, red dotted line depicts expected preference if performance was at chance level, higher preference value indicates preference for the novel item).

Taken together, infants' looking preferences in test suggest that three blocks of familiarisation are needed for infants to express evidence of category formation. Interestingly, infants seem to express a preference for the category prototype in the first category formation test block suggesting that infants find category prototype less familiar. As infants express a novelty preference in the novelty preference trial that immediately followed, it is possible that they formed one category. Another plausible explanation is that infants' looking was driven by a familiarity preference, suggesting that the process of category formation was in progress at the time when test block was presented.

Infants' exploration strategies in training

Familiarising infants with four items simultaneously provides them with an opportunity to compare and contrast multiple items while all of them are visually accessible. Experiment 2 revealed that, when presented with items in

pairs, infants with higher shifting rates were faster to demonstrate a novelty preference. To explore whether more shifting leads to better performance when infants are presented with four items simultaneously, two aspects of exploration strategies were analysed - shifting rates over the course of familiarisation trials and shifting strategies, i.e. whether shifts were more frequent for within- and between-categories.

Shifting rates

Mean number of shifts was calculated for each trial of the three familiarisation blocks. As depicted in Figure 3.31, the shifting rate decreased over the course of the experiment. This was further confirmed by a growth curve analysis in which mean number of shifts across the experiment was modelled with fixed effect of a Trial (1-12) and a random effect of *Participants* on the intercept and slope using a linear growth curve model (Mirman, 2014) in R (lme4 package, version 1.1-12). There was a significant effect of *Trial* on the slope, showing a decrease in shifting rates over the course of the experiment ($\chi^2(6) = 14.21,\ p < .001$). On average there was a decrease of .29 (SE=.07) shifts per trial.

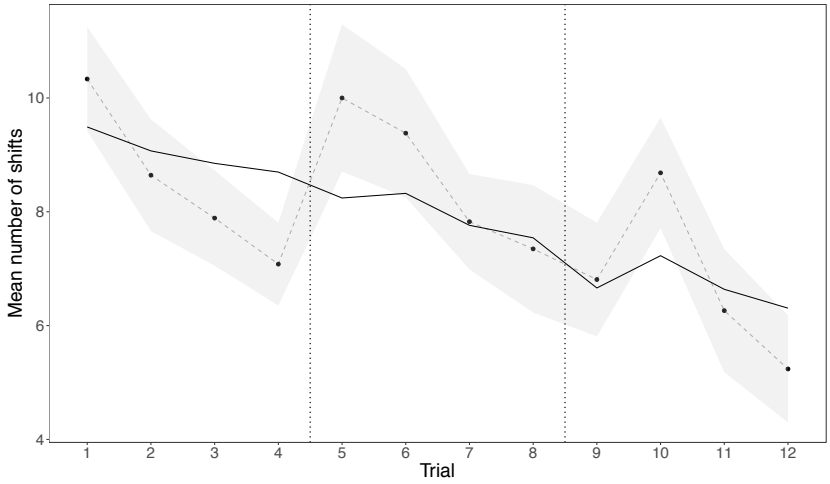


Figure 3.31. Experiment 3: Mean number of shifts across infants in each trial of the three familiarisation blocks (dashed line depicts mean number of shifts in each trial, solid line represents model values, vertical dotted lines represent time points when test blocks were presented, grey envelope depicts standard error).

Shifting strategies: within- and between-categories shifts

With four items simultaneously available during familiarisation, infants may engage different strategies to explore and compare these items. One strategy may be searching for similarities among items, in which case we would expect infants to make more shifts to items from the same category than to items coming from different categories. An alternative strategy infants might employ is contrasting, that is searching for differences between objects. In case the contrast-focused strategy is more prevalent, more shifts to items from different categories are expected.

To explore whether either of the two strategies was present, characteristics of infants' shifting patterns were further explored using Markov chains (MC). MCs are stochastic systems that assume a definite set of states that a system can be in (Grinstead & Snell, 2003). The system is dynamic and moves from one state to another. Based on a history of these steps between states, using

MC it is possible to calculate transitional probabilities, i.e. probabilities of transitioning from one state to another (Grinstead & Snell, 2003). MC describe states (all potential values a system can take) and probabilities of transitioning from one state to another. The discrete time Markov chain is a type of MC that represents a sequence of states with the Markov property, which means that the probability of moving to the next state depends on the present state and not on the previous states (Spedicato, 2017).

One frequently used illustration of Markov chain implementations is a weather forecast. For example, lets assume there are two potential outcomes of the weather forecast (the two potential states of the system) - dry and rainy weather. The question is whether it is possible to predict the weather on the next day based on the weather on the current day. Based on the history of the weather over a period of time, Markov chains provide an information about the probability of the next day to be dry or rainy based on today's weather. If there are two potential states, using MCs it is possible to calculate four transitional probabilities - probability of staying in one state (if it is dry today, it will be dry tomorrow), probability of staying in the other state (if today is rainy, tomorrow will also rain), probability of transitioning from state one to another (if it is dry today, the probability it will rain tomorrow) and lastly, probability of transitioning from the second to the first state (probability of having a dry weather tomorrow, if it was raining today).

I implemented MCs because of its potential to provide a fine grained measure of infants' exploration patterns during familiarisation and quantify

frequency of shifting between items from the same and from different categories.

The Markov chain was implemented using the "markovchain" package (Spedicato 2017, version 0.6.9.8-1) in R. This package was selected because it allows to fit a discrete time MC estimating the transitional matrix from a sequence of data. Only trials with an equal number of items from both categories were selected for this analysis (total of 130 trials). The model was defined with a space of two potential states, each state representing one of the two categories items can come from (Figure 3.32).

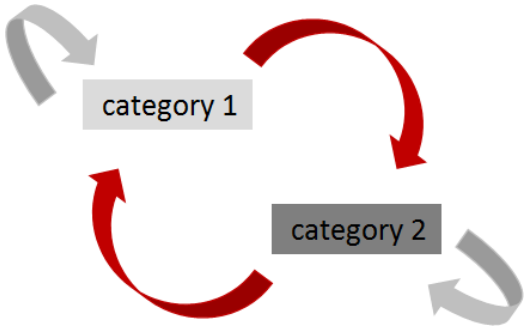


Figure 3.32. Markov chain with two states and four potential transitions between the states.

First, a transition path, i.e. a record of a looking sequence was calculated for each trial (for instance, sequence 1 2 2 1 is a trajectory of an infant who first looked at an item from category 1, then switched to category 2, then to the second item from the category 2, and then looked again at one of the items from category 1). In the following step, the Markov chain was fit to the obtained transition paths of each trial for each participant. As a result, we obtained mean transitional probabilities within categories (mean transitional probability of staying in category 1 and staying in category 2) and between

categories (mean probability of switching from category 1 to category 2 and from category 2 to category 1), then averaged for each familiarisation block (see Figure 3.33).

Since within- and between-category TPs are complementary (i.e. add up to 1), the analysis was performed only on between-category TPs. To test whether one of the shifting strategies (within- or between-category) was dominant, mean TPs for each block was compared against chance. Performance at chance would mean that infants were equally likely to shift to an item from the same category as to an item from a different category. With four items simultaneously presented, once an item is fixated, an infant could shift to one of the three remaining items. One of those items belonged to the same category, whilst two items were from a different category. Therefore, infants were by chance twice as likely to make a *between-category* shift than a *within-category* shift. This between-category TPs were compared against a chance level set to 0.66.

The analysis revealed that the transitional probability of between-category shifts did not deviate from chance-level performance in any of the three training blocks ($all\ p_s > .05$). There results suggest that none of the strategies was dominant in infants' exploration behaviour.

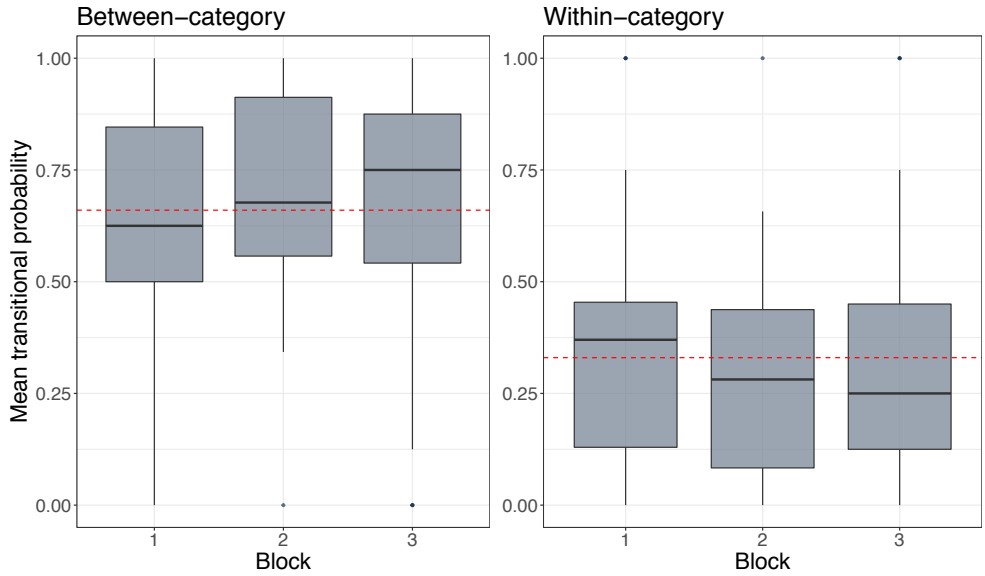


Figure 3.33. Mean transitional probabilities for *between-* and *within-category* shifts in the three familiarisation blocks (red dotted line depicts expected transitional probability if performance was at chance level).

Relationship between shifting rates and transitional probabilities

In the following step, the relationship between the shifting rates (mean number of shifts) and transitional probabilities (probability of shifting to an item from the same or from a different category) was examined. As probabilities of transitioning within category and between categories are complementary (add up to one), here we present an analysis looking at the relationship between shifting rate and the *between-category* transitional probabilities. Mean number of shifts and mean transitional probability of between categories shifts was calculated for each participant. No systematic relationship between the two measures of visual exploration were found ($r = -0.166$, $p > .05$, see in Figure 3.34), suggesting that probabilities of between category comparisons were not just a mere consequence of higher shifting rates in general.

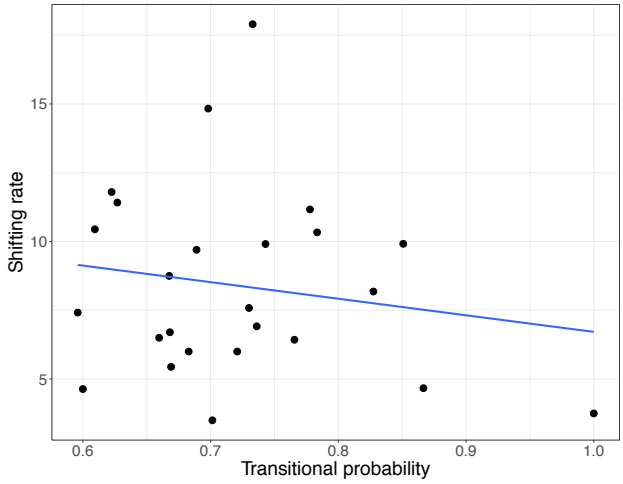


Figure 3.34. Relationship between mean number of shifts and between-category transitional probabilities.

The relationship between exploration strategies during training and performance in test

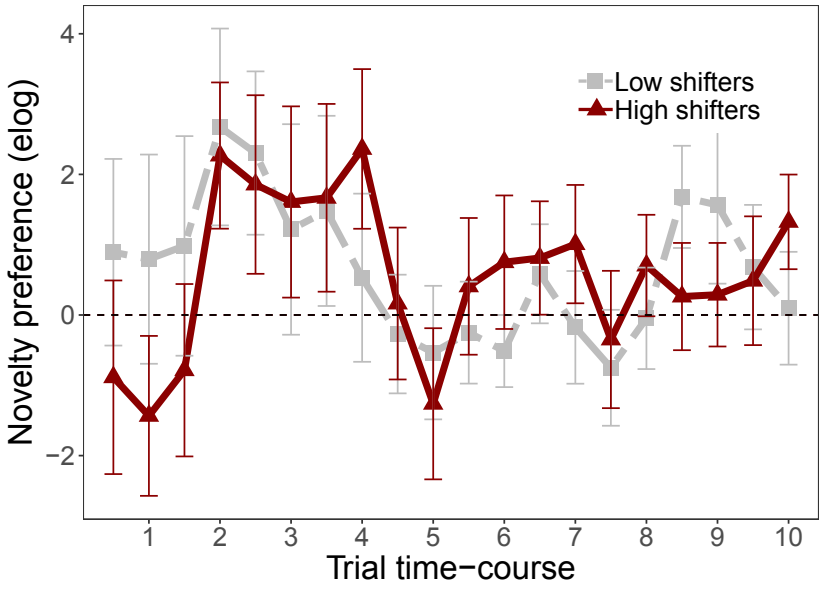
Results of Experiment 2 showed that higher shifting rates during familiarisation resulted in a faster novelty preference. To investigate whether higher shifting rates lead to stronger novelty preference when four items are presented simultaneously, the present analysis focused on the relationship between exploration strategies and performance in the first trial of the final test block.

Eye-tracking data was down-sampled into 500 ms bins by aggregating 60 raw samples (data was recorded with a 120 Hz sampling frequency, i.e. raw bins were 8.33 ms). For each bin, proportion of time the inter-category item was fixated was calculated by dividing number of fixations to the inter-category item by the total number of fixations. Then, proportions were elog transformed (corrected logit transformation, in order to account for 0 and 1 values, as recommended by Dink & Ferguson 2015). Infants were divided

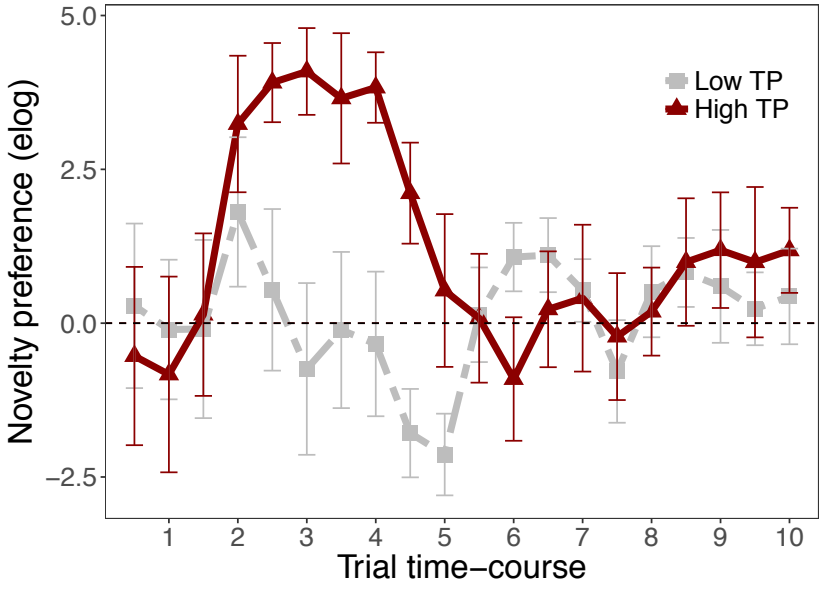
into low- and high-shifters based on median split, ($Mdn = 7.46$). Infants in the low-shifting group expressed on average 5.29 shifts ($SD = 1.57$), whereas infants in the high-shifting group performed 10.47 shifts ($SD = 2.86$). Figure 3.35a depicts time course of the looking preference during the trial separately for infants with low and for infants with high shifting rates in training. There seemed to be an early novelty preference effect in the first half of the trial, present for both low- and high- shifters. To further explore this, a mixed ANOVA was conducted with the between-subjects factor *Shifting Rate* (low- and high-shifters) and the within-subjects factor *Trial Part* (first and second half of the trial). No main effects nor interactions were significant ($all\ p_s > .5$).

As shifting rates and transitional probabilities seem to be independent indices of exploration patterns, an analogous analysis was conducted to explore an effect of different levels of transitional probabilities on the magnitude of the novelty preference. Infant were split into two groups based on the median split ($Mdn = .70$): infants that expressed low probabilities of between-category shifts (hereafter, low-TP, $M = .65,\ SD = .04$) and high probabilities of making transitions (hereafter, high-TP, $M = .79,\ SD = .08$) between items from different categories. As shown in Figure 3.35b, the two groups show different dynamics of the novelty preference over the course of the trial. Whereas the high-TP group exhibits a preference in the first half of the trial, the low-TP group expresses a later and weaker novelty preference. Only trials from participants who accumulated looking time

throughout the entire trial were included in the analysis (no gaze missing, number of included participants: N=17). A mixed ANOVA with the within-subjects factor *Trial Part* (first and second half) and between-subjects factor *TP Group* (low- and high-TP) revealed a significant interaction of *Trial Part* and *TP Group* ($F(1, 15) = 6.71,\ p = .02,\ ges = .14$) which was driven by the difference between the groups in the first part of the trial ($t(15) = -2.68,\ p = .017,\ d = 1.03$) with the high-TP group expressing a stronger novelty preference.



(a)



(b)

Figure 3.35. Time-course of the novelty preference in the low- and high-shifting groups (a) and in the low- and high-transitional probability groups (b), (error bars depict standard error).

3.5.4 Discussion

The present study investigated infants' category learning in a familiarisation regime where four items were presented at once. Infants' category formation was probed after each familiarisation phase. Again, as in Experiments

1 and 2, having familiarisation and test blocks interleaved allowed for investigation of the role of familiarity in infants' category formation. The results showed that after one block of familiarisation infants express a preference for the category prototype. There are two potential explanations for the observed result. One is that infants formed one category comprising of all items and their looking behaviour is driven by a novelty preference, which is also noted in the novelty preference trial in the same test block. Alternatively, infants preference for the category prototype could be driven by a familiarity preference. This would imply that infants already formed some rudimentary category representations, but are still in the early stage of learning and thus familiarity preference is a dominant looking strategy. This is in line with a classic model of infants' habituation (Hunter & Ames, 1988). Since the category formation is still in progress, infants express familiarity preference. However, in the immediately following novelty preference test, infants do recognise an out-of-category item as different to the emerging category representations and therefore express a novelty preference.

Manipulating the number of items immediately available during familiarisation put different demands in terms of perceptual and memory load. Having four items simultaneously available during familiarisation provided infants with an opportunity to compare freely these items while all items are visually available. This reduces memory demands since infants were able to directly compare items. At the same time, having multiple items simultaneously presented increases the amount of information and poses higher

attention demands - an infant needs to actively search and select relevant information while ignoring other information. In order to compare items, infants must disengage from one item and shift their gaze to another. Our results suggest that the number of shifts decreases over the course of the familiarisation blocks. Interestingly, better performance in test was not related to overall shifting rates, but shifting focused on differences (between-categories shifting) was related to a stronger novelty preference in test.

3.6 Habituation rates across experiments

The three experiments presented in this chapter utilised an identical experimental design that enabled comparison of infants' performance. Despite differences in a trial structure (number of exemplars simultaneously presented), overall familiarisation time was identical in the three experiments. In this section we explore whether there are differences in the habituation rates found in the three experiments. As habituation is an indicator of learning, we assumed that differences observed in test performance will also be present in characteristics of the habituation phase.

To compare habituation rates, the proportion of looking time was calculated by dividing the accumulated looking time by the trial duration. Experiment 1 consisted of 24 familiarisation trials, while Experiments 2 and 3 of 12 familiarisation trials. For Experiment 1 proportions of looking time for each two trials were averaged (e.g. trial 1 and 2, trial 3 and 4, etc.), thus enabling a direct comparison to Experiment 2 and 3. As shown in Figure 3.36 habituation rate was lower in Experiment 1 as compared to Experiment 2 and 3. This was confirmed by a growth curve analysis in which looking time was modelled with fixed effects of Trial (1-12) and Experiment (1, 2 and 3) and random effect of Participant on the intercept and slope. There were significant effects of Trial and Experiment (Experiment 1 was used as a baseline, $\chi^2(8) = 17.96$, $p < .001$, for model parameters see Table 3.7).

This result is in line with infants' performance observed in the test blocks. The lowest habituation rate is present in Experiment 1, which is also when

little evidence of category formation is found. Faster habituation rates are present in Experiment 2 and 3, also conditions in which infants do succeed in forming categories.

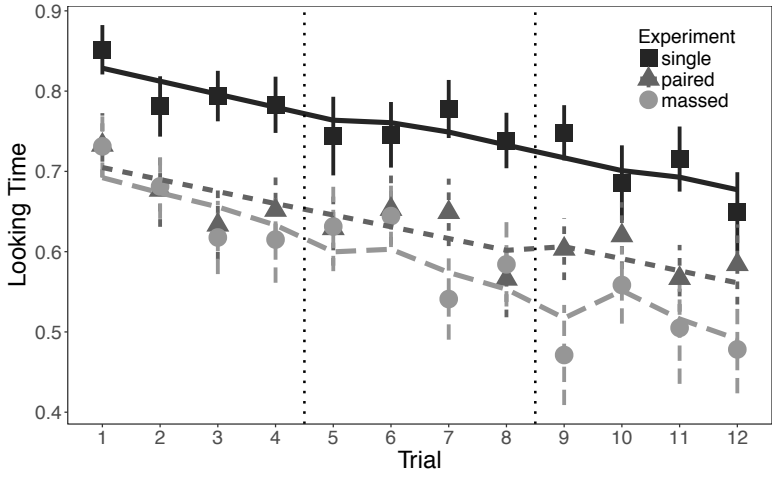


Figure 3.36. Mean looking time in three experiments (squares, triangles and circles depict mean looking time in single, paired and massed experiment, respectively; solid, dashed and dotted lines represent model values for each experiment; vertical dotted line represents time points when test blocks were presented).

Table 3.7. Habituation in Experiment 1, 2 and 3: Growth Curve Analysis (parameter estimates).

	Estimate	Std. Error	df	t value	p
Intercept	0.85	0.03	116.52	31.56	**<.001**
Trial	-0.02	0.00	73.47	-6.57	**<.001**
Experiment 2	-0.12	0.04	82.50	-3.31	**<.001**
Experiment 3	-0.15	0.04	86.30	-4.16	**<.001**

3.7 Chapter 3: Discussion

Experiments presented in this chapter investigated the role of perceptual load and memory demands in infants' category learning by varying the number of items presented simultaneously during familiarisation. In Experiment 1, infants were familiarised with one item at a time, in Experiment 2 items were presented in pairs, whereas in Experiment 3 infants were presented with four items simultaneously. The amount of familiarisation time was kept constant across the three experiments. Also, the test blocks were identical across the three experiments, which enabled me to compare performance across the three experiments.

Category learning is expressed by a novelty (or familiarity) preference in test blocks. Infants expressed no looking preferences when presented with one item at a time, even after three familiarisation blocks, suggesting that they failed to learn categories. However, when the same items were presented in pairs, infants showed evidence of category learning already after one block of familiarisation. When presented with four items simultaneously, infants succeeded but it took them three blocks of familiarisation to succeed.

Taken together, experiments presented in this chapter highlight how characteristics of a learning situation shape infants' ability to learn categories. Changing from single-item to paired presentation increased the perceptual load and attention demands, but at the same time decreased memory demands. Having two exemplars simultaneously present enables infants to directly compare them on the same representational level, whereas single-item

presentation demands a comparison of an available item with a retained presentation of a previously presented exemplar. This decrease in memory demand accompanied with an increase in perceptual load seems to facilitate infants' extraction of category relevant information. Further increase in perceptual complexity by presenting four exemplars simultaneously however slowed down the learning process, suggesting that there is an optimal perceptual complexity for learning, which in the case of the results presented in this chapter is a paired presentation.

One way of viewing these results is the perceptual load hypothesis (Plunkett, 2010). Plunkett (2010) used this hypothesis to explain cross-modal effects of labels on categorisation, the perceptual load hypothesis proposes that the outcome of computations underlying category formation is dependent on the load that individual stimuli place on the perceptual system (Plunkett, 2010). The perceptual load hypothesis further assumes that category formation is sensitive to the perceptual load according to the inverted U-shaped function, where too little and too much load interfere with, whereas an optimal load enhances category formation.[4] (Lavie, 1995).

Here, we propose that the perceptual load hypothesis can be a useful tool for interpreting the results obtained in Experiment 1, 2 and 3. The current version of the hypothesis can account for the results after one block of familiarisation: paired presentation presents an optimal perceptual load,

[4]This concept is related to the theory of perceptual load, influential but highly debated theory that proposes that the level of perceptual and cognitive load determine success or failure of selective attention, i.e. efficiency of distractor rejection. This issue is further dicsussed in the final chapter of this thesis

whereas single and presentation of four items simultaneously fall outside of this optimal load region and no learning is seen (Figure 3.37). However, introducing additional familiarisation blocks revealed that infants are able to learn categories when familiarised with four items, but need more training to achieve this result. We propose an extension to the perceptual load hypothesis to account for this effect, by introducing another axis to the model. This dimension refers to the amount of familiarisation, i.e. the amount of experience with items. One of the questions future studies should address is what would happen if three items are presented during familiarisation. This would further determine whether presenting items in pairs is an optimal context or whether three is still within the optimal load range for infants to form categories.

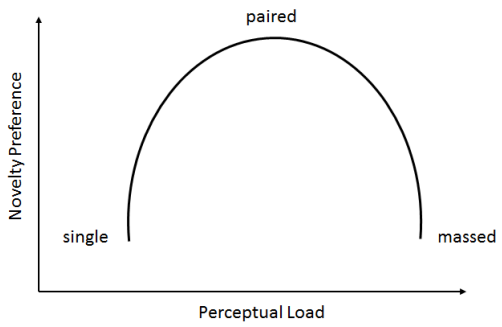


Figure 3.37. Perceptual load hypothesis: Novelty preference as a function of perceptual load.

What makes paired familiarisation an optimal setting for extracting category-relevant information? In addition to reducing memory demands and increasing perceptual load, presenting two exemplars along side each other during familiarisation invites infants to actively compare and contrast the two exemplars. Infants who adopted this strategy and made more comparisons also

performed better in test as compared to infants who tended to compare less, potentially reflecting individual differences in some more general abilities, such as the speed of processing (Kovack-Lesh et al., 2008).

Another important advantage of the paired familiarisation regime from a methodological point of view is the fact that in this experimental design familiarisation and test trials have the same visual format - two items presented on the screen. Namely, whereas in both sequential and simultaneous conditions the shift from learning to test was always accompanied with a change in the number of objects presented on the screen, there is no such shift in the paired condition. Thus, novelty preference, if expressed, cannot be ascribed to the change in configuration of presented objects, but due to the differences in characteristics of stimuli presented.

There are several limitations to the experiments presented in this chapter that need to be pointed out. One of the main limitations of the experimental design implemented in the experiments presented in this chapter is that category formation is probed using only two trials. Increasing the number of test trials would be beneficial in terms of obtaining a richer and potentially fine-grained insight into a structure of the formed categories. However, by presenting test items multiple times, they are not novel any more. Another related potential issue with the current studies is the interleaved design. While the majority of studies used a one-block design, the interleaved design was introduced in order to probe categorisation at different levels of familiarisation. However, a question is whether a design based on a novelty

preference can be used multiple times. The pattern of results observed in the three experiments presented in this chapter suggests that infant do express novelty preference even after being presented with novel items several times.

To further explore the relationship between perceptual load and learning, future studies could benefit from employing a variable trial duration driven by infants' looking behaviour. Instead of a fixed trial duration (used in the experiments presented here, as well as in a typical experimental design used to address category learning in infants), presenting items for as long as an infant is expressing an interest in it, would provide an additional information as to how infants habituation rates are dependent on the perceptual load.

The present chapter explored the role of perceptual complexity and memory load in infants' category learning. All experiments reported here focused on visual perceptual category learning. However, infants rarely encounter purely visual categories, but are rather faced with a multimodal learning situation in which exemplars of the to-be-learned categories are dynamic and often are named. In the next chapter we explore some aspects of the multimodal category learning by focusing on the role of motion and labels.

Chapter 4

Motion and labels in infants' category learning

This chapter investigates the role of dynamic visual features and labels in category learning. While trying to organise the incoming information about newly learnt objects into categories, infants rarely encounter silent static objects in isolation. Instead, they are immersed into a dynamic environment. Furthermore, the information coming into their visual stream whilst they are observing these objects is often accompanied by language or some other kind of auditory input. Therefore, in order to form categories, infants must encode, organise and bind information from different modalities simultaneously. As pointed out by Rakison (2004), sensitivity to the correlations between dynamic and static features is crucial for infants' developing knowledge about motion properties of a wide range of categories. Therefore, understanding the role of motion and labels, which are highly salient features, in the pro-

cess of category learning seems to be vitally important to uncovering the mechanisms underlying infants' category formation.

There is a large body of existing research that explores the effects of language on infants' category formation. There is substantial evidence to suggest that labels may profoundly shape categorisation (Althaus & Mareschal, 2014; Althaus & Plunkett, 2015; Plunkett, 2010; Plunkett et al., 2008), although the mechanisms underlying these effects are under debate. Some authors claim that the label is a feature (Gliozzi et al., 2009; Plunkett et al., 2008) while others propose that language has a special status and serves as a category marker (Waxman & Markow, 1995). Plunkett et al. (2008) explored the impact of labels on the formation of correlation-based categories and found that infants are successful in learning two novel categories when exemplars are presented in silence or if accompanied with two category-specific labels. However, when these very same exemplars (from both categories) are accompanied by only one label, infants tend to group all the items together into one category; and when there is no systematicity in the label-category mapping, infants form no categories. These results suggest that labels may have a strong impact on category formation. Several studies have suggested that labels achieve this by directing infants' attention to commonalities between items that share the same label (Althaus & Mareschal, 2014; Althaus & Plunkett, 2015).

In contrast, considerably fewer studies investigated the role of motion in category learning. This seems surprising considering that motion represents

a highly salient visual feature. Preference for moving over static objects is present already in newborns and very young infants (Dannemiller, 2000; Slater, 1989). However, when infants become able to integrate static and dynamic information is under debate. Some studies show that already at seven months of age, infants are able to integrate information about static and dynamic properties of objects (Pauen & Träuble, 2009) and map a motion to a specific object (Markson & Spelke, 2006), suggesting that very early on infants have the ability to combine static and dynamic visual information to learn about objects. In contrast, some studies found that integration of static and dynamic visual information is a difficult task for infants (Burnham et al., 1988; Rakison, 2004; Rakison & Poulin-Dubois, 2002). When examining infants' ability to memorise colour-shape compounds that were either static or dynamic, Burnham et al. (1988) discovered that infants aged seven to ten months were successful in remembering the static, but failed to remember the moving compounds. Similarly, Rakison & Poulin-Dubois (2002) explored infants' ability to detect correlations between dynamic and static features. They familiarised infants with two objects, each characterised by a distinct motion trajectory and found that 18, but not 10 month-old, infants are able to detect correlations among these features. In a related study, Rakison (2004) found that only 22 month-old infants were able to generalise the learned feature-motion relations to novel instances.

Nonetheless, a recent study by Deng & Sloutsky (2015) explored the role of motion in learning categories and found that motion facilitates category

learning. They familiarised 8 to 12 month-old infants with exemplars of a novel category. Exemplars were accompanied either with a label or a motion (moving feet). When tested for category formation, only infants exposed to motion demonstrated evidence of learning the category whereas those exposed to labels failed in doing so (Deng & Sloutsky, 2015).

Taken together, the existing research offers conflicting hypotheses about the role of motion in category learning. On the one hand, there is evidence that motion is a highly salient feature from early on in development (Dannemiller, 2000; Slater, 1989) and that it supports category formation (Deng & Sloutsky, 2015). On the other hand, a separate set of studies suggests that encoding correlations among static and dynamic visual features is a difficult task that infants fail to succeed in until later in their development (Burnham et al., 1988; Rakison, 2004; Rakison & Poulin-Dubois, 2002). The observed difficulties in binding static and dynamic features of objects might be due to difficulties in integrating information processed in different parts of the visual stream. Occlusion studies have suggested that infants fail to integrate information processed separately in the dorsal and ventral cortical pathway (Mareschal & Johnson, 2003; Mareschal et al., 1999).

The aim of the experiments presented in this chapter was to reconcile these diverging results and determine the role of dynamic features in category learning. Another aim was to compare the impact of motion and labels on category formation. If motion facilitates category formation, we expected that infants who learn moving categories will perform better than infants

exposed to labelled categories. However, if infants have difficulty encoding correlations among static and dynamic features, we expected that motion might inhibit the learning process. In the light of findings presented in the previous chapter, it seems that adding more information and increasing complexity beyond an optimal perceptual load might actually have detrimental effects on learning. Therefore, it is possible that both types of information will lead to suboptimal performance and hinder learning. Together with the findings of Experiment 3, the present set of experiments aims to provide a better understanding of how increasing processing load by adding different type of information, such as label and motion, affects category learning.

4.1 Overview of the studies

To explore the role of motion and labels in category formation, four experiments were conducted. Experiments 4 and 5 explore the role of motion in category learning. Experiment 4 tests the effects of motion associated with one category; meanwhile, in Experiment 5 both categories have a category-specific motion associated with them. The effects of labels are investigated in Experiments 6 and 7. In Experiment 6 only exemplars from one category are labelled, and the other category remains unlabelled. Experiment 7 explores how having two distinct labels for two categories affects the process of category formation. All four experiments employ a standard familiarisation-novelty preference procedure. After being familiarised with a set of training items, infants' category formation was measured as a preference for one of

the two items presented at test. Again, as in the experiments presented in the previous chapter, three familiarisation blocks were interleaved with test blocks, enabling a direct comparison of infants' performance across experiments.

One important novelty the present set of experiments introduced is a gaze-contingent familiarisation. Instead of a pre-determined schedule presentation of labels or motion, in the present experiments items would only start moving or be labelled once fixated. A gaze-contingent familiarisation provides a more realistic scenario of learning - infants have a more active role and design their own learning experience. It has been shown that infants quickly discover that they are in control in the gaze-contingent learning setting and that this has positive effects on their performance (Wang et al., 2012). In addition to positive effects on infants' engagement in the task, this experimental design has a potential to provide more insight into individual differences in exploration strategies infants develop. However, the use of gaze-contingent design in this set of experiments means that direct comparisons with results of experiments presented in the previous chapter are not going to be straightforward. An outline of the experiments presented in this chapter is given in Figure 4.1.

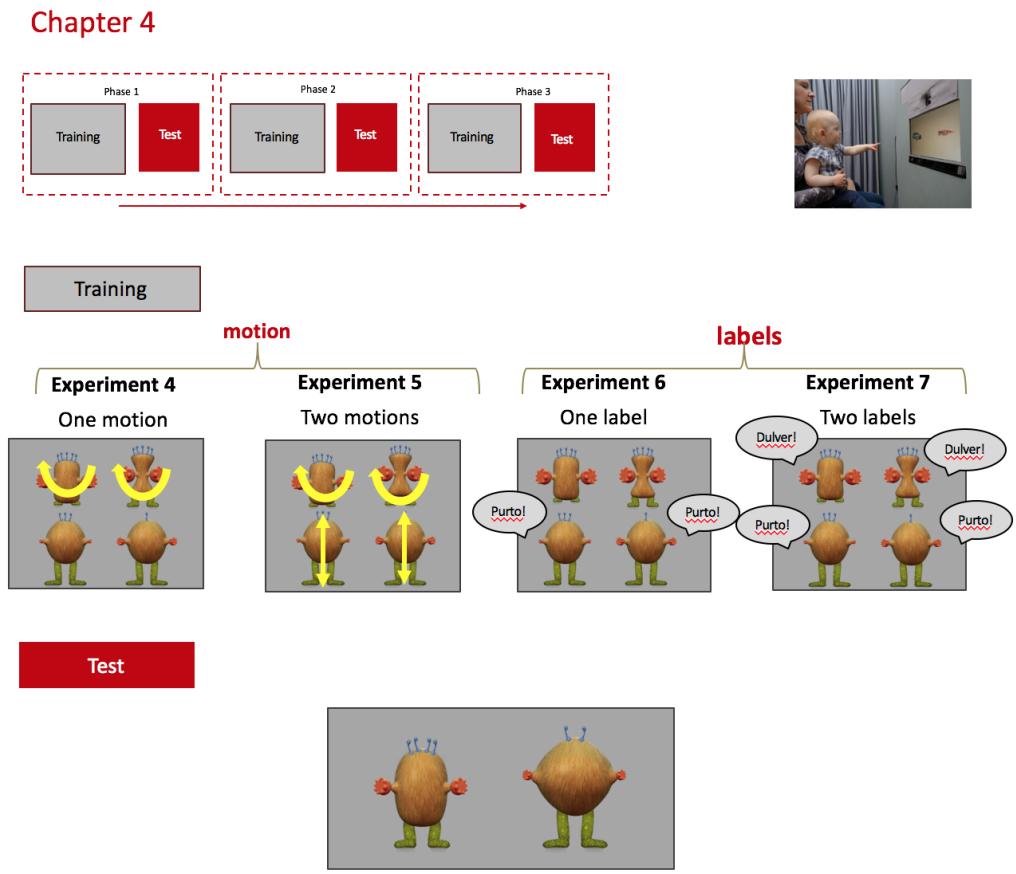


Figure 4.1. Outline of the chapter: The experiments presented use an identical experimental design with interleaved training and test trials (grey boxes refer to training blocks, red boxes depict test blocks). Experiment 4 and 5 explore the role of motion in category learning, whereas Experiment 6 and 7 investigate the role of labels in categorisation. All experiments used a gaze-contingent familiarisation paradigm and an identical test procedure based on novelty preference (illustrated in the bottom part of the figure).

4.2 Methods

To be able to compare the performance of infants learning categories accompanied by motion and labels, four experiments presented in this chapter have the same method which is described below, with the exception of some aspects of the familiarisation phase which are separately described for each experiment.

4.2.1 Stimuli

The same stimulus set as in Experiment 1, 2 and 3 was used in all experiments presented in this chapter (for a detailed description of the stimulus set see Methods Section of Chapter 3). A total of 24 Sukis was presented during familiarisation. In addition, prototypes of the two categories, an inter-category item (i.e. containing average values on all features) and a novel out-of-category item were presented at test.

4.2.2 Experimental design

The experimental setup was identical to that described in Chapter 2. Experiments consisted of three phases, each comprising of a training and test block (see Figure 4.2).

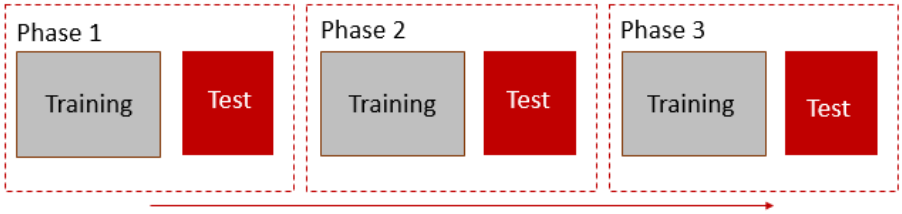


Figure 4.2. Schematic depiction of the experimental design: three familiarisation blocks interleaved with three test blocks.

Training

Each training block consisted of four familiarisation trials. Each trial was initiated with an animated star in the central location of the screen. Once infants fixated the star, four training items were simultaneously presented for 12,000 ms. Keeping total familiarisation duration identical across all four experiments made results comparable across experiments as well as to results of the experiments presented in the previous chapter.

Familiarisation trials were gaze-contingent. Depending on the experiment, infants were presented with an additional feature when landing a fixation on an item. Once fixated, the item would either become dynamic (perform a particular motion) or it would be labelled. In the motion condition (Experiments 4 and 5), exemplars from the same category would start moving after being fixated. Two category-specific motions were used: familiarisation items could either rock (angle change of 2 degrees every 16ms, ranging from -10 to + 10 by 2 steps of degrees) or jump (vertical displacement ranging from -25 to +25 with steps of 5 degrees every 16ms - refresh rate). In the label condition (Experiment 6 and 7), exemplars were immediately labelled on fixation. Two novel labels "purto" and "dulver" were used

(these labels have previously been used in category learning experiments in the Oxford BabyLab). In order to keep other factors as similar as possible, the motion and labelling effects had the same duration (approximately 400 ms). Label-category association was balanced. One category was associated with the label "purto" for half of the participants, and with "dulver" for the other half of participants". The same was the case for the motion condition, where half of the participants saw one category jumping and the other half saw the same category rocking.

Presentation sequences were identical to those of Experiment 3. These sequences were carefully selected to minimise any potential confounding effects of presentation order (for a detailed explanation on how sequences were selected, see the Methods Section of Chapter 3). The positioning of the categories on the screen was controlled, i.e. exemplars of both categories were presented in the various location across trials; thus, location could not be used as a cue for category membership. Keeping presentation sequences identical across experiments was introduced in order to enable the closest possible comparison across different experiments, as well as to provide an opportunity to use the findings of Experiment 3 as a baseline performance against which the results of Experiments 4-7 can be compared.

Test

Each test block consisted of four trials. Two trials were category formation test trials in which prototypes of two categories (items 2222 and 6666) were presented paired with an inter-category item (item 4444, see Figure 4.3

on page 128). Unlike Experiments 1-3, where the same pair was presented in two category formation trials, the two category formation test trials contain both prototypes, one in each trial. We reasoned that it is important to test for category formation for both categories as, in Experiments 4 and 6, it was expected that there would be differences in speed of learning of animated/labelled and silent categories. In the third test trial, the two category prototypes were presented side by side to test whether infants express a preference for either of the two categories. The order of the three trials was balanced using a Latin square. The final test trial was always a novelty preference trial. A novel, out-of-category item was presented together with a previously presented training item to validate whether infants' behaviour had indeed been driven by novelty preference. Each trial was initiated with a presentation of an animated star for 2,000 ms, followed by a pair of test items that remained on screen for 6,000 ms.

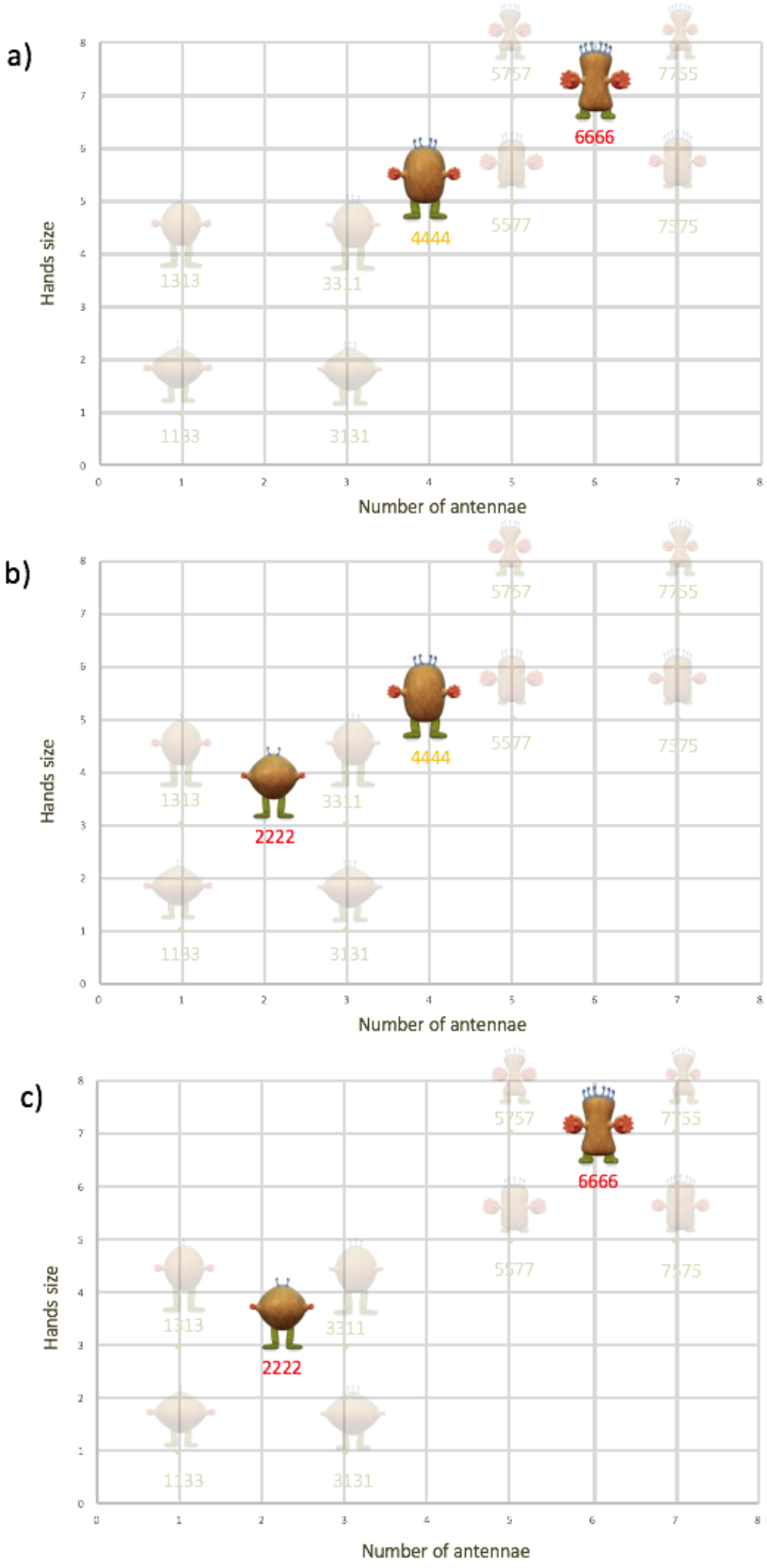


Figure 4.3. Schematic depiction of the test trials: category formation (a and b) and category preference test trial (c), (novelty preference test not depicted).

4.2.3 Participants

A total of 109 10-month-old infants took part in these experiments. All participants were full-term babies with no known health conditions, who had been recruited at the local maternity ward. Infants who refused to look at the screen were excluded from analysis, as were those whose total looking time in a training block was more than three standard deviations below the mean looking time of all infants in the same experiment (a total of 16 infants). The study was approved by the University of Oxford Medical Sciences Ethics Committee.

4.2.4 Statistical analysis

Training

All statistical analyses were performed in R (R Core Team, 2013). The mean and total looking time accumulated in each training block was calculated for each participant. Infants whose accumulated looking time in each training block was more than three standard deviations below the mean were excluded from subsequent analysis. To explore whether there was a decrease in the amount of looking time over the course of training, a standard ANOVA with a within-subjects factor *Block* (1, 2 and 3) was conducted. In addition, to obtain information about the time course of potential habituation effects, a growth curve analysis was conducted using lme4 (Bates et al. 2015, version 1.1-12) and lmerTest package (Kuznetsova et al. 2017, version 2.0-33). Mean looking time was modelled using a linear growth model with a fixed effect

of Trial and random effect of *Participant* on the intercept and slope terms. The fixed effect of *Trial* was added to the base model and the effect was evaluated using model comparisons (Mirman, 2014).

To characterise infants exploration strategies during training, mean shifting rates and transitional probabilities based on Markov chains were calculated in the same way as for Experiment 3 (for more details see the Method section of Experiment 3).

Test

Infants' category formation was probed in test trials and their looking preferences for one of the two presented objects used as an index of category formation. Each test block consisted of four types of trials. Two trials assessed category formation by presenting prototypes of one of the two categories alongside an inter-category item. If infants formed two categories, it is expected they will spend more time looking at the inter-category item; however, if they formed one category, it is expected that infants will find the category prototype more novel and thus look at it for longer. First, looking time accumulated in each trial was calculated for each infant. Infants whose looking time was less than 3 standard deviations below the mean were excluded from analysis of that trial. Following this, preference scores were calculated by dividing the time spent looking at the inter-category item by the total time spent looking at both objects. The third trial tested whether infants show a preference for one of the two categories by presenting prototypes of the two categories together. The order of these three trials was

counterbalanced. Finally, the fourth trial was a novelty preference trial in which one of the previously familiarised training items was presented in a pair with a novel, out-of-category item. The novelty preference test was always presented as the last trial in a block and served as a control to assess whether the infants' behaviour was indeed driven by a novelty preference.

For each test (category formation, category preference and novelty preference test) a separate repeated-measures ANOVA with factors Block (1, 2 and 3) was performed to determine whether there had been changes in the preferences over the course of the three test blocks (Greenhouse-Geisser corrections for F-statistics were applied where sphericity assumptions were violated). Planned comparisons against chance were performed for each block to determine whether infants expressed preference for one of the test items in each test block (t-tests or non-parametric tests when data was not normally distributed).

Generalised eta-squared was reported as a measure of the effect size for repeated-measures ANOVAs as it is invariant across different research designs, thus allowing for comparison of effect sizes across different analyses (Olejnik & Algina, 2003) and Cohen's d for t-tests using lsr R package (Navarro 2015, version 0.5).

4.3 Experiment 4: Contrasting dynamic and static categories

4.3.1 Introduction

The aim of this experiment was to determine the role of visual dynamic features in infants' category learning. As mentioned in the introduction of this chapter, motion represents a highly salient visual feature and even newborns show preference for dynamic objects over static objects (e.g. Slater, 1989). However, the exact role of dynamic features in young infants' category learning is not well understood. While one study discovered that infants fail to encode correlations among static and dynamic visual features in a categorisation context until they are 22 months of age (Rakison, 2004), another study found that motion facilitates category learning already in 8-12 months-old infants (Deng & Sloutsky, 2015).

To address the question of whether dynamic features facilitate or hinder category learning, in this experiment we introduce a novel design in which one category has motion as an additional feature, while the other category remains static. Formation of both dynamic and static categories is tested after each of the three familiarisation blocks. If motion facilitates extraction of the category relevant information, we expect that the dynamic category will be formed faster than the static one. If, by contrast, motion hinders learning, we expect that the static category will be formed more efficiently than the dynamic one.

4.3.2 Method

Participants

Twenty-nine 10-month-old infants took part in this experiment (13 girls and 16 boys, $M_{age} = 10.12$, age range: $9.60 - 10.9$). Five infants were excluded due to insufficient looking time accumulated in the first training phase. In addition, one participant started crying after the first phase and therefore provided no data for the second and third phase.

Stimuli and experimental design

The stimulus set and experimental design used in this experiment were described earlier in the General Methods section of this chapter.

As in Experiment 3, each of the three training blocks consisted of four familiarisation trials. Four items were presented simultaneously in each trial. Each trial was initiated with an animated star (an attention getter) that remained on the screen until infants fixated it. Following this, four items were presented for a fixed duration of 12,000 ms. Exemplars from one category were animated, whereas exemplars from the other category remained static (Figure 4.4). Which category was animated was counterbalanced (for half of the participants one category was dynamic whereas for the other half of participants this category was static). For a detailed explanation of the animation and the design of test blocks, see Section 4.2.2 of this chapter.

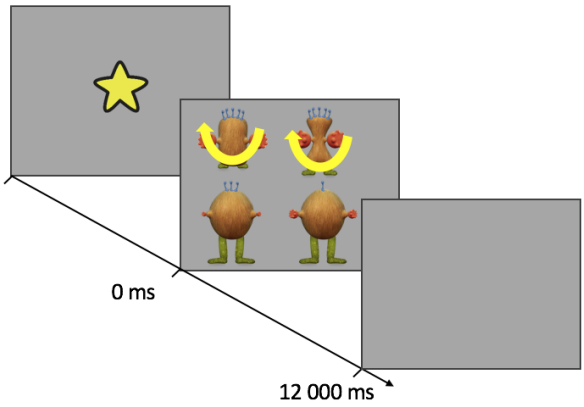


Figure 4.4. Experiment 4: Structure of familiarisation trials.

4.3.3 Results

Training

Figure 4.5 depicts the amount of looking time infants accumulated for the static and dynamic category in each training block. Infants dedicated more time to exploring exemplars from the dynamic category across all three training blocks. This was confirmed by a repeated-measures ANOVA with within-subject factors *Block* (1, 2, and 3) and *Category Type* (static and dynamic) that revealed a significant main effect of *Category Type* ($F(1, 21) = 98.30$, $p < .001$, $\eta_g^2 = .52$).

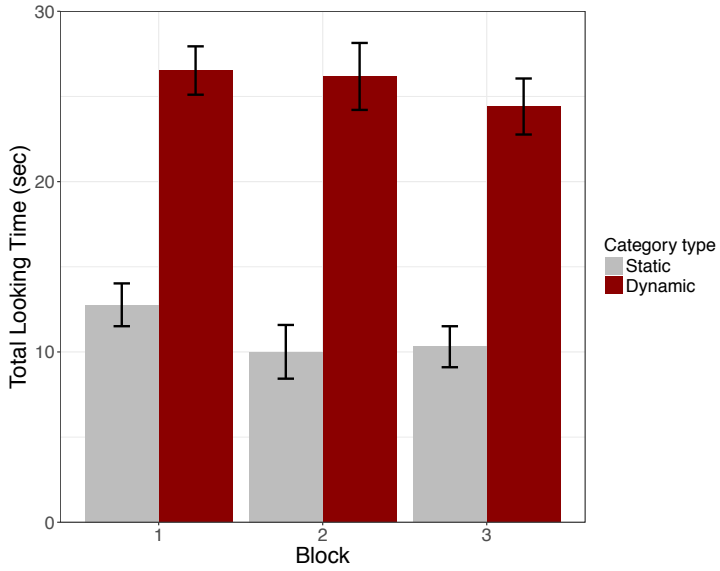


Figure 4.5. Experiment 4: The mean amount of looking time to dynamic and static category exemplars in each training block (error bars depict standard error).

To further explore looking patterns over the course of familiarisation trials, growth curve analysis was conducted. Mean looking time was modelled using a linear growth model with fixed effects of *Trial* (1 to 12) and *Category Type* (dynamic and static) and random effect of *Participant* on the intercept and slope terms. There was a decline in the mean looking time over the course of familiarisation (mean looking time decreased 91.87 ms per trial on average, see Figure 4.6 and Table 4.1). In addition, there was a significant effect of *Category Type* ($\chi^2(7) = 200.73$, $p < .001$). Infants spent on average 3618.39 ms more time looking at the dynamic category exemplars as compared to static category exemplars.

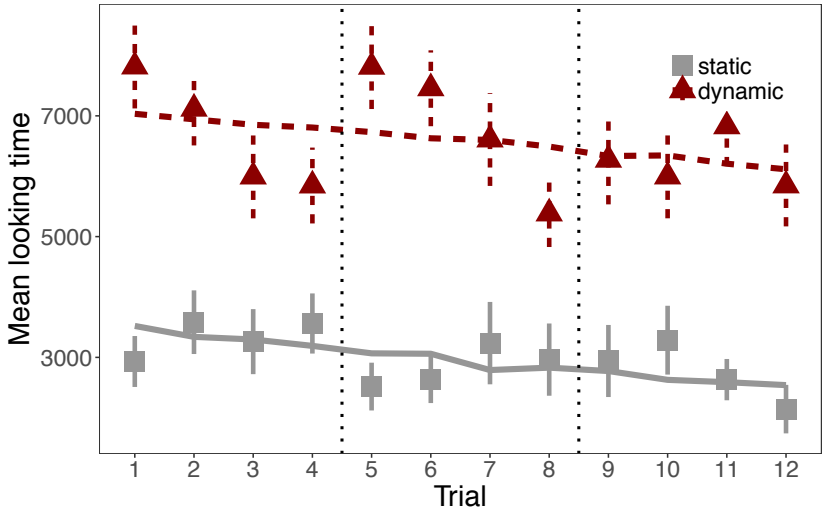


Figure 4.6. Experiment 4: Mean looking time for static and dynamic category exemplars over the course of familiarisation (squares and triangles depict mean looking times in each trial for static and dynamic category, respectively; solid and dashed line represents model values for static and dynamic category; error bars depict standard error).

Table 4.1. Experiment 4: Growth Curve Analysis. Parameter estimates.

	Estimate	Std. Error	df	t value	p value
Intercept	3571.72	339.63	37.86	10.52	**.00**
Trial	-91.87	33.53	529.81	-2.74	**.01**
Category Type	3618.39	230.90	948.87	15.67	**.00**

To explore whether there was a recovery in the looking time between blocks, the looking time in the last two trials of the first block was averaged and compared to the average of the looking time in the first two trials in the block 2. The same was done for the transition between the second and the third block. The averaged values were compared using a paired t-test. No significant effects were found.

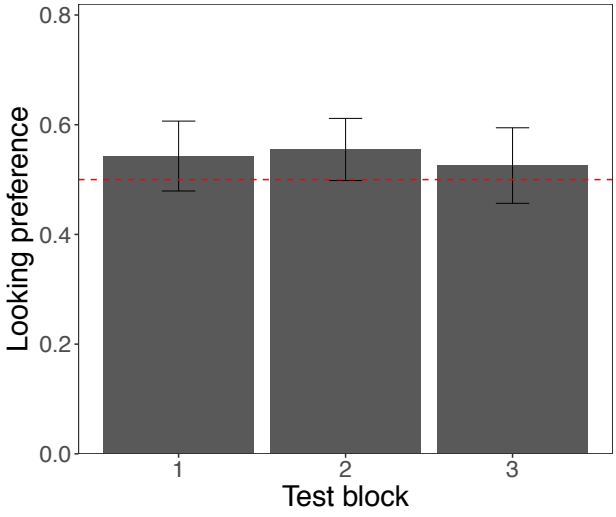
Taken together, these results indicate that there was a familiarisation effect expressed as a decrease in the amount of looking time over the course of the experiment. There was a difference in the amount of time infants

dedicated to exploring exemplars from the two categories as infants spent more time looking at the exemplars from the dynamic category.

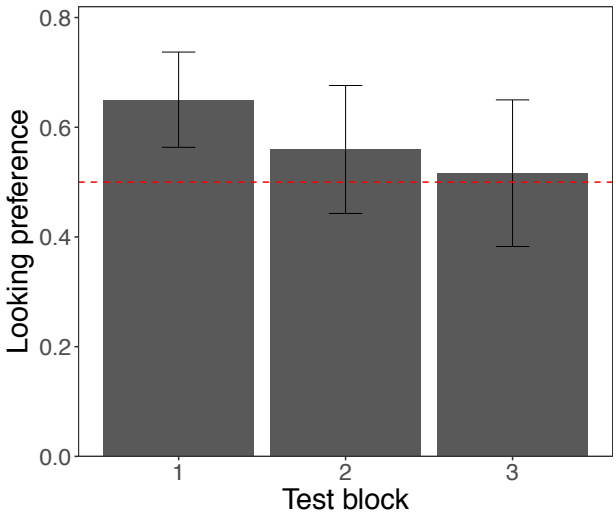
Test - Dynamic category formation

Figure 4.7a depicts mean looking preferences in the dynamic category formation test. A repeated-measures ANOVA with a factor *Block* (1, 2 and 3) on preference scores yielded no significant effects. Planned comparisons of the mean preference in each test block against chance also revealed no significant effects ($all\ p_s > 0.5$), therefore suggesting that infants failed to learn the dynamic category.

To control for the potential confounding effects of the trial presentation sequence, the order of test trials was counterbalanced: the dynamic test was presented as the first, second and third trial an equal number of times in each test block. To explore potential effects of the presentation order, an additional analysis was conducted on a subset of dynamic test trials presented first in the test block. As shown in Figure 4.7b, the overall pattern of looking preferences is similar as in the previous analysis. An ANOVA with *Block* as a between-subjects factors revealed no significant effects, but these results need to me taken with caution due to low number of participants per group.



(a)



(b)

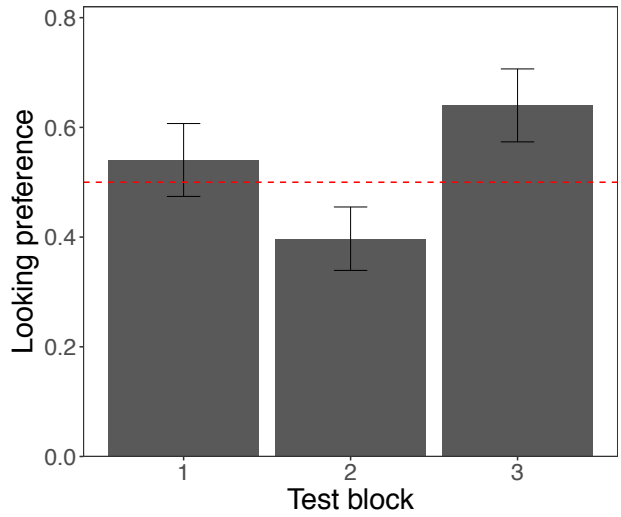
Figure 4.7. Experiment 4: Looking preferences in the dynamic category formation test (a), and when presented as the first test trial (b), (values higher than .5, depicted by a horizontal dashed red line, suggest a preference for the inter-category item, error bars depict standard error, error bars depict standard error).

Test - Static category formation

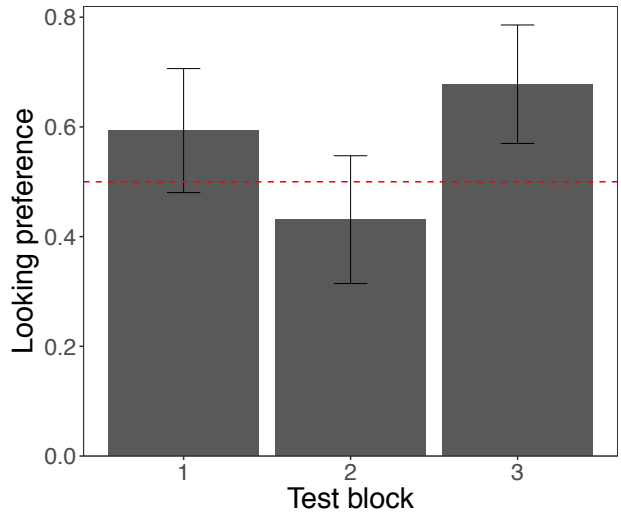
Preference scores were submitted to a repeated-measures ANOVA with a factor *Block* (1, 2 and 3). No main effect of *Block* was present, but planned comparisons of the mean preference in each test block against chance revealed a significant preference for the inter-category item in the final test block ($t(19) = 2.12$, $p = 0.049$, $d = .47$) and a trend towards preference for

the static category prototype in the second test block ($t(21) = -1.78, p = 0.089, d = .38$, Figure 4.8a). These results suggest that infants formed a static category after three training blocks.

An overall pattern of preferences in a subset of trials which were presented first in the test block followed the overall pattern of the results (see 4.8b). This pattern of results suggests there was familiarity to novelty transition over the course of the experiment.



(a)



(b)

Figure 4.8. Experiment 4: Looking preferences in the static category formation test (a) and when presented as the first test trial (b), (values higher than .5, depicted by a horizontal dashed red line, suggest a preference for the inter-category item error bars depict standard error).

Test - Between categories preference

To test whether there was a preference for one of the two categories, in this trial both the prototype of the dynamic and the prototype of the static category were presented along side each other. Preference scores were calculated by dividing the time spent looking at the static prototype by the total looking time for both the dynamic and static prototypes (see Figure 4.9a). A repeated-measures ANOVA with a factor *Block* (1, 2 and 3) revealed a main effect of *Block* ($F(2, 40) = 5.21,\ p < .01,\ \eta_g^2 = .15$). Planned comparisons showed there was a significant preference for the static category exemplar in the final test block ($t(22) = 2.13,\ p = .044,\ d = .44$). In addition, there was a trend towards the effect in the first test block ($t(22) = 1.76,\ p = .09,\ d = .36$) and an opposite trend towards the preference for the dynamic category in the second test block ($t(23) = -1.98,\ p = .059,\ d = .40$). These results suggest that infants treated the categories differently, showing more interest for the static category in the final test block.

Identical pattern of looking preferences was observed when the analysis was performed on a subset of trials that were presented first in the test blocks (see Figure 4.9b).

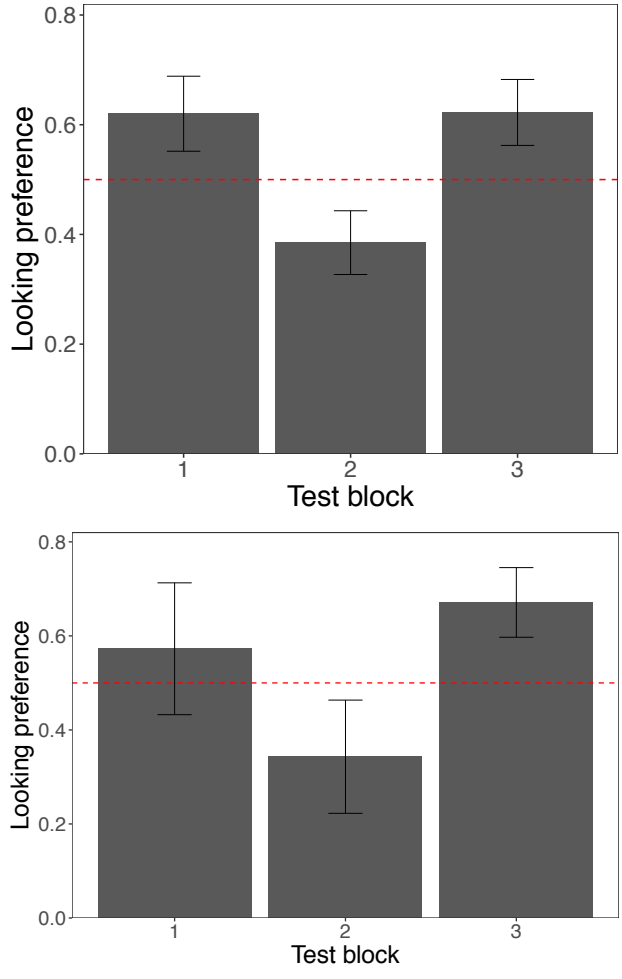


Figure 4.9. Experiment 4: Looking preferences in the between categories preference test (a) and when presented as the first trial (b), (values higher than .5, depicted by a horizontal dashed red line, suggest a preference for the static category prototype, error bars depict standard error).

Test - Novelty preference

To determine whether infants' looking preference was indeed driven by a novelty preference, the final trial of each test block was a novelty preference test. Infants' time spent looking at the novel item was divided by the total looking time accumulated in this trial. Mean preferences are depicted in Figure 4.10. A one-way ANOVA with a within-subjects factor *Block* (1, 2, and 3) revealed a significant main effect of block ($F(2, 38) = 3.73$, $p <$

.05, $\eta_g^2 = .11$). Planned comparisons of mean preference in each block was tested against chance (.5). A significant novelty preference was present in the first and the last test block (block 1: $t(22) = 2.59$, $p < .05$, $d = .53$; block 3: $t(21) = 2.52$, $p < .05$, $d = .54$).

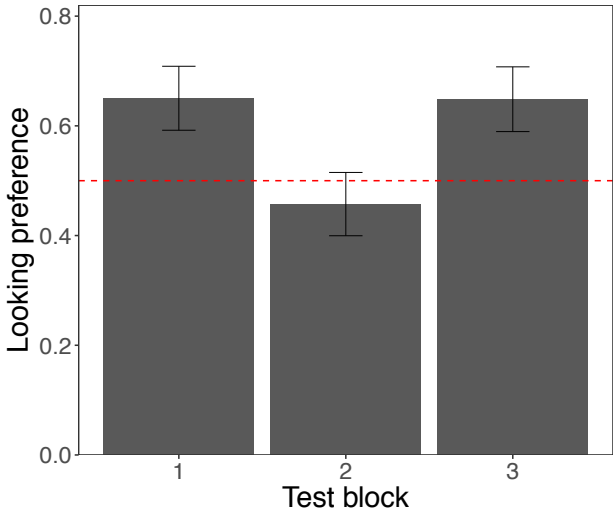


Figure 4.10. Experiment 4: Looking preferences in the novelty preference test (values higher than .5, depicted by a horizontal dashed red line, suggest a preference for the inter-category item, error bars depict standard error).

To summarise, the results suggest that, despite accumulating more looking towards the dynamic category, infants failed to learn it, while it took them three blocks of training to learn the static category.

The between-categories preference test showed that over the course of the test blocks, infants expressed preference for both categories at different points of the experiment, suggesting that they treated the categories differently and detected some of the defining characteristics, but failed to discover a more robust boundary between the categories.

Exploration strategies during training

Shifting

Presenting four items simultaneously during training provides infants with an opportunity to develop various exploration strategies. In addition, gaze-contingent familiarisation creates a more active learning context where effects are contingent upon where an infant looks (e.g. an item will move only once it's fixated). To investigate characteristics of exploration strategies infants engage, we first explored mean shifting rates, i.e. the mean number of shifts infants performed in each block. Three mean rates were calculated: for between-category shifts, for shifts within the dynamic category and for shifts made within the static category (see Figure 4.11a).

Each trial consisted of two exemplars from each of the two categories; therefore, between-category shifts were by chance twice more likely than within-category shifts. To determine whether either of these shift types were more frequent, proportion of the between-category shifts was calculated for each trial by dividing the number of the between-categories shifts by the total number of shifts in that trial. The same was done for the within-category shifts (see Figure 4.11b). A repeated-measures ANOVA with a within-subject factor *Block* (1, 2 and 3) was performed on mean between-category shifting rates (total of 23 participants). As there was no significant main effect of *Block*, the means were averaged for the three test blocks. The mean between-category shifting proportion was compared against a chance level (.66) and the analysis revealed that infants made significantly fewer between-category

shifts than expected by chance ($t(23) = 13.66,\ p < .001,\ d = 2.79$). Within-category shifts were more frequent for the dynamic category than for the static category as revealed by a paired t-test ($t(23) = 4.86,\ p < .001,\ d = .99$, Figure 4.11c).

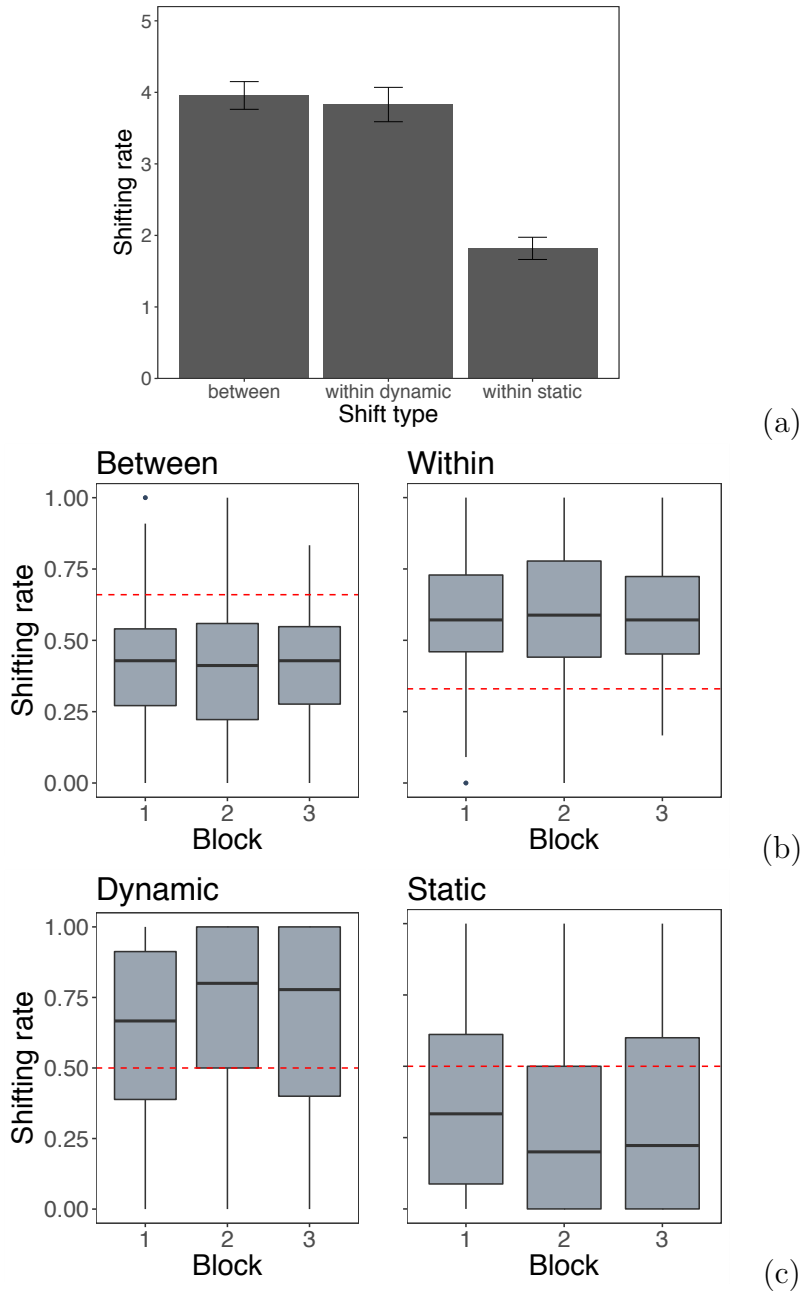


Figure 4.11. Experiment 4: Mean shifting rates (a); Proportions of the between-category and within-category shifts (b); proportion of shifts within the dynamic category and within the static category (c), (dotted line depicts shifting rates at chance).

Transitional probabilities

To investigate infants' exploration strategies in more detail, transitional probabilities of different shifts were calculated using Markov chains. As outlined in Chapter 3, this enables us to obtain a fine grained measure of exploration patterns that provided additional information about the nature of shifts infants made during familiarisation. We implemented an identical analysis to that performed in Experiment 3. The Markov chain was implemented using the markovchain package (Spedicato 2017, version 0.6.9.8-1) in R.

The model was defined with a space of two potential states, each state representing one of the two categories items can come from (dynamic or static). Four transition types were possible: two within-category transitions (from one dynamic exemplar to another, from one static exemplar to another) and two between-categories transitions (dynamic to static exemplar and static to dynamic exemplar). Probabilities for each of the four transition types were calculated for each infant.

In Experiment 3, which had an identical design except that no categories were animated (non-gaze-contingent paradigm), TPs for within and between shifts were averaged. However, since in the present experiment two categories were not identical, i.e. one had an additional feature of motion, in this case no averaging of TPs was done, as they carry unique information about exploration behaviour.

Mean transitional probabilities for the four types of transition are depicted in Figure 4.12. When making transitions from a dynamic category

exemplar, infants were more likely to shift to another dynamic exemplar ($t(23) = 10.59,\ p < .001,\ d = 2.16$, see Figure 4.12a) than to static category exemplars ($t(23) = -10.28,\ p < .001,\ d = 2.08$). When fixating static category exemplar, both shifts to a dynamic and to a static exemplars were equally likely (*all* $p_s > .05$, see Figure 4.12b).

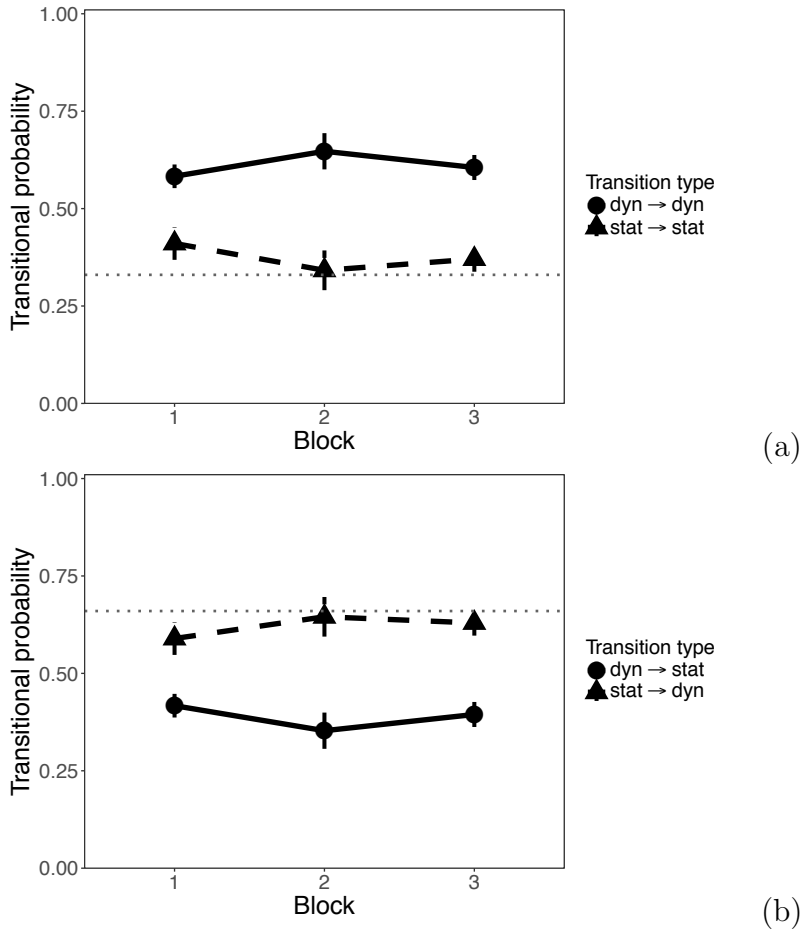


Figure 4.12. Experiment 4: Transitional probabilities in each training block for within-category shifts (a) and for the between-category shifts (b), (dotted line depicts performance at chance, error bars depict standard error).

The relationship between exploration strategies during training and performance in test

The set of analyses presented in this section explored whether different exploration strategies during training lead to different performance at test. Two main types of strategies infants engaged were the similarity-focused strategy, expressed as a higher probability of making transitions within categories, and the contrast-focused strategy, defined by a higher transitional probability of making shifts between categories. The analysis focused on the performance in the final test block, which is the block where some evidence of category learning was found.

The mean preference scores at test trials and mean transitional probability were calculated for each participant. No correlation between the overall preference scores in test and the transitional probabilities was found. In the next step, we compared the time course of looking preferences of infants who engaged different exploration strategies. First, eye-tracking data was downsampled into 200 ms bins by aggregating 24 raw samples (data was recorded with a 120 Hz sampling frequency, i.e. raw bins were 8.33 ms). For each bin, the proportion of time the inter-category item was fixated was calculated by dividing the number of fixations to the inter-category item by the total looking time to both items. Then, proportions were elog transformed (corrected logit transformation, in order to account for 0 and 1 values, as recommended by Dink & Ferguson 2015).

Based on the transitional probability scores in training, infants were di-

vided into two groups based on a median split (Mdn = .60) into a similarity-focused group (higher probabilities of shifting within the dynamic category, M = .69, SD = .05) and a contrast-focused group (higher probabilities of shifting from the dynamic to the static category, M = .50, SD = .10). In the next step, looking patterns of the similarity- and contrast-focused groups were compared for each test type, i.e. for the dynamic category formation test, the static category formation test, between categories preference test and the novelty preference test. There was no difference in performance of the similarity- and contrast-focused groups in any of the four test trials (see Figure 4.13).

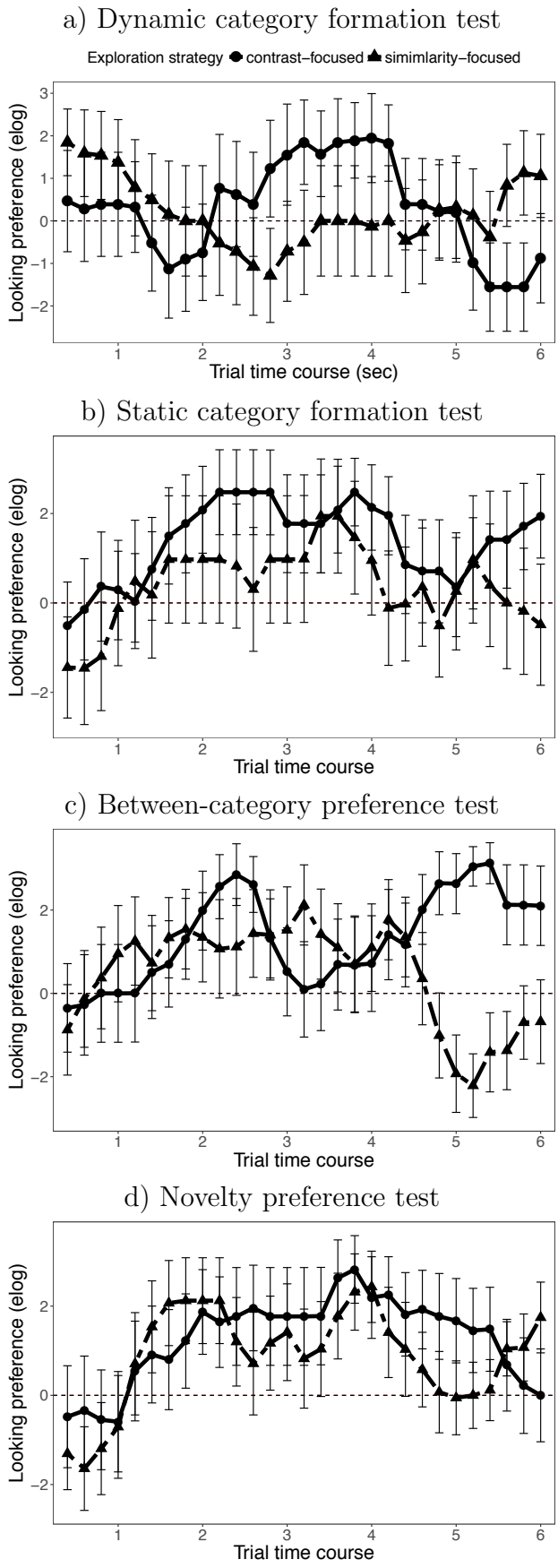


Figure 4.13. Experiment 4: Performance of infants engaged in similarity-focused and contrast-focused exploration strategies (error bars depict standard error).

4.3.4 Discussion

This experiment explored the role of motion in infants' category learning. It focused on infants' ability to form correlation-based novel categories where one category had an additional dynamic feature, while the other category remained static. During familiarisation infants were presented with four exemplars at once. Exemplars of one category would start moving once an infant landed a fixation on them, whereas exemplars from the other category had no additional dynamic feature associated with them. There were three familiarisation blocks and category formation was probed after each. The results revealed that infants show a preference for the inter-category item in the final static category formation test, thus providing some evidence that the static category was learned after three blocks of familiarisation. This is in line with the results of Experiment 3, showing that this amount of familiarisation experience is needed to learn the same categories when there is no dynamic features associated with the categories. Despite longer looking times for dynamic category exemplars in in all familiarisation blocks, infants failed to form the dynamic category.

There are several possible explanations for the observed result. One possibility is that infants have difficulties coding correlations between static and dynamic information, as proposed by some studies (Burnham et al., 1988; Rakison, 2004), and hence, infants succeed in forming the static category but fail to form the dynamic category.

Another possibility is that infants are able to encode correlations among

static and dynamic features, but this is not evident in a category formation test as there are two competing mechanisms driving infants attention. One mechanism is a novelty preference, which predicts preference for the inter-category item if infants formed categories. At the same time, there might be a violation of expectations driving infants' looking. Since infants learnt that the dynamic category moves and both items in test trials are static, they might look longer at the dynamic category prototype because it violated their expectation that it should start moving once fixated. In order to test for this hypothesis, it is necessary to introduce a dynamic test with both test items being dynamic. In case violation of expectations was the dominant drive behind infants' looking preferences in the current experiment, thus masking category learning, having a test where both prototypes are dynamic would rule out this possibility.

Another possibility is that infants struggle to form categories because the two categories are defined by a different number of features. Whereas one category is defined by four visual features, the second category is also defined by an additional dynamic feature. Most studies so far have focused on categories that have the same number of features. Therefore, it is possible that the lack of symmetry in structures of categories is posing difficulties for defining of the categories, and not only a difficulty in encoding static-dynamic correlations.

4.4 Experiment 5: Dynamic categories

4.4.1 Introduction

The previous experiment revealed that infants have difficulties forming categories when one of the categories has an additional dynamic feature, suggesting that infants fail to encode correlations among static and dynamic features. However, it is also possible that asymmetry in category structure caused the observed effects: one category had five features (four visual features and a dynamic feature), whereas the other category had four (only four visual features).

The aim of the present experiment was to explore infants' ability to distinguish two categories in a learning context in which both categories have category-specific dynamic features associated to them. Using an identical design to that of the previous experiment, infants were familiarised with four items simultaneously, with an exception that in the present experiment exemplars from both categories would move in a category-specific way once fixated.

Another issue we addressed in this experiment was whether static test trials were appropriate for probing the formation of dynamic categories. Experiment 4 revealed that infants have difficulty forming dynamic categories. However, an alternative possibility is that infants formed the categories, but that no evidence of learning was found due to using static test trials. It could be the case that motion is a highly salient feature of the category representation and omitting this feature, i.e. presenting a static test, prevented them

from expressing evidence of category learning. To test for this possibility, in addition to the category formation tests used in previous experiments, at the end of the final test infants' category formation was tested with dynamic test trials in which both test items moved.

4.4.2 Method

Participants

Twenty-seven 10-month-old infants took part in this experiment (13 girls and 14 boys, $M_{age} = 10.37$, age range = 9.49 - 11.07). Two infants were excluded due to insufficient looking time accumulated during training. One additional participant was excluded from the analysis of the final block as their session was ended at the beginning of the third block due to fussiness.

Stimuli and experimental design

The stimulus set and experimental procedure used in this experiment were identical to that of Experiment 4, except that both categories were animated during the familiarisation phase. As presented in Figure 4.14, exemplars of one category would rock and exemplars from the other category would jump while the infant was looking at them (for an explanation of the animations, see Section 4.2.2).

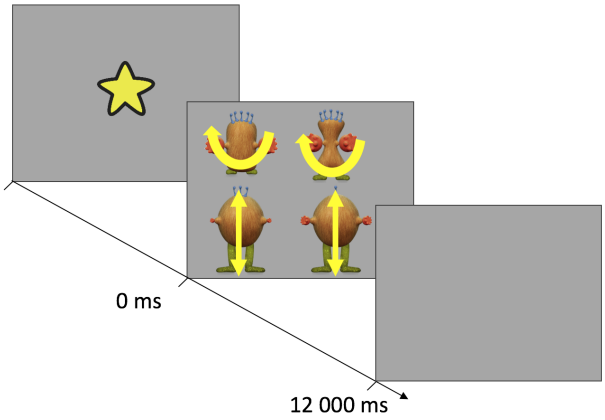


Figure 4.14. Experiment 4: Structure of familiarisation trials.

After each training block, infants were presented with a test block (also described in Section 4.2.2). An additional test block was presented at the very end of the experiment. In this block, category formation was tested with two trials in which the two prototypes were presented side-by-side. Unlike in previous test blocks, both prototypes were animated from the onset of the trial. In one trial, items were animated with a movement specific for one of the categories, while in the other trial they moved in a fashion associated with the other category. Hence, in both trials one item had a category-congruent motion and the other item moved in a way incongruent to its category. Each trial started with an animated star, which remained on screen until infants looked at it; then it was replaced by a presentation of the two prototypes for 6,000 ms. The position of the two prototypes and movement type were balanced across participants.

4.4.3 Results

Training

To explore whether there was a decrease in the amount of looking time in the three familiarisation blocks a repeated-measures ANOVA with factors *Block* (1, 2 and 3) and *Category* (1 and 2) on the looking time was conducted. There were no main effects of *Block* nor *Category*, but there was a significant interaction of *Block* and *Category* ($F(2, 46) = 7.47$, $p < .05$, $\eta_g^2 = .1$). Post-hoc test revealed that there was more looking time for one of the categories only in the first training block ($t(24) = 2.99$, $p < .01$, $d = .97$), whereas no difference was present in the later two training blocks (Figure 4.15). The association of the motion type (jumping or rocking) and category was counterbalanced (one category was jumping for half of the participant and rocking for the other half). It is unclear why infants expressed a preference for one category during the first familiarisation block.

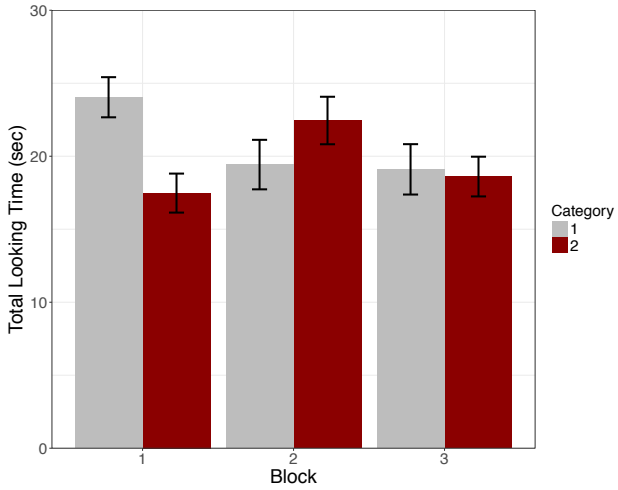


Figure 4.15. Experiment 5: Mean looking time in each training block (category 1 was a set of stimuli defined by low values on feature range, whereas category 2 was defined by high values on the feature range, error bars depict standard error).

The growth curve analysis revealed that there was a near significant improvement in model fit when Trial was added to the base model, suggesting that a weak habituation effect was present during familiarisation, though the model fit was poor (see table 4.2 and Figure 4.16).

Table 4.2. Experiment 5: Growth Curve Analysis. Parameter estimates

	Estimate	Std. Error	df	t value	p value
Intercept	6013.24	331.72	105.87	18.13	**.00**
Trial	-79.76	40.13	108.16	-1.99	**.05**
Category	-359.94	268.83	443.38	-1.34	.18

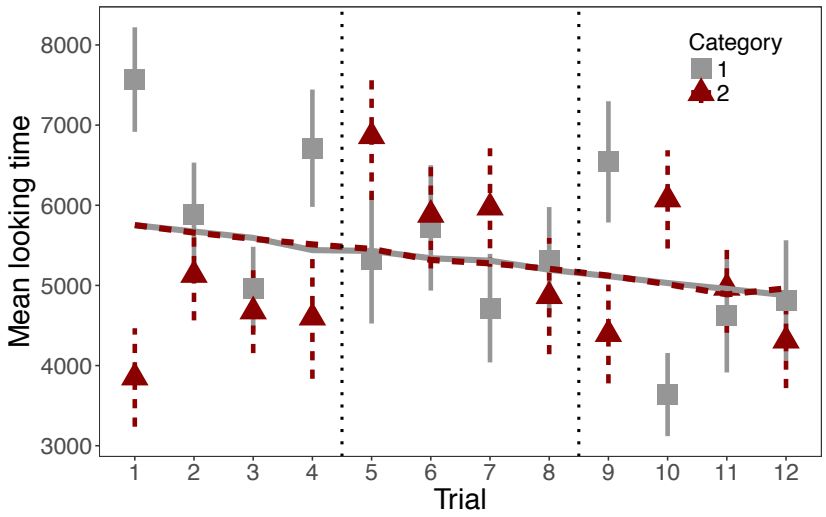


Figure 4.16. Experiment 5: Mean looking time for exemplars from the two categories over the course of familiarisation (squares and triangles depict the two categories, solid and dashed line represent model values, vertical dotted lines represent time points when test blocks were presented, error bars depict standard error).

Test - Category formation

Following each training block category formation was tested with two trials. Preference for the inter-category items was calculated by dividing looking time to the inter-category item by the total looking time, and averaged for the two trials. As shown in Figure 4.17, infants showed no preferences in any

of the three test blocks, hence suggesting that they failed to form the two categories (*all* $p_s > .05$).

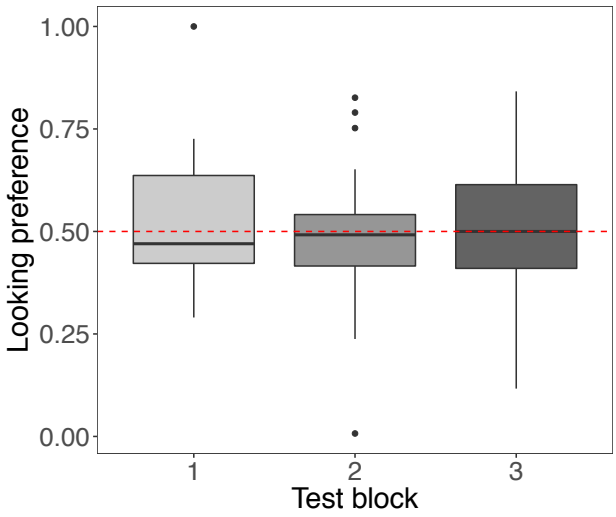


Figure 4.17. Experiment 5: Mean looking preferences in category formation test (higher preference value means that they prefer the inter-category item, whereas lower preference value indicates preference for the category prototype).

Test - Between categories preference

In one of the test trials infants were presented with the two prototypes to test if there was a preference for one of the two categories. No systematic preference was observed. This was expected, as both categories had dynamic visual features (*all* $p_s > .05$, Figure 4.18).

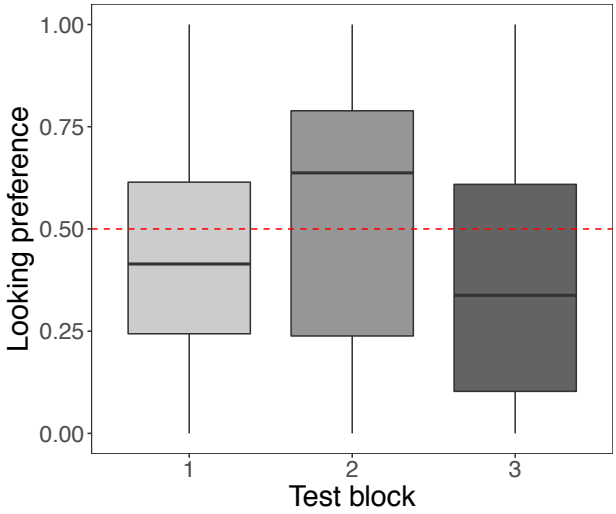


Figure 4.18. Experiment 5: Mean looking preferences in the between category formation test

Test - Novelty preference

To validate that infants' behaviour in test was driven by novelty preference, the final trial in each test block was a novelty preference trial. Infants expressed novelty preference only in the first test block (Wilcoxon test: $Mdn = .734,\ ,Z = -2.52,\ V = 257,\ p < .05,\ d = .51$, Figure 4.19).

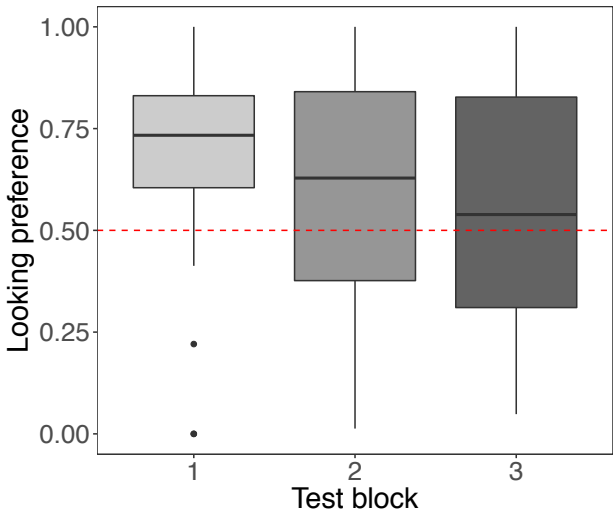


Figure 4.19. Experiment 5: Mean looking preferences in the novelty preference test (higher preference value higher than .5 indicates preference for the novel item; value of 0 indicates familiarity preference).

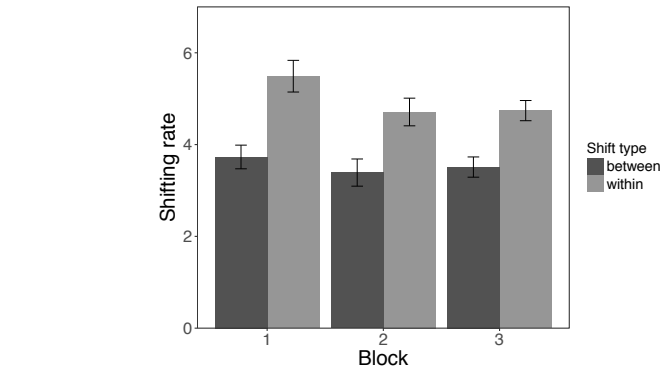
4.4.4 Infants' exploration strategies in training

Presenting multiple items simultaneously provides infants with an opportunity to develop their own exploration path of a visual display. At the same time, in order to explore one of the several items, infants must be able to choose which one to focus on while simultaneously ignoring the remaining items. In addition, they can decide how long to explore an item for, when to disengage from it and which item to explore next. To determine the characteristics of the exploration strategies infants engaged in the present experiment, we focus on mean shifting rates and transitional probabilities. As Experiment 3 revealed that these two parameters of exploration are not correlated, we present analysis of the two parameters and their relationship to test performance separately.

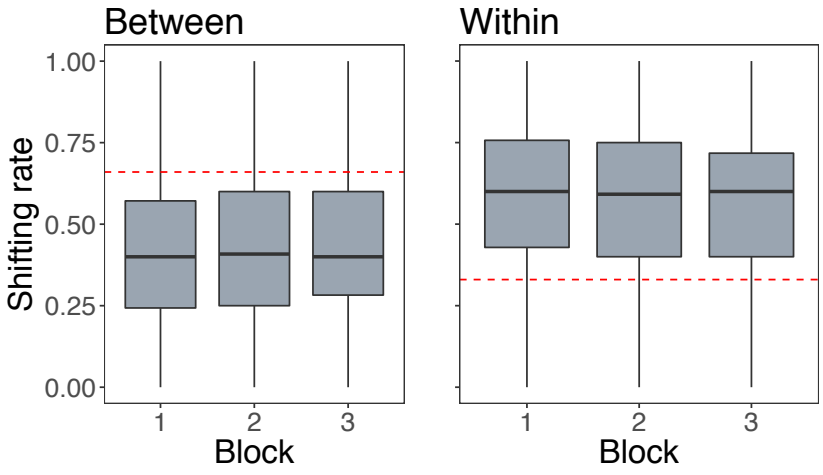
Shifting

Figure 4.20a depicts mean shifting rates in the three familiarisation blocks. Infants consistently made more within-category than between-categories shifts, even though between-categories shifts were twice as likely to occur by chance. To control for these differences, proportions of within- and between-category shifts were calculated by dividing the number of within-shifts with the total number of shifts and the number of between-category shifts by the total number of shifts. As the two measures are related (i.e. they sum up to one), a repeated measures ANOVA with a within-subject factor *Block* and the proportion of between-category shifts as a dependent variable revealed no significant effects; thus the mean proportions are averaged across blocks

and a one-sample t-test against chance (0.66) revealed that infants made significantly less between-category shifts ($t(24) = -19.89,\ p < .001,\ d = 3.78$) and engaged more in the within-category shifts (see Figure 4.20b).



(a)



(b)

Figure 4.20. Experiment 5: Mean shifting during familiarisation blocks (a) and proportion of within-category and between-category shifts (dotted line depicts performance at chance).

Transitional probabilities

The probabilities of making transitions within a category and between categories were calculated using discrete Markov Chains in the same way as described in Chapter 3. Mean probabilities were calculated for within-category and between-categories transitions. In each trial, when fixating an exemplar from one category, an infant may shift to one of the three other

items: one from the same category (*within-category* shift) or to the two items, exemplars of the other category (*between-category* shifts). Therefore, the probability of shifting within-category by chance is .33, whereas the probability of making a between-category shift by chance is .66. The analysis revealed that infants were more likely to make *within-category* transitions ($t(24) = -11.9,\ p < .001,\ d = 2.25$), than between-category transitions ($t(24) = 11.28,\ p < .001,\ d = 2.38$, see Figure 4.21).

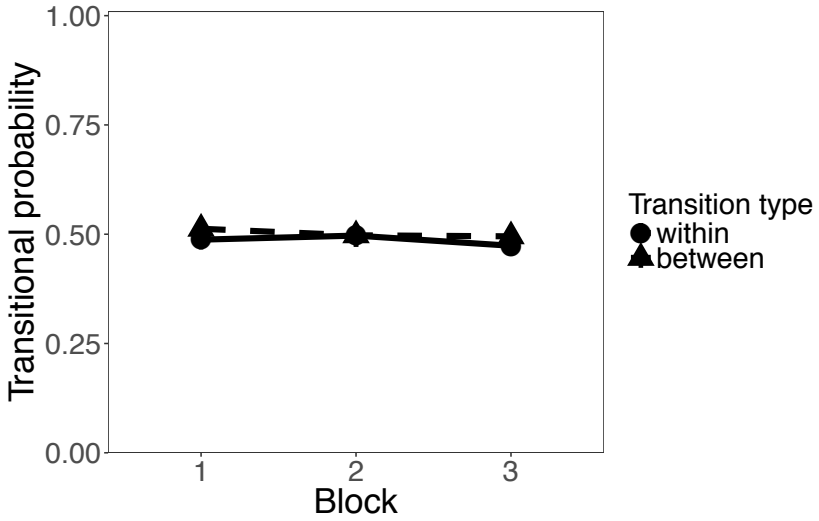


Figure 4.21. Experiment 5: Mean transitional probabilities during familiarisation (circles and triangles depict mean transitional probabilities for within- and between shifts, respectively; error bars depict standard error).

4.4.5 Exploration strategies in training and performance in test

In the next step, the analysis explored a relationship between the exploration strategies infants adopted in training and performance at test. Two separate analyses were conducted, one focusing on links between shifting rates during training and looking preferences in test, while the second one determined

whether transitional probabilities predict performance in test. For both parameters (shifting rates and transitional probabilities), infants were split into two groups based on a median split and performance of the two groups was compared. There was no difference between low- and high- shifters, nor between infants who engaged similarity-focused strategy (i.e. low transitional probabilities) and infants who used a contrast-focused strategy (i.e. high transitional probability).

4.4.6 Moving objects test

After completing three interleaved training and test blocks, infants were presented with an additional test block. This block consisted of three trials. The first two trials comprised of the prototypes of the two categories, and both items were dynamic. In one trial both items moved in a manner specific for one of the two categories, and in the other trial each moved in a manner specific to the other category. The final trial was again the novelty preference trial. No preferences were present in any of the three trials. Infants expressed no evidence of category learning even when test items were dynamic. However, since the experiment was very long, only 16 participants successfully completed this block, providing a small data set and limited statistical power.

4.4.7 Discussion

The present study explored infants' ability to learn dynamic visual categories. Infants were familiarised with four items simultaneously, with familiarisation

items moving while they were fixated. Each category was associated with a specific motion: one category with jumping and the other rocking. Category formation tests followed each training block and revealed no evidence of category learning whatsoever; this was further confirmed by a lack of preference in dynamic tests of category formation that took place at the very end of the study. These results indicate that infants aged 10 months fail to encode correlations among static and dynamic features. Moreover, the presence of dynamic features seems to hinder category formation since the static information alone was sufficient to extract categories (as discovered in Experiment 3).

These findings are in contrast with a recent study which found that motion facilitates extraction of categories in infants of the same age (Deng & Sloutsky, 2015). However, there are several important differences in the experimental designs of the two studies that might be responsible for the differing results. The first key difference between the two studies is the number of items infants are simultaneously presented with. In our study infants were presented with four items in each familiarisation trial, whereas in the study of Deng and Sloutsky (2015) infants were familiarised with one item at a time. Consequently, high perceptual load could have been one of the reasons infants failed to learn categories in the present study. Presenting multiple items that consist of several static and dynamic features might pose a sub-optimal perceptual load that does not allow learning to occur.

Another key difference between the two studies is that, in the present

study, infants were familiarised with exemplars from two categories; Deng and Sloutsky (2015) familiarised infants with exemplars from only one category. Thereupon, the two experiments pose different tasks: splitting similar looking objects into two categories requires infants to draw a boundary between the two categories, whereas learning one category might draw on a different mechanism and instead search for similarities. In addition, there is a decisive difference in the nature of the dynamic features in the two studies. In the present study motion is a general feature, i.e. an entire item moves, while in Deng and Soutsky's study only one feature of an item (creature's feet) moves. Thus motion represented a feature of a feature.

Another important difference in the experimental design of the two studies that might have driven the observed differences is the timing of the motion. In our study, an item would start moving as soon as infants landed a fixation on it, whereas in the study by Deng and Sloutsky (2015) there was a delay, infants first heard a phrase ("Look at this one!") and only then would the feet of an item start moving. This difference in timing indicates that there might be some sort of a "motion overshadowing" taking place when the static and dynamic features to be learned are presented simultaneously. This would parallel the notion of auditory overshadowing which suggests that simultaneous presentation of visual information and labels can have detrimental effects on learning (Robinson & Sloutsky, 2007; Sloutsky & Robinson, 2008).

4.5 Experiment 6: Contrasting labelled and silent categories

A large body of research suggests that language has an important role in concept formation and that category labels can impact category formation (Althaus & Mareschal, 2014; Althaus & Plunkett, 2015; Plunkett et al., 2008; Sloutsky & Robinson, 2008). Typically, in these studies each familiarisation exemplar was named, either with one general label or accompanied by a category-specific label. Most often, a silent condition (that is, when there is no label) is conducted as a control experiment with a separate group of infants. Comparison of the performance of the two groups is used to test for the effect of labels on categorisation. Yet, no studies have explored whether labelled and silent categories be learnt in the same context.

The aim of the present experiment was to bring these two conditions together and investigate infants' ability to learn categories in a context where one category was accompanied by a label, while the other category had no labelled attached. In addition to providing a novel contribution to the literature on the role of labels in categorisation, this experiment also provides further insight into the effect of unbalanced categories, i.e. categories that are defined by a different number of features on category learning. Following an identical experimental design to the previous two experiments presented in this chapter, infants are familiarised with four items simultaneously. Exemplars from one category are labelled while fixated, while exemplars from another category remain silent.

Based on the insights of existing studies, we expect that infants will form the labelled category more efficiently than the silent one. However, if adding a label increases perceptual load above the level optimal for learning, the opposite is expected - the silent category will show an advantage as compared to the labelled category.

4.5.1 Method

Participants

Twenty-four 10-month-old infants took part in this experiment (10 girls and 14 boys, M_{age} = 10.25, age range = 9.60 - 10.7). Two infants were excluded due to fussiness and crying.

Stimuli and experimental design

The stimulus set and experimental procedure are given in the Methods section of this chapter. During familiarisation, exemplars of one category were labelled, whereas exemplars from the other category remained silent (see Figure 4.22).

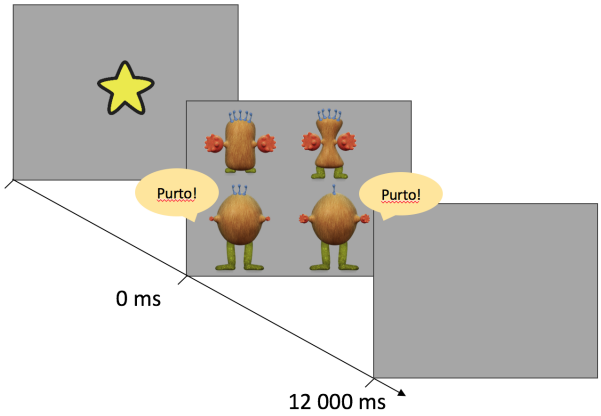


Figure 4.22. Experiment 6: Structure of familiarisation trials.

Test blocks were identical to those in Experiment 4 and 5 (for a description of test trials see Section 4.2.2).

4.5.2 Results

Training

Figure 4.23 depicts mean looking time for labelled and silent categories in the three familiarisation blocks. A repeated-measures ANOVA with factors *Block* (1, 2 and 3) and *Category Type* (labelled and silent) revealed a main effect of *Block* ($F(2, 40) = 15.66,\ p < .001,\ \eta_g^2 = .44$) suggesting there was a decrease in the amount of looking time over the course of familiarisation. A main effect of Category Type and an interaction terms failed to reach significance.

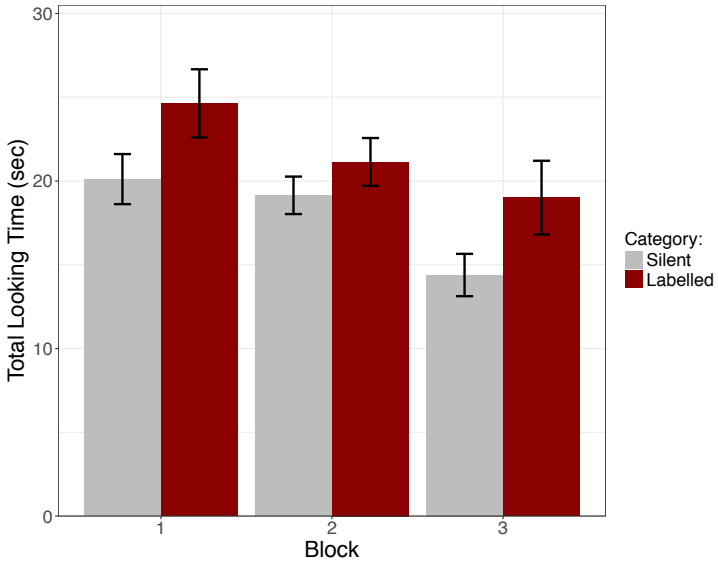


Figure 4.23. Experiment 6: Mean looking time in the three familiarisation blocks (error bars depict standard error).

Growth curve analysis however revealed that, in addition to the main effect of Trial, there was also a significant effect of Category Type. Mean

looking times were significantly higher to to the labelled category than to the silent category (see Table 4.3 and Figure 4.24).

Table 4.3. Experiment 6: Growth Curve Analysis. Parameter estimates

	Estimate	Std. Error	df	t value	p value
Intercept	5768.65	340.73	35.91	16.93	**.00**
Trial (Slope)	-168.78	36.29	308.06	-4.65	**.00**
Category Type	839.19	246.62	989.06	3.40	**.00**

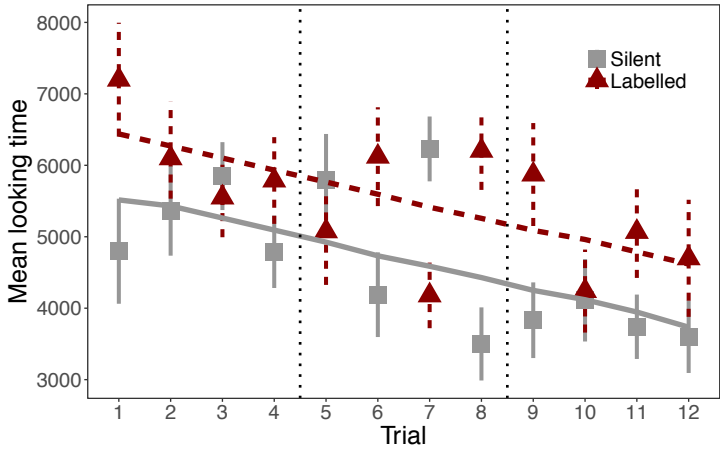


Figure 4.24. Experiment 6: Mean looking time across infants over the course of familiarisation (red triangles and grey squares depict mean looking times for labelled and silent categories, respectively; dashed red line and grey solid line represent model values for labelled and silent category, respectively; vertical dotted lines represent time points when test blocks were presented, error bars depict standard error).

Test - Labelled Category formation

Mean looking preferences were submitted to a one-way ANOVA with a within-subjects factor *Block* (1, 2 and 3). The analysis revealed no significant effect, nor did planned comparisons of performance in each block against chance (*all* $p_s > .05$, Figure 4.25a). An exploratory analysis performed on a subset of trials presented first in each test block suggested that there was a trend towards a novelty preference in the second test block

($t(7) = 3.27$, $p < .05$, $d = 1.16$, Figure 4.25b); however, this effect should be taken with caution due to the small sample size in the subset (N=8).

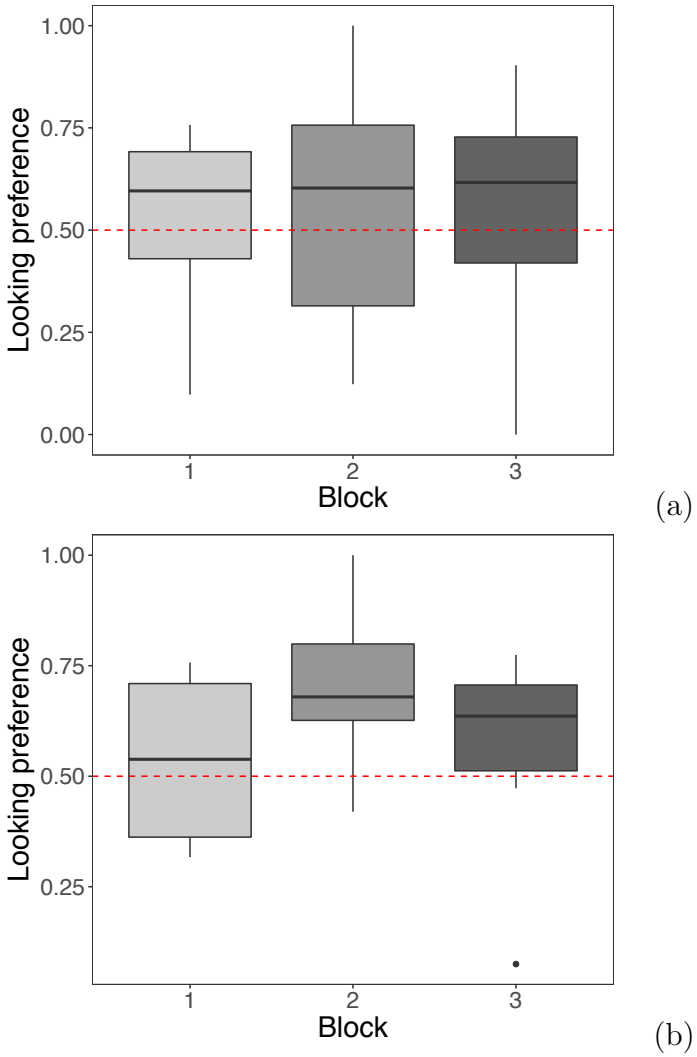


Figure 4.25. Experiment 6: Mean looking preferences in the labelled category formation test (a) and in the first test trial only (b), (values higher than .5 (horizontal dashed red line) suggest a preference for the inter-category item).

Test - Silent Category formation

No evidence of forming the silent category was found as a one-way ANOVA with a within-subjects factor *Block* (1, 2 and 3) yielded no effect, nor did planned comparisons against chance of performance in each block (*all* $p_s >$ .05, Figure 4.26a). When the analysis was performed on the subset of trials

that were presented as first trials in each block, there was a significant novelty preference in the final test block ($t(6) = 2.66,\ p < .05,\ d = 1$, see Figure 4.26b). This provides some evidence that infants were able to form the silent category, however, due to this small sample size, this result needs to be taken with caution.

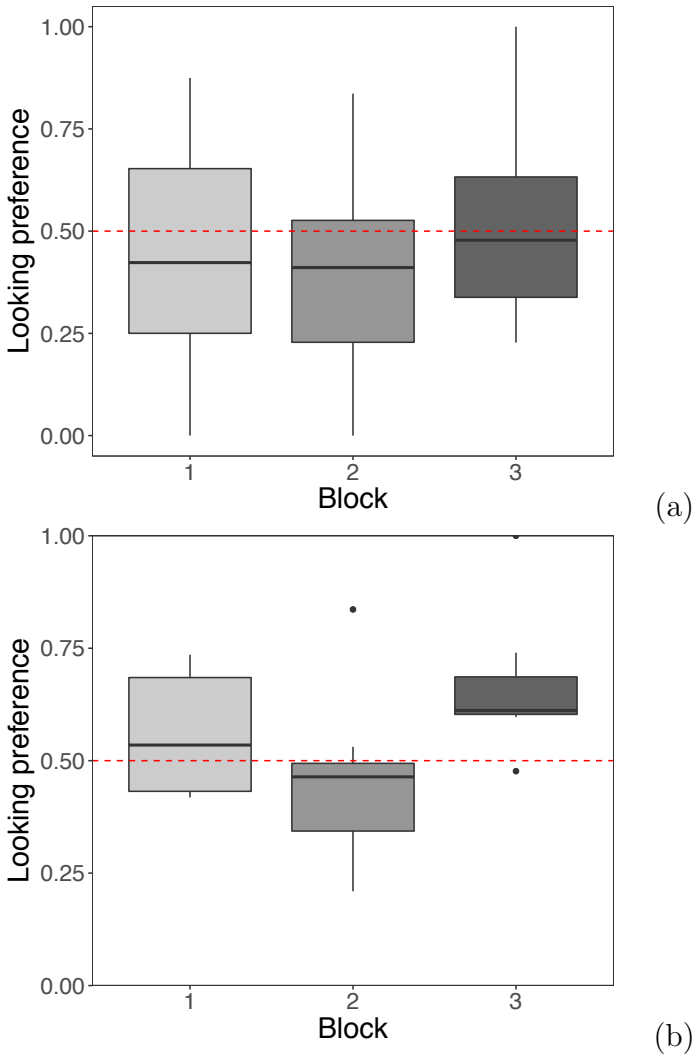


Figure 4.26. Experiment 6: Looking preferences in the silent category formation test (a) and in the subset of trials presented first in each test (b), (values higher than .5 (horizontal dashed red line) suggest a preference for the inter-category item).

Test - Between categories preference

The between categories preference trial tested whether infants express a preference for one of the two categories. Analysis revealed that no significant preference for either of the two categories was present in any of the three test blocks (*all* $p_s > .05$, Figure 4.27a). No preferences were found in the subset of trials presented first in each test block (Figure 4.27b).

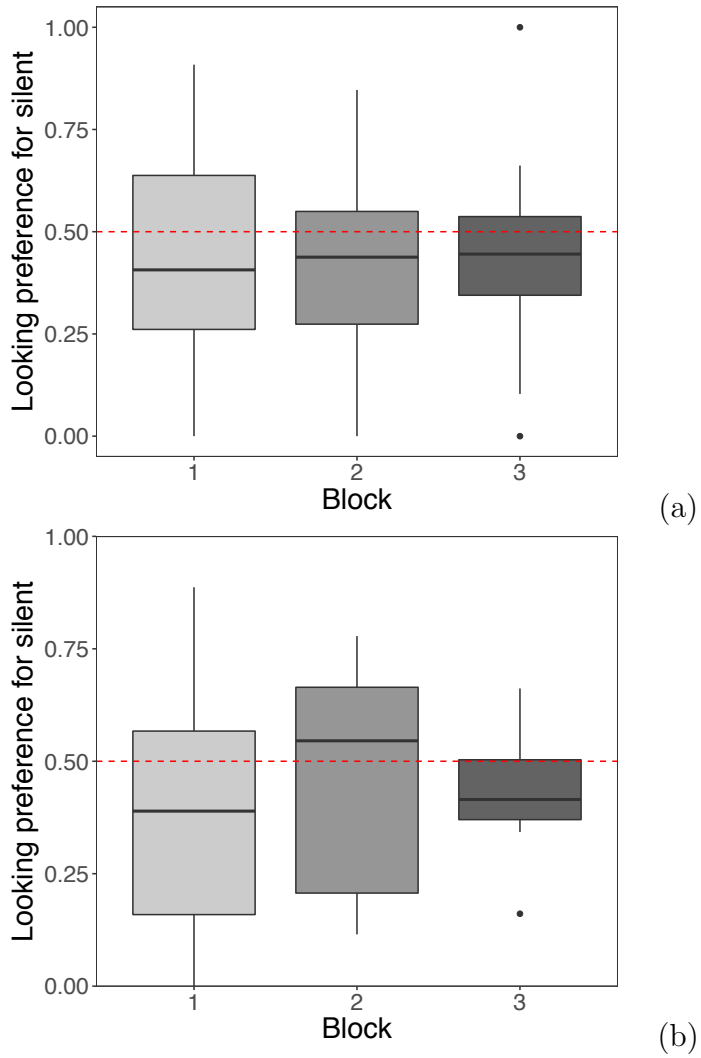


Figure 4.27. Experiment 6: Looking preferences in the between categories preference test (a) and in the subset of trials presented first in each test (b) (values higher than .5 (horizontal dashed red line) suggest a preference for the silent category).

Test - Novelty Preference

A one-way ANOVA with a within-subjects factor *Block* on mean preference scores revealed no effects, but planned comparisons against chance showed a significant novelty preference in the third test block ($t(20) = 2.42$, $p < .05$, $d = .53$, Figure 4.28).

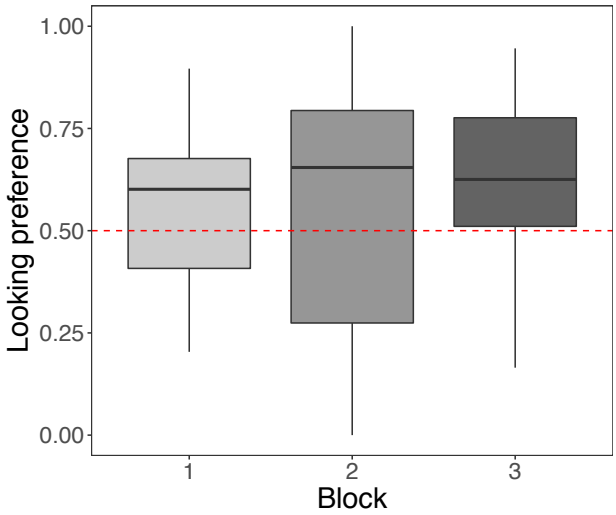


Figure 4.28. Experiment 6: Looking preferences in the novelty preference test (values higher than .5 (horizontal dashed red line) suggest a preference for the novel object)

To summarise, the results suggest that infants showed no evidence of forming either of the two categories, nor did they express a preference for any of the two categories in the between-category preference test..

Exploration strategies during training

To investigate characteristics of the exploration strategies infants engaged in during familiarisation, mean shifting rates and transitional probabilities were analysed.

Shifting

The mean proportions of between-category shifts were submitted to a repeated-measures ANOVA with *Block* (1, 2 and 3) as a within-subjects factor. No significant effects were present and the proportions were averaged over the three blocks. One-sample t-test against chance level (set at 0.66) revealed that infants performed significantly fewer between-category shifts ($t(21) = -9.11,\ p < .001,\ d = 1.94$, Figure 4.29b). There were no systematic differences in shifts within the labelled and within the silent category ($p > .05$, see Figure 4.29c).

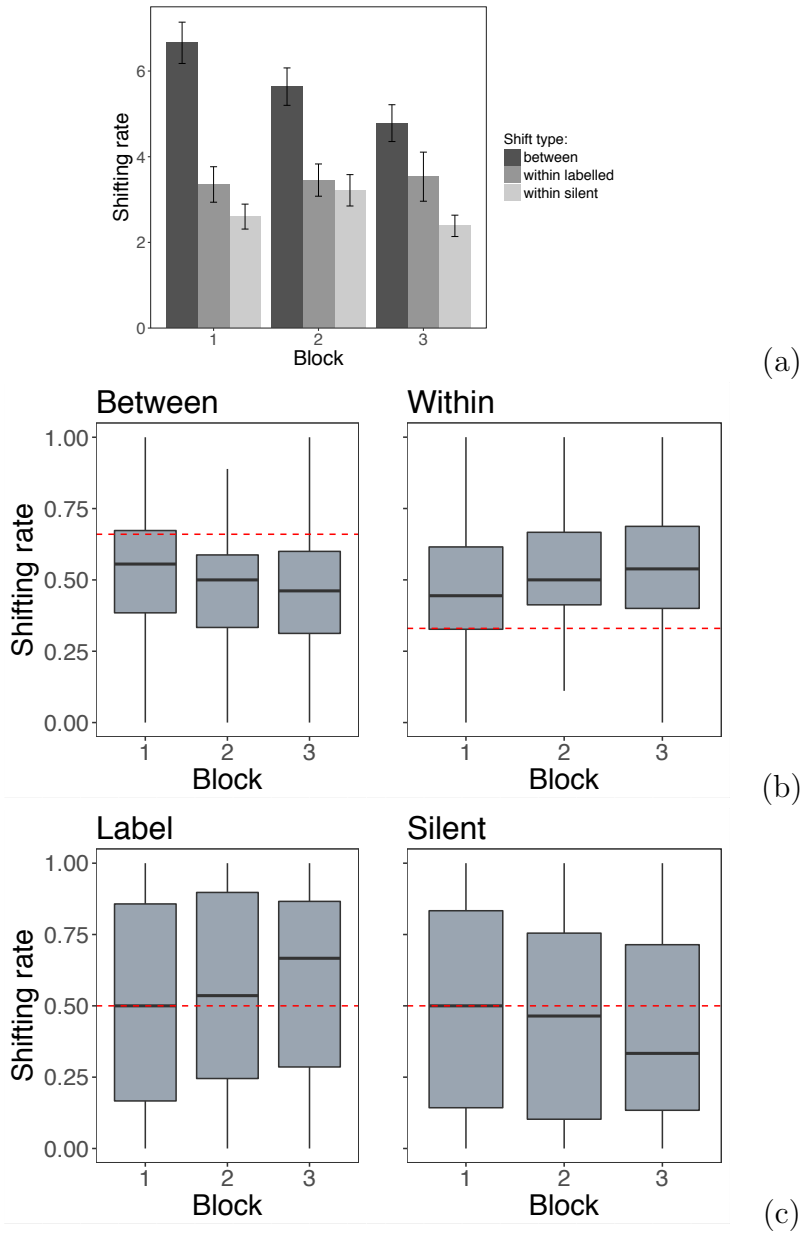


Figure 4.29. Experiment 6: Mean shifting rates in the three familiarisation blocks (a); Proportions of the between-category and within-category shifts (b); proportion of shifts within the labelled category and within the silent category (c), (dotted line depicts shifting rates at chance).

Transitional Probabilities

Visual exploration of a multiple-item display consists of several shifts between the presented items. When looking at one item, infants must decide which item to look at next. If they were looking at a labelled exemplar, the

infant's next look can be directed towards another exemplar from the same category or to an exemplar from the silent category; vice versa, looking at an exemplar from a silent category can be followed by fixations to another silent exemplar or to one from the labelled category. To explore whether certain types of shifts were more likely, again we implemented discrete Markov chains to calculate probabilities of all possible types of transitions (the approach used was identical to those in previous experiments. For a detailed explanation, see the Methods section of Chapter 3).

Overall, infants expressed a preference for similarity focused exploration strategies (see Figure 4.30). That is, *within-category* transitions were more likely than predicted by chance (.33) (within labelled category: $t(21) = 4.33,\ p < .001,\ d = .924$; within silent category: $t(21) = -3.043,\ p < .01,\ d = .65$) and fewer between-category transitions (silent-to-labelled: $t(21) = 3.5,\ p < .01,\ d = .75$; labelled-to-silent: $t(21) = -4.455,\ p < .001,\ d = .95$).

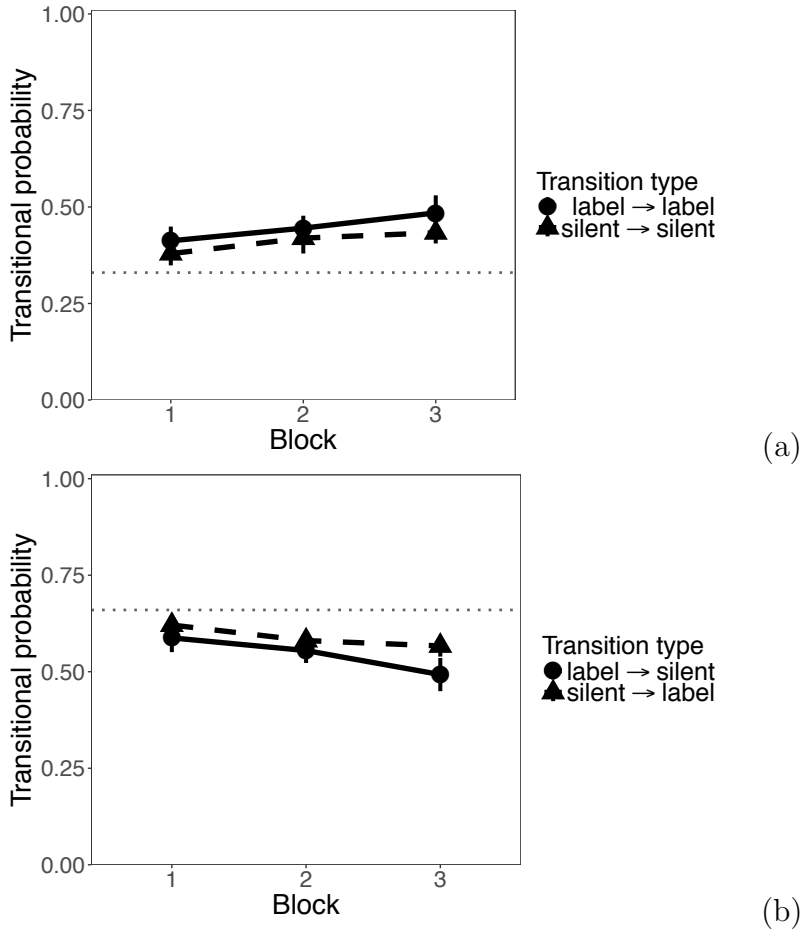


Figure 4.30. Experiment 6: Transitional probabilities in each training block for within-category shifts (a) and for the between-category shifts (b), (dotted line depicts performance at chance, error bars depict standard error, however, not visible due to small values).

4.5.3 Exploration strategies in training and performance in test

Shifting and test performance

Infants were split into two groups: low- and high-shifters based on a median split (Mdn=4.12, SD=.92) and their performance was compared between the groups. As shown in Figure 4.31, there were no consistent patterns, except that in the first test block the low shifters (Mdn=4.08, SD=.94) ex-

pressed weaker preference for the inter-category item as compared to the high shifters (Mdn=4.17, SD = .92) in the silent category formation test ($t(7) = -2.55,\ p < .05,\ d = .9$).

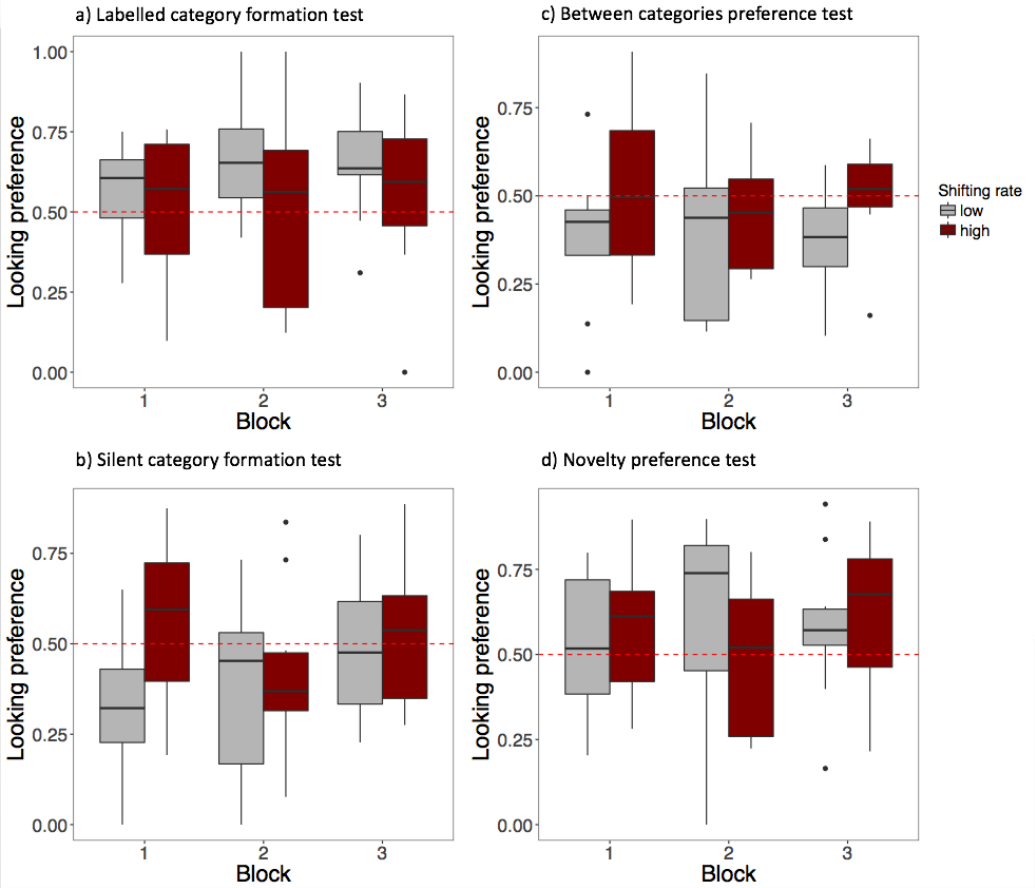


Figure 4.31. Experiment 6: Performance of low- and high-shifters in test.

Transitional probabilities and test performance

Since probabilities of possible transitions from a labelled and from a silent exemplar are independent, two analyses were conducted, one focusing on the effect of strategies present when transitioning from a labelled exemplar and the other one for transitions from a silent category exemplar.

For transitional probabilities of moving from a labelled to a silent category exemplar we found that infants who are more likely to transition from

labelled to silent exemplars than to another labelled exemplar perform better in the labelled category formation test (Figure 4.32a, $t(9) = 2.45$, $p < .05$, $d = .77$). The same effect was present in the novelty preference test (Figure 4.32d). Therefore, the ability to disengage from a highly salient items (labelled category exemplars) seems to lead to a better performance in labelled category formation test.

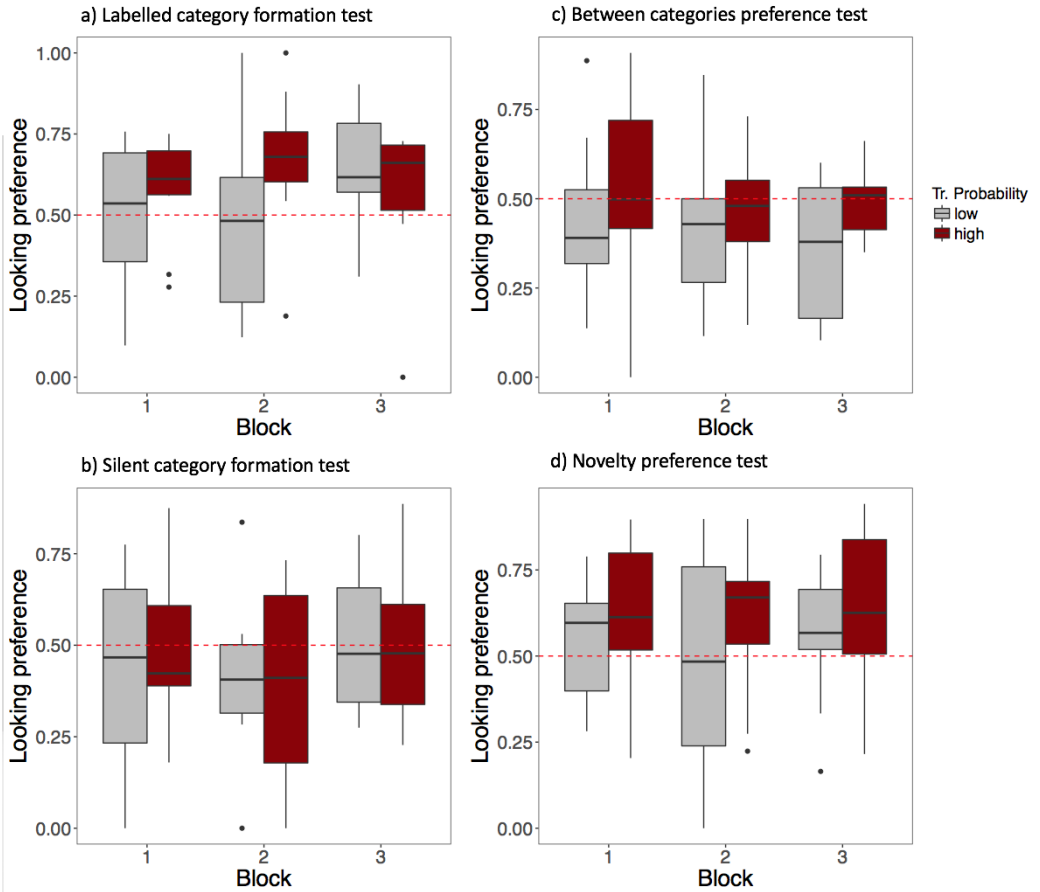


Figure 4.32. Experiment 6: Test performance of infants engaging contrast-focused (high probability of label-to-silent transition) and similarity-focused strategies (low label-to-silent transitional probability).

When split into low-TP (transitional probability) and high-TP groups based on transitions from silent category exemplars and when performance in test is compared, lower probability of transitioning from silent to labelled

exemplar was linked to better performance in the labelled category formation test (block 1: $t(10) = 3.31$, $p = .01$, $d = .9$; block 2: $t(9) = 2.41$, $p < .05$, $d = .76$, Figure 4.33). This suggests that the ability to disengage from a highly-salient item and explore less salient items contributed to the performance in the later test.

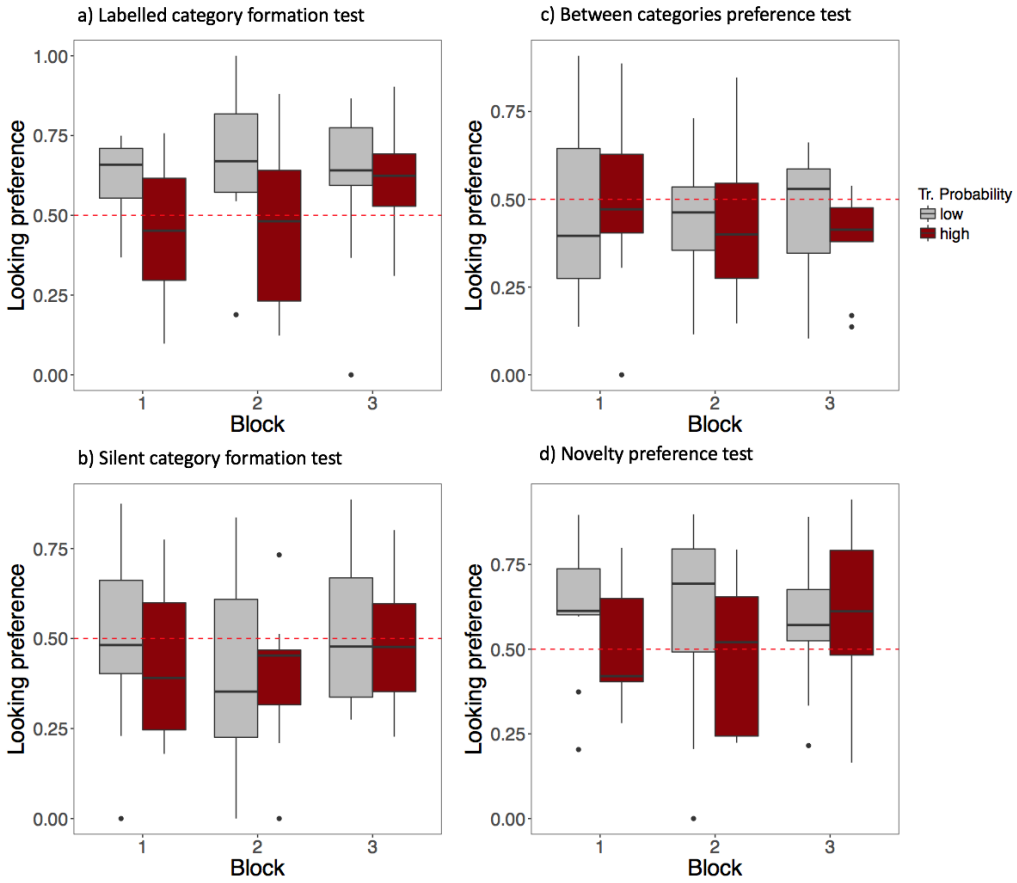


Figure 4.33. Experiment 6: Test performance of infants with low probability of silent-to-label transition and infants with high silent-to-label transitional probability.

4.5.4 Discussion

The present study investigated infants' ability to learn two novel categories when exemplars from one category were accompanied by a category-specific

label while another category had no label associated with it. The aim was to test two proposed hypotheses. One was that the labelled category will be formed faster because of the facilitative effects of labels. An alternative hypothesis was that labels would inhibit category formation due to a high perceptual load and that the silent category would therefore show an advantage. The obtained results provided no support for either of the two hypotheses and suggested that infants learn neither of the two categories. There is some tentative evidence that infants were able to form the silent category in the final test block; this effect was present only if the silent category formation was the first test trial; therefore this finding needs to be replicated and further explored before drawing further conclusions.

In Experiment 3 we showed that infants were able to learn these categories when presented in silence. The only difference between the present experiment and Experiment 3 is that in the present experiment one of the categories was accompanied by a label. Why did adding a label to one category hinder category learning? One possibility is that categories have to be balanced: that is, they should each have an equal number of features in order for infants to learn to distinguish them and successfully establish a boundary between the two categories. In the present experiment, one category was defined by four visual features and a label, whereas the other category was defined only by four visual features. It might be the case that one-to-one mapping it needed: that is, each feature in one category can be contrasted with a feature of the other category in order to distinguish categories and

identify the category boundary. Another possibility is that adding a label hinders learning because it sets a high processing load and therefore impairs infants' ability to extract information relevant to category formation. However, if this was the case, one would expect that infants would still be able to form a silent category, which was not the case.

4.6 Experiment 7: Labelled categories

4.6.1 Introduction

It is well known that language has an important role in shaping categories (Althaus & Plunkett, 2015) and labels can facilitate or even overwrite formation of perceptual categories (Plunkett et al., 2008). Experiment 6 showed that in some circumstances labels can also hinder category learning as it revealed that having one category labelled completely hinders learning of not only the labelled category, but the silent category as well.

The present experiment explored infants' learning of two novel categories, which were both accompanied by a category-specific label. As in Experiment 6, infants were trained using a gaze-contingent eye-tracking paradigm in which four items were presented simultaneously and a label was presented as soon as an infant looked at an item. Unlike Experiment 6, where only one category was labelled, in the present experiment all items were accompanied by category-specific labels.

Based on the existing literature, I hypothesised that labels would facilitate category learning and that categories would be formed faster than when those categories are to be learnt in silence (Experiment 3). However, taking into account the results of Experiments 4, 5 and 6, an alternative hypothesis is that adding an additional feature to both categories will completely impede category learning due to high perceptual complexity and processing demands.

4.6.2 Method

Participants

Twenty-nine 10-month-old infants took part in this experiment (12 girls and 17 boys, M_{age} = 10.22, age range = 9.69 - 11.04). Three participants were excluded from the analysis due to insufficient looking time accumulated in the first training block. Three additional participants were excluded from analysis of the third block due to low looking time (one participant) and not completing the experiments (two participants). Of the 26 participants in the final sample, only 23 contributed to the data for the final testing block.

Stimuli and experimental design

Explanation of the stimulus set and experimental procedure is given in the Methods section of this chapter (Section 4.2.2). The key experimental manipulation in this experiment was that exemplars from both categories were labelled during familiarisation (see Figure 4.34).

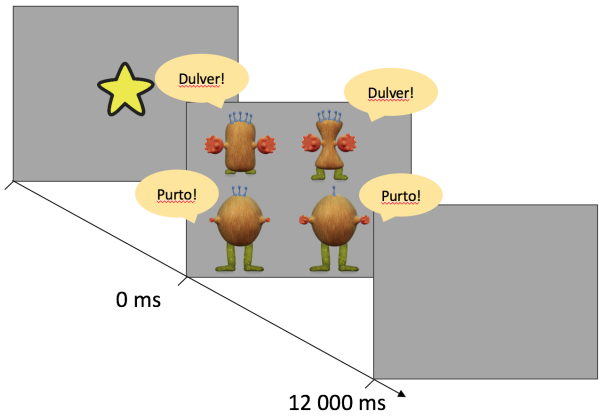


Figure 4.34. Experiment 7: Structure of familiarisation trials.

Test blocks were identical to those in previous experiments and described in Section 4.2.2 of this chapter.

4.6.3 Results

Training

As depicted in Figure 4.35, there was a decrease in the amount of looking time over the course of the experiment and infants spent similar amounts of time exploring the two categories. This was confirmed by a two-way ANOVA on the looking time with within-subjects factors *Block* (1, 2 and 3) and *Category* (1 and 2) revealing a main effect of *Block* ($F(2, 42) = 3.51,\ p < .05,\ \eta_g^2 = .19$), while there was no effect of Category nor an interaction.

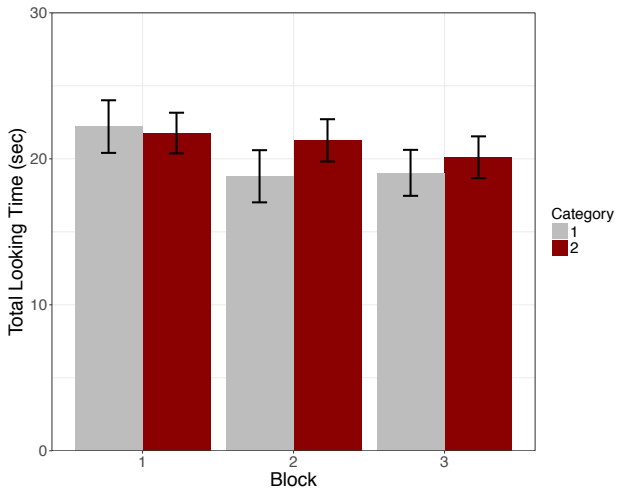


Figure 4.35. Experiment 7: Mean looking time in three familiarisation blocks (red and grey bars represent looking time accumulated for the two categories separately, error bars depict standard error).

Growth curve analysis confirmed there was a habituation effect as adding Trial to the base model improved the model fit. Adding Category (1 and 2) to the model provided no model improvements, therefore confirming infants spent a similar amounts of time exploring both categories (see Table 4.4 and Figure 4.36).

Table 4.4. Experiment 7: Growth Curve Analysis. Parameter estimates.

	Estimate	Std. Error	df	t value	p value
Intercept	5835.10	303.92	39.80	19.20	**.00**
Trial	-129.32	36.86	508.38	-3.51	**.00**

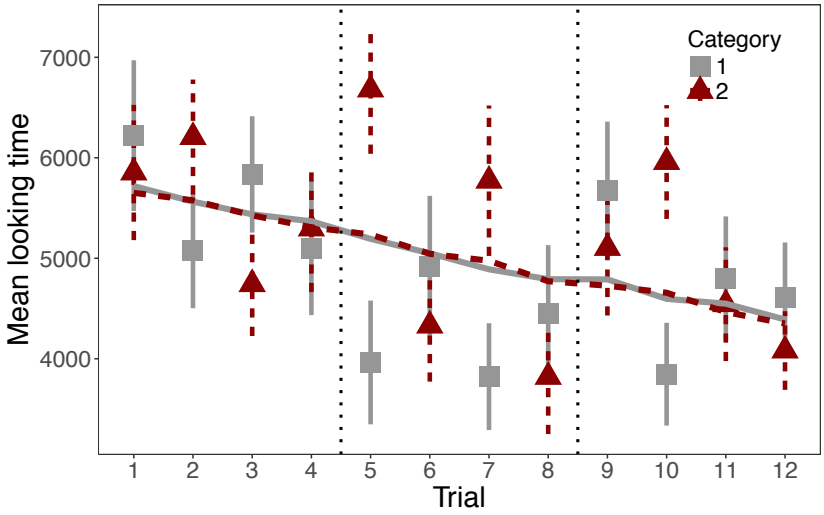


Figure 4.36. Experiment 7: Mean looking time across infants in each training block (red triangles and grey squares depict mean looking times for the two categories; dashed red line and grey solid line represent model predictions; error bars depict standard error).

Test - Category formation

Preference scores were calculated by dividing the time spent looking at the inter-category item by the total time spent looking at both objects. As there was no reason to assume different performance for the two categories, preference scores in the two trials were averaged. A one-way ANOVA with within-subject factor *Block* (1, 2 and 3) revealed no effect. Planned comparisons of the mean preference in each test block against chance (0.5) revealed evidence of category formation in all three test blocks (block 1: $t(25) = 2.45$, $p < .05$, $d = .48$; block 2: $t(25) = 2.16$, $p < .05$, $d = .42$; block 3: $t(21) = 2.36$, $p < .05$, $d = .50$, see Figure 4.37) .

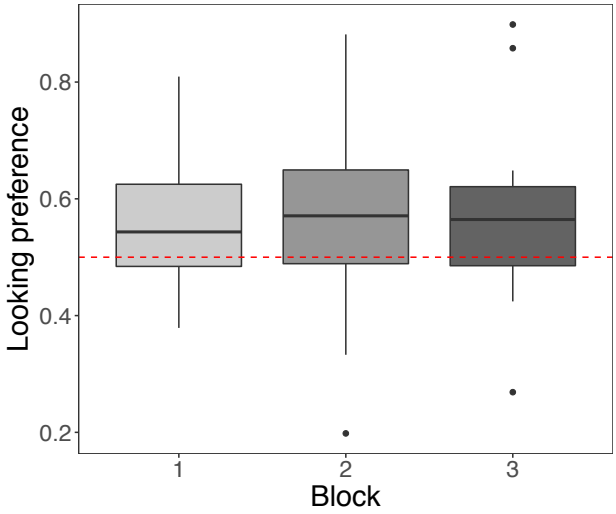


Figure 4.37. Experiment 7: Looking preferences in the category formation test (higher preference value indicates preference for the inter-category item).

Test - Between categories preference

This test was introduced to determine whether infants express a preference for one of the two categories. A one-way ANOVA with within-subject factor *Block* (1, 2 and 3) and planned comparisons of the mean preference in each test block against chance showed no systematic preference was present for any of the two categories in any of the three test blocks (Figure 4.38).

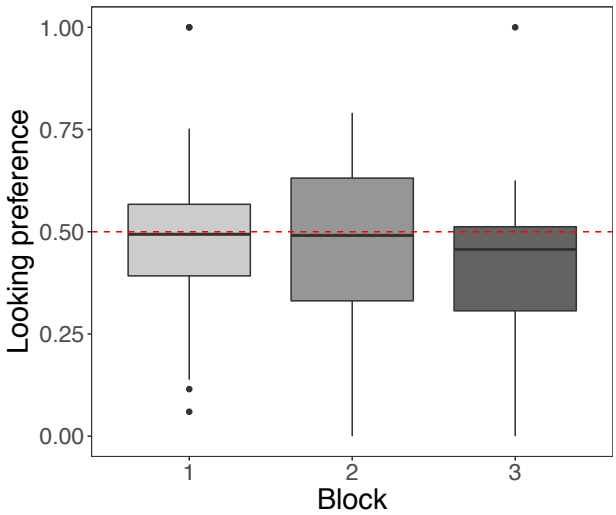


Figure 4.38. Experiment 7: Looking preferences in the between-categories preference test.

Test - Novelty preference

Mean looking preferences in the novelty preference trials were submitted to a one-way ANOVA with within-subject factor *Block* (1, 2 and 3) and no effect was present. Planned comparisons of the mean preference in each test block against chance showed novelty preference was present in all three blocks ((block 1: $t(25) = 1.74$, $p = 0.09$, $d = .34$; block 2: $t(25) = 2.49$, $p < 0.05$, $d = .49$; block 3: $t(21) = 3.20$, $p < 0.01$, $d = .68$, Figure 4.39).

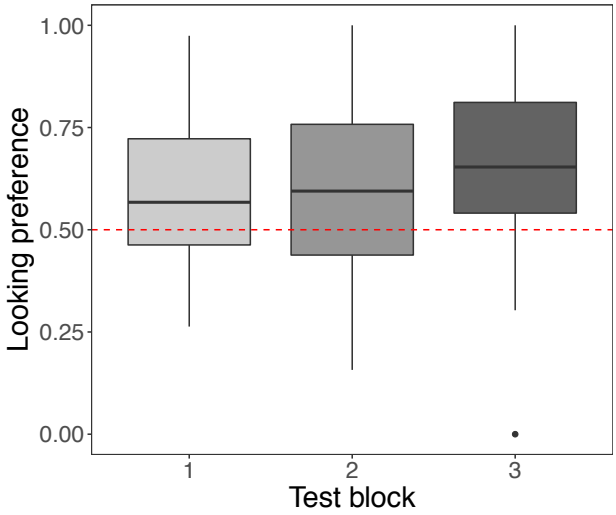


Figure 4.39. Experiment 7: Looking preferences in the novelty preference test (higher preference value indicates novelty preference).

To summarise, the analyses of looking preferences in the test showed that infants learnt labelled categories already after one block of training and express the same pattern of preferences in all test blocks.

Exploration strategies during training

Two parameters of exploration behaviour during familiarisation were analysed: mean shifting rates and transitional probabilities.

Shifting

Figure 4.40a depicts mean shifting rates for within- and between-category shifts in the three familiarisation blocks. Infants make more within-category shifts in all three blocks, suggesting they engage a similarity-focused exploration strategy.

A repeated-measures ANOVA with a within-subject factor *Block* (1, 2 and 3) on the proportion of between-category shifts revealed no significant effects and mean proportions were averaged across the three blocks. A one-sample t-test against chance (0.66) revealed that infants made fewer between-category shifts than expected by chance ($t(28) = -14.37,\ p < .001,\ d = 2.67$, Figure 4.40b) and engaged similarity-focused explorations strategies.

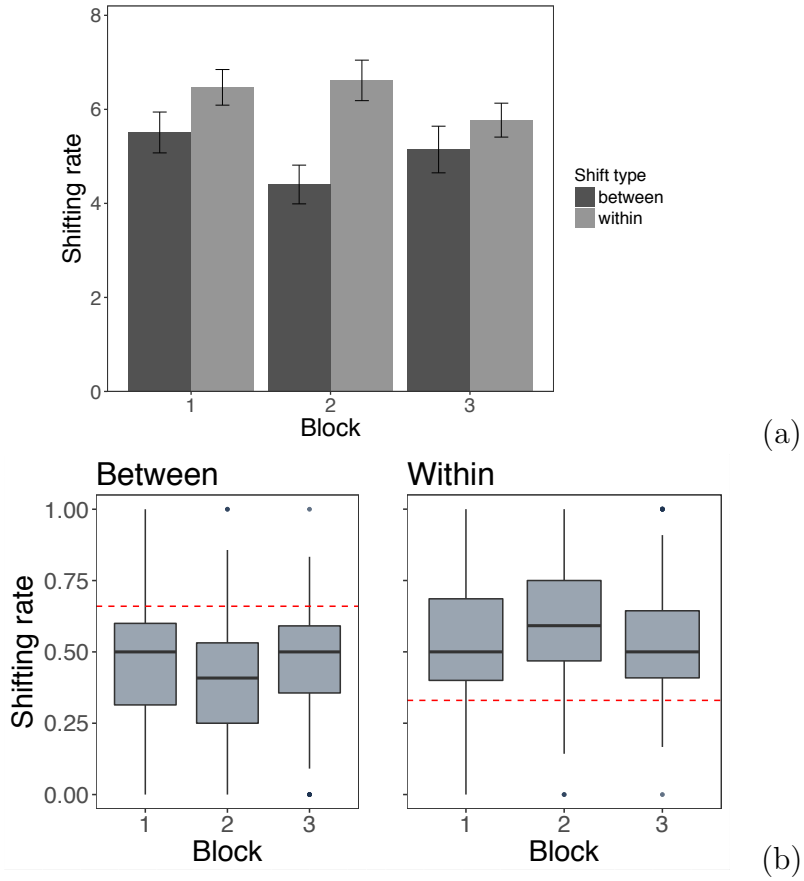


Figure 4.40. Experiment 7: Mean shifting rates in the three familiarisation blocks (a) and mean shifting proportions for between and within-category shifts (b).

Transitional probabilities

As in previous experiments, we calculated the respective probabilities of the different transitions possible during looking exploration. As both categories were labelled, transitional probabilities were averaged for transitions within categories and for the between-categories transitions. Within-category transitions were significantly more likely than was predicted by chance ($t(27) = 12.79,\ p < .001,\ d = 2.32$), whereas between-category transitions were less probable ($t(27) = -12.03,\ p < .001,\ d = 2.27$). Together with the shifting analysis, these results show that infants engaged similarity-

focused exploration strategies (Figure 4.41).

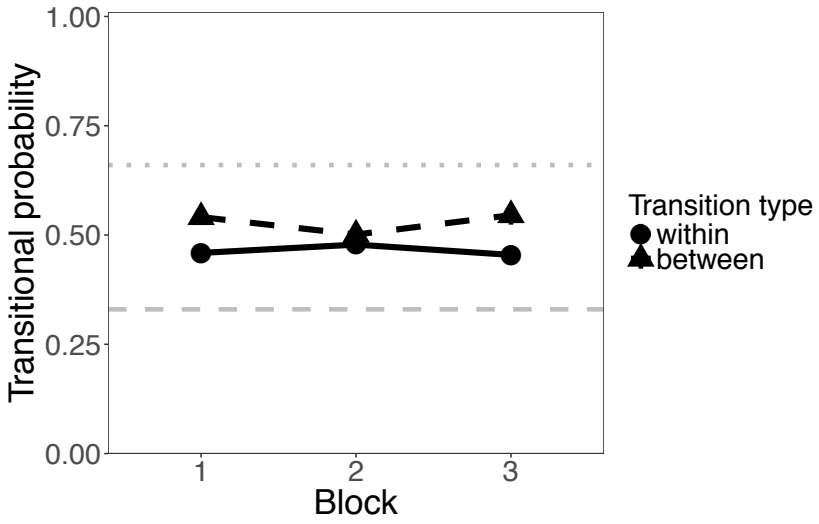


Figure 4.41. Experiment 7: Transitional probabilities for the within- and between-category shifts in the three training blocks (dashed grey line depicts chance level performance for *within-category* and dotted grey line for *between-category* transitions).

4.6.4 Exploration strategies in training and performance in test

In the next step we explored whether there is a difference in the test performance of infants who engage in different strategies during learning. Separate analyses were conducted for shifting and transitional probabilities scores. For each of the two exploration measures, infants were divided into two groups based on a median split and the performance of the two groups was compared. The analyses revealed no differences between low- and high-shifters, nor between low and high-TP groups.

4.6.5 Discussion

The present experiment explored whether infants are able to form two novel categories based on feature correlations in a situation when the categories were labelled. Results revealed that infants are successful in forming two categories after only one block of familiarisation and maintain this preference at test throughout the three blocks.

This result shows that introducing labels as additional features does not hinder category learning. On the contrary, adding labels facilitates the learning process. Namely, infants in the present experiment show evidence of category formation already after only one block of training, whereas when these categories were presented in silence, it takes three times as much exposure for infants to succeed in this (Experiment 3).

Another interesting result of the present experiment is that infants continuously show preference in all three test blocks. This is in contrast to the findings of the set of studies presented in the previous chapter, which suggested that infants tend to disengage from the task once they learn the categories. Hence, it seems that labels not only facilitate learning, but also help to keep infants engaged in the task.

4.7 Chapter 4: Discussion

This chapter was set out to explore the role of motion and language in infants' category learning. Four experiments were conducted to explore how these highly salient additional features affect the process of encoding correlations among features and using these correlations as a basis for category formation. Previously, we found that the efficiency with which infants learn these categories is dependent on the perceptual load. Presenting two items simultaneously during familiarisation facilitates category learning as compared to presenting the same items in succession. However, presenting four items at the same time slows down the learning process and it takes three times as much familiarisation experience for infants to learn the same categories. One aim of the present set of studies was to extend this finding and determine how increasing the input complexity by introducing both motion and labels affect the learning process. Experiment 4 explored how adding a dynamic feature to one category affects category formation and in Experiment 5 both categories had category-specific motions. Experiment 6 and 7 tested whether adding labels affects category learning, first when only one category was labelled (Experiment 6), and secondly when both categories have labels attached (Experiment 7).

The results of the four experiments revealed that the effects of additional features are different in the case of motion and labels. Adding dynamic motion features to both categories (Experiment 5) impeded the process of category formation. In contrast, adding category names to both categories

facilitated category formation (Experiment 7). These findings suggest the relationship between perceptual complexity and learning is not linear. Instead, it is dependent on the nature of the features comprising the category. Infants at 10-months of age seem to struggle with binding static and dynamic visual information which may lead to a failure in extracting the category relevant information. This result is particularly interesting considering that encoding the dynamic feature is not necessary when forming categories (Experiment 3 showed that infants are able to learn these categories). This might be a result of young infants' limited ability to bind static and dynamic information, as suggested by several studies (Burnham et al., 1988; Rakison, 2004; Rakison & Poulin-Dubois, 2002) or some sort of "motion overshadowing" in which processing dynamic visual information affects the processing of static visual features (as a parallel to the concept of auditory overshadowing proposed by Sloutsky & Robinson 2008).

In contrast to the detrimental effects of dynamic features, adding the label as an additional feature facilitated the process of category formation (Experiment 7). This is further supported by comparing the results of the Experiment 7 (two labels) to the findings of the Experiment 4 (presented in the previous chapter). In Experiment 4, infants were familiarised with the same categories as in Experiment 7, but no labels were presented and three familiarisation blocks were needed in order for infants to succeed in the category learning task. In contrast, when categories were accompanied by category-specific labels, infants express evidence of category learning after

just one familiarisation block.

This facilitative effect is line with an existing body of research showing a compelling impact of labels on categorisation (Althaus & Mareschal, 2014; Althaus & Plunkett, 2015; Plunkett, 2010; Plunkett et al., 2008; Waxman & Markow, 1995). Mechanisms driving the observed effects are yet to be fully understood. Some researchers see labels as features (Gliozzi et al., 2009; Plunkett et al., 2008), while others give them a special status of category markers (Waxman & Markow, 1995). The present study clearly demonstrates the facilitative effect of labels. Whether these effects are due to labels having a different status than other features, or they are features that drive the observed effects by directing and maintaining infants' attention is to be addressed in future research.

Experiments 4 and 6 explored whether introducing an additional feature to only one of the two categories affects infants' ability to form a boundary between the two categories. The results suggest that for both motion and label this has negative effects on category formation. Infants do show some evidence of having learned the static category, but fail to learn the dynamic one. In the case of labels, however, they learn no categories whatsoever. To some extent this suggests that categories need to be symmetrical, i.e. have identical number of features and that each category can be defined by each feature (which is not possible for the silent/static category) in order for the categories to be established. Alternatively, the failure to form the dynamic category can be explained by difficulties in encoding dynamic-static com-

pounds, but provides no explanation for the lack of learning in the one-label context. One possible way to address these questions is to determine whether infants are capable of learning categories in a semi-supervised fashion, that is, when some exemplars of a category are labelled while some are not.

Lastly, the present set of experiments used a gaze-contingent familiarisation paradigm in which items would only be labelled or start moving once fixated. This provided infants with a more active role and design their own learning experience. At the same time, using gaze-contingent labelling implies that it is not possible to completely separate effects of the experimental manipulation in interest (i.e., the presence of labels) and an effect of gaze-contingency. Strictly speaking, this means that the facilitative effects of labels discovered are actually gaze-contingent labelling effects. The fact that motion does not provide similar effects suggest that it is labels that impact learning, but the extent to which gaze-contingency also contributed to these effects needs to be addressed in future studies. One possible route is to test for the specific contributions of the gaze-contingency by introducing a more neutral gaze-contingent effect (instead of labels or motion), such as luminance change. For instance, when fixated, an item could change luminance. This will enable us to better understand the unique contributions of labels in the observed effects.

The present chapter explored the roles of motion and language in separate experiments. However, infants usually encounter all this information within the same context. Therefore, future studies should investigate how these

different types of information interact. Taking all experiments presented in the previous and present chapters together, defining a learning context in terms of task difficulty and perceptual load is not a matter of just adding up the information and the number of features - it is a more complex process. Experiments aiming to understand aspects of these multimodal interactions would provide a better understanding how the complex and fascinating abilities of categorisation infants as young as 10-months develop. One possible route is to test for the combined effects of both labels and dynamic visual information in the same learning context.

Part II

Category Learning in Adults

Chapter 5

Implicit categorisation in adults

5.1 Introduction

The ability to spontaneously organise objects into categories is seen as a pillar of the conceptual system. The process of implicit categorisation has been extensively studied in infants (Bomba & Siqueland, 1983; Eimas & Quinn, 1994; Mareschal & Quinn, 2001; Younger, 1985), but the focus of categorisation research in adults has been on the explicit, supervised category learning, while less attention has been given to implicit categorisation (Carvalho & Goldstone, 2014; Goldstone et al., 1978; Reed et al., 1999). In addition, there have been very few attempts to compare infants' and adults' category learning processes and to understand whether there are mechanisms of spontaneous conceptual organisation that are shared across the lifespan.

According to the Competition Between Verbal and Implicit Systems model (COVIS, Ashby et al. 1998), the implicit category learning system relies on the neural substrates that become mature during the first year of life and

operates in a similar way at different points of development (Livingston & Andrews, 2005; Minda et al., 2008). Another integrative account for infant and adult category learning was offered by a computational model (Gureckis & Love, 2004; Love et al., 2004). Initially developed as a model of adult categorisation, the *Supervised and Unsupervised Stratified Adaptive Incremental Network* (SUSTAIN) proposes that the mechanisms underlying infant and adult categorisation are not substantially different. The model assumes a continuous trajectory of conceptual development and describes developmental differences as a result of limitations in memory and stimulus encoding during infancy. Surprisingly, not many attempts have been made to test these hypotheses empirically. Given the differences in cognitive and behavioural capacities of infant and adults, a major challenge is designing experimental paradigms that would allow for certain comparisons to be made.

The main purpose of this experiment was to explore whether there is a way to engage adults in the processes of implicit category learning in the same way the novelty preference tasks do in infants. To achieve this, a novel experimental paradigm, based on the *1-back task* was developed. Participants were presented with a series of items in succession and for each item asked to judge whether it was identical or different from the previous item. Some comparisons involved items from the same category, while others were between-category comparisons. No information about categories and no feedback was given to participants, therefore providing a way to test the process of implicit unsupervised learning. The main aim was to test whether this

task invokes categorisation processes in the same way as the novelty preference task does in infants. Another goal of this experiment was to determine whether the stimulus set developed to investigate infants' categorisation (the Suki set) invokes categorisation processes in adults, just as it does in infants. Developing a task that engages implicit categorisation processes in adults, whilst using identical stimuli as in infants would represent the first step towards empirically addressing the question of shared mechanisms underlying categorisation in infancy and adulthood.

Understanding how categories are formed based only on the perceptual input, in the absence of instructions and feedback is closely related to the phenomenon of perceptual learning which provides a substantial evidence for categorical perception, i.e. better discrimination between objects coming from different categories, as compared to discriminating objects that come from the same category, even when perceptual difference is identical in both cases.

Categorical perception is a clear demonstration of the interaction between perception and cognitive processes, such as categorisation. Several studies have shown that perception cannot be seen merely as a provider of fixed building blocks later used by categorisation, but on the contrary, there is a dynamic interaction and categorisation can influence products of perception (Schyns et al., 1998).

Impact of categorisation on perception has been demonstrated in a series of studies by Goldstone and colleagues (Goldstone, 1994; Goldstone et

al., 1978, 2001). For instance, Goldstone (1994) conducted a series of experiments in which participants first took part in a category learning task. Then, in a perceptual task participants were presented with the same set of items and asked to make same-different judgements. Goldstone found that features relevant in the categorisation task were better discriminated in the subsequent perceptual discrimination task and proposed that perception supports categorisation by adapting it to the requirements imposed by the task (Goldstone, 1994).

The novel experimental paradigm implemented in the present experiment enabled us to extend these findings and explore whether the observed perceptual effects are also present when there are no such task requirements. The present task did not require participants to make any category-related judgements, and thus any observed categorical effects should reflect spontaneous organisation of the perceptual input. Furthermore, whereas the perceptual task in these studies followed a category learning task, the novel task implemented in the present experiment only consists of a perceptual task, and categorisation effects are be an outcome of the interaction between perception and categorisation.

Taken together, existing research suggests that category learning results in making objects from the same category become more similar and objects organised into different categories become increasingly dissimilar (Goldstone et al., 2001) and adjusting relevance of perceptual features in response to experience (Sigala et al., 2002). Therefore, if categorisation processes were

engaged during the task, we expected to observe a difference in the time needed to respond to *within-* and *between-category* comparisons emerging over the course of the task, even when perceptual similarity is controlled for. More specifically, we predicted that *within-category* comparisons would become harder and therefore elicit slower responses as compared to the *between-category* comparisons. In contrast, if the task did not trigger categorisation processes, no systematic differences in responses to *within-* and *between-category* comparisons are expected.

5.2 Method

5.2.1 Participants

Twenty five participants, undergraduate and graduate students at the University of Oxford, took part in this study and received course credit or were reimbursed for their participation (age: $M = 23,\ SD = 2.54$). All participants were right handed and had normal or corrected-to-normal vision with no known neurological impairments. The study was approved by the University of Oxford Medical Sciences Ethics Committee. One participant did not complete the study due to fatigue and hence was not included in the analysis.

5.2.2 Stimuli

The Suki stimulus set, previously used to investigate categorisation in infants (Experiments 1-7), was used in the present study (for more details about the stimulus set, see Chapter 3).

5.2.3 Experimental design

The novel task we introduced was an adaptation of the *1-back task*. Participants were presented with a series of items in succession and for each item asked to judge whether it was identical to or different from the previous item. To balance the number of *same* and *different* responses, participants were asked to respond only following cued items, i.e. when a red probe was presented in the middle of the screen overlayed on the actual stimulus item.

The study consisted of seven blocks. Twenty four training exemplars and six test exemplars were presented in each block. Each block contained 36 training trials, followed by 10 test trials. The stimuli for the training and test trials were drawn from the same set; thus, the transition between the training and test phases was not evident to the participants.

Training

The order of presentation was pseudorandomised and different within each block. Each trial started with a presentation of a fixation cross for 500 ms, followed by a single exemplar item presented for 1000 ms and an inter-trial interval of 500 ms. Participants were instructed to indicate whether the current item was identical to or different from the previous item by pressing

a button on a response box as soon as possible after the item appeared (see Figure 5.3). In order to balance the number of *same* and *different* responses, participants were instructed to respond only for cued trials, i.e. when a red dot appeared in the centre of the screen overlaying the item. Therefore, there were 16 *same-trials*, 16 *different-trials* and four *no-response* trials in the training phase of each block. Half of the *different-trials* compared *between-category* items and the other half compared *within-category* items (see Figure 5.2a).

To familiarise participants with the task, the study started with a warm-up block that was not included in the analysis. This block had an identical structure to the main study, though participants were presented with exemplars from an existing natural category (frogs).

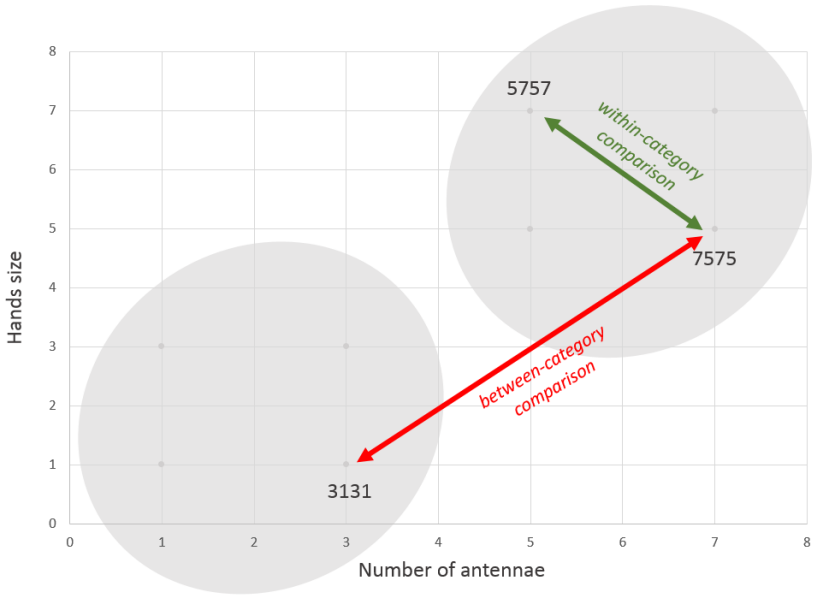
Test

In the training, there could be an advantage discriminating *between-category* pairs not only due to category membership, but also due to differences in the overall visual similarity between items. Items from different categories also look more different from each other than items from the same category, therefore making it easier to distinguish them.

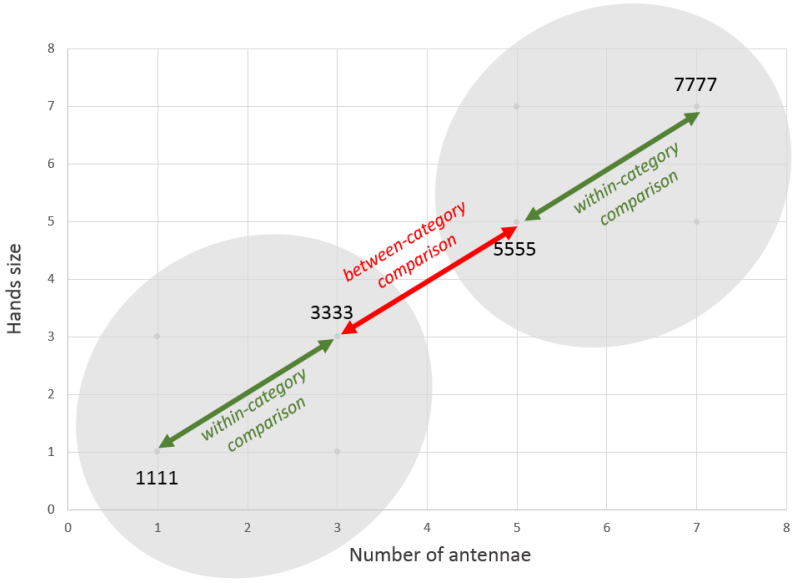
In the test phase, for any two items for which a response was required the visual similarity between them (as measured by the Euclidean distance) was kept constant (see Figure 5.1 and for more details about similarity calculations see Chapter 3). While having the same visual similarity, one pair crossed the category boundary (3333-5555), one pair was borderline (4444-

6666) and two pairs were within the same category (1111-3333, 5555-7777). Out of ten test trials, four were *same-trials*, five were *different-trials* (two *within-category*, two *between-category* and one *borderline*) and one was a *no-response* trial (see Figure 5.2b). In this way, we tested for the categorisation effects while controlling for the effects of the visual similarity.[1]

[1]Euclidean distances were used to determine visual similarities between items; subjective similarity measures are yet to be obtained and compared to the used measures.

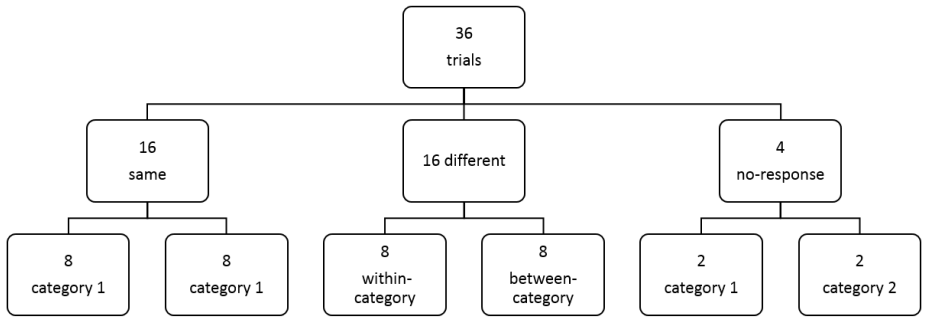


(a)

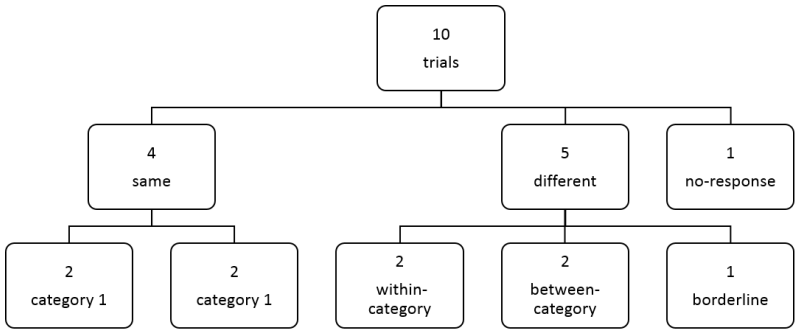


(b)

Figure 5.1. Visual similarity for *within-* and *between-category* comparisons was variable in the training blocks (a), but kept constant in the test blocks (b).



(a)



(b)

Figure 5.2. Experiment structure in training (a) and test blocks (b).

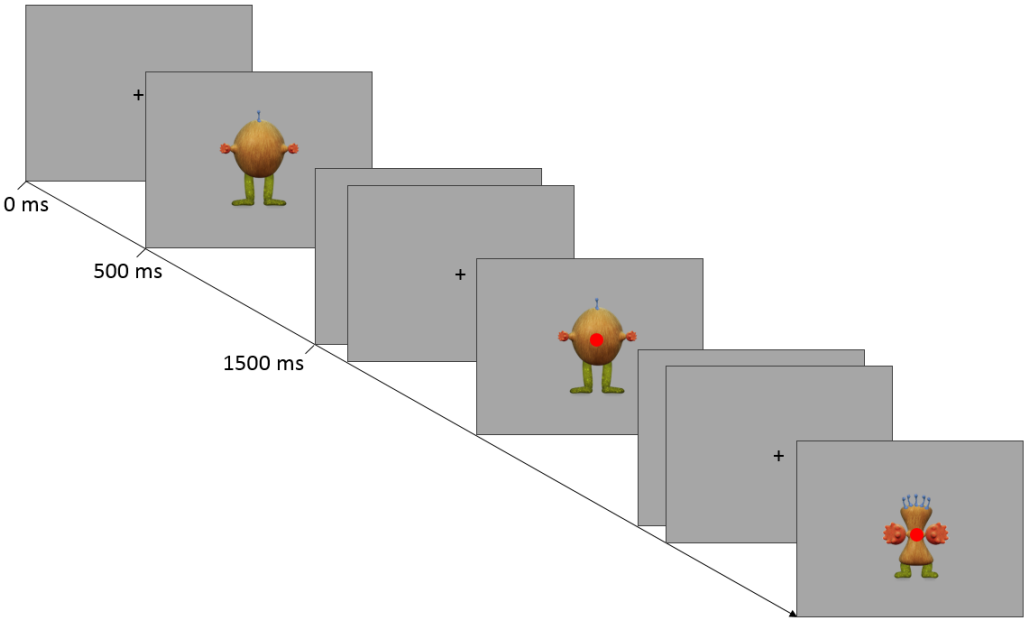


Figure 5.3. Example of the trial sequence.

5.2.4 Equipment and experimental procedure

Stimuli of 500 x 500 pixels each were presented against a 5% grey background in the middle of a 1920x1080 inch screen. The study was run with custom Matlab scripts based on the Psychophysics Toolbox. Participants responded by pressing one of two buttons on a response box (Model RB-540, Cedrus Inc., San Pedro, CA, www.cedrus.com). Response-button assignment was counter-balanced across participants. Upon arrival, participants had the procedure explained and signed the consent form. The experiment was conducted in a sound-attenuated test booth. Short breaks were given after every two blocks.

Upon completing the task, there was a short debriefing session. Participants were asked whether they think items they saw were organised into categories.

5.2.5 Analysis plan

The same statistical analysis of reaction time was performed for both training and test blocks. Only trials with correct responses were included in the analysis. All analyses were carried out in R using the lme4 and lmerTest packages (version 1.1-12 and 2.0-33, respectively). None of the participants reported being aware that items came from two categories, so all participants entered the analysis.

To determine if there was a difference in reaction times for different types of judgements over the course of the experiment, mean reaction time was

modelled using a linear growth model (Mirman, 2014) with fixed effects of *Block (1-7)* and *Comparison (levels: same, borderline, within-category, between-categories)*. A random effect of *Participant* on the intercept and slope terms was included in the model to account for individual differences in initial speed and rate of change. The fixed effect was added to the base model and its effect on model fit was evaluated using model comparisons.

In addition, for the test trials, *within-* and *between-category* trials were directly contrasted using the same approach. Mean reaction times were modelled using growth curve analysis with fixed effects of *Block* and *Comparison*. Model specifications for the growth curve analysis were as follows:

base model: $RT \sim Block + (1 + Block \mid ID)$

model 1: $RT \sim Block + Comparison + (1 + Block \mid ID)$

model 2: $RT \sim Block * Comparison + (1 + Block \mid ID)$

5.3 Results

A total of 3182 training and 1166 test trials were entered into the analysis, see Tables 5.1, 5.2 and 5.3 for details. Only correct trials were analysed and trials were exluded from analysis if the reaction time was above two standard deviations from the mean. There was no speed-accuracy trade off ($t(22) = 1.60,\ p = .12$).

Table 5.1. Trial statistics (trials were excluded from analysis if the reaction time was above two standard deviations from the mean).

	Total	Response Trials	Correct	Outliers	Analysed
Training	6048	4032	3183	1	3182
Test	1680	1363	1169	3	1166

Table 5.2. Mean accuracy (proportion of correct response) and standard deviation in seven blocks of training (upper section) and test (lower section of the table).

	Block							
	1	2	3	4	5	6	7	Overall
Training								
Mean	0.82	0.83	0.85	0.81	0.89	0.82	0.84	0.83
SD	0.38	0.37	0.36	0.39	0.32	0.38	0.37	0.37
Test								
Mean	0.89	0.91	0.89	0.92	0.89	0.89	0.90	0.90
SD	0.31	0.28	0.31	0.26	0.31	0.31	0.30	0.30

5.3.1 Training

A growth curve analysis revealed a decrease in the mean reaction time over the course of the training blocks (see Figure 5.4a). In addition, including the fixed effect of *Comparison* with *same* judgements as a baseline improved the model fit ($\chi^2(9) = 36.32,\ p < .001$, Table 5.4a). Participants were faster for *same* responses when compared to *different* responses. In addition, with increased training, *between-category* comparisons were faster than *within-category* comparisons (see Table 5.4a).

5.3.2 Test

In the first step, the growth curve analysis included all four types of comparisons (*same, within-category, between-categories* and *borderline* comparison). Figure 5.4b depicts mean reaction times for the four types of comparisons

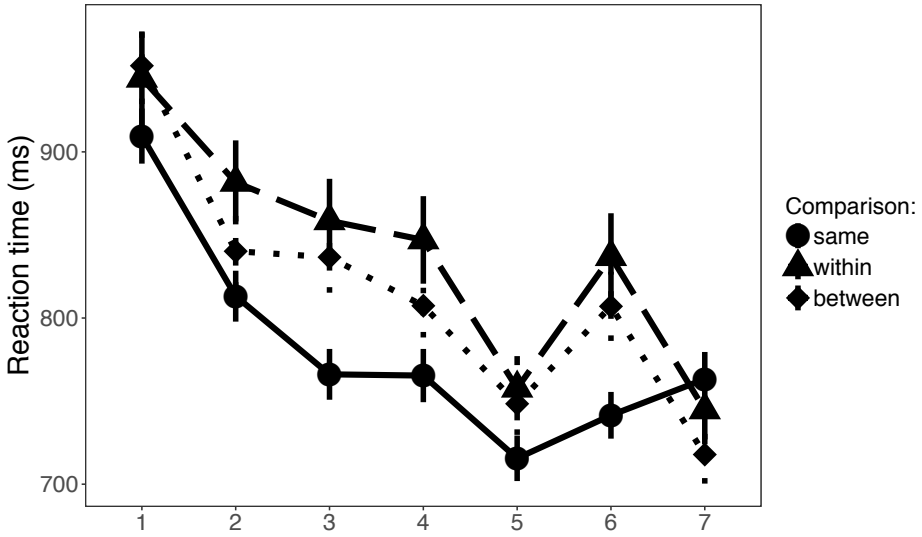
Table 5.3. Descriptive statistics for reaction time in training and test (in milliseconds).

	Mean	SD	min	max
Training	800.42	241.19	337.80	1515.20
Test	801.78	229.19	305.39	1498.47

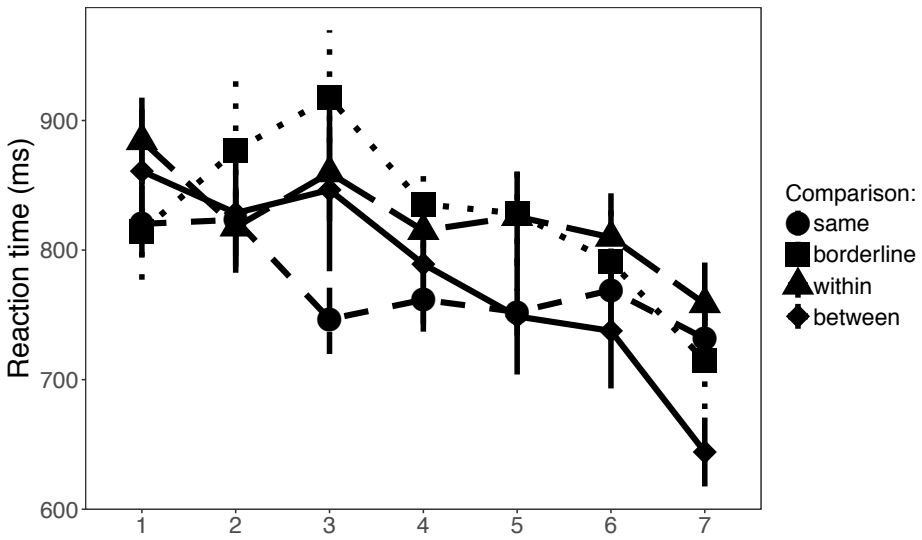
over the course of the seven blocks. The analysis confirmed that adding a fixed effect of *Comparison* term to the model significantly improved model fit ($\chi^2(9) = 18.67,\ p < .001$). Model parameters are given in Table 5.4b. The significant intercept term shows that mean reaction time in the initial block was unsurprisingly not zero. A significant Block term shows that there is a significant decrease in the reaction time over the course of the experiment. The significant effect of *Comparison* is revealed by effects of comparison types.

In the second step, the analysis focused on directly contrasting the two main conditions of interest: reaction times to *within-* and *between-categories* judgements. The growth curve analysis revealed that adding *Comparison* to the model significantly improved model fit, showing that the *within-category* judgements were overall slower than the *between-category* judgements ($\chi^2(1) = 5.68,\ p < .05$). To explore whether there is a change in the magnitude of this difference over the course of 7 blocks, another model including an interaction term *Block* $\times$ *Comparison* was tested. The results revealed that the interaction term marginally improved model fit ($\chi^2(1) = 3.33,\ p = .068$, see Figure 5.4c and Table 5.4c).

(a) Training blocks



(b) Test block: all comparison types



(c) Test blocks: only *within-* and *between-category* comparisons

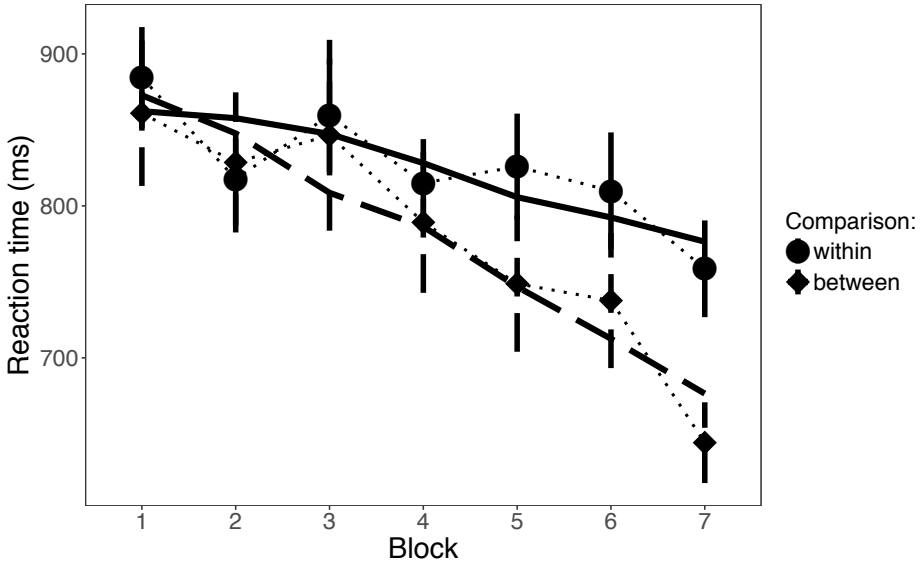


Figure 5.4. Mean reaction time over the course of the seven blocks for different comparisons in training blocks (a) and in test blocks (b and c), (error bars depict standard error).

Table 5.4. Growth Curve Analysis: Parameter estimates.

(a) **Training**

	Estimate	Std. Error	df	t value	p value
Intercept (*same*)	887.98	22.84	25.44	38.88	**0.00**
Block	-26.89	3.53	23.68	-7.62	**0.00**
Within Category	54.81	10.13	3132.92	5.41	**0.00**
Between Categories	38.05	8.84	3128.49	4.30	**0.00**

(b) **Test**: All comparisons

	Estimate	Std. Error	df	t value	p value
Intercept (*borderline*)	892.61	30.64	39.60	29.13	**0.00**
Block	-17.14	4.49	24.10	-3.81	**0.00**
Same	-52.22	17.60	1127.58	-2.97	**0.00**
Within Category	1.61	19.02	1127.74	0.08	0.93
Between Categories	-44.42	22.00	1126.89	-2.02	**0.04**

(c) **Test**: Within- and between-category comparisons

	Estimate	Std. Error	df	t value	p value
Intercept (*between*)	891.49	39.48	31.59	22.58	**0.00**
Block	-16.46	7.37	37.76	-2.23	**0.03**
Comparison	23.73	42.59	423.90	0.56	0.58
Block×Comparison	-17.32	9.48	424.39	-1.83	0.068

5.4 Discussion

The present experiment investigated implicit category learning in adults. Participants were presented with a series of items in succession and, for each cued item, asked to judge whether it was identical to or different from the previous item. This design enabled creating a learning situation in which participants were unaware of the categorical nature of the stimuli or the aim of the experiment, thus resembling experiments with infants. Debriefing session revealed that indeed no participants were aware of the two categories.

The main goal of this study was to explore the implicit category learning in adults. More specifically, the aim was to test whether adults are sensitive to statistical regularities in the input and whether they organise items into categories using these regularities when no explicit instructions were given, as infant do. Our hypothesis was that, if the task engaged implicit categorisation processes, increased experience with category exemplars would result in altering category space, making items from the same category look more similar, and items from different categories more dissimilar. Therefore, we expected that increased training would make *between-category* judgements easier, and therefore faster, as compared to *within-category* judgements.

The results confirmed this hypothesis and revealed that increased exposure to exemplars from different categories leads to faster judgements, as revealed by a decrease in the mean reaction time over the course of the seven blocks. The increase in speed was stronger in responding to pairs coming from different categories, as compared to when the items were members of

the same category. This suggests that items from different categories become easier to distinguish with more categorisation experience. This is consistent with the existing research suggesting that category learning alters object processing as objects grouped in the same category become more similar, whereas objects organised into different categories become increasingly dissimilar (Goldstone et al., 2001).

Further evidence for the role of categorisation in altering object representations comes from the results of the test trials. The results revealed slower reaction times to *within-category* judgements as compared to *between-category* judgements. Items in the test phase were designed such that perceptual similarity is identical for both *within-* and *between-category* judgements. The only difference was in the category membership of the items. Therefore, the observed effects were not driven by the visual similarity (since it was kept constant), but were likely to result from the categorisation process.

There are two main contributions of the present study. Firstly, it demonstrated it is possible to use the same visual categories to explore the process of category learning in both infants and adults. Secondly, by implementing a novel task in which participants were unaware of the categorical nature of the stimuli, we created a learning situation analogous to the one in studies with infants. We believe that this represents the first step towards understanding implicit category learning mechanisms over the course of the human development. In combination with insights provided by computational models of categorisation, this provides a promising route for further research.

The present study suggests that categorisation may affect object perception, making items from the same category more similar and items from different categories more dissimilar. Future studies should explore in which way implicit categorisation expertise modulates the perceived visual similarity among items. One possible route would be to obtain similarity ratings from participants prior to and following the implicit category learning task. To better understand what is driving the changes in perceived similarity, further research should explore whether these effects are accompanied by modulations in object processing on a neural level. One aspect of this question was tackled in an experiment that explored the neural signatures of categorisation presented in the following chapter.

Taken together with findings of Experiment 1, the results of the current study demonstrate that infants and adults showed the ability to compute statistical regularities from a stream of visually presented items and further use this information as a basis to spontaneously group objects into categories. Further studies are needed to determine which aspects of the learning process are altered by increasing cognitive capacities and how these interact with the learning process.

Chapter 6

Neurophysiological signatures of category learning

6.1 Introduction

Chapter 5 demonstrated that adults are able to extract statistical regularities from a stream of visually presented items and organise these items into categories. Reaction time measures showed that within-category discrimination becomes more difficult than between-category discrimination, indicating that category learning alters object representations making objects grouped in the same category become more similar, whereas objects organised into different categories become increasingly dissimilar (also see Goldstone et al. 2001; Sigala et al. 2002).

Different experimental paradigms and various behavioural measures, such as similarity ratings, reaction times and looking preferences are used to investigate characteristics of category learning. In infants, for instance, more

experience with categories is reflected by changes in their looking preferences. In adults, a decrease in time needed to respond in behavioural tasks is modulated by category familiarity. In this experiment we use the event-related potential technique (thereafter, ERP) to investigate how such an organisation alters the neural signatures of object processing, by comparing object processing before and after extensive category learning.

Existing ERP studies on categorisation suggest that the same object is processed differently depending on the level of categorisation. For instance, Tanaka et al. (1999) compared ERPs to the same items when categorised on subordinate (e.g. Labrador), basic (e.g. dog) and superordinate levels (e.g. animal). They found that subordinate categorisation elicited enhanced early negative deflection (N1) in posterior sites as compared to the basic-level categorisation. In contrast, superordinate-level categorisation elicited more negative potential in the frontal region when compared to the basic-level categorisation. Kiefer (2001) investigated differences between categorising items presented as images and as written words. He presented participants with images or names of items from natural categories (such as animals and plants) and artificial categories (e.g. furniture and tools). Participants were presented with a category name followed by a target object. The target object was presented either as a picture or a word. Participants' task was to judge whether the target object was a member of the previously presented category or not. Kiefer (2001) found category-specific differences in perceptual processing around 160-200 ms in the visual modality. In addition to the

early effects, he found that category-specific effects were also present at a later stage (300-500 ms) and were similar across modalities.

Curran et al. (2002) compared effects of categorisation and recognition memory on ERPs using novel categories. In this study participants learned to categorise a set of visual shapes and then performed categorisation and recognition tasks involving these categories. Curran et al. (2002) found that early visual processing is sensitive to category membership, whereas the later window (300-500 ms) was sensitive to both category membership and recognition, suggesting a temporal transition from processes sensitive to categorisation to later processes sensitive to recognition and discrimination. However, in this study participants' EEG was recorded only after an extensive training has been completed, therefore providing no information about how the training itself modulated ERPs.

The existing research has demonstrated that categorisation shapes objects processing at the neural level. However, most studies investigating neural markers of categorisation have focused on existing, real world categories, or on the outcomes of the extensive categorisation training. In contrast, not much is known about how the process of category formation modulates ERP signatures of the to-be-categorised objects. The aim of the present study is to fill this gap and explore neural signatures of the ongoing category formation process. To achieve this, an EEG signal was recorded while participants performed a category learning task. The learning task was interleaved with test blocks in which no feedback was provided. Based on the results of Ex-

periment 8, our hypothesis was that increasing categorisation expertise will modulate behavioural responses, with more training resulting in faster reaction times to within-category priming as compared to between-category priming. Moreover, we expect that these behavioural effects to be accompanied by modulations in the event-related neural activity, in particular the early ERP components sensitive to categorisation processes (Curran et al., 2002; Kiefer, 2001). To minimise potential effects of low-level perceptual characteristics on the EEG signal, simple drawings of novel creatures were used. The stimuli were created to resemble a set that was used in several categorisation studies with infants (Mather & Plunkett, 2011; Plunkett et al., 2008; Younger & Cohen, 1986).

6.2 Method

6.2.1 Participants

Twenty eight participants took part in this study. All participants were undergraduate students at the University of Oxford and received course credit for their participation. All participants were right handed, had normal or corrected-to-normal vision and no known neurological impairments. The study was approved by the University of Oxford Medical Sciences Ethics Committee. Data files containing data from two participants were damaged, therefore unusable, resulting in 26 participants entering the analysis.

6.2.2 Experimental design

Electroencephalography data (EEG) was recorded while participants learned to categorise novel objects into two categories. The study consisted of eight training blocks interleaved with eight test blocks.

Training

Training was designed as a supervised category learning task. Participants were instructed to imagine they had just landed on a new planet and their task is to learn to distinguish two alien species that inhabit the planet. One species is dangerous and should be avoided, while the other species is good and can be approached. This experimental design was adapted from a study by Lupyan et al. (2007).

A total of 24 alien exemplars were presented in each training block (12 from each category). Each trial started with a fixation cross for 500 ms, followed by an exemplar (i.e. an alien) from one of the two categories. After 800 ms, an agent (explorer) appeared on the left or on the right of the exemplar (position of the explorer was counterbalanced). Participants' task was to indicate as fast as possible whether the explorer should approach or avoid the alien by directing the explorer towards or away from it. Participants indicated their response by pressing one of two buttons on a response box. For example, if an explorer appeared on the left side of an alien that should be approached, they should press a button with an arrow pointing to the right, i.e towards the alien. An immediate auditory feedback followed the

response to indicate whether it was correct or not (a bell sound for correct and a beep sound for incorrect response). The explorer and alien remained on screen for an extra 1200 ms after the response was given (see Figure 6.1).

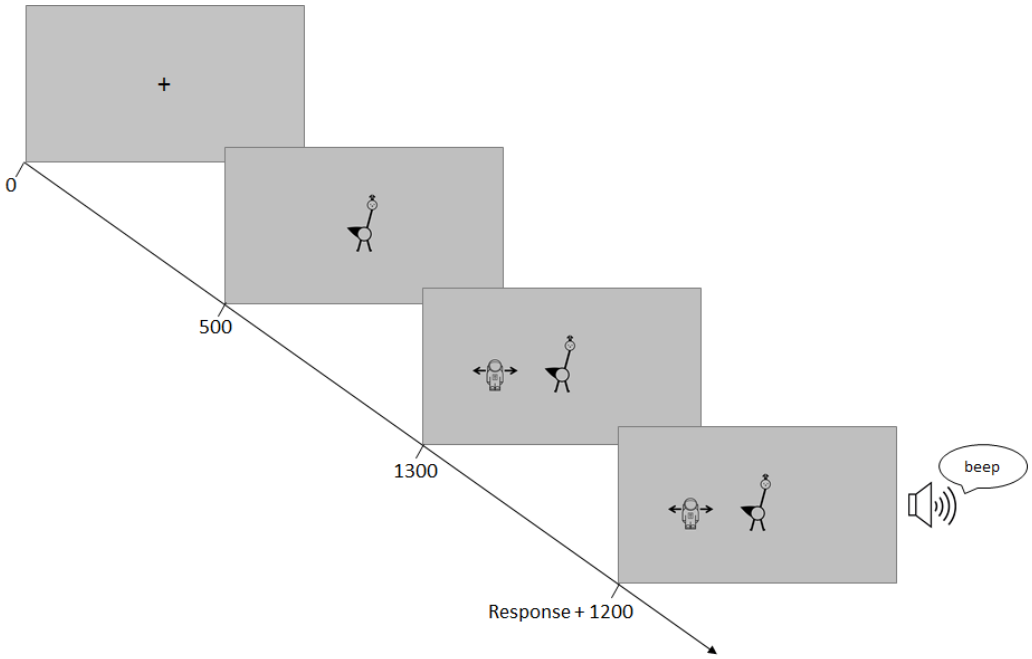


Figure 6.1. Training phase: Trial timeline.

Test

In the test trials, participants were presented with two aliens in succession and their task was to indicate whether they belonged to the same category (*within-category* condition) or were from different categories (*between-category* condition). The first alien was a prime and the second alien was a target. Each test block consisted of 36 trials presented in a random order. The same target was primed with both *within-* and *between-category* primes. 24 trials contained pairs of previously presented training exemplars (training items). The other 12 trials comprised of novel previously unseen exemplars (test items). There was an equal number of same and different trials. Fol-

lowing a fixation cross presented for 300 ms, the first alien (prime) appeared on the screen for 800 ms. Then, following an interstimulus interval (a blank screen for 200 ms), the second alien (target) was presented in the centre of the screen. After 800 ms, a prompt screen containing two words, "same" and "different", appeared (see Figure 6.2). To prevent preparatory motor response activity, the location of the response choices was randomly assigned to left and right positions. Response was then given on a response box by pressing a button alligned with one of the two response words. No feedback was provided for test trials. The procedure was adapted from Kietzmann et al. (2016).

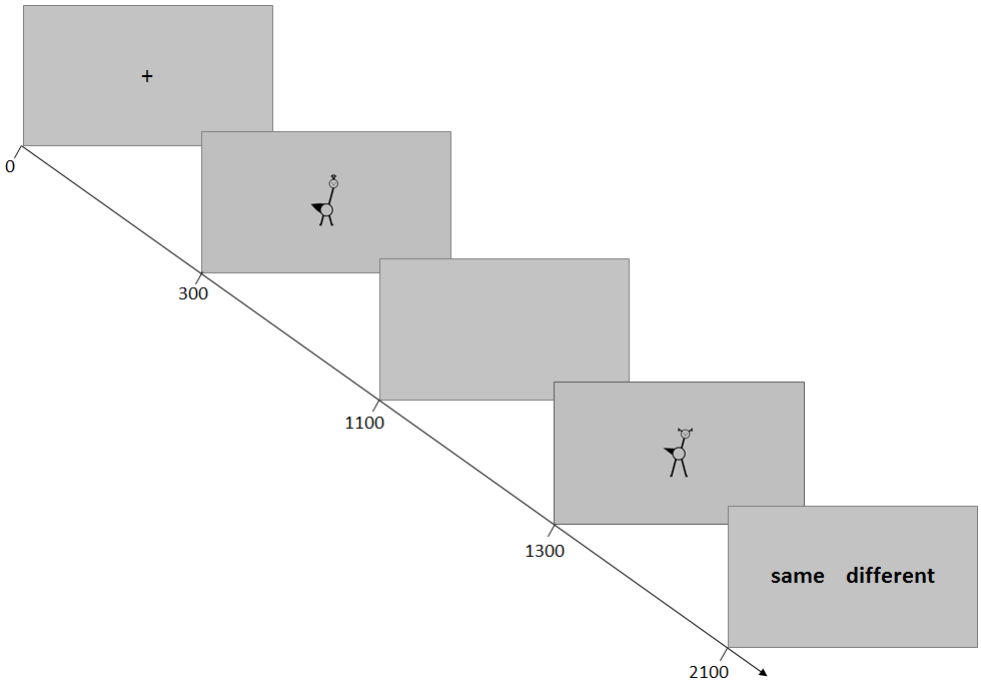


Figure 6.2. Test phase: Trial timeline.

6.2.3 Stimuli

The stimulus set consisted of simple line drawings of novel animal-like creatures that resembled the stimuli used in several studies exploring category formation in infants (Mather & Plunkett, 2011; Plunkett et al., 2008; Younger & Cohen, 1986). Each item had four features: legs of varying length, neck of varying length, ears varying in their separation and a tail of varying thickness. The values of one feature were predictive for values of other features. For instance, items with short legs always had long necks, while thick tails always accompanied spread ears. As in the experiments with infants presented in previous chapters of this thesis, each feature could take one of seven potential values, as opposed to the original studies which had used a 5-point range. Examples of stimuli are given in Figure 6.3a and the structure of all stimuli is given in Appendix D.

A total of 24 training items were created (12 from each of the two categories). The structure of each stimulus can be described with four numbers, each number representing a level of a particular feature: the first number stands for the length of legs, the second for the neck size, the third for the ear separation and the fourth for the tail thickness. Four additional test items were created and presented only in test blocks (see Figure 6.3b). All stimuli were presented against a 5% grey background. A drawing of a space explorer was presented to participants in training trials (see Figure 6.1). Two auditory stimuli (a beep for incorrect responses and a bell sound for correct responses) were presented to provide feedback to participants. Both stimuli

were stereo, 500 ms long and matched in pitch.

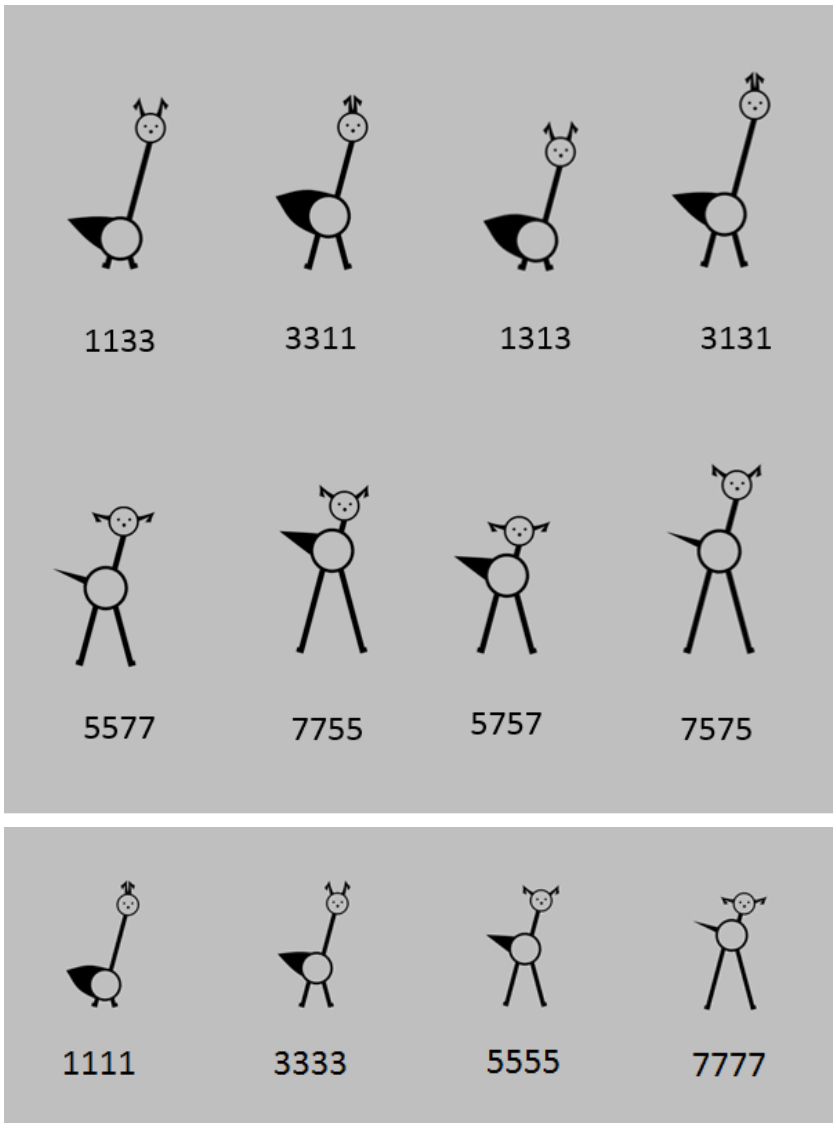


Figure 6.3. Stimuli used in the training phase (a) and in the test phase (b), (numbers below each item represent the value of each for the four features).

6.2.4 Equipment

Visual stimuli were presented centrally on a 32-inch LCD monitor. Audio stimuli were presented via two speakers positioned on the left and right of the screen. Stimulus presentation was done with Presentation (Version 18.3, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). Partici-

pants responded by pressing one of two buttons on a response box (Model RB-540, Cedrus Inc., San Pedro, CA, www.cedrus.com).

For details about the EEG signal recording setup see Chapter 2 and Figure 6.4.

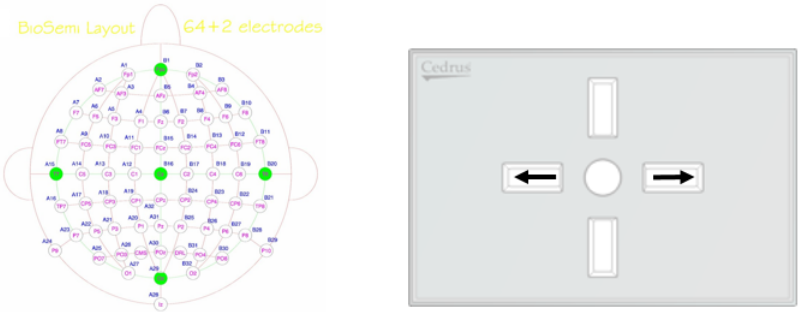


Figure 6.4. BioSemi ActiveTwo 64-channel position outline used for data acquisition (retrieved from *BioSemi* (n.d.)) and response box used to record behavioural responses (retrieved from *Cedrus* (n.d.)).

6.2.5 Experimental procedure

Upon arrival, participants were explained the procedure and signed the consent form. The experiment was conducted in an electrically shielded and sound-attenuated test booth. Participants were instructed to minimise blinks and movements. After the EEG cap with electrodes was placed and impedances checked, to familiarise participants with the task, the study started with a warm-up block that was not used in the analysis. This block had an identical structure as the main study, consisting of a training and a test block, though participants were asked to categorise exemplars from two existing categories, "fruit" and "vegetable". After the warm-up block, impedances were checked once again and the experiment started as soon as participants were ready.

Short breaks were given after each test block.

6.3 Experimental measures

6.3.1 Behavioural measures

Reaction time and accuracy measures were recorded for participants' responses in each training and test trial. Trials with reaction times outside the 350-4000 ms range were excluded from the analysis. In addition, reaction times higher than three standard deviations above the mean were also excluded.

6.3.2 EEG measures

The EEG signal was acquired at 2048 Hz and processed using custom made Matlab routines and the eeglab toolbox, version 14-1-1b (Delorme & Makeig, 2004). Signals were first filtered using zero-phase filters with a high-pass filter at 0.5 Hz and a low-pass filter at 30 Hz. Bad channels (kurtosis measure) were removed and interpolated using neighbouring channels by spherical interpolation. Channels were then re-referenced to the left mastoid. The signals were epoched from the onset of visual presentation for a duration of 800 ms. Each epoch was baseline corrected with respect to the average signal in the 200 ms prior to the visual presentation. Epochs with artifacts were removed from analysis as follows:

1. Peak-to-peak amplitude larger than 200 μV;

2. Absolute amplitude larger than 100 μV;

3. Blinks as defined by a difference in voltage in EOG channels (Left EOG Low - Left EOG High) larger than 75 μV.

In addition, participants who contributed less than ten out of the 24 trials in each training block were excluded from the analysis. To be included in the analysis of the test blocks, at least half of participant's 12 trials had to be valid (i.e. containing no artifacts) in a block.

6.4 Analysis plan

6.4.1 Behavioural analysis plan

Response accuracy

Mean accuracy was calculated for each block for each participant. To determine whether there was a change in accuracy over the course of the experiment, a growth curve analysis was conducted (Mirman, 2014) in R using the lme4 and lmerTest packages (version 1.1-12 and 2.0-33, respectively). Mean accuracy was modelled using a linear growth model with a fixed effect of *Block* (1-8) and random effects of *Participant* on the intercept and slope terms. The fixed effect of *Block* was added to the base model and the effect was evaluated using model comparisons. An improvement in the model fit when the effect of *Block* was added would indicate a significant change in accuracy over the course of the experiment.

Reaction time

Mean reaction time was calculated for each block for each participant. Only trials with correct responses were included in the analysis of the reaction time. To determine whether there was a change in the response speed over the course of the experiment, the mean reaction time was modelled using a linear growth model with a fixed effect of *Block* (1-8) and random effects of *Participant* on the intercept and slope terms. The fixed effect of *Block* was added to the base model and the effect was evaluated using model comparisons.

Mean reaction times in test blocks were also analysed using growth curve analysis. To test whether judgement type influenced the response speed, a fixed effect of *Comparison Type* (*within-category* and *between-category*) was added to the model. Novel items were analysed separately using the same approach.

Model specifications for the growth curve analysis were as follows (model 2 was tested only for analysis of reaction times in the test):

base model: $RT \sim 1 + (1 + Block \mid ID)$

model 1: $RT \sim 1 + Block + (1 + Block \mid ID)$

model 2: $RT \sim 1 + Block + Comparison\ Type + (1 + Block \mid ID)$

6.4.2 EEG analysis plan

Signals acquired from 64 electrode sites were grouped in six regions of interest. The scalp was divided into three bands with coronal orientation, each of which was subdivided by hemisphere (see Figure 6.5).

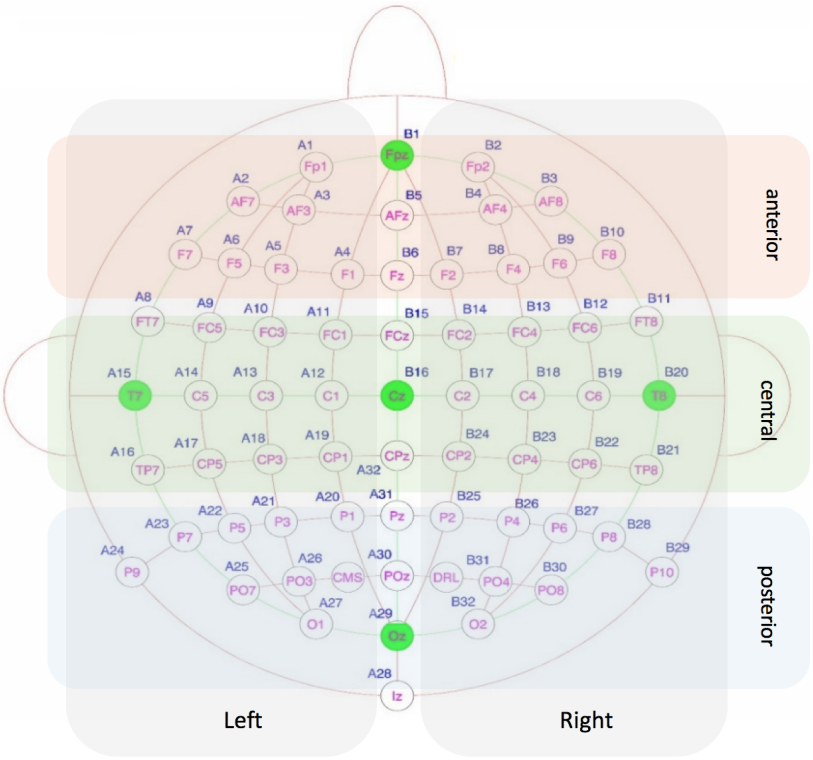


Figure 6.5. Electrode sites included in the analyses: three bands with coronal orientation subdivided by hemisphere: anterior left (Fp1, AF3, AF7, F1, F3, F5, F7), anterior right (Fp2, AF4, AF8, F2, F4, F6, F8), central left (FC1, FC3, FC5, FT7, C1, C3, C5, T7, CP1, CP3, CP5, TP7), central right (FC2, FC4, FC6, FT8, C2, C4, C6, T8, CP2, CP4, CP6, TP8), posterior left (P1, P3, P5, P7, P9, PO3, PO7, O1), and posterior right (P2, P4, P6, P8, P10, PO4, O2).

Analyses were performed on mean amplitude ERP values calculated over fixed time windows. The three time windows of interest were (also shown in Figure 6.6)):

1. early effects (N200-like, 150-250 ms post-stimulus);

2. middle latency effects (P300 window, evoked positivity potential peaking around 300 ms post-stimulus, 250-400 ms post-stimulus);

3. late effects (400-600 ms post-stimulus).

Training: To test for the effect of extensive category training on event-related potentials, mean ERPs in the first two blocks were averaged and compared to the averaged ERPs of the final two blocks. Thus, the *initial* and the *final* phase were contrasted. Mean ERP amplitude in each time window was submitted to a repeated-measures ANOVA with factors *Phase* (*initial* and *final*) x *Band* (*anterior, central* and *posterior*) x *Hemisphere* (*left* and *right*). Greenhouse-Geisser corrections were applied where required and Bonferroni corrections to p-values were implemented for post-hoc tests.

Test: For test blocks, initial analysis included both priming conditions, whereas subsequent analysis was conducted separately for *within-* and *between-category* priming. It is important to note that the same target was primed with both *within-* and *between-category* prime. We compared ERPs to targets primed by items from the same category (*within-category* priming). Then, the same analysis was conducted for primes from a different category (*between-category* priming). The main aim of the analyses was to determine whether there is a difference in processing identical target items when they are primed with items from the same or with primes from another category. A key question was whether these effects are dependent on the level of expertise, i.e. the amount of exposure in training. For each window of interest, two sets of analyses were conducted. Initially, mean ERP amplitudes

were submitted to a repeated-measures ANOVA with within-subjects factors *Condition* (*within* and *between*), *Phase* (*initial* and *final*), *Band* (*anterior, central* and *posterior*) and *Hemisphere* (*right* and *left*). In the second step, mean ERP amplitudes were submitted to a repeated-measures ANOVA with within-subjects factors *Phase* (*initial* and *final*), *Band* (*anterior, central* and *posterior*) and *Hemisphere* (*right* and *left*), separately for *within-category* and *between-category* priming conditions.

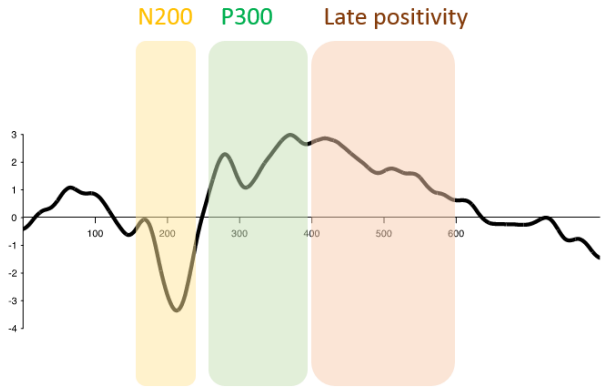


Figure 6.6. Mean amplitude was analysed for three time windows: N200, P300 and late positivity time window.

6.5 Results: Behavioural data

A total of 7449 trials were entered into the analysis (out of the 7776 that had been run overall). The mean reaction time was 737.11 ms (SD = 354.60). On average, each participant contributed 34.65 trials per block ($M = 34.65,\ SD = 1.79,\ range : 27 - 36$).

6.5.1 Training

Response accuracy

The mean accuracy was calculated for each of the eight blocks for every participant. The accuracy was high across all blocks. The growth curve analysis revealed that there was no change in the accuracy over the course of the training blocks (Figure 6.7a).

Reaction time

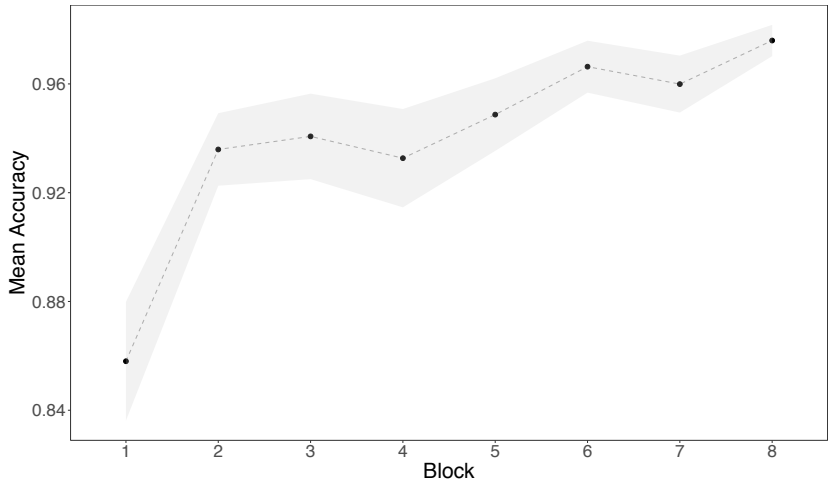
Only the trials with the correct responses were included in the analysis (4712 out of 5016 trials). The mean reaction time was calculated for each training block. As depicted in Figure 6.7b, the mean reaction time gradually decreased over the course of the training blocks.

The growth curve analysis revealed that this decrease was significant, therefore suggesting that participants had become faster in making categorisation judgements with the training ($\chi^2(1) = 15.40,\ p < .001$, see Table 6.1). The significant intercept term shows that the mean reaction time in the initial block was non zero. The significant slope term shows that there is

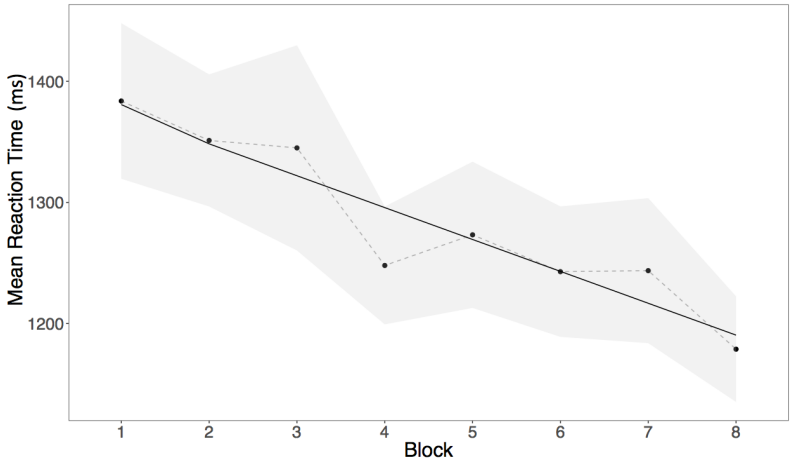
a significant decrease in the reaction time over the course of the experiment. The reaction times decreased by, on average, 26.46 ms per trial ($SE = 5.89$).

Table 6.1. The growth curve analysis on the mean reaction time in training.

	Estimate	Std. Error	df	t value	p value
Intercept	1407.23	57.63	27.90	24.42	**0.00**
Slope (Block)	-26.46	5.89	125.98	-4.50	**0.00**



(a)



(b)

Figure 6.7. Mean accuracy (a) and mean reaction time (b) across participants over the course of training blocks (the dashed line depicts the mean accuracy in each trial; the solid line represents the model prediction; grey envelope depicts standard error).

6.5.2 Test

Response accuracy

Participants were highly accurate in the test blocks. Overall, the mean accuracy was 0.855 (SD = 0.077). The mean accuracy in each block is shown in Table 6.2. The mean accuracy for *within-category* judgements was .87 (SD = .06) and .84 (SD = .089) for *between-category* judgements. The growth curve analysis showed that there was no change in the accuracy rates over the course of the test blocks.

Table 6.2. Mean accuracy in test blocks.

Block	Accuracy	SD
1	0.84	0.37
2	0.84	0.36
3	0.86	0.35
4	0.86	0.35
5	0.86	0.35
6	0.87	0.33
7	0.86	0.35
8	0.85	0.35

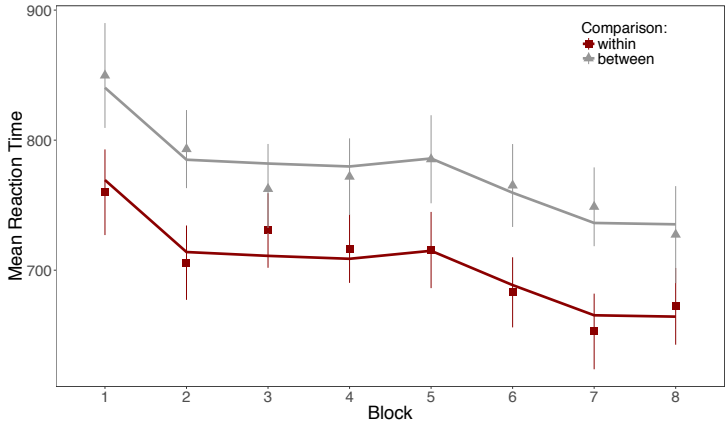
Reaction time

The test trials consisted of items previously presented in the training and a new set of items, included to test whether there are generalisation effects. The familiar and novel test items were analysed in separate analyses.

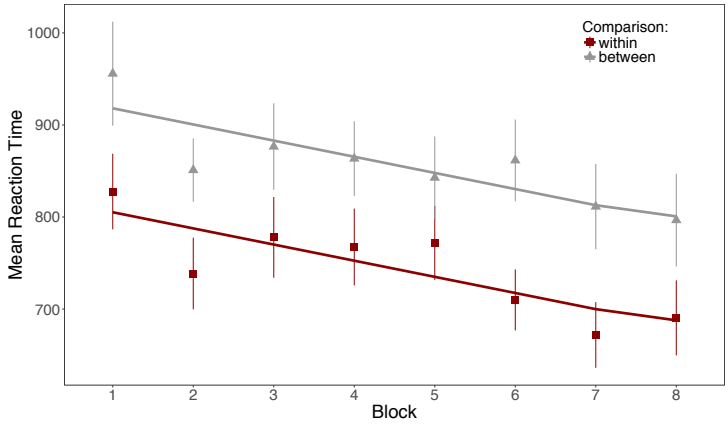
The growth curve analysis tested whether there is a change in the speed over the course of training and whether this is affected by a prime type, that is, whether there is a difference between the *within-* and *between-category* judgements. The analysis for the familiar items revealed that there is a

decrease in the mean reaction time and that the participants made faster responses for the *within-category* judgements (see Figure 6.8a; model estimates are given in Table 6.3a). The significant intercept term showed that the mean reaction time in the initial block was not zero. The significant effect of *Block* showed that there was a decrease in the reaction time over the course of the experiment. The significant effect of *Prime* shows that this was dependent on the type of comparison. No interaction was found.

A separate analysis was conducted with the novel-item test trials. An identical pattern of results was observed: there was a decrease in the mean reaction times over the course of the eight blocks and participants responded faster when items were primed with within-category members as compared to when the prime was from another category (see Figure 6.8b. Model parameter estimates are given in Table 6.3b). The significant intercept term shows that the mean reaction time in the initial block was not zero. The *Block* term shows that there was a decrease in the reaction time over the course of the experiment and the effect of *Prime* shows that this was dependent on the type of comparison.



(a)



(b)

Figure 6.8. Mean reaction time for trained-item test trials (a) and for novel-item test trials (b), (solid line represents model prediction, error bars depict standard error).

Table 6.3. Growth curve analysis on mean reaction time across test blocks.

(a) **All test trials (including novel items)**

	Estimate	Std. Error	df	t value	p value
Intercept	729.66	25.95	39.72	28.12	**.00**
Block	-10.13	2.17	402.89	-4.67	**.00**
Prime	70.55	9.90	402.83	7.13	**.00**

(b) **Novel-item test trials**

	Estimate	Std. Error	df	t value	p value
Intercept	759.03	40.96	30.40	18.53	**.00**
Block	-13.87	5.38	27.29	-2.58	**.02**
PrimeType	167.95	19.64	374.45	8.55	**.00**

6.6 Results: Event-related potentials

6.6.1 Training

The mean number of trials in the initial block was 40.52 ($SD = 5.89$, range: 26-48) and 39.96 ($SD = 7.3$, range: 17-47) in the final block.

1. Early effects (N200)

The analysis revealed that there were significant effects of *Phase* and *Band* (Phase: $F(1, 25) = 5.85$, $p = .023$, $\eta_g^2 = .19$; Band: $F(1.13, 28.27) = 5.79$, $p = .020$, $\eta_g^2 = .19$). In addition, there was a significant *Phase* $\times$ *Hemisphere* interaction ($F(1, 25) = 9.90$, $p = .004$, $\eta_g^2 = .28$). Averaged ERP waves, mean amplitudes and a difference topomap are given in Figure 6.9. The post-hoc paired t-test with Bonferroni corrections revealed that these effects were driven by the difference in the right hemisphere ($p = .004$), whereas no difference was present in the left hemisphere ($p = .12$). This suggests that the categorisation training led to a decrease in the early negative deflection in the right hemisphere in the visual event-related potentials.

2. Middle latency effects (P300)

The analysis revealed significant effects of *Phase*, *Band* and *Hemisphere* and a *Phase* $\times$ *Band* interaction (Phase: $F(1, 25) = 7.51$, $p = .011$, $\eta_g^2 = .23$; Band: $F(1.19, 29.80) = 10.76$, $p = .002$, $\eta_g^2 = .30$; Hemisphere: $F(1, 25) = 5.22$, $p = .031$, $\eta_g^2 = .17$; interaction Phase x Band: $F(1.15, 28.87) = 12.99$, $p < .001$, $\eta_g^2 = .34$, see Figure 6.10). The post-hoc t-tests with the Bonferroni corrected p-values revealed that these effects were present in

the anterior ($p = .002$) and central ($p = .004$), but not in the posterior region ($p = .48$). Therefore, categorisation training resulted in enhanced positivity in the P300-like window in frontal and central sites.

3. Late effects

The analysis of amplitudes in the late time window revealed a significant main effect of *Hemisphere* and a significant *Phase* × *Band* interaction (Hemisphere: $F(1, 25) = 10.21$, $p = .004$, $\eta_g^2 = .29$; interaction: $F(1.25, 31.23) = 4.17$, $p = .042$, $\eta_g^2 = .14$). The post-hoc analysis revealed that the difference between initial and final block was significant in posterior ($p = .022$), but not in other regions (anterior: p=.64, central $p = .66$), with categorisation training resulting in decreased positivity in posterior sites. For more details see Figure 6.11.

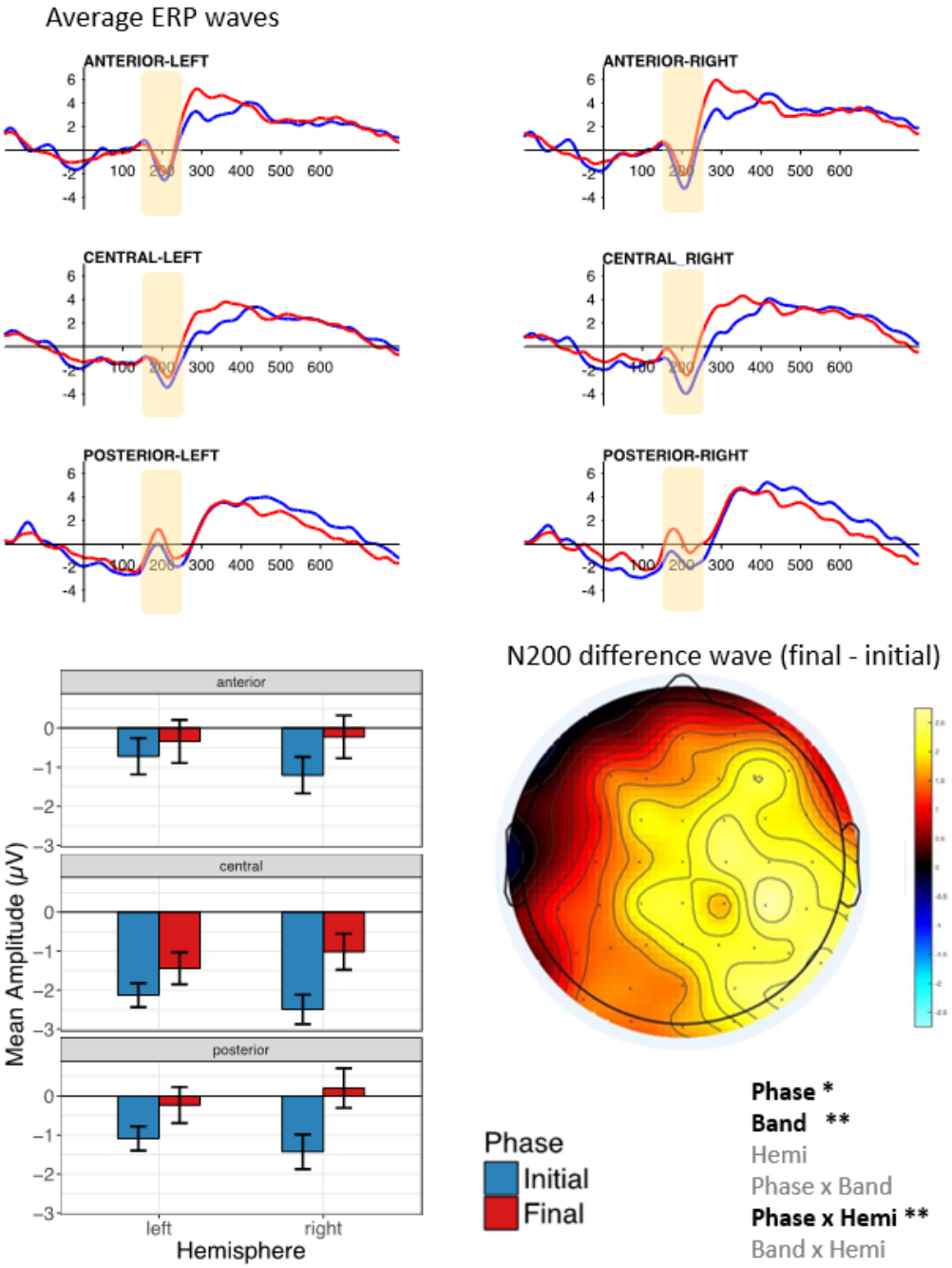


Figure 6.9. N200 amplitudes in training blocks. (*) $p < .07$; * $p < .05$; ** $p < .01$.

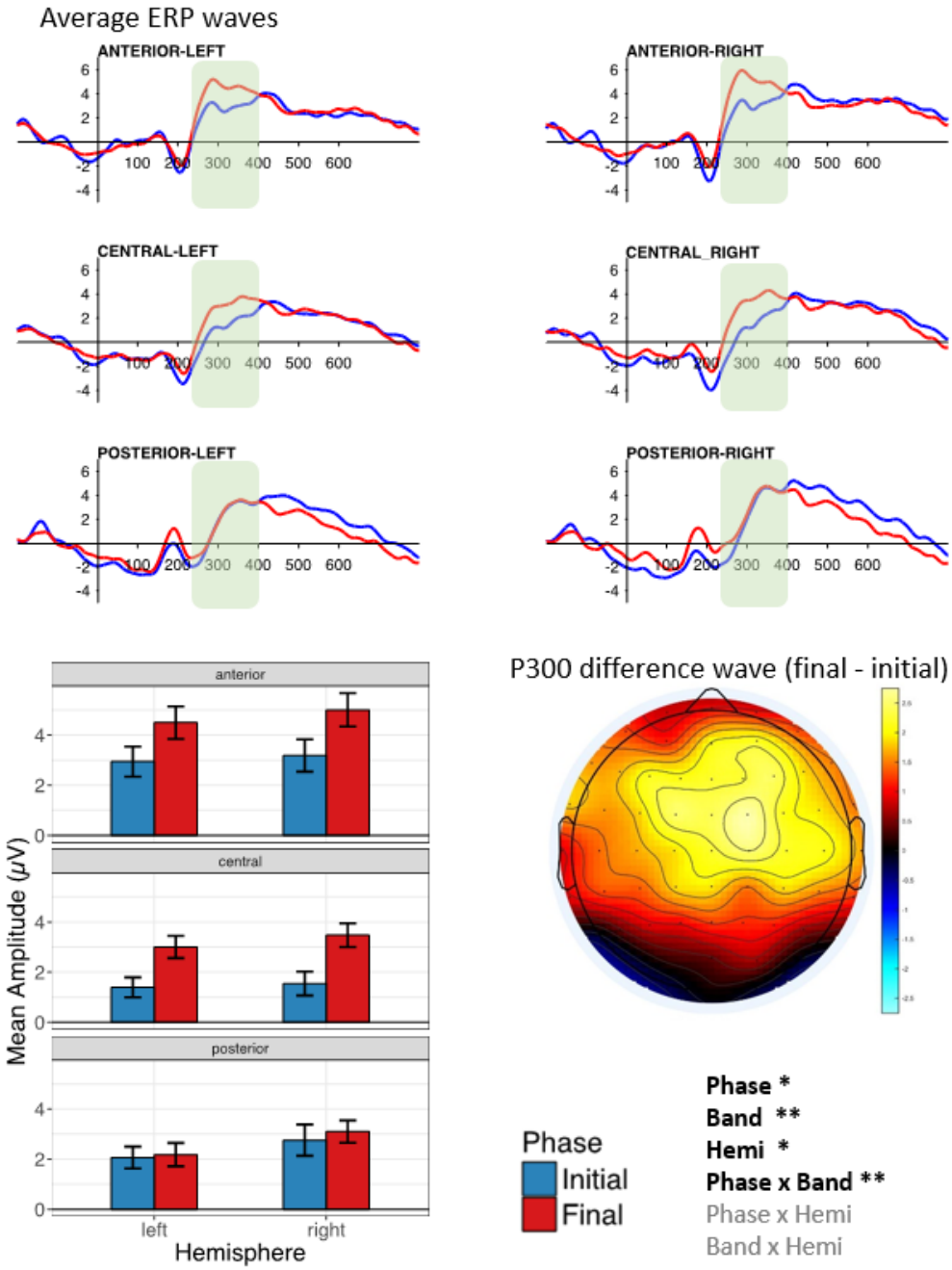


Figure 6.10. P300 amplitudes in training blocks. (*) $p < .07$; * $p < .05$; ** $p < .01$.

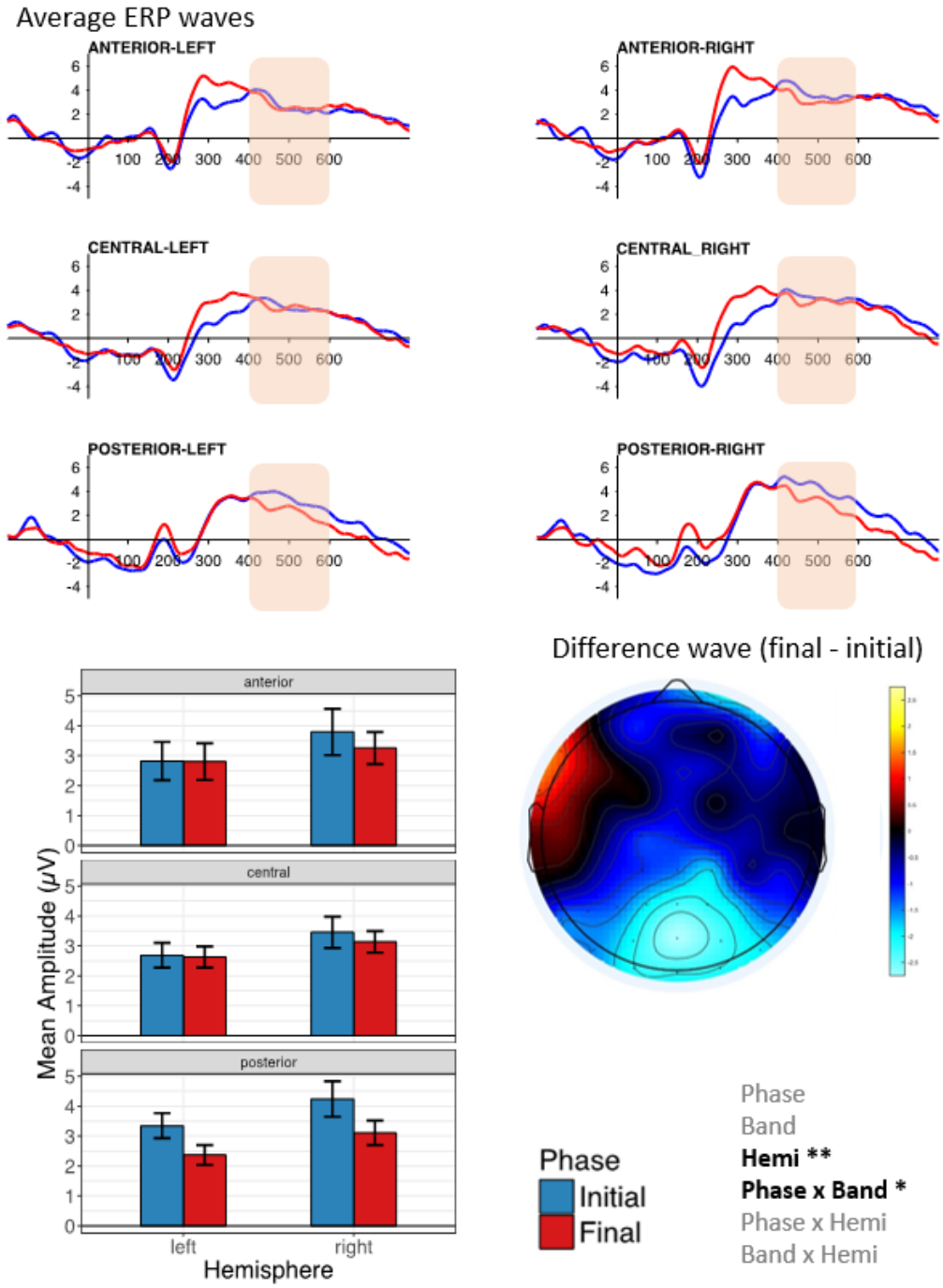


Figure 6.11. late positivitu: Mean amplitudes in the training blocks. (*) $p < .07$; * $p < .05$; ** $p < .01$.

6.6.2 Test

Twenty participants were included in the analysis (six were excluded due to accumulating less than six trials per block).[1] The mean number of trials in the initial test block was 19.91 ($SD = 4$, range: 10-2) and 20.64 ($SD = 3.79$, range: 11-24) in the final test block.

Analysis including both priming conditions

In the initial step, mean ERP amplitudes were submitted to a repeated-measures ANOVA with within-subjects factors *Condition* (*within* and *between categories*), *Phase* (*initial* and *final*), *Band* (*anterior, central* and *posterior*) and *Hemisphere* (*right* and *left*) for each window of interest.

Early effects (N200). The analysis revealed a significant effect of *Band* ($F(1.33, 25.23) = 21.89,\ p < .001,\ \eta_p^2 = .54$) and *Condition* × *Band* ($F(1.11, 21.00) = 4.84,\ p = .036,\ \eta_p^2 = .20$) and and *Condition* × *Hemisphere* interactions ($F(1, 19) = 28.67,\ p < .001,\ \eta_p^2 = .60$).

Middle latency effects (P300). The analysis revealed significant effect of *Condition* ($F(1, 19) = 7.23,\ p = .015,\ \eta_p^2 = .28$) and *Condition* × *Band* interaction ($F(1.06, 20.12) = 14.97,\ p < .001,\ \eta_p^2 = .44$).

Late latency effects. The analysis revealed significant main effects of Band and Hemisphere (Band:$F(1.09, 20.76) = 37.08,\ p < .001,\ \eta_p^2 = .66$; Hemisphere: $F(1, 19) = 48.62,\ p < .001,\ \eta_p^2 = .72$), and *Condition* × *Band* ($F(1.12, 21.32) = 14.43,\ p < .001,\ \eta_p^2 = .43$) and *Condition* × *Phase* ×

[1] An identical analysis was run on all 26 participants and revealed the same pattern of results

Band interaction ($F(1.36, 25.86) = 4.23,\ p = .039,\ \eta_p^2 = .18$).

Within-category priming: early effects (N200)

The analysis revealed significant effects of *Band* and *Hemisphere* (Band: $F(1.28, 24.29) = 8.93,\ p = .004,\ \eta_g^2 = .32$; Hemisphere: $F(1, 19) = 10.47,\ p = .004,\ \eta_g^2 = .36$). No effect of *Phase* was found, suggesting that *within-category* priming had no effects on the early stages of visual ERPs. For more details see Figure 6.12.

Within-category priming: middle latency effects (P300)

The analysis revealed a near significant effect of *Phase* ($F(1, 19) = 3.92,\ p = .062,\ \eta_g^2 = .17$) and a *Band* $\times$ *Hemisphere* interaction ($F(2, 38) = 3.84,\ p = .030,\ \eta_g^2 = .17$). Post-hoc revealed that there was a significant effect of *Phase* only in the anterior band (p=0.048), but not in the central and the posterior bands (central: $p = 0.078$; posterior: $p = 0.22$, see Figure 6.13). Thus, the categorisation expertise had modulated *within-category* priming in the anterior region in the middle latency window.

Within-category priming: late effects

ANOVA on mean amplitudes in the late time window yielded significant main effects of *Hemisphere* and *Band* (Hemisphere: $F(1, 19) = 40.50,\ p < .001,\ \eta_g^2 = .03$; Band: $F(1.10, 20.86) = 23.88,\ p < .001,\ \eta_g^2 = .23$). No effects of Phase nor interactions were found (Figure 6.14), therefore suggesting within-category priming was not modulated by the categorisation expertise in the late window of the visual ERPs.

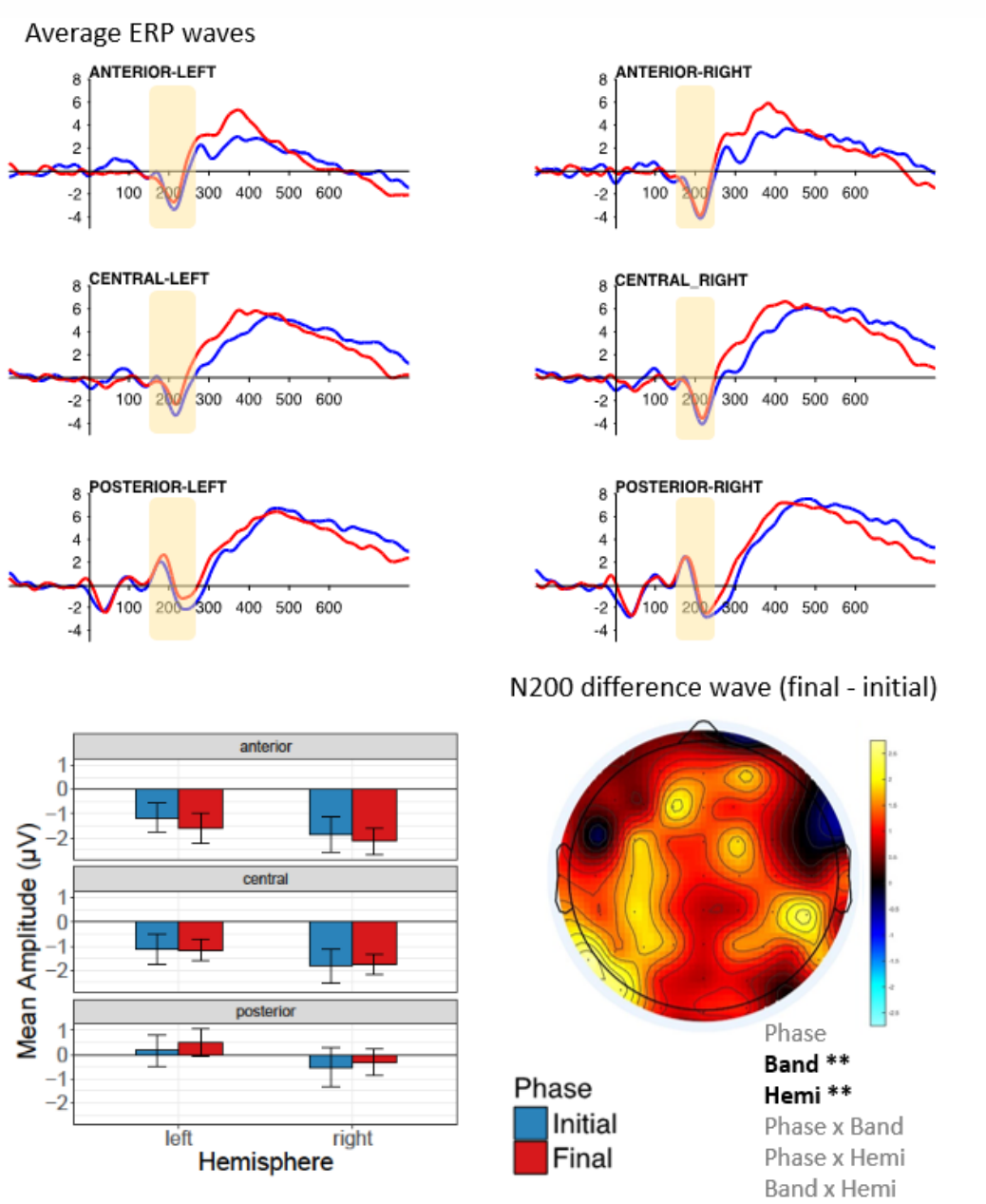


Figure 6.12. N200 amplitudes in within-category priming trials in test blocks. (*) $p < .07$; * $p < .05$; ** $p < .01$.

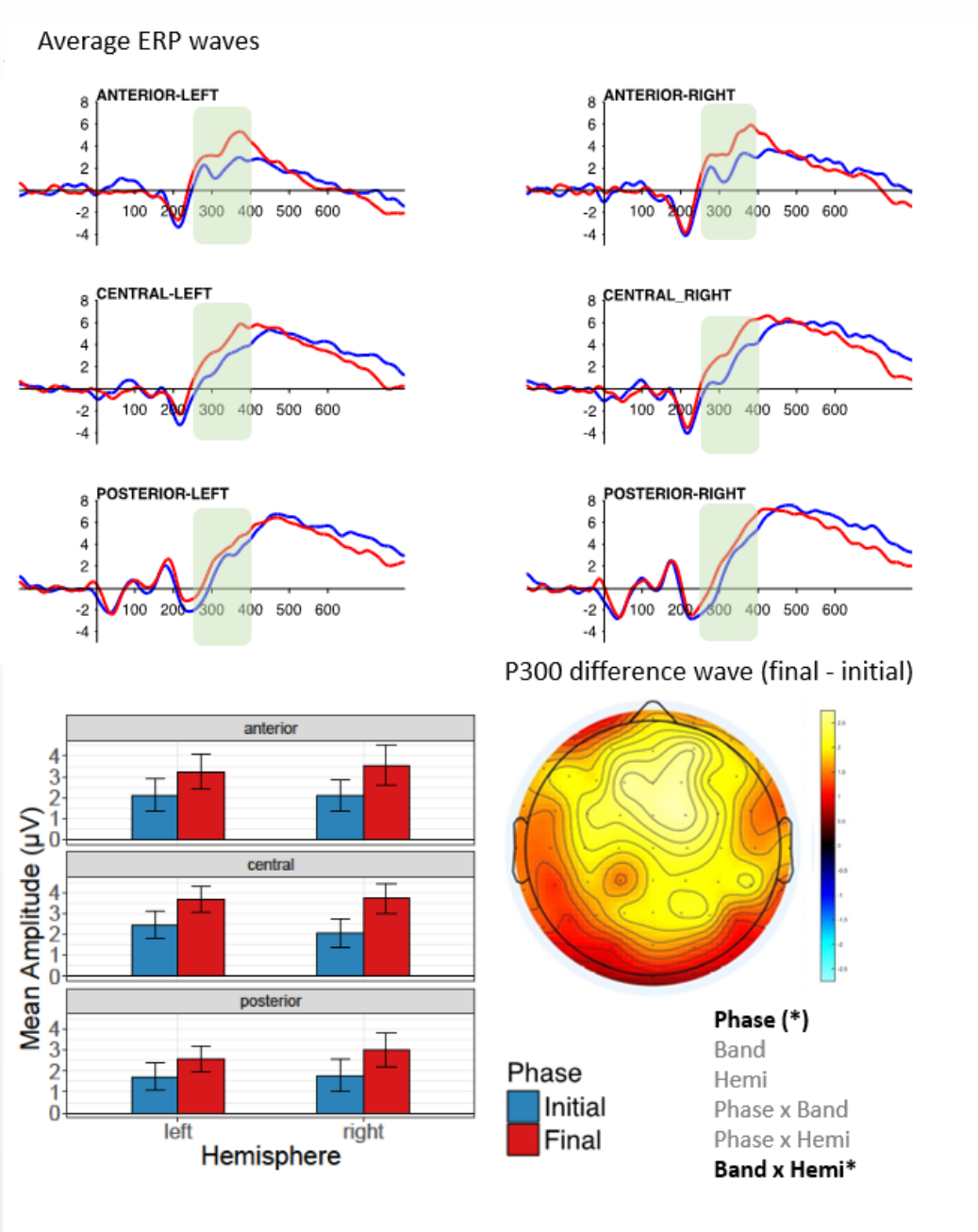


Figure 6.13. P300 amplitudes for within-category priming trials in test blocks. (*) $p < .07$; * $p < .05$; ** $p < .01$.

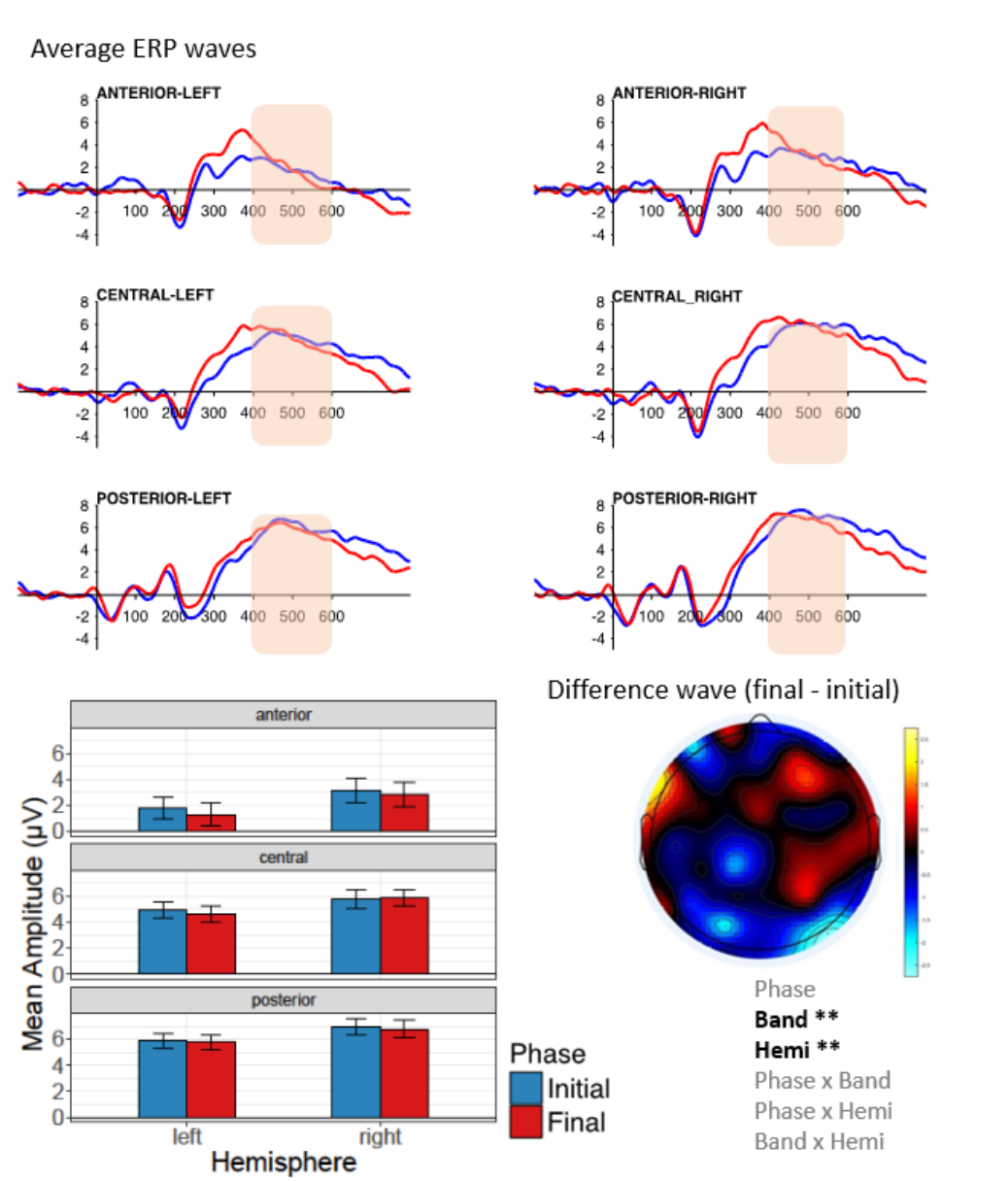


Figure 6.14. Late positivity amplitude differences for within-category priming trials in test block. (*) $p < .07$; * $p < .05$; ** $p < .01$.

Between-category priming: early effects (N200)

There was a significant main effect of *Band* ($F(1.26, 23.90) = 31.57,\ p < .001,\ \eta_g^2 = .62$) and a near-significant *Phase* × *Band* interaction ($F(1, 19) = 3.81,\ p = .066,\ \eta_g^2 = .17$). No other main effects nor interactions were significant (see Figure 6.15).

Between-category priming: middle latency effects (P300)

Here, we compared whether between category priming effects change with increased categorisation training. There was no effect of Phase, but there were significant effects of *Hemisphere* and *Band* (hemisphere:$F(1, 19) = 17.08,\ p < .001,\ \eta_g^2 = .47$; band: $F(1.09, 20.64) = 4.50,\ p = .043,\ \eta_g^2 = .19$). Nonetheless, none of the comparisons were significant in the post-hoc tests (see Figure 6.16).

Between-category priming: late effects

The analysis revealed a significant main effect of *Band* and *Hemisphere* (Band: $F(1.10, 20.98) = 44.14,\ p < .001,\ \eta_g^2 = .70$; Hemisphere: $F(1, 19) = 23.91,\ p < .001,\ \eta_g^2 = .56$). In addition, there was a significant *Phase* × *Band* interaction ($F(1.41, 26.74) = 5.99,\ p = .013,\ \eta_g^2 = .24$), driven by differences in the posterior band ($p = .012$, see Figure 6.17).

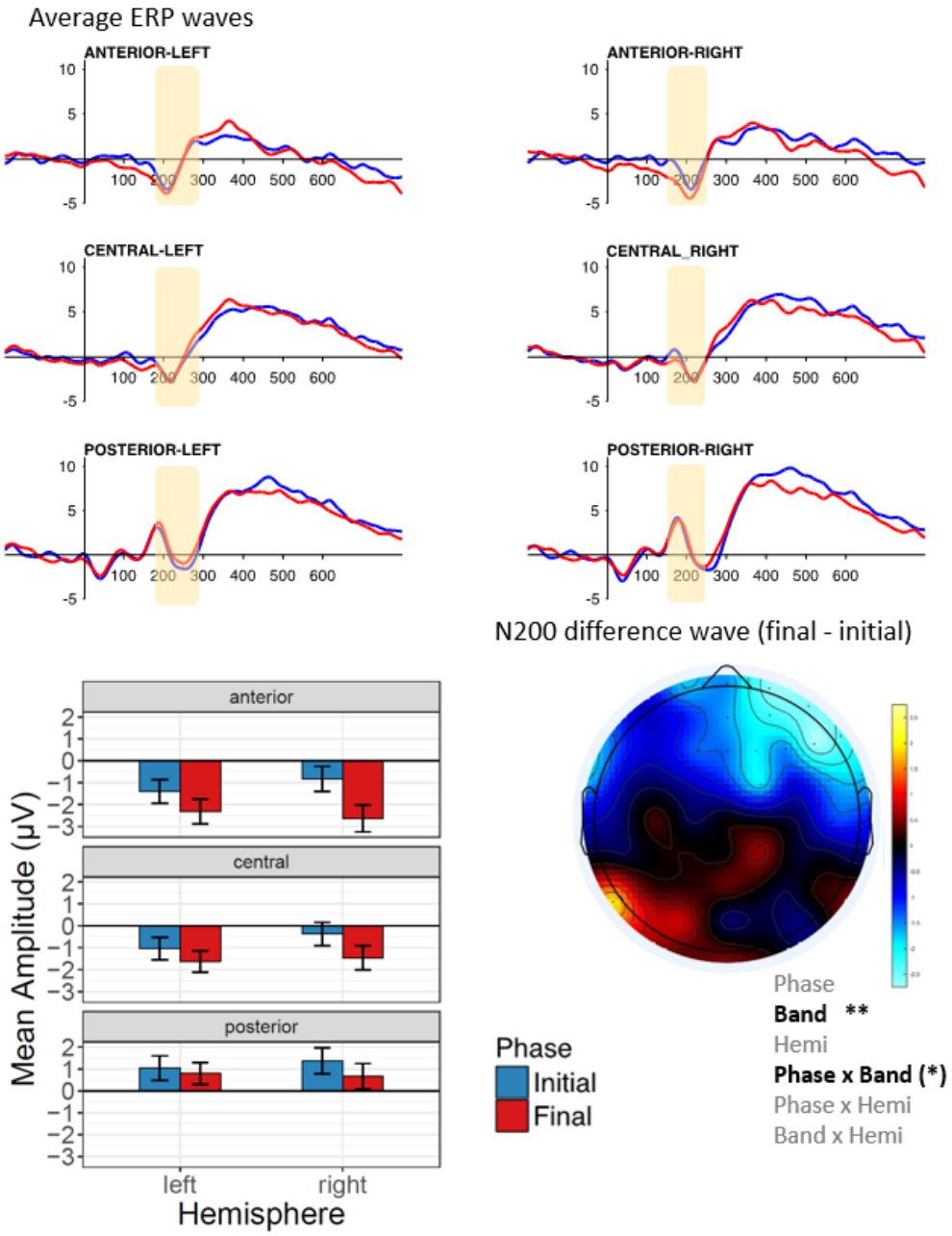


Figure 6.15. N200 amplitudes for between-category priming in test blocks. (*) $p < .07$; * $p < .05$; ** $p < .01$.

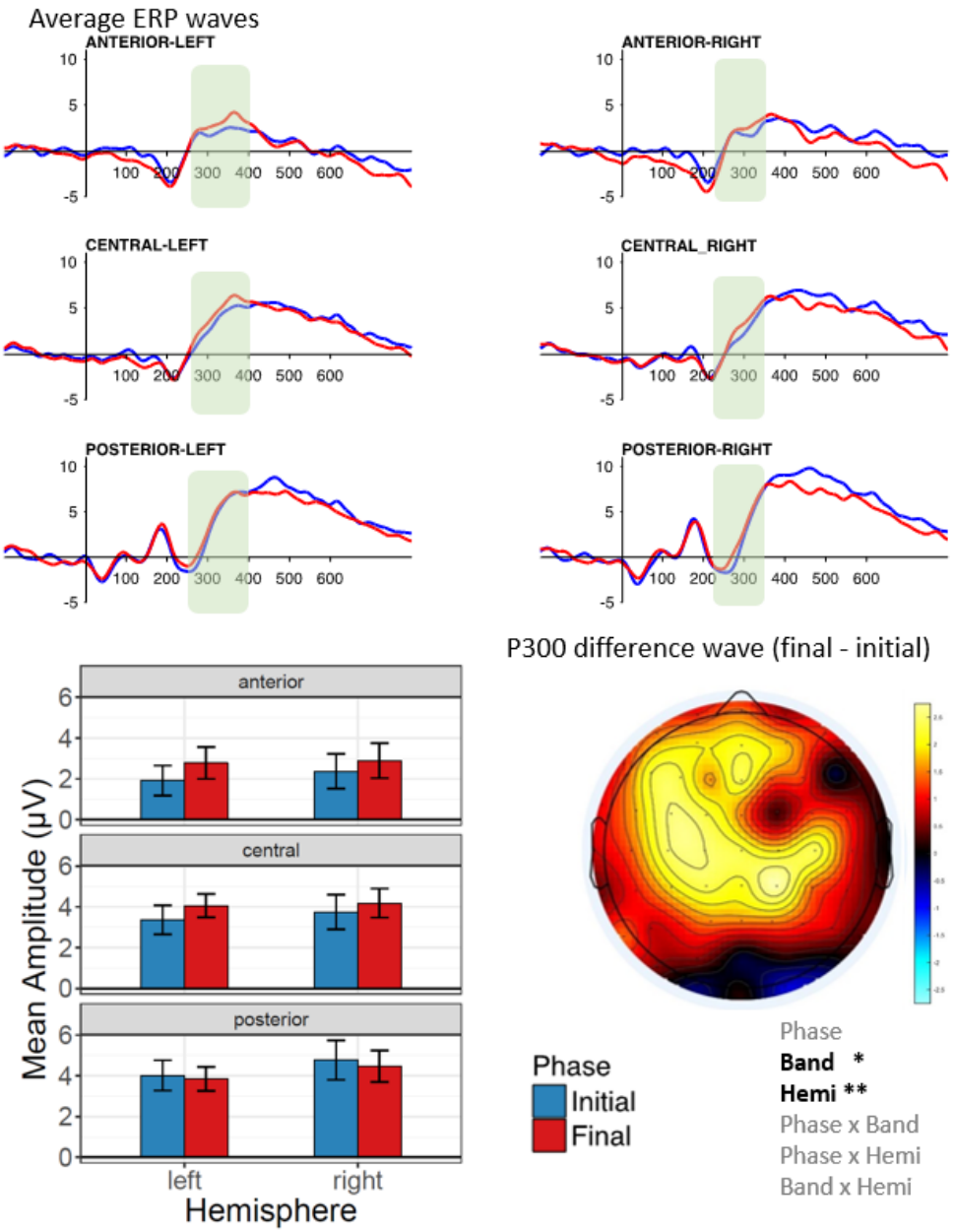


Figure 6.16. P300 amplitudes for between-category priming in test blocks. (*) $p < .07$; * $p < .05$; ** $p < .01$.

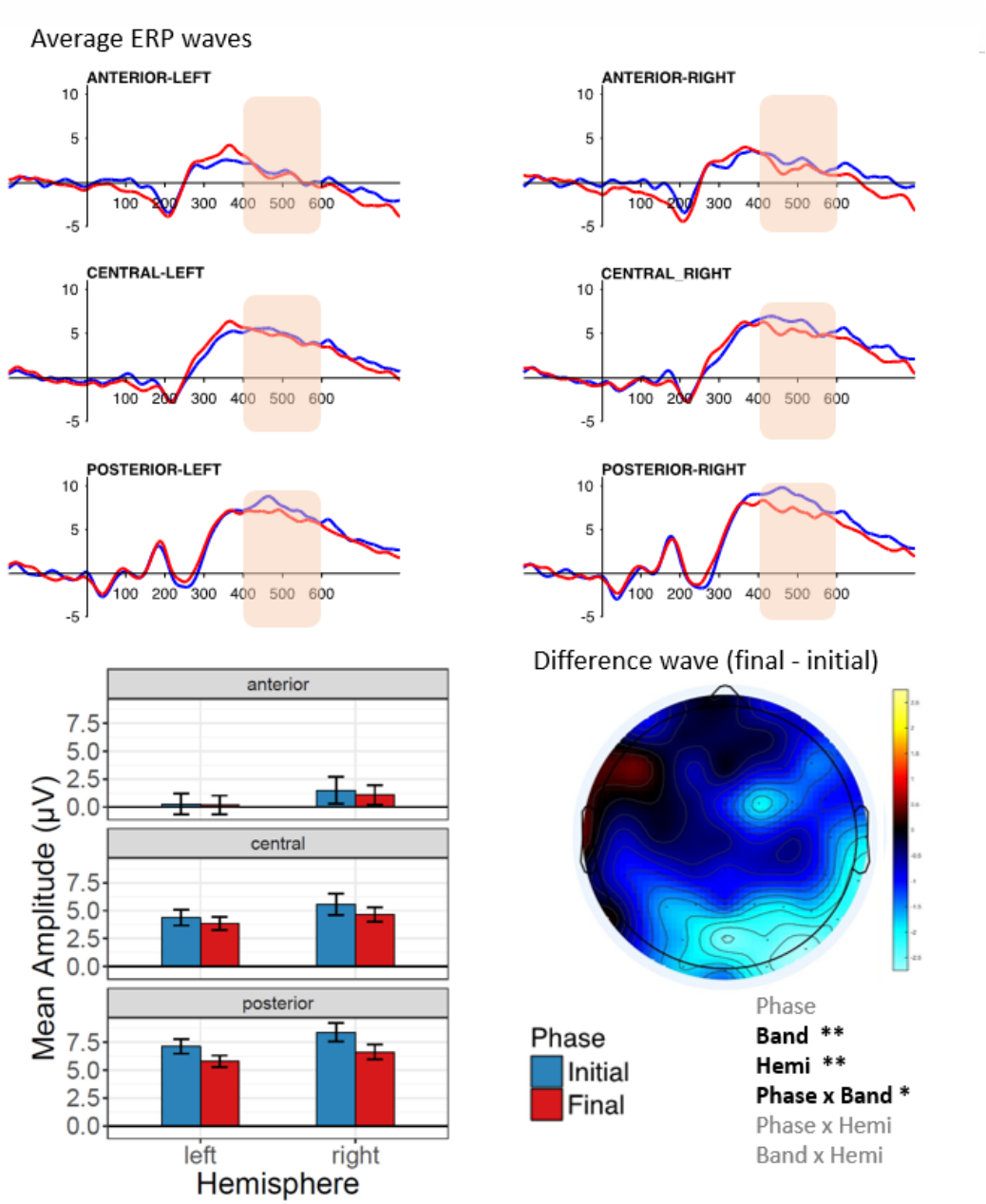


Figure 6.17. Late positivity effects for between-category priming in test blocks. (*) $p < .07$; * $p < .05$; ** $p < .01$.

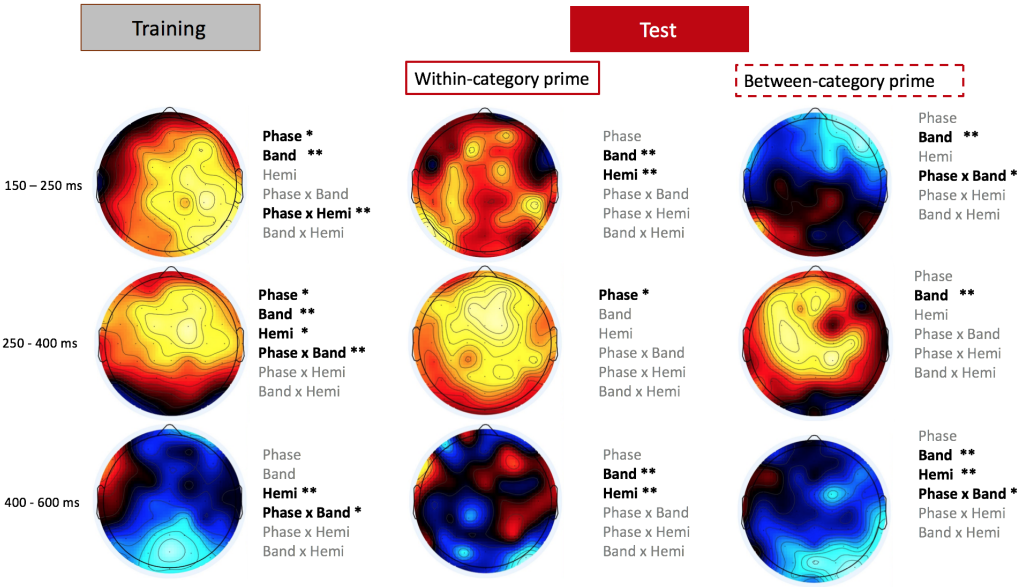


Figure 6.18. Summary of the ERP analyses: (significance levels: (*) $p < .07$; * $p < .05$; ** $p < .01$).

6.7 Discussion

The present experiment explored how categorisation training modulates visual processing of category members. Participants learned to categorise a set of novel exemplars into two categories. An extensive training consisted of eight training blocks, in which participants received feedback for their responses. Interleaved with training, test blocks comprised pairs of items presented in succession (prime followed by target), where participants had to indicate whether the items come from the same category or not. No feedback was provided.

Analysis of behavioural data revealed that participants successfully learned categories, as indexed by improvements in reaction times and accuracy over

the course of eight training blocks. In test blocks, when asked to judge whether two successively presented items come from the same or from different categories, participants consistently performed faster when responding to *within-category* pairs. This result is in line with a standard priming literature that shows, for instance, that participants are faster to respond to a word if it is primed with another, semantically related word (for example, a faster reaction time to the word "doctor" if it is primed with "nurse", see Weingarten et al. (2016) for a recent review of priming effects).

Analyses of ERPs in training blocks revealed that category training modulates the visual evoked potentials elicited by the trained exemplars. These modulations were most prominent in two windows. Firstly, extended training elicited a decrease in an early negativity in the right hemisphere, peaking around 200 ms post-stimulus. Secondly, training elicited an increase in positivity in frontal and central regions in the later time window, starting around 300 ms after stimulus presentation.

The analysis of the test phase revealed that training differentially modulates visual evoked potentials to the target item depending on category membership of the prime. In the case of *within-category* priming, there was a strong positive enhancement in the late time window across the entire scalp in the final training block as compared to the initial stage. In the case of *between-category* priming, i.e. when prime and target belong to different categories, increased categorisation expertise leads to a more negative early component (peaking around 200 ms) in frontal regions and reduction in the

late-window positive deflection in the occipital areas. These results clearly demonstrate that priming effects differ depending on the amount of categorisation experience and that there is a prime-specific (*within* or *between-category*) modulation of visual object processing.

Effects of categorisation training in the early time window (150 - 250 ms) suggest there is a modulation in the N1 component. The more pronounced negative deflection after the initial training, as opposed to a reduced effect that followed an extensive categorisation training, suggests that increased expertise reduces the N1 component. An increased N1 is typically linked to an increase in perceptual processing (Kiefer, 2001; Tanaka et al., 1999) and heightened attention to visual stimuli (Mangun & Hillyard, 1991). The decrease in N1 following an extensive amount of training reflects the fact that initially, while starting to learn categories, participants paid more attention to features of items in order to extract category relevant information; later on and with extensive experience perceptual processing is less demanding. Another possibility is that the observed modulations are an effect of an increased familiarity with these items. However, results of the test block contradict this account: namely, cross-category priming also modulated this component, while no such effect was present for within-category priming. This suggests that these effects cannot be only due to familiarity, but that that they are driven by the categorisation processes.

Early visual ERP components are known to be very sensitive to perceptual properties of objects, where higher perceptual complexity elicits more

negativity (Handy et al., 2001; Hillyard & Anllo-Vento, 1998; Mangun & Hillyard, 1991; Rorden et al., 2008). This sometimes represents a confounding factor as some categories are naturally more complex than others (for instance animals versus vehicles), making it hard to distinguish contributions of perceptual and category related factors to N1 modulations. However, we suggest that this is not the case in the present study, since the stimulus set was selected to rule out this possibility. Exemplars of both categories were simple line drawings consisting of identical features. In order to distinguish categories, participants had to learn relations between features, not the features themselves.

Categorisation experience also modulated visual evoked potentials in a later time window, between 300 and 600 ms. Increased expertise was linked with increased positivity in this time window. This effect was present in training blocks, i.e. when participants were asked to categorise individual items, and also in the test. Interestingly, this enhancement was present only in the case when targets were primed with items from the same category, but no effect was present when primed with *between-category* primes. These results are in line with the literature on the P300 component, which is assumed to reflect categorisation processes (Azizian et al., 2005; Kutas et al., 1977; Mecklinger & Ullsperger, 1993).

Taken together, these results clearly show that there are differences in brain ERPs elicited during visual processing of objects depending on the categorisation experience.

The observed results are in line with the existing categorisation ERP literature (Curran et al., 2002; Kiefer, 2001; Tanaka et al., 1999). The two time windows in which we observed categorisation-dependent modulations are the same time windows where the effects were reported in studies that explored how different level of categorisation (subordinate, basic and superordinate) affect the processing of objects. However, whereas these studies mostly focused on items from existing categories processed on different level of abstraction, in the present study the focus was on processing and learning novel categories, keeping the same categorisation level throughout.

The present study revealed that categorising objects alters their neural representations. As compared to real-world categories, where one is exposed to a large number of exemplars, the current study showed that even eight training blocks are already sufficient to modulate ERPs to objects. Behavioural studies show that certain categorisation effects are present in behaviour even after a much shorter exposure. For instance, infants show evidence of category learning after being exposed to less than a dozen category exemplars. This suggests that some sort of "on-line" categories are formed during this short period of time. Future studies should explore whether these on-line categories are coupled with certain transient effects at the neural level.

Chapter 7

General Discussion

The concluding chapter provides an overview of the main findings of this thesis and its key contributions to our understanding of the category learning process. Finally, it discusses the limitations of this research and proposes directions for future research.

7.1 Overview of main findings

7.1.1 Main findings part I - category learning in infants

Part I examined novel category learning in infants focusing on two aspects: how different parameters of the learning situation shape the category formation process, and the role of dynamic features, such as labels and motion in the category formation process.

Optimising category learning context

Experiments 1 - 3 presented in Chapter 3 systematically explored how characteristics of the learning context shape category learning (see Figure 7.1 for a summary).

Memory demands and the perceptual load were manipulated by a systematic variation the number of items presented at the same time during the familiarisation phase. Changing from single to multiple-item familiarisation reduced memory demands since multiple items were available. At the same, this increased the perceptual load, i.e. the amount of visual information simultaneously available. Importantly, multi-item presentation enabled infants to directly compare and contrast familiarisation exemplars. Using a standard familiarisation-novelty preference procedure, infants were first presented with a set of training items, followed by a test phase in which category formation was assessed with the presentation of two novel items - one within and one outside the category item. Preference for the novel out-of-category object was used as an index of category formation. Familiarisation was interleaved with test blocks in order to explore category formation at different levels of familiarity and evaluate the dynamics of the learning process.

To determine how memory demands and perceptual load affect category learning, infants were familiarised with one item at a time (Experiment 1 and 1a), with a pair of items (Experiment 2) and with four items simultaneously (Experiment 3). The results revealed that paired presentation provides an optimal balance of memory and perceptual demands, therefore, the ideal

learning context, since infants showed evidence of category learning already after one familiarisation block. Presenting four items simultaneously during familiarisation had a detrimental effect on learning; infants needed three times more familiarisation experience to form categories. In contrast, when presented with one item at a time during familiarisation, infants failed to show any evidence of category learning, even after three blocks of familiarisation.

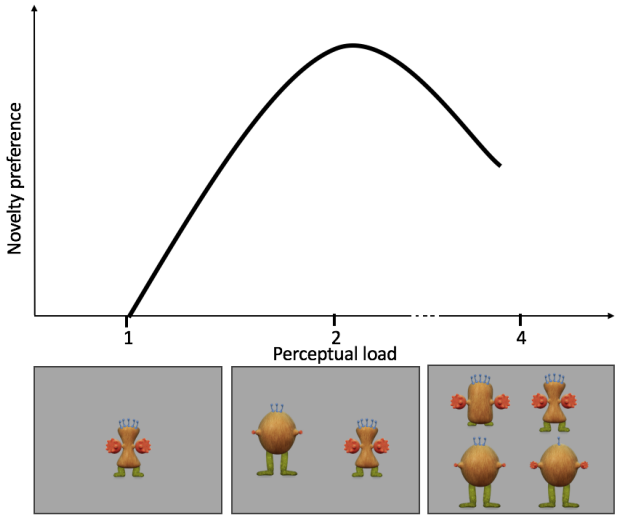


Figure 7.1. Thesis Part I: Chapter 3 demonstrated that paired familiarisation represents optimal learning context for category learning in 10-month-old infants.

Dynamic features in category learning

Experiments 4 - 7 explored the role of dynamic features in category learning. I focused on two particular types of features: labels and motion (see Figure 7.2 for a summary). Here I used a gaze-contingent adaptation of the standard familiarisation-novelty preference procedure. After being familiarised with a set of training items, infants' preference for one of the two items presented

in the test was used as an index of category formation. Three familiarisation blocks were interleaved with test blocks, enabling a direct comparison of infants' performance across experiments. One important novelty that this set of experiments introduced was a gaze-contingent familiarisation; an item was labelled or moved only when infants fixated on them. This allowed infants to take a more active role in learning and guide the dynamic of the familiarisation. More fixations lead to more events, whereas less shifting and longer fixations resulted in fewer labelling and motion events. The only difference between the categories used in this set of experiments and those in Chapter 3 is that I attached additional features to these categories: labels or motions. This enabled comparison of the two sets of experiments, and it facilitated a systematic exploration of how adding features to visual information affects the learning process. The results revealed that adding motion as a category feature hinders category learning. However, attaching a label as an additional feature facilitates category learning. Interestingly, this is only the case when both categories are accompanied by a category-specific label. When only one category is labelled, no learning can be observed. These results suggest that labels and motion do not have the same effect on visual category learning in 10-month old infants: while labels facilitate learning, motion hinders infants' ability to learn categories.

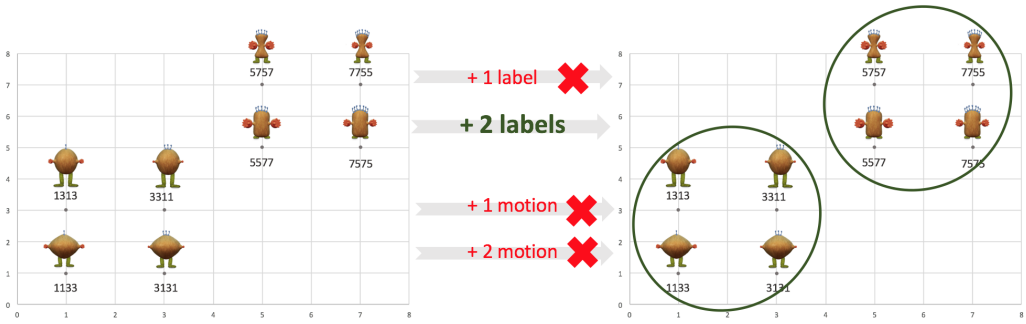


Figure 7.2. Thesis Part I: Chapter 4 demonstrated that motion hinders and labels facilitate the process of visual category learning in 10-month-old infants.

7.1.2 Main findings part II - category learning in adults

Part II examined implicit category learning in adults and how category learning shapes the visual processing of objects at the neural level.

Implicit categorisation in adults

Experiment 8 explored the process of implicit, unsupervised category learning in adults. I developed a novel paradigm to test whether the characteristics of implicit category learning observed in infants are similar to those in adults. Participants were presented with a fast-paced presentation of category exemplars and were asked to judge whether the two exemplars presented in succession were identical or different. There was no explicit mention of categories, and participants were not aware of the main aim of the task. The categories would only be formed if adults were sensitive to the feature correlations across objects. Response time was used to assess whether there were differences between within- and between-category judgements. The results revealed that with increased exposure to the exemplars of the two categories, participants became faster in judging between-categories pairs as compared

to the within-category pairs. Importantly, this effect was also true in the case of test blocks, in which similarity between exemplars in pairs was constant for both within- and between-category pairs. This demonstrated that adults, as infants, use statistical regularities to spontaneously organise a stream of visually presented items into categories (see Figure 7.3).

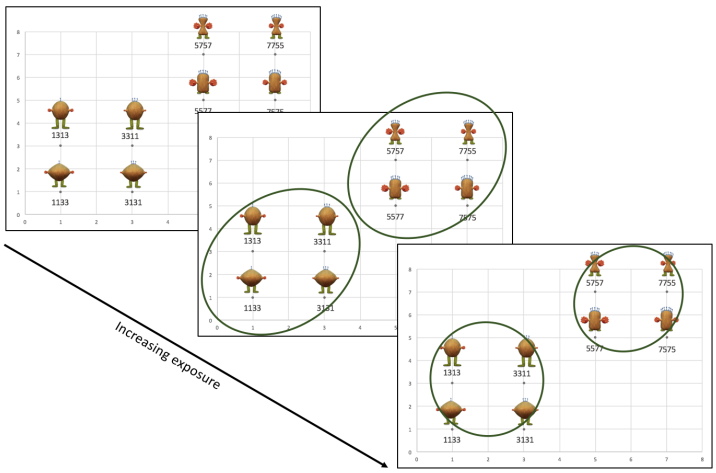


Figure 7.3. Thesis Part II: Chapter 5 demonstrated that adults are able to extract category structure from a stream of visually presented objects.

Neural signatures of categorisation

Experiment 9 investigated neural signatures of category learning in adults. Participants performed a novel category learning task whilst their brain waves were recorded (Figure 7.4). By comparing neural signatures of visual object processing at the very beginning of the category learning process and after an extensive categorisation practice, I tested how categorisation experience alters the way in which exemplars are processed. The results revealed that categorisation experience is linked to a modulation of several ERP components: extended training elicited a decrease in N200 in the right hemisphere

and enhanced positivity in the frontal and central regions in the later time window, starting around 300 ms after stimulus presentation.

In addition, priming effects were found to be category-dependent: in the case of within-category priming, increased categorisation experience led to a strong positive enhancement in the late time window across the entire scalp. In the case of cross-category priming, increased expertise led to stronger early negative component (peaking around 200 ms). Taken together, this experiment revealed that categorising objects alters their neural representations. It showed that even a small set of training blocks (as compared to the extensive experience with real-world categories) modulated event-related potentials to objects, in particular the early components related to visual processing.

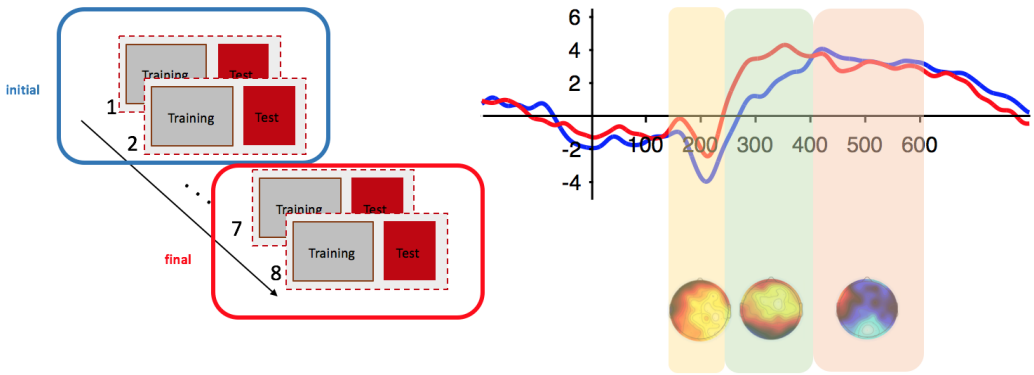


Figure 7.4. Thesis Part II: Chapter 6 demonstrated that categorisation experience was linked to a modulation of several ERP components.

7.2 Implications

7.2.1 Optimising the category learning context

The results presented in Chapter 3 demonstrate that infants' ability to form categories is shaped by the characteristics of the learning situation.

A simple manipulation of the basic characteristics of the learning context – varying the number of items presented at the same time during familiarisation led to vastly different learning outcomes. This manipulation resulted in the change of two basic parameters of the learning context: memory demands and perceptual load. Increasing the number of items presented simultaneously decreases memory demands and increases the perceptual load. After discussing the contributions of this optimisation in the following paragraph, I turn to another property that a multi-item leaning context offers - a unique opportunity to directly compare and contrast simultaneously presented items; and I discuss the role of this property in driving the learning process.

As pointed out by Aslin (2010), limited working memory abilities inhibit fully effective statistical learning and therefore also infants' ability to learn correlation-based categories. Decreasing working memory demands by presenting multiple items at the same time enable infants to more efficiently extract category relevant information. The discovery that switching from single-item to paired presentation has beneficial effects on category learning is in line with this account. However, my finding that further increase in the number of items does not facilitate, but rather slow down the learning

suggest that there are other factors that determine the learning dynamics.

One such factor is the perceptual load. Increasing perceptual load poses higher demands as more information is simultaneously available and needs to be filtered and processed. If the relationship between perceptual load and learning efficiency was linear, minimal perceptual load would represent an optimal learning context. However, this is not the case. The results show that in the current set of experiments paired presentation is the optimal load. The perceptual load hypothesis can be a useful framework for understanding these effects (Plunkett, 2010). This hypothesis proposes that the outcome of computations underlying category formation is dependent on the load that individual stimuli place on the perceptual system. It further assumes that category formation is sensitive to the perceptual load according to the inverted U-shaped function, where too little or too much load interfere with, but an optimal load enhances category formation. The original version of the hypothesis (Plunkett, 2010) can account for the results after one block of familiarisation (Experiment 2): paired presentation presents an optimal perceptual load, whereas the presentation of single and multiple items (four, Experiment 3) falls outside of this optimal load region; thus, no learning can be observed. However, introducing additional familiarisation blocks revealed that infants are able to learn categories when familiarised with four items, but need more training to achieve this task. In order to account for this effect, I introduce another factor to the model. This factor takes account of the amount of familiarisation, i.e. the amount of experience with items,

and enables integrating results of the perceptual load and the familiarisation level into one model.

Even though the perceptual load hypothesis may provide a useful account for interpreting the present set of data, it is important to mention that the theory of perceptual load has been a subject to a debate due to the inherent circularity of the term *perceptual load*. As pointed out by Murphy et al. (2016), perceptual load is not defined by explicit definitions, but instead by parameters of paradigms, such as set size. Thus, perceptual load is lacking a clear operational definition which poses serious limitations to its explanatory power.

To summarise, as shown in Figure 7.5, paired familiarisation regime represents an optimal learning context because it achieves a balance of several parameters, such as working memory demands and the perceptual load and other relevant factors that had not been experimentally varied in these experiments, such as attention demands and inhibition, to name just a few.

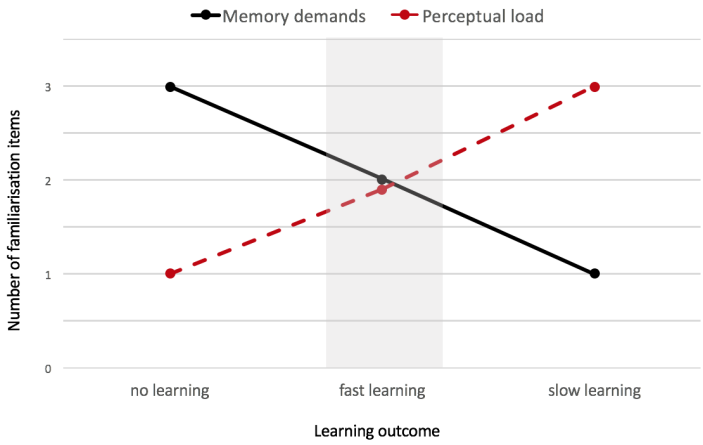


Figure 7.5. The role of memory and perceptual load in category learning: paired presentation represents the optimal learning context (area depicted in grey).

As discussed in Chapter 3, presentation of multiple items during familiarisation provides a unique opportunity to compare and contrast the presented items and lead to uncovering the structure of the categories. Since visual information is constantly available, infants can immediately examine similarities and differences between the two items. Furthermore, infants are at liberty to decide for themselves how to allocate attention. In contrast, when presented with one item at a time, direct comparisons are not possible. Infants can only compare a visually presented item with a working memory representation of the previously presented item. Inevitably, this leads to comparisons performed across different representational levels. The results of this study suggest that more direct lower-level comparisons are more useful for the process of category formation in infants. This is consistent with the notion that comparison and alignment of two representations highlights common structures and relational knowledge (Gentner & Namy, 1999), and therefore leads to more efficient category learning.

Moreover, these results have several implications for a wider research community. They suggest that the number of pictures presented on the screen is an important parameter to consider when designing eye-tracking studies with infants.

These results also have translational potential. For instance, in the case of storybooks for young children, the number of visual items presented on one page can be designed in a way to optimise the perceptual load and attentional demands of young readers.

7.2.2 Labels and motion in category learning

Experiments 4 - 7 demonstrated that visual category formation is affected by adding dynamic features to these categories. In order to form categories, visual information was sufficient (as shown in Experiment 3); this means that these additional features were redundant. Despite their redundant nature, both motion and labels had a strong impact on category formation. Whereas motion impeded the process of category formation, labels facilitated learning.

These results clearly demonstrate that labels and motion have a different status in category learning in 10 month-old infants. Possible reasons for the detrimental effects of motion on visual category learning might be related to infants limited ability to bind static and dynamic information (Burnham et al., 1988; Rakison, 2004; Rakison & Poulin-Dubois, 2002), to an underdeveloped ability to integrate information processes in the dorsal and ventral cortical pathway (Mareschal & Johnson, 2003; Mareschal et al., 1999) or to motion overshadowing (Sloutsky & Robinson, 2008).

The observed detrimental effects of motion on category learning are in contrast with a recent study which found that motion facilitates extraction of categories in infants of the same age (Deng & Sloutsky, 2015). However, there are several important differences in the experimental designs of the two studies that might be responsible for the opposing results. Firstly, the number of items in familiarisation. In my study infants were presented with four items in each familiarisation trial, whereas Deng and Sloutsky (2015) presented infants with one item at a time. Secondly, in my study, infants

were meant to learn to distinguish between two categories; in the study of Deng and Sloutsky (2015) infants learned only one category. Thirdly, there is a decisive difference in the nature of the dynamic features in the two studies. In my study motion is a general feature, i.e. an entire item moves, while in Deng and Soutsky's study only one feature of an item (the creature's feet) moves, making motion a feature of a feature.

In addition, the two studies differed in terms of the timing of the motion. As opposed to immediate motion in my study (an item would start moving as soon as infants fixated it), in the study by Deng and Sloutsky (2015), there was a delay: infants first heard a phrase ("Look at this one!") and only then there would be a motion event. It is possible that a certain type of "motion overshadowing" took place when the static and dynamic features to be learned are presented simultaneously. This would parallel the notion of auditory overshadowing which suggests that simultaneous presentation of visual information and labels can have detrimental effects on learning (Robinson & Sloutsky, 2007; Sloutsky & Robinson, 2008). Experiments that systematically manipulate the timing of the motion and the number of categories are needed to resolve these open questions.

This is the first study to directly demonstrate that labels facilitate category learning. A substantial research has demonstrated a compelling impact of labels on categorisation (Althaus & Mareschal, 2014; Althaus & Plunkett, 2015; Plunkett, 2010; Plunkett et al., 2008; Waxman & Markow, 1995), but none of these studies provided a direct evidence of the facilitation. Two

properties of the present set of experiments directly show that labels facilitate category learning. By comparing these experiments with the results of Experiment 3, I show that there is a difference between learning categories in silence (Experiment 3) and with labels (Experiment 7). Using an interleaved familiarisation-test design enabled me to determine the amount of familiarisation necessary for learning with labels and in silence. When accompanied by labels, one familiarisation block is sufficient for learning; when taught in silence, infants need three familiarisation blocks to succeed.

One of the questions that remains open is related to the nature of these facilitative effects; whether labels have a special status (Waxman & Markow, 1995), or they are features (Gliozzi et al., 2009; Plunkett et al., 2008) that drive the observed effects by directing and maintaining infants attention.

Taken together, Experiments 1 - 7 suggest that in a multimodal learning context, the relationship between load and performance is not linear. The effects of the additional features are not uniform, but depend on the nature of the information. Adding highly salient dynamic motion information completely hinders learning. In contrast, increasing overall perceptual complexity by including category-specific labels facilitates learning, but only if both categories are accompanied by labels.

These results imply that when it comes to categorisation development, our primary focus should extend beyond an affirmation of infants' ability to form categories, by exploring the *conditions* under which this success can be achieved.

7.2.3 Implicit category learning in adults

In Experiment 8, I explored whether the same type of rapid implicit category learning is present in adults, as in infants. The novel experimental paradigm revealed that adults indeed spontaneously organise objects into categories when presented with a visual stream of items. These behavioural effects correspond to looking preferences observed in eye-tracking experiments with infants. This experimental paradigm represents a promising route for further understanding the basic mechanisms underlying statistical category learning in infants and adults. This paradigm is particularly suitable for testing predictions of the models that propose a shared implicit category learning system over the life span (Ashby et al., 1998; Gureckis & Love, 2004). These models assume that differences in the performance of infants and adults are due to memory and encoding demands - this can be directly tested using our task by varying the task parameters such as the presentation time and response requirements.

7.2.4 Neural signatures of category learning

The final experiment of this thesis provided evidence that increased categorisation experience alters object processing at the neural level. Modulations in the early ERP components suggest that categorisation experience leads to a decreased perceptual engagement in processing, whereas enhancement in later components reflects categorisation-related processes. This suggests that the processing demands that individual objects set decrease as category

expertise increases. Thus, the optimal perceptual load is a dynamic value: the same stimuli might become perceptually less complex with more category related experience. Further studies are needed to map this dynamics.

Another important insight this study provides is that even a limited categorisation expertise (as compared to the large amount of exposure to real-word categories) alters neural representations of objects. One of the key questions to be considered in future studies is whether analogous effects can be observed in infants. Experiments using looking preferences suggest that there are categorisation-dependent modulations to object processing. Studies combining eye-tracking and EEG would offer a better understanding of the ways in which looking preferences are driven by modulations in neural processing.

7.3 Methodological limitations

As we have seen, studies presented in Chapters 3 and 4 used looking preferences to explore infants' learning. Even though looking behaviour represents a good measure of underlying cognitive processes in infants (Luna et al., 2008), it is still an indirect measure. The standard familiarisation-novelty preference paradigm used to investigate category formation in infants assumes that behaviour is driven by the novelty preference. If infants express preference for a novel item, this indicates that they have learned categories. Overall preference score is a reliable measure used for several decades; however, it might not be sensitive enough to capture some less prominent effects

and finer differences in infants' behaviour. In addition, looking preferences are dynamic in their nature and it is to be expected that there are individual differences in the speed of reaching the novelty preference stage. However, the standard statistical approach to quantifying these behaviours does not take this into account. It is necessary to introduce finer statistical tools into the field of looking behaviour research. For instance, using stochastic modelling approach, such as Markov models to quantify underlying states driving looking preferences is one way we could obtain a more robust assessment of the states underlying the observed looking preferences.

In these experiments, infants' category formation was tested using only one or two trials. Thus, one of the aims of future studies would be to increase the number of test trials. This would contribute to the robustness of the observed effects. At the same time, this introduces additional complexity since infants also learn from these trials (i.e. there is a limitation in the number of times an item can be presented and be considered as novel). Overall, my study employs a methodology in which category learning is tracked indirectly (using eye-movements) with an indirect index of learning (using novelty preference). Adding the noisy nature of infant data, this poses limitations in terms of the power of the utilised measures to capture potentially less prominent, but still relevant effects.

Studies presented in Chapters 3 and 4 explored infants' ability to learn categories based on statistical regularities. There are two additional methodological limitations to the present set of findings that need to be discussed.

Firstly, any potential prior preference for any of the test items was not investigated. The main reason infants were not tested for prior preference was to keep test items completely novel until the test block. If infants had a preference for any of the items before the training, that would compromise results of test blocks. However, which item was presented in test was counterbalanced across participants and therefore, the prior preference cannot account for the obtained set of results. To completely rule out the possibility that any prior preference might have influenced infants' looking preferences in test, in future studies this preference should be tested with a separate set of infants that will not take part in the learning task. The present set of experiments also didn't test whether infants were able to discriminate between individual training items. Having in mind that I used increased perceptual variability of features (each feature could have taken one of the seven potential values), future studies using this stimuli set need to include tests of item discrimination to rule out this possibility.

7.4 Theoretical limitations

One of the aims of the present work was to determine whether similar mechanisms underlie implicit category learning in infants and adults. The obtained set of results provided some evidence that there are indeed similarities in computations underlying category formation in infants and adults. There are, however, important limitations to making claims about the developmental trajectories of the postulated mechanisms because the present work focused

on two age groups, 10 month old infants and adults. The first age group was selected because previous studies have identified this age as very important in development of statistical visual category learning. Adults were tested to examine characteristics of the category learning at the stage when cognitive abilities are fully developed. Therefore, the present work compared category learning at two very specific points in development; at an early point and at a point of a fully operational functioning. To map out developmental trajectories of the investigated processes, future studies need to include a wider age range.

7.5 Future directions

As outlined before, the optimal learning context is determined by several factors, such as the amount of information simultaneously available during learning, the level of familiarisation (i.e. expertise level) and the nature of the categories. Future studies should use a longitudinal within-subject design to explore the role of individual and developmental differences in determining the optimal load.

The present work focused on the ability to learn correlation-based categories. This type of learning can be seen as a particular case of a more general mechanism of statistical learning. Moreover, it would be interesting to explore possible relations between statistical category learning and statistical learning in other domains, such as language. For example, comparing infants' performance in the correlation-based visual category learning and

artificial grammar learning task. This would bring us closer to understanding whether there is a domain general mechanism responsible for different instances of learning or rather several domain specific mechanisms.

This thesis demonstrated that both infants and adults show a remarkable ability to efficiently extract statistical regularities from a visual input and use this information to learn novel categories on extremely short time-scales. One of the key questions that future research needs to address is how these fast-paced, on-line categories are retained and translated into the long-term knowledge, and identify the conditions that make this transfer possible.

7.6 Conclusion

This thesis explored how the human mind formulates abstract, general principles of organisation to form novel categories. Part I demonstrated that infants' category learning is a dynamic process shaped by different factors which determine outcomes of the learning situation. To compare and contrast items when multiple items are presented during learning provides a unique opportunity to the uncover structure of the categories, with additional category features, such as labels and motion having a crucial role in determining the dynamics of learning. Part II examined analogous kinds of implicit, unsupervised category learning in adults and showed that category formation is reflected in reaction time measures and in modulations of object processing at the neural level. Ultimately, this thesis provides evidence for the view that categorisation is a complex dynamic process where fundamen-

tal principles are shared across the life span. This takes us one step further in mapping the complex route that the *mindbrarian* takes in deciphering the structure of human knowledge.

Appendix A

Experiment 1

Table A.1. Experiment 1: Stimuli structure used in three familiarisation blocks (each number represents a value on each of the four features).

Block	Category	Antenna	Hands	Body	Legs
1	1	1	1	3	3
	1	1	3	1	3
	1	3	1	3	1
	1	3	3	1	1
	2	5	5	7	7
	2	5	7	5	7
	2	7	5	7	5
	2	7	7	5	5
2	1	1	1	1	3
	1	1	1	3	1
	1	3	3	3	1
	1	3	3	1	3
	2	5	5	7	5
	2	5	5	5	7
	2	7	7	5	7
	2	7	7	7	5
3	1	1	3	1	1
	1	1	3	3	3
	1	3	1	1	1
	1	3	1	3	3
	2	5	7	5	5
	2	5	7	7	7
	2	7	5	5	5
	2	7	5	7	7

Training phase - model specification

Model specifications for the growth curve analysis of the looking time in familiarisation trials:

base model: $LT \sim 1 + (1 + Trial \mid ID)$

model 1: $LT \sim 1 + Trial + (1 + Trial \mid ID)$

As shown in the equations, in the base model looking time was modelled with an intercept and a random effects of participants on the intercept and slope. Model 1 was a nested model with an effect of trial added to the base model. As recommended in (Mirman, 2014), the model comparisons were performed in order to evaluate whether there was a significant improvement in the model fit when adding additional parameters.

Appendix B

Experiment 1a

Table B.1. Experiment 1a: Stimuli structure used in three familiarisation blocks (each number represents a value on each of the four features).

Block	Category	Antenna	Hands	Body	Legs
1	1	1	1	7	7
	1	1	7	1	7
	1	3	3	5	5
	1	3	5	3	5
	2	5	3	5	3
	2	5	5	3	3
	2	7	1	7	1
	2	7	7	1	1
2	1	1	1	1	7
	1	1	1	7	1
	1	3	3	3	5
	1	3	3	5	3
	2	5	5	3	5
	2	5	5	5	3
	2	7	7	1	7
	2	7	7	7	1
3	1	1	7	1	1
	1	1	7	7	7
	1	3	5	3	3
	1	3	5	5	5
	2	5	3	3	3
	2	5	3	5	5
	2	7	1	1	1
	2	7	1	7	7

Appendix C

Experiment 2

Experiment 2: Training phase - model specification

Model specifications for the growth curve analysis of the looking time in familiarisation trials.

base model: $LT \sim 1 + (1 + Trial \mid ID)$

model 1: $LT \sim 1 + Trial + (1 + Trial \mid ID)$

Table C.1. Experiment 2: Growth curve analysis for the decrease in the amount of looking time over the course of the experiment. Parameter Estimates.

	Estimate	Std. Error	df	t value	p value
Intercept	6.07	0.45	197.26	13.60	0.00
Trial	-1.61	0.55	289.61	-2.92	0.00

Experiment 2: Training phase - model specification

Model specifications for the growth curve analysis of the looking time in familiarisation trials.

base model: $elog \sim (ot1 + ot2 + ot3) + ((ot1 + ot2 + ot3) \mid ID)$

model 1: $elog \sim (ot1 + ot2 + ot3) + ShiftCat + ((ot1 + ot2 + ot3) \mid ID)$

model 2: $elog \sim (ot1 + ot2 + ot3) * ShiftCat + ((ot1 + ot2 + ot3) \mid ID)$

Table C.2. Experiment 2: Parameter Estimation for the Growth Curve Analysis - modelling novelty preference in low- and high-shifters).

	Estimate	Std. Error	df	t value	p value
Intercept	0.20	0.19	30.96	1.08	0.29
ot1	0.32	0.60	283.97	0.53	0.60
ot2	-0.30	0.72	35.76	-0.42	0.68
ot3	0.92	0.63	128.30	1.47	0.14
Shifting Category (high)	0.57	0.25	33.30	2.25	0.03

Appendix D

Experiment 9

Table D.1. Structure of the stimulus set used in the ERP experiment.

Set	Category	Leg length	Neck length	Tail width	Ear separation
1	1	1	1	3	3
	1	1	3	1	3
	1	3	1	3	1
	1	3	3	1	1
	2	5	5	7	7
	2	5	7	5	7
	2	7	5	7	5
	2	7	7	5	5
2	1	1	1	1	3
	1	1	1	3	1
	1	3	3	3	1
	1	3	3	1	3
	2	5	5	7	5
	2	5	5	5	7
	2	7	7	5	7
	2	7	7	7	5
3	1	1	3	1	1
	1	1	3	3	3
	1	3	1	1	1
	1	3	1	3	3
	2	5	7	5	5
	2	5	7	7	7
	2	7	5	5	5
	2	7	5	7	7

As in Younger & Cohen (1986) leg length varied from short (1) to long (7), neck length from long (1) to short (7), tail width from thick (1) to thin (7) and ear separation from close (1) to spread (7).

References

Althaus, N., & Mareschal, D. (2014). Labels direct infants attention to commonalities during novel category learning. *PLOS ONE*, *9*, 1-10.

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

Ankowski, A. A., Vlach, H. A., & Sandhofer, C. M. (2013). Comparison versus contrast: Task specifics affect category acquisition. *Infant and Child Development*, *22*(1), 1–23.

Ashby, F. G., Alfonso-Reese, L. A., Turken, U., & Waldron, E. M. (1998). A Neuropsychological Theory of Multiple Systems in Category Learning. *Psychological Review*, *105*(3), 442–481.

Aslin, R. (2010). Perceptual Constraints on Implicit Memory for Visual Features: Statistical Learning in Human Infants. In L. M. Oakes (Ed.), *Infant perception and cognition: Recent advances, emerging theories, and future directions* (pp. 111–124). New York, NY: Oxford University Press.

Aslin, R. (2012). What's in a look? *Developmental Science*, *10*(1), 48–53.

Azizian, A., Freitas, A. L., Watson, T. D., & Squires, N. K. (2005). Electrophysiological correlates of categorization: P300 amplitude as index of target similarity. *Biological Psychology*, *71*(3), 278–288.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Biosemi. (n.d.). https://www.biosemi.com/headcap.htm. (Accessed: 2018-04-23)

Bomba, P. C., & Siqueland, E. R. (1983). The nature and structure of infant form categories. *Journal of Experimental Child Psychology*, *35*(2), 294 - 328.

Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 3–170). Lawrence Elbaum Associates.

Burnham, D. K., Vignes, G., & Ihsen, E. (1988). The effect of movement on infants' memory for visual compounds. *British Journal of Developmental Psychology*, *6*(4), 351–360.

Carvalho, P. F., & Goldstone, R. L. (2014). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory & Cognition*, *42*(3), 481–495.

Cedrus. (n.d.). https://cedrus.com/rb_series/. (Accessed: 2018-04-23)

Chatfield, C., & Collins, A. (1981). *Introduction to multivariate analysis.* Taylor & Francis.

Clapper, J. P. (2006). When more is less: Negative exposure effects in unsupervised learning. *Memory & Cognition*, *34*(4), 890–902.

Clapper, J. P., & Bower, G. H. (1994). Category Invention in Unsupervised Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(2), 443–460.

Clapper, J. P., & Bower, G. H. (2002). Adaptive Categorization in Unsupervised Learning. *Journal of Experimental Psychology: Learning Memory and Cognition*, *28*(5), 908–923.

Curran, T., Tanaka, J. W., & Weiskopf, D. M. (2002). An electrophysiological comparison of visual categorization and recognition memory. *Cognitive, Affective, & Behavioral Neuroscience*, *2*(1), 1–18.

Daltrozzo, J., & Conway, C. M. (2014). Neurocognitive mechanisms of statistical-sequential learning: what do event-related potentials tell us? *Frontiers in Human Neuroscience*, *8*, 437.

Dannemiller, J. L. (2000). Competition in Early Exogenous Orienting between 7 and 21 Weeks. *Journal of Experimental Child Psychology*, *76*(4), 253–274.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21.

Deng, W. S., & Sloutsky, V. M. (2015). Linguistic labels, dynamic visual features, and attention in infant category learning. *Journal of experimental child psychology*, *134*, 62–77.

Dink, J. W., & Ferguson, B. (2015). *eyetrackingr: An r library for eye-tracking data analysis.* www.eyetrackingr.com.

Eckstein, M. K., Guerra-Carrillo, B., Miller Singley, A. T., & Bunge, S. A. (2017). Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Developmental Cognitive Neuroscience*, *25*, 69–91.

Eimas, P. D., & Quinn, P. C. (1994). Studies on the formation of perceptually based basic-level categories in young infants. *Child Development*, *65*(3).

Fagan, J. F. (1978). Facilitation of Infants' Recognition Memory. *Source: Child Development*, *49*(4), 1066–1075.

Fantz, R. (1961). The origin of form perception. *Scientific American*, *204*(5), 66-72.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(24), 15822–6.

Gentner, D., & Namy, L. L. (1999). Comparison in the Development of Categories. *Cognitive Development*, *14*(4), 487–513.

Gliozzi, V., Mayor, J., Hu, J. F., & Plunkett, K. (2009). Labels as features (not names) for infant categorization: A neurocomputational approach. *Cognitive Science*, *33*(4), 709–738.

Goldstone, R. L. (1994). Influences of Categorization on Perceptual Discrimination. *Journal of Experimental Psychology: General*, *123*(2), 178–200.

Goldstone, R. L., Kersten, A., & Carvalho, P. (1978). Categorization and concepts. In J. Wixted (Ed.), *Steven's handbook of experimental psychology and cognitive neurocience: Language and thought* (pp. 275–317). New Jersey: Wiley.

Goldstone, R. L., Lippa, Y., & Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition*, *78*(1), 27–43.

Gredebäck, G., & Johnson, S. (2009). Eye Tracking in Infancy Research. *Developmental neuropsychology*, *35*(1), 1–19.

Grinstead, C. M., & Snell, J. L. (2003). *Introduction to probability*. AMS.

Gureckis, T. M., & Love, B. C. (2004). Common Mechanisms in Infant and Adult Category Learning. *Infancy*, *5*(2), 173–198.

Handy, T. C., Soltani, M., & Mangun, G. R. (2001). Perceptual load and visuocortical processing: Event-Related Potentials Reveal Sensory-Level Selection. *Psychological Science*, *12*(3).

Hillyard, S. A., & Anllo-Vento, L. (1998). Event-related brain potentials in

the study of visual selective attention. *Proceedings of the National Academy of Sciences*, *95*(3), 781–787.

Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development*, *13*(4), 341–348.

Huang-Pollock, C. L., Todd Maddox, W., & Karalunas, S. L. (2011). Development of Implicit and Explicit Category Learning. *Journal of Experimental Child Psychology*, *109*(3), 321–335.

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. In L. L. Rovee-Collier C (Ed.), *Advances in infancy research* (Vol. 6, pp. 69–95). Ablex: Norwood, NJ.

Hurley, K., & Oakes, L. M. (2012). Experience and distribution of attention: Pet exposure and infants' scanning of animal images. *Journal of Cognition and Development*, *100*(2), 130–134.

Jankowski, J. J., Rose, S. A., & Feldman, J. F. (2001). Modifying the distribution of attention in infants. *Child Development*, *72*(2), 339–351.

Johnson, S. P., Slemmer, J. A., & Amso, D. (2004). Where infants look determines how they see: Eye movements and object perception performance in 3-month-olds. *Infancy*, *6*(2), 185–201.

Jones, S. S., & Smith, L. B. (1993). The place of perception in children's concepts. *Cognitive Development*, *8*(2), 113–139.

Káldy, Z., & Leslie, A. M. (2005). A memory span of one? Object identification in 6.5-month-old infants. *Cognition*, *97*(2), 153–177.

Kemler Nelson, D. G. (1984). The Effect of Intention on What Concepts Are Acquired. *Journal of Verbal Learning and Behavior*, *23*, 734–759.

Kiefer, M. (2001). Perceptual and semantic sources of category-specific effects: Event-related potentials during picture and word categorization. *Memory & Cognition*, *29*(1), 100–116.

Kietzmann, T. C., Ehinger, B. V., Porada, D., Engel, A. K., & König, P. (2016). Extensive training leads to temporal and spatial shifts of cortical activity underlying visual category selectivity. *NeuroImage*, *134*, 22–34.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*, *83*(2), 35–42.

Kotovsky, L., & Gentner, D. (1996). Comparison and Categorization in the Development of Relational Similarity. *Child Development*, *67*, 2797–2822.

Kovack-Lesh, K. A., Horst, J. S., & Oakes, L. M. (2008). The cat is out of the bag: The joint influence of previous experience and looking behavior on infant categorization. *Infancy*, *13*(4), 285–307.

Kovack-Lesh, K. A., McMurray, B., & Oakes, L. M. (2014). Four-month-old infants' visual investigation of cats and dogs: Relations with pet experience and attentional strategy. *Developmental Psychology*, *50*(2), 402–413.

Kovack-Lesh, K. A., & Oakes, L. M. (2007). Hold your horses: How exposure to different items influences infant categorization. *Journal of Experimental Child Psychology*, *98*(2), 69–93.

Kovack-Lesh, K. A., Oakes, L. M., & McMurray, B. (2012). Contributions of Attentional Style and Previous Experience to 4-Month-Old Infants' Categorization. *Infancy*, *17*(3), 324–338.

Kurtz, K. J., Boukrina, O., & Gentner, D. (2013). Comparison promotes learning and transfer of relational categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(4), 1303–1310.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: the p300 as a measure of stimulus evaluation time. *Science*, *197*(4305), 792–795.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(3), 451–468. doi: 10.1037/0096-1523.21.3.451

Lin, E. L., & Murphy, G. L. (1997). Effects of Background Knowledge on Object Categorization and Part Detection. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(4), 1153–1169.

Livingston, K. R., & Andrews, J. K. (2005). Evidence for an age-independent process in category learning. *Developmental Science*, *8*(4), 319–325.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A Network Model of Category Learning. *Psychological Review*, *111*(2), 309–332.

Luck, S. (2005). *An introduction to the event-related potential technique*. Cambridge, Mass.: MIT Press.

Luna, B., Velanova, K., & Geier, C. F. (2008). Development of eye-movement control. *Brain and cognition*, *68*(3), 293–308.

Lupyan, G., Rakison, D. H., & Mcclelland, J. L. (2007). Language Is Not Just for Talking Redundant Labels Facilitate Learning of Novel Categories. *Psychological Bulletin*, *18*(12), 1077–1083.

Machery, E. (2009). *Doing without concepts*. Oxford University Press.

Mandler, J. M. (2000). Perceptual and conceptual processes in infancy. *Journal of Cognition and Development*, *1*, 3–36.

Mandler, J. M. (2004a). *Foundations of mind: Origins of conceptual thought*. Oxford University Press.

Mandler, J. M. (2004b). Thought before language. *Trends in Cognitive Sciences*, *8*(11), 508–513.

Mangun, G. R., & Hillyard, S. A. (1991). Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-

spatial priming. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(4), 1057–1074.

Mareschal, D., & Johnson, M. H. (2003). The what and where of object representations in infancy. *Cognition*(88), 259–276.

Mareschal, D., Plunkett, K., & Harris, P. (1999). A computational and neuropsychological account of object-oriented behaviours in infancy. *Developmental Science*, *2*(3), 306-317.

Mareschal, D., & Quinn, P. C. (2001). Categorization in infancy. *Trends in Cognitive Sciences*, *5*, 443-450.

Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin*, *129*(4), 592–613.

Markson, L., & Spelke, E. S. (2006). Infants' rapid learning about self-propelled objects. *Infancy*, *9*(1), 45–71.

Mather, E., & Plunkett, K. (2011). Same items, different order: Effects of temporal variability on infant categorization. *Cognition*, *119*(3), 438 - 447.

Mecklinger, A., & Ullsperger, P. (1993). P3 varies with stimulus categorization rather than probability. *Electroencephalography and clinical Neurophysiology*, *86*, 395–407.

Medin, D. L. (1989). Concepts and Conceptual Structure. *American Psychologist*, *44*(12), 1469–1481.

Minda, J. P., Desroches, A. S., & Church, B. A. (2008). Learning Rule-Described and Non-Rule-Described Categories: A Comparison of Children and Adults. *Journal of Experimental Psychology: Learning Memory and Cognition*, *34*(6), 1518–1533.

Mirman, D. (2014). *Growth curve analysis and visualization using r*. CRC Press.

Murphy, G. (2002). *The big book of concepts*. Boston, Mass.: MIT Press.

Murphy, G., Groeger, J. A., & Greene, C. M. (2016). Twenty years of load theory-where are we now, and where should we go next? *Psychonomic Bulletin & Review*, *23*(5), 1316–1340.

Navarro, D. (2015). Learning statistics with r: A tutorial for psychology students and other beginners. (version 0.5) [Computer software manual]. Adelaide, Australia.

Nyström, M., & Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior Research Methods*, *42*(1), 188–204.

Oakes, L. M. (2012). Advances in eye tracking in infancy research. *Infancy*, *17*(1), 1–8.

Oakes, L. M., Kovack-Lesh, K. A., & Horst, J. S. (2009). Two are better than one: Comparison influences infants' visual recognition memory. *Journal of Experimental Child Psychology*, *104*(1), 124–131.

Oakes, L. M., Plumert, J. M., Lansink, J. M., & Merryman, J. D. (1996). Evidence for Task-Dependent Categorization in Infancy. *Infant Behavior and Development*, *19*, 425–440.

Oakes, L. M., & Ribar, R. J. (2005). A comparison of infants' categorization in paired and successive presentation familiarization tasks. *Infancy*, *7*(1), 85–98.

Olejnik, S., & Algina, J. (2003). Generalized Eta and Omega Squared Statistics: Measures of Effect Size for Some Common Research Designs. *Psychological Methods*, *8*(4), 434–447.

Pascalis, O., & de Haan, M. (2003). Recognition mamory and novelty preference: What model? In H. Hayne & J. Fagen (Eds.), *Progress in infancy research* (Vol. 3, p. 95-120). New Jersey: Lawrence Erlbaum Associates.

Pauen, S., & Träuble, B. (2009). How 7-month-olds interpret ambiguous motion events: Category-based reasoning in infancy. *Cognitive Psychology*, *59*(3), 275 - 295.

Plunkett, K. (2010). The Role of Auditory Stimuli in Infant Categorization Infant Categorization. In L. Oakes, C. Cashon, M. Casasola, & D. Rakison (Eds.), *Infant perception and cognition: Recent advances, emerging theories, and future directions* (pp. 203–221). Oxford University Press.

Plunkett, K., Hu, J., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition*, *106*(2), 665 - 681.

Posner, M. I., & Keele, S. W. (1970). The retention of abstract ideas. *Journal of Experimental Psychology*, *83*(2), 304–308.

Quinn, P. C. (1987). The categorical representation information. *Cognition*, *27*, 145–179.

Quinn, P. C. (2002). Category Representation in Young Infants. *Current Directions in Psychological Science*, *11*(2), 66–70.

R Core Team. (2013). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org/`

Rakison, D. H. (2004). Infants' sensitivity to correlations between static and dynamic features in a category context. *Journal of Experimental Child Psychology*, *89*(1), 1 - 30.

Rakison, D. H., & Lupyan, G. (2008). Developing object concepts in infancy: An associative learning perspective. *Monographs of the Society for Research in Child Development*, *73*(1), 1–127.

Rakison, D. H., & Poulin-Dubois, D. (2002). You go this way and ill go that way: Developmental changes in infants detection of correlations among static and dynamic features in motion events. *Child Development*, *73*(3), 682–699.

Reed, J. M., Squire, L. R., Patalano, A. L., Smith, E. E., & Jonides, J. (1999). Learning about categories that are defined by object-like stimuli

despite impaired declarative memory. *Behavioral Neuroscience*, *113*(3), 411–419.

Robinson, C. W., & Sloutsky, V. M. (2007). Visual processing speed: effects of auditory input on visual processing. *Developmental Science*, *10*(6), 734–740.

Roder, B. J., Bushneil, E. W., & Sasseville, A. M. (2000). Infants' preferences for familiarity and novelty during the course of visual processing. *Infancy*, *1*(4).

Rorden, C., Guerrini, C., Swainson, R., Lazzeri, M., & Baylis, G. C. (2008). Event related potentials reveal that increasing perceptual load leads to increased responses for target stimuli and decreased responses for irrelevant stimuli. *Frontiers in Human Neuroscience*, *2*, 4.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic Objects in Natural Categories. *Cognitive Psychology*, *8*, 382–439.

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2003). Infant visual recognition memory: independent contributions of speed and attention. *Developmental Psychology*, *39*(3), 563–571.

Ross-Sheehy, S., Oakes, L. M., & Luck, S. J. (2003). The development of visual short-term memory capacity in infants. *Child Development*, *74*(6), 1807–22.

Rugg, M. (2001). Event-related/Evoked Potentials. In *International encyclopedia of the social & behavioral sciences* (pp. 4962–4966). Elsevier.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926-1928.

Schyns, P. G., Goldstone, R. L., & Thibaut, J. P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences*, *21*, 1–54.

Sigala, N., Gabbiani, F., & Logothetis, N. K. (2002). Visual categorization and object representation in monkeys and humans. *Journal of Cognitive Neuroscience*, *14*(2), 187–198.

Slater, A. (1989). Visual memory and perception in early infancy. In *Infant development.* (pp. 43–71). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Sloutsky, V. M., & Robinson, C. W. (2008). The Role of Words and Sounds in Infants' Visual Processing: From Overshadowing to Attentional Tuning. *Cognitive Science*, *32*, 342–365.

Smith, E. E. (2008). The case for implicit category learning. *Cognitive, Affective and Behavioral Neuroscience*, *8*(1), 3–16.

Sokolov, E. (1963). *Perception and the conditioned reflex.* Oxford, UK: Pergamon Press.

Sophian, C. (1980). Habituation is not enough: Novelty preferences, search,

and memory in infancy. *Merrill-Palmer Quarterly of Behavior and Development*, *26*(3), 239–257.

Spedicato, G. A. (2017). Discrete Time Markov Chains with R. *The R Journal*.

Tanaka, J. W., Luu, P., Weisbrod, M., & Kiefer, M. (1999). Tracking the time course of object categorization using event-related potentials. *Neuroreport*, *10*(4), 829–835.

Wang, Q., Bolhuis, J., Rothkopf, C. A., Kolling, T., & Knopf, M. (2012). Infants in Control: Rapid Anticipation of Action Outcomes in a Gaze-Contingent Paradigm. *PLoS ONE*, *7*(2), 30884.

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.

Weingarten, E., Chen, Q., McAdams, M., Yi, J., Hepler, J., & Albarracín, D. (2016). From primed concepts to action: A meta-analysis of the behavioral effects of incidentally presented words. *Psychological Bulletin*, *142*(5), 472–497.

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

Younger, B. A. (1992). Developmental change in infant categorization: the perception of correlations among facial features. *Child Development*, *63*(6), 1526–1535.

Younger, B. A. (2010). Categorization and concept formation in human infants. In D. Mareschal, P. C. Quinn, & S. E. G. Lea (Eds.), *The making of human concepts* (p. 245-263). New York: Oxford University Press.

Younger, B. A., & Cohen, L. B. (1986). Developmental Change in Infants' Perception of Correlations among Attributes. *Source: Child Development*, *57*(3), 803–815.

Younger, B. A., & Fearing, D. D. (1999). Parsing items into separate categories: developmental change in infant categorization. *Child Development*, *70*(2), 291–303.

Younger, B. A., & Furrer, S. D. (2003). A comparison of visual familiarization and object-examining measures of categorization in 9-month-old infants. *Infancy*, *4*(3), 327–348.

Younger, B. A., & Mekos, D. (1992). Category Construction in Preschool-Aged Children: The Use of Correlated Attributes. *Cognitive Development*, *7*, 445–466.

Zeithamova, D., & Maddox, W. T. (2009). Learning mode and exemplar sequencing in unsupervised category learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*(3), 731 – 741.