UNIVERSITY OF OXFORD



THE INTERACTION BETWEEN COGNITIVE AND LINGUISTIC CATEGORISATION IN EARLY WORD LEARNING

LOUKIA TAXITARI
WOLFSON COLLEGE

A thesis submitted for the degree of

DPHIL IN EXPERIMENTAL PSYCHOLOGY

Acknowledgements

I would like to express my deepest gratitude to my supervisor Professor Kim Plunkett whose help, support and expert guidance throughout the past four years proved valuable in the completion of this thesis.

I would also like to thank all the people who have been part of the Babylab Team at some point over the past few years: Vanja, Suzy, Nivi, Natalia, Emily, Julien, Jon-Fan, Valentina, my friend Rebecca, and Dr Gert Westermann, for their support, care and friendship, and our entertaining lunches.

I am also grateful to the Leventis Foundation for partially covering the expenses of my studies in Oxford.

My deepest love and gratitude is also due to my family and friends both in Oxford and in Cyprus for their never-ending encouragement during my time in Oxford. A special note of thanks is due to my housemates for putting up with the crankiness of the last stage of my write-up, and to Andreas Athenodorou who has been brilliant in every sense when putting up with me while finishing my thesis and for pushing me a little bit further when my patience was running out.

A very special thanks goes out to all those babies and mums who participated in my experiments and made my research a cheerful and sweet enterprise.

Any errors of fact or interpretation are, of course, entirely my own.

This thesis is dedicated to my parents, who have given me life and taught me how to live it, and Andreas, who has turned my Ithacas into charming, delightful journeys.

Short Abstract

The Interaction Between Cognitive And Linguistic Categorisation In

The Early Stages Of Word Learning

Loukia Taxitari, Wolfson College

Degree Of Doctor Of Philosophy, Michaelmas Term 2008

This thesis investigates the strategies infants use to generalise labels to different objects in the early stages of lexical development. It aims to directly test the assumption that a taxonomic bias exists which guides infants to extend words to categories of objects instead of individual instances of them, against the hypothesis that infants discover the extension of words through exposure to multiple naming instances of different objects.Experiments One and Two attempted to teach two object-label pairings to infants at the end of their first year of life, and test generalisation of those labels to new objects from the same adult linguistic categories. This aim failed because infants showed evidence for prior knowledge of the words. Experiments Three and Four employed a more infant-controlled procedure using a habituation task during training; in the former a single exemplar from each adult category was used, whereas in the latter multiple exemplars from each category were used. In both Experiments evidence for word learning was provided at test, but infants failed to generalise the labels to other objects. Experiment Five used a training phase identical to Experiment Four but tested infants for perceptual categorisation in the absence of any labels. Some infants showed evidence for their ability to create such categories on the basis of the training set, suggesting that the inability to generalise in Experiments Three and Four was not due to a perceptual limitation.

These findings suggest that infants at the end of their first year do not seem to be guided by any linguistic biases in their generalisation of labels. This thesis concludes that 10-month-old infants seem to have more advanced linguistic abilities than has previously been thought and constraint-like behaviour in later stages of lexical development might be a result of experience instead of a qualitative shift in cognitive processes.

Long Abstract

The Interaction Between Cognitive And Linguistic Categorisation In The Early Stages Of Word Learning

Loukia Taxitari, Wolfson College

Degree Of Doctor Of Philosophy, Michaelmas Term 2008

This thesis investigates the strategies infants use to generalise labels to different objects in the early stages of lexical development. Principles-and-constraints accounts suggest that infants are guided by the taxomonic constraint, a bias to extend a label to a class of objects, and that they should use their perceptual categorisation abilities during this process to generalise newly-learned labels to different objects. Exemplar-based accounts on the other hand suggest that infants build their linguistic categories in the same manner as they build their perceptual categories; initially they limit a word to a specific object and only gradually do they include more objects in the reference of their word, on the basis of experience and similarity with different within-category objects.

In order to investigate the mechanisms by which infants generalise labels to new objects at the early stages of lexical development, this research project focuses on infants at the end of their first year of life (10 months). The experimental task involves teaching infants two (real) object-label pairings and then testing their generalisation of labels to new exemplars from the same adult categories using an Intermodal Preferential Looking (IPL) task. On the basis of the theoretical approaches to categorisation (perceptual and linguistic) a decisive factor was manipulated in the design: amount of exposure to different naming instances. Additionally, similarity

and prototypicality of the objects were used to guide the selection of the exemplars, as defined by ratings from adult native monolingual speakers of British English.

Experiments One and Two were a first attempt to investigate word learning abilities of infants before their first birthday. In these experiments, infants were trained with two word-object pairs and then tested for word learning using the same typical exemplars presented during training as well as for word generalisation using new typical exemplars from the same adult category. A standard IPL task was administered at test; two pictures were simultaneously presented to infants along with a single word appearing exactly half way through the picture presentation time. Half of the infants were presented with a pretest block in order to test prior knowledge of the to-be-learned labels. Evidence for word comprehension was found in the pretest block, but in the post-training test blocks evidence for word learning was limited (and mainly in the form of distracter looking). This is the first evidence for the ability of 10-month-old infants to exhibit some preliminary categorisation ability during pretest, which allows them to recognise de-contextualised, stationary pictures of non-salient items as referents of a known label. However, the main aim of the study, namely to test generalisation of the labels to new objects, was made impossible to investigate.

Experiments Three and Four also used real objects, but less familiar to infants, and a more infant-controlled procedure during training. They only differed in whether infants were presented with a single exemplar (Experiment Three) or multiple exemplars (Experiment Four) from each of two linguistic categories ("cow" and "horse"). Principles-and-constraints accounts predict that infants should immediately generalise a newly-learned label to a whole category of objects; infants should, therefore, be willing to generalise labels in both Experiments. Exemplar-based accounts suggest that infants need experience with multiple naming instances with different objects before they begin generalising a label to different objects; infants should, therefore, generalise a newly-learned label in Experiment Four, but not in Experiment Three.

During test an IPL task was used. In Condition A infants were tested for word learning using the training exemplars, whereas in Condition B infants were tested for word generalisation using new within-category typical exemplars. In both Experiments infants spontaneously divided into two groups on the basis of whether they reached a pre-defined habituation criterion; these two groups exhibited different behaviours both during training and test. Evidence for word learning was found in both Experiments for infants who have habituated during training, but no evidence for word generalisation was found. This finding lends partial support to exemplar-based accounts, because infants seem to be unaware of the taxonomic constraint in their comprehension of words. However, was this failure of infants to generalise the newly-learned labels a linguistic "limitation", that is a lack of understanding that labels refer to classes of objects instead of individual instances of them, or rather a perceptual "limitation", that is an inability to perceive the training and testing exemplars as being the "same kind of thing"?

In order to disambiguate between these two explanations and in an effort to evaluate whether infants formed one or two categories during test, Experiment Five was designed. In this Experiment the training phase of Experiment Four was used, but during the post-training test blocks a novelty preference task without any labels was administered. Test Block A tested whether infants formed any categories on the basis of the training and test set; in each trial a new within-category ("cow" or "horse") and a new out-of-category ("tiger" or "bear") animal was presented. Test Block B tested whether one or two categories were formed; a task similar to Younger (1983) was used, during which a Modal stimulus containing the extreme values from each category (atypical exemplar) and an Average stimulus consisting of a combination of features from the two categories were presented. As in the previous experiments infants divided into two habituation groups. In Test Block A non-Habituators showed evidence for a novelty-preference by looking more at the

out-of-category exemplars, confirming that 10-month-olds are in position of using the training set for forming perceptual categories. In Test Block B Habituators preferred to look more at the Modal stimuli; the interpretation of this preference in Test Block B is rather problematic, since Modal and Average stimuli can only be conventionally defined when using real categories. It could be that these infants are showing a familiarity preference or that they perceive the Modal stimuli as different from the objects used in the familiarisation phase.

One perplexing finding was the inability of Habituators in Experiment Five to show evidence for category formation in Test Block A. In Experiment Four infants who habituated were those who exhibited evidence for word comprehension at test. In Experiment Five infants who habituated did not exhibit a novelty preference at test, and those who did not habituate showed evidence for discriminating between the within- and out-of-category novel objects. It is suggested that this pattern could be pointing to a period in development when infants are just beginning to discover the relationship between words and objects. Those infants who have already discovered this relationship could be attending to the word learning task during training, and possibly ignoring the categorisation part. These infants are those who have reached criterion in Experiments Four and Five. Infants who have not reached criterion might be those who have not yet discovered that there is something special about words and objects and are, therefore, finding the visual categorisation task more appealing, showing a novelty preference in Experiment Five and no word learning in Experiment Four. It is concluded that the inability of infants in Experiments Three and Four to extend the newly-learned labels to new within-category objects might be due to processing limitations, which prevent infants at this stage in development from exhibiting taxonomic behaviour in language tasks.

The objects used for testing word learning and word generalisation were all typical exemplars of the corresponding adult categories. In the training set of Experiment

Four objects of varying typicality were used. In order to choose these objects in a controlled manner three studies with adults were also conducted; native speakers of British English were asked to rate the objects for typicality and visual similarity. These ratings were taken into account when choosing the stimuli for the training and test blocks, and are described in Chapter Six.

This thesis has shown that infants at the end of their first year, and the early stages of lexical development, have started associating words to objects and are in position to identify newly-encountered, de-contextualised, (pictures of) objects as referents of familiar labels, indicating their ability to detach from the original word learning setting and exhibit evidence for "true" word learning. Additionally, these infants exhibited the ability to quickly form word-object associations in the laboratory over the course of a brief experimental session. These findings suggest that we should be very cautious when referring to biases and constraints during word learning: among the very first words of infants one can find both taxonomic generalisations and also failure to generalise, possibly depending on the amount of experience with the objects and their labels. This thesis concludes that it appears that infants at the end of their first year have more sophisticated linguistic abilities than has previously been thought and that lexical development might not proceed through qualitative shifts, but instead through the accumulation of experience with language and its use.

Contents

1 General Introduction **1**
- 1.1 Lexical Categories . . . 3
  - 1.1.1 Category Structure . . . 4
  - 1.1.2 Development of Lexical Categories . . . 13
- 1.2 Categorisation Abilities in Infancy . . . 22
  - 1.2.1 Non-Linguistic Categorisation . . . 23
  - 1.2.2 The Impact of Labels on Category Formation . . . 26
  - 1.2.3 Linguistic Categorisation . . . 30
- 1.3 Thesis Outline . . . 36

2 Experiments One and Two **38**
- 2.1 Introduction . . . 38
  - 2.1.1 Infants' First Words . . . 38
  - 2.1.2 Word Learning in the Laboratory . . . 39
  - 2.1.3 Methods for Investigating Word Learning . . . 41
  - 2.1.4 Current Research . . . 42
- 2.2 Experimental Design . . . 44
  - 2.2.1 Participants . . . 44
  - 2.2.2 Stimuli . . . 44
  - 2.2.3 Design . . . 48
- 2.3 Procedure . . . 49
- 2.4 Predictions . . . 51
- 2.5 Measures . . . 52
- 2.6 Scoring . . . 55
- 2.7 Analysis . . . 57
- 2.8 Results . . . 60
  - 2.8.1 Experiment One . . . 60
  - 2.8.2 Experiment Two . . . 79
- 2.9 General Discussion . . . 91

3 Experiment Three - Single Exemplar **95**
- 3.1 Habituation . . . 96
  - 3.1.1 Infant Attentional Preferences and Habituation . . . 96
  - 3.1.2 Habituation In the Laboratory . . . 98
- 3.2 Experimental Design . . . 100

3.2.1 Participants . . . 100
3.2.2 Stimuli . . . 100
3.2.3 Design . . . 103
3.2.4 Procedure . . . 104
3.2.5 Predictions . . . 108
3.2.6 Measures . . . 109
3.2.7 Scoring . . . 112
3.2.8 Analysis . . . 112
3.2.9 Training Block . . . 112
3.2.10 Test Blocks . . . 113
3.3 Results . . . 115
3.3.1 Training Phase . . . 115
3.3.2 Test Blocks . . . 117
3.4 Discussion . . . 133

4 Experiment Four - Multiple Exemplars 137
4.1 Creating Categories in the Laboratory . . . 138
4.2 Experimental Design . . . 139
4.2.1 Participants . . . 139
4.2.2 Stimuli . . . 140
4.2.3 Design . . . 141
4.2.4 Procedure . . . 142
4.2.5 Predictions . . . 145
4.2.6 Measures . . . 145
4.2.7 Scoring . . . 146
4.2.8 Analysis . . . 147
4.2.9 Training Block . . . 147
4.2.10 Test Blocks . . . 149
4.3 Results . . . 150
4.3.1 Training Phase . . . 150
4.3.2 Test Blocks . . . 153
4.4 Discussion . . . 160

5 Experiment Five - Categorisation 164
5.1 One or Two Categories: A Crucial Test . . . 166
5.2 Experimental Design . . . 167
5.2.1 Participants . . . 167
5.2.2 Stimuli . . . 167
5.2.3 Design . . . 169
5.2.4 Procedure . . . 171
5.2.5 Predictions . . . 172
5.2.6 Measures . . . 172
5.2.7 Scoring . . . 173
5.2.8 Analysis . . . 174
5.2.9 Training Block . . . 174

5.2.10 Test Blocks . . . 175
5.3 Results . . . 176
5.3.1 Training Phase . . . 176
5.3.2 Test Blocks . . . 179
5.4 Discussion . . . 184

6 Adult Rating Studies **187**
6.1 Experiments One and Two: Typicality Ratings . . . 188
6.1.1 Participants . . . 189
6.1.2 Stimuli . . . 189
6.1.3 Procedure . . . 190
6.1.4 Measures . . . 191
6.1.5 Analysis . . . 191
6.1.6 Results . . . 192
6.1.7 Discussion . . . 196
6.2 Experiments Three and Four: Typicality and Similarity Ratings . . . 197
6.2.1 Typicality Rating Study . . . 197
6.2.2 Similarity Rating Study . . . 206
6.2.3 Similarity Ratings and Typicality Ratings . . . 216

7 General Discussion and Conclusions **219**
7.1 Introduction . . . 219
7.2 Overview of Experiments . . . 220
7.2.1 Experiments One and Two . . . 220
7.2.2 Experiments Three and Four . . . 222
7.2.3 Experiment Five . . . 224
7.3 (True) Word Learning . . . 225
7.4 Word Generalisation . . . 228
7.4.1 Disambiguating between Different Accounts . . . 228
7.4.2 Conclusions . . . 230
7.5 The Impact of Labels on Category Formation . . . 231
7.5.1 Conclusions . . . 233
7.6 Attentional Preferences and IPL . . . 233
7.6.1 Target versus Distracter Looking . . . 233
7.6.2 Habituation and its Implications . . . 236
7.6.3 Conclusions . . . 237
7.7 Measuring Infant Attention in an IPL Task . . . 238
7.8 Evaluation . . . 238
7.9 Final Conclusions . . . 241

A Visual Stimuli in Infant and Adult Studies **243**
A.1 Infant Studies . . . 243
A.1.1 Experiments One and Two . . . 243
A.1.2 Experiment Three . . . 244
A.1.3 Experiments Four and Five . . . 245

A.2 Adult Studies 247
A.2.1 Typicality Rating Study One 247
A.2.2 Typicality Rating Study Two 250
A.2.3 Similarity Rating Study 250

B Communicative Development Inventory **251**

C Category Experience Questionnaire **260**
C.1 Introduction 260
C.2 The Questionnaire 263
C.3 Infant Data 273
C.3.1 Participants 273
C.3.2 Scoring 273
C.3.3 Analysis 273
C.3.4 Results 274
C.3.5 Discussion 282

References **285**

List of Figures

1.1 Tree-like taxonomy with different category levels. 5
1.2 Geometrical patterns used in Rosch (1973) to study categorisation of form. 9
1.3 Made-up animal sketches used by Younger (1985) in their categorisation study with 10-month-old infants. 25

2.1 Longest Look Difference for the two Trial Phases (pre- and post-naming), in the three test blocks, separately for Conditions A (Objects) and B (Animals). 61
2.2 Proportion of infants looking at the target in each 40ms window of analysis. 63
2.3 Proportion of infants fixating the target over the course of the trial in the first post-training test block. 65
2.4 Proportion of infants fixating the target over the course of the trial in the second post-training test block. 66
2.5 LLD in the two trial phases separately for each Comprehension group in the three test blocks. 69
2.6 Scatter-Plot showing the relationship between LLD and TCS in pretest block of Experiment One. 70
2.7 Latency of Shift to Target and Distracter in the Pretest Block. 71
2.8 TLT in the two phases of the trial separately for each test block in Experiment One. 73
2.9 Proportion of Target Looking before and after naming separately for each post-training test block in Experiment Two. 80
2.10 Proportion of infants fixating the target picture over the course of the trial in Test Block A. 83
2.11 Proportion of infants fixating the target picture over the course of the trial in Test Block B. 84
2.12 Proportion of target looking in the two trial phases separately for each test block and Comprehension group. 86
2.13 Total Looking Time to screen (TLT to target + TLT to distracter) for Experiment Two. 87

3.1 On-Screen Training Trial 107
3.2 Mean Looking Time presented separately for each target picture in Condition A (Word Learning) and Condition B (Word Generalisation). 116

3.3 Mean Looking Time for each Habituation group separately for each training sub-block. . . . 117
3.4 PTL for the two habituation groups in each test block and Condition separately. . . . 118
3.5 Longest Look Difference for the pretest block of Condition A for each target picture separately. . . . 121
3.6 Proportion of infants fixating the target picture as the trial unfolds in the test block of Condition A. . . . 122
3.7 LLD for the post-training test block in Condition A showing the two pictures separately. . . . 124
3.8 LLD for the two trial phases of the pretest block in Condition B showing the two target pictures separately. . . . 126
3.9 LLD for the two trial phases of the post-training test block in Condition B showing the two target pictures separately. . . . 127
3.10 Scatter-Plot showing the relationship between LLD and TCS in the test blocks of Experiment Three. . . . 130
3.11 Scatter-Plot showing the relationship between LLD and CEQ Score in the test blocks of Experiment Three. . . . 132

4.1 Mean Looking Time separately for each Category in the two Conditions of Experiment Four. . . . 151
4.2 Mean Looking Time presented separately for each Target Picture in the two Conditions of Experiment Four. . . . 152
4.3 Mean Looking Time for each training sub-block presented separately for the two Habituation groups. . . . 152
4.4 LLD for each test block separately for the two Conditions. . . . 153
4.5 LLD for each Habituation group in each test block and Condition separately. . . . 156
4.6 Scatter-Plot showing the relationship between LLD and CEQ score in the test blocks of Experiment Four. . . . 159

5.1 Average stimulus used in Test Block B of Experiment Five. . . . 170
5.2 Mean Looking Time separately for each Target Category. . . . 177
5.3 Mean Looking Time presented separately for each Habituation group in the two Conditions. . . . 178
5.4 Mean Looking Time for each familiarisation sub-block presented separately for the two Habituation groups. . . . 179
5.5 LLD for each test block and Habituation group separately. . . . 180
5.6 Scatter-Plot showing the relationship between LLD and CEQ score in the test blocks of Experiment Five. . . . 183

6.1 Mean Typicality Ratings for the animal and object categories separately. 193
6.2 Scatter-plot showing the relationship between the pre-defined typicality groups and the clusters produced by the K-means cluster analysis. 195
6.3 An example of a trial from the Typicality Rating Study. . . . 199

6.4 Relationship between Typicality Ratings and Reaction Times. 202
6.5 An example of a trial in the Similarity Study. 207
6.6 Similarity Ratings for pairs of pictures from the same category (Within) and from different categories (Across). 212
6.7 Similarity Ratings for the two categories, "cow" and "horse". 213
6.8 Relationship between Similarity Ratings and Reaction Times. 214

C.1 Percentage contribution of each super-ordinate category to the total CEQ score. 275
C.2 Relationship between CEQ Score and CDI Score (LEX2005 Database) 283

List of Tables

2.1 Percentages of infants reported to understand the words used in the Experiment according to the LEX2005 database. 46
2.2 Attention Phrases used in each phase of Experiments One and Two. . 47
2.3 Means and Standard Deviations for both PTL and LLD measures separately in the two phases of the trial for the pretest block. 62
2.4 Means and Standard Deviations for both PTL and LLD measures in the two phases of the trial for Test Block A. 67
2.5 Means and Standard Deviations for both PTL and LLD measures in the two phases of the trial for Test Block B. 67
2.6 Means and Standard Deviations for the TLT (Total Looking Time) measure separately in the two phases of the trial for the three test blocks. 74
2.7 Percentages of infants reported to understand the test labels in the LEX2005 database and the Oxford CDI word list. 77
2.8 Means and Standard Deviations for both PTL and LLD measures separately in the two phases of the trial and the two Conditions (Animals and Objects) in Test Block A. 81
2.9 Means and Standard Deviations for both PTL and LLD measures separately in the two phases of the trial and the two Conditions (Animals and Objects) in Test Block B. 82
2.10 Percentages of infants reported to know the test labels in the LEX2005 database and the Oxford CDI word list. 90

4.1 Total Looking Time (in seconds) for each Category and Habituation group in the two Conditions. 148
4.2 Number of presentations for each exemplar of Cow and Horse in the two Conditions. 148

5.1 Total Looking Time (in seconds) for each Category and Habituation group. 174
5.2 Number of presentations for each exemplar of Cow and Horse. 175

6.1 Mean Typicality Ratings and Reaction Times for the different predefined typicality groups (aggregated data from the six categories). . 193
6.2 Final Centres for the three-cluster grouping produced by the K-means analysis. Animals and Objects are presented separately. 195

6.3 Mean Typicality Ratings for the seven exemplars in the two categories. 202
6.4 Three Clusters provided from the K-means analysis, separately for each category. 204
6.5 Similarity Ratings for each same-category pair of objects. 217
6.6 Similarity Ratings for each across-category pair of objects. 218

C.1 Mean CEQ score for each Super-Ordinate Category in the questionnaire presented for each gender separately. 276
C.2 The object categories in the CEQ listed on the basis of their mean in descending order. 282
C.3 Top 10 object categories with the highest CEQ Score for the two genders separately. 283

If we had to name anything as the life of the sign, we would have to say that it was its use. Wittgenstein

Chapter 1

General Introduction

Imagine you find yourself in a foreign country, the language of which you do not understand (apart from a few words maybe). You are walking across a village with a native - who does not speak *your* language - when suddenly a rabbit hops in front of you chased by a dog. Your friend points to the poor little animal and its pursuer and cries "Gavagai"! You are happy you have just learned a new word and note it down in your dictionary of new words; gavagai = rabbit!

You have probably decided that the word should mean "rabbit", because you had already come across the word for dog - which is "Maranai" - and, therefore, that was immediately ruled out. You were pretty confident you found the real meaning of the word. But what could this exotic word mean? In fact, the word "Gavagai" could have an infinite number of meanings: "rabbit", "certain breed of dog", "my pet dog", "rabbit being chased by a dog", "white rabbit", "white rabbit being chased by a dog", "white rabbit being chased by a dog at this very moment", "rabbit's ear", "small furry animal", "something running before our eyes", "hunting", etc., etc.

The philosopher Quine uses this example to refer to the indeterminacy of trans-

lation (Quine, 1960); however, it nicely applies to the problem infants need to solve during word learning. Besides the infinite number of meanings a new word could have, infants manage to reach its correct meaning as shared by the linguistic community they form part of. So how is it that these naive and uninitiated word learners reach the meaning of words in the mist of this plethora of semantic interpretations? This is what has been termed as the "induction problem" in language.

This thesis is concerned with the way infants generalise newly-learned words to different objects during early lexical development and the mechanisms used to build one's lexical categories. By "lexical categories" we mean classes of objects to which a word refers. How are perceptual categories brought to bear in the word-learning task? What is the role of perceptual similarity in this process? In turn, what is the role of language in conceptual development? All these questions are hot potatoes in the area of cognitive development and on-going debates with experimental findings from different theoretical backgrounds abound.

It should become evident in this chapter that when talking about "categorisation" the different theories often do not differentiate between linguistic and perceptual/conceptual categorisation. This is due to two reasons: because labels appear to have a powerful impact on the formation of categories and concepts and because in research with adults it is very difficult to investigate categorisation abilities separately from language. When describing categorisation, one gets the feeling that most human categories are encoded linguistically through the use of labels - and this is very often the case - but this is not the full story. Humans are able to create categories "on the go" and without necessarily a label to go with them, like for example "things thrown all over the floor in the living room". Even though we are able to create such casual categories, most of the categories in our heads go together with words,

and therefore research in adult categorisation often does not separate one from the other. However, this separation between lexical and non-lexical categories is more pronounced in infancy, because infants in their first years do not always have any words available to go with the categories they form, although some of their perceptual categories are available from the first months of life.

In order to better understand the issue of how babies generalise words in their first year of life, we need to take a closer look at the theoretical views and experimental data on lexical categories and their development and their interaction with other cognitive categorisation abilities of infants in their first year. The first part of this chapter is devoted to "lexical categories", their structure, characteristics and development, and a presentation of the prominent theoretical accounts aiming to explain it. The second part of the chapter offers a description of infant categorisation abilities, both linguistic and non-linguistic in nature. The final part of this chapter provides a layout of the thesis.

1.1 Lexical Categories

Categorisation is a fundamental ability underlying much of cognition. It constitutes a means to reduce the load on memory and other cognitive processes (Rosch & Mervis, 1975): "Categorisation is assumed to occur when observers respond in a equivalent manner to discriminably different stimuli" (Quinn, Westerlund, & Nelson, 2006) or exhibit "a recognised equivalence among stimuli, objects or events that are discriminably different" (Cohen, 2003). Evidently, these definitions emphasise the fact that responses to *different exemplars* of a category are similar even when these items are *discriminable* from each other. This ability allows humans, adults and infants

alike, to organise efficiently the world around them into meaningful groups, create expectations about the objects/events they come across and make inferences about newly-encountered objects/events. This considerably reduces the processing load and frees memory from having to retain redundant information about each individual instance of a category.

Speakers of a linguistic community seem to share multiple conventions about the words of the language and their usage; there seems to be an agreement between them about which objects in the world can be called "dogs" and which objects cannot. Words - with the exception of proper nouns - typically refer to classes of objects and not just to individual instances of them. The word "dog" for example might be used to denote one's pet, but it is also used to refer to dogs in general.

1.1.1 Category Structure

Categories of objects in the real world have been shown to be highly determined and structured (Rosch & Mervis, 1975; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976); they form complex hierarchical nets (*taxonomies*) with different levels of informativeness. Such a taxonomy is depicted in Figure 1.1.1. At the highest branches of these tree-like taxonomies, categories of sufficient abstraction exist, so that they have few, if any, attributes common to all members (such as "animal", "vehicle", "furniture", etc) - categories at this level are called *superordinate*. The higher level branches down to *basic-level* categories; these have been shown to be the most informative (in terms of sharing the most attributes across items within a category) and also the most differentiated from one another (in terms of sharing the least common attributes with items from other categories, for example "dog", "bus", "table", etc). Finally basic-level categories further branch down to *subordinate* categories, which are characterised by high homogeneity among members, and thus

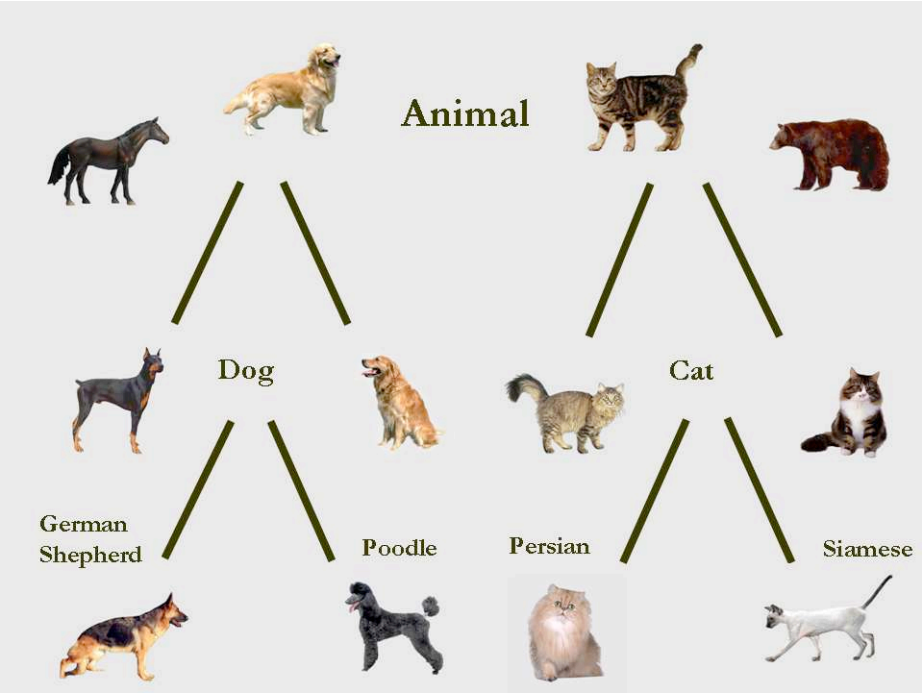


Figure 1.1: Tree-like taxonomy with different category levels: the top branch corresponds to the *superordinate* level, which is often characterised by high abstraction and variability among its members - such as "animal". Further down the tree, we find *basic-level* categories, which are the most informative in terms of category membership and the most differentiated from one another - such as "cat" and "dog". Finally the lowest branch is the *subordinate* level, which contains categories with high homogeneity among their members - such as "Siamese cats" or "German Shepherd dogs".

narrow down the set of items which belong to them even more (such as "Labrador", "German Shepherd", "Akita Inu" for the basic-level "dog" category) (Rosch & Mervis, 1975; Rosch et al., 1976).

Basic-level categories have been shown to bear a special psychological status; this level of categorisation has been shown to be the most inclusive level at which a generalised shape of category exemplars is identifiable and imaginable, at which different exemplars elicit similar motor movements and of which the names are first acquired by children (Rosch et al., 1976). However, research on conceptual hierarchies

has also shown that this special status of the basic-level can be modified on the basis of experience; that is, experts and novices perform differently in categorisation tasks and the former appear to use the subordinate level as their basic-level of categorisation (Tanaka & Taylor, 1991; Chi, Feltovich, & Glaser, 1981; Lynch, Coley, & Medin, 2000; Mervis, Johnson, & Scott, 1993). It, therefore, appears that categorisation and knowledge of the external world can be continually reshaped and altered by learning and experience throughout human development.

Categories are assumed to be composed of a *core meaning* which consists of the best examples, the "clearest cases", which are "surrounded" by other category members of decreasing similarity to that core meaning. Thus the items comprising the category differ in terms of *prototypicality*, of their perceived fitness in the category. If people are asked to rate the "goodness" of different members of a category (say the category "dog"), they indicate that some instances (such as "Labrador") are more representative of the category than others (such as "Chihuahua"), although they all qualify for membership.

According to prototype-based accounts of categorisation, if one thinks about categories as sets of objects, *typical* items are perceived as clustering around the centre of the category, whereas *atypical items* spread toward the periphery. It has been shown that typical items within a category share the most attributes with each other while typical items across categories have the least attributes in common with items from other categories (Rosch & Mervis, 1975; Rosch et al., 1976). Even natural kinds, where categories should consist of essential features and have clear boundaries (an animal can either be a dog or not, and it cannot have an intermediate status), they still exhibit prototypicality effects. An item is therefore categorised as A, and not B, if it is more similar to A's best representation - it's "prototype" - and not to B's (Taylor & Gelman, 1989). Different prototype-based theories have different definitions for what a prototype is: it could be an abstract mental representation, the mean of

attributes that have a metric in a category (Rosch & Mervis, 1975; Rosch, 1998; Murphy & Medin, 1985; Quinn & Oates, 2004; Rips & Medin, 2005), or a summary representation of the attributes characteristic of the category (Hampton, 1979, 1987; R.M. & Zaki, 2002; Smith & Medin, 1981) - for a discussion see (Murphy, 2002). More recently an alternative solution has been given to what constitutes a prototype with the introduction of *schemata* to represent concepts. *Schemata* are structured representations of concepts which divide up the features of objects into dimensions and values on those dimensions. This allows to differentially weight the different features and place restrictions on what values they can take, providing a more strict control over what constitutes a prototype for a category (Markman, 1999).

Exemplar-based accounts are one line of thought within the prototypicality approach; they assume that an object is categorised as A, and not B, if it is more similar to the items belonging to category A than it is to those belonging in category B. According to this view and in agreement with the basic premise of prototype-accounts, categories are structured entities; the important dimension however is the distance of the members from each other (or from the most typical exemplar), and not their distance from an abstract mental representation. Category members which are highly similar to each other cluster together in the centre of the category space (and not around an abstract representation); otherwise items spread toward the periphery of the category if they are more distant in terms of similarity. The prototype of the category becomes the exemplar which falls in the centre of the category on the basis of its similarity to the rest of the category members (Brooks, 1978; Medin & Smith, 1984; Medin & Schaffer, 1978; Medin & Barsalou, 1987).

Whether based on an abstract mental representation or built around (a) specific object(s) of the outside world (exemplars), prototype-based accounts emphasise one important dimension of category membership: a perceived prototypicality of category members exists and this gives rise to different degrees of *goodness* for membership in

a category. This phenomenon is crucially dependent on the high *similarity* of within-category items with each other, and increased dissimilarity with items from other categories (Rosch, 1998).

Sources of Prototypicality

For a limited number of categories prototypicality appears to be a property of the human perceptual system: Rosch's work on colour and form categories makes exactly this point; even categories describing sensations (such as colour, or certain geometrical forms or spatial orientations) exhibit graded degrees of membership. Figure 1.1.1 shows a set of geometrical patterns used in Rosch, 1973, study on form: participants came from a stone-age tribe whose language lacks words for geometrical forms. They were trained with form categories created around different prototypes and were then tested for category membership, but were also asked to indicate which exemplar they thought was more typical. Participants both learned the categories easier when they were organised around the first square and they chose the first square as the most typical example of the category, even when the category was not organised around that first exemplar. This was suggested to be due to the fact that certain exemplars are more salient in perceptual terms and thus end up acquiring the status of a prototype. Other members which deviate from this perceptually salient pattern are perceived as less representative of the category, and thus become peripheral, or atypical, members (Berlin & Kay, 1969; Heider, 1970, 1971, 1972; Rosch, 1973; Rosch & Mervis, 1975; Rosch et al., 1976).

This cannot be the case with artifact categories though; there is nothing inherent in the perceptual system which makes certain tables or cars more typical than others.

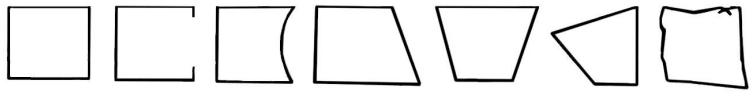


Figure 1.2: Geometrical patterns used in Rosch (1973) to study categorisation of form. Participants both learned the categories easier when they were organised around the first square and they chose the first square as the most typical example of the category, even when the category was not organised around that first exemplar. This was interpreted as the outcome of perceptual salience, which gives an exemplar the status of a prototype.

However, Rosch suggests that, even though a natural prototype is absent from those categories, they could still be organised around an artifact prototype. In this case, factors other than perceptual salience have been suggested for creating prototypicality effects (Taylor, 1995):

- *Frequency*: typical items appear more often (and we use their names more often as well) and thus become more central in a category (Rosch et al., 1976).

- *Order of learning*: typical items are not only the most frequent instances of a category but are also those encountered first. Therefore the category evolves around those first exemplars which thus become central (Rosch et al., 1976).

- Prototype instances might be those embodying the *mean values of variable attributes* which are important for a given category. They therefore become more central because they constitute the *average stimulus* for that category (Rips, Shoben, , & Smith, 1973).

- Prototype instances might be those embodying the *most salient attributes*: some features are more salient than others either in cultural or perceptual terms. This makes exemplars with these attributes more typical than those without (Wierzbicka, 1985).

- *Efficiency of prototype-categories*: cognition could benefit greatly from a flexible system of categorisation which allows new items to be incorporated in existing

categories as peripheral members, even if they fail to satisfy all criteria for membership in the category. Even if the world has a highly correlated structure, the huge amount of information presented to us daily often exhibits attributes far from perfectly correlated. Therefore classical criterial theories of categorisation could come across great difficulties and increased redundancy if they had to make a different decision every time an item did not satisfy all criteria for category membership (Taylor, 1995).

- *Formation of Concepts*: instead of being retrieved from memory to represent categories, concepts might be constructed "on the go" on the basis of goals, context and experience, by retrieving generic and episodic information from long-term memory and creating temporary concepts in working memory. This means that the structure of categories is unstable and different concepts might temporarily represent the same category in different occasions (Barsalou, 1987) or that the cognitive models people build about the world might not match the world perfectly giving rise to prototypicality effects (Fillmore, 1982).

Perceptual, Conceptual and Lexical Categorisation

In discussions about categorisation one often gets the feeling that categories are not differentiated between different kinds; this is sometimes true about adult categories, because by then it is very difficult to separate perceptual, conceptual and linguistic information in already established categories. In developmental research, however, it is important to determine the nature of categories we are investigating, as well as the kind of information available to infants at each stage in development. A brief description of the types of categories follows, along with a brief discussion about the main debates in the field.

Concepts are defined as mental representations of classes of things, which are formed on the basis of different features, perceptual, linguistic, functional, cultural, etc,

whereas *categories* are used to refer to the classes themselves (Murphy, 2002; Cohen, 2003).

Perceptual categories are groups of objects formed on the basis of perceptual similarity of the objects. Perceptual similarity is defined by the physical appearance of objects resulting from (combinations of) perceptual features (e.g. shape, colour, texture, smell, etc.) (Mandler, 2000; Cohen, 2003).

Lexical categories are groups of objects which share a common name, e.g. "dog". These categories could be based on any kind of information, perceptual or conceptual. In the adult stage, these categories usually correspond to concepts - dog, cat, room, freedom, etc, although there can be concepts which are not lexicalised (Taylor, 1995).

A major debate in the developmental literature concerns the relationship between conceptual and perceptual categorisation; namely, are they different kinds of categorisation, available to infants at different stages (or even in parallel) or do they involve a single process which matures and becomes more sophisticated? According to Mandler (1993,2000) perceptual and conceptual categorisation involve two independent and fundamentally different processes: perceptual categorisation computes the perceptual similarity of objects on the basis of their physical appearance, creating perceptual schemas of what objects are like. Conceptual categorisation on the other hand involves redescription of the perceptual information into image schemas on the basis of the roles of objects in events, in a theory-like way (See Section 1.1.2). This abstraction of the perceptual information involves deliberation, meaning and interpretation and gives rise to forms of conceptual primitives, such as animacy, causality and containment. These two types of categorisation coexist from early in life and act in parallel; thus, infants early concepts are supposed to be global (corresponding to super-ordinate categories), such as "animal", and only later, as infants begin to use more feature dimensions in their concepts, do they become more specific and differentiated (corresponding to basic-level categories), such as "dog",

"cat", etc. (Mandler, 2000; Mandler & McDonough, 1993; Cohen, 2003).

Other researchers, however, have argued that it is unnecessary to invoke such a dual system of categorisation, which is at large unparsimonious; instead, it can be assumed that abstract conceptual information is derived from perceptual representation. According to this view, the two kinds of categorisation emerge as a result of changes in infants' cognitive abilities (linguistic, motor, attentional, cognitive, etc), which provide with new types of information which in turn serve to modify early perceptual categories. Thus, conceptual development is suggested to begin at a concrete - or basic - level (e.g. dogs) and then gradually, as features become abstracted, concepts begin to be generalised to various instances building superordinate categories (e.g. cats and other animals) (Rosch, 1998; Younger, 2003; Eimas & Quinn, 1994; Rakison, 2003; Madole & Oakes, 1999; Quinn & Eimas, 1997).

Because this is an on-going debate, this thesis will not be concerned with the kinds of categorisation available to infants during lexical development; what is of interest here is how cognitive categorisation abilities, whether perceptual or conceptual, interact with lexical development in the early stages of word learning. Thus the term "cognitive categorisation" will be here used to denote any kind of categorisation, other than lexical, infants might be using during lexical development, unless we need to specify which type of categorisation is being referred to.

The second debate concerns the relationship between cognitive categorisation and language, and it is the focus of this thesis. What types of cognitive categorisation are available to infants at each stage in lexical development? How do the two interact? Does cognitive categorisation guide lexical development? And does lexical categorisation affect cognitive development? These issues will be discussed in detail below.

1.1.2 Development of Lexical Categories

Speakers of a linguistic community seem to share implicit conventions for words, their meaning and usage. Adult native speakers of a language have common agreements about what constitutes the meaning of a word, about the properties which need to exist or the criteria which need to be satisfied before an item is considered to be an instance of a given linguistic category. This is what, in Linguistics, is referred to as the *intension* of a word. If we take the now classic example of the word "bachelor", the intension of the word is: "unmarried", "man" and "of marriable age"; one needs to be a grown-up man and not married to qualify as a "bachelor".

Speakers of a language also share conventions about the objects (or events) a word refers to; in other words, they (roughly) agree as to the set of things (or events) to which a word applies. This is what, in Linguistics, is referred to as the *extension* of a word. Speakers of a language agree which things can be called dogs and which cannot, even if they cannot describe the properties or criteria that make them belong to the "dog" category. Often the intension of words cannot be consciously reflected upon; one can identify an object as being an X or a Y, but might not be able to explain why this is so. This is more evident with some types of words (for example, colour terms or words for other sensations) than others (for example, object words, which seem to be the easiest and most unanimously defined by speakers of a language).

Wittgenstein emphasises this point when describing the meaning of the word "game": although speakers of a language have no problem deciding which instances belong to the category, it might become extremely difficult to define what the criterial attributes for belonging to that category are. This becomes more evident if we consider the different instances of the word "game": board-games, card-games, ball-games, Olympic games, etc. There are no features which are common to them all, but there is instead a *complicated network of similarities overlapping and criss-crossing: sometimes overall similarities, sometimes similarities of detail* (Wittgenstein, 1978).

During development, it seems that infants occasionally "mess up" with their usage of words, which often does not agree with what adult words are used for. For example, an infant might use the word "dog" to refer to only her pet dog, while another infant (or even the same one at different stages during language development) might use the word "dog" to refer to both cats and dogs, "committing" what we (linguists, psychologists, developmentalists etc) call under- and over-extension errors, respectively. However, eventually children figure it out and gradually they build a semantic system in a way which agrees with the adult (semantic) representations of the world.

But how do infants discover the extension of words? Different theories make clearly different predictions about how infants extend labels to objects throughout lexical development. Classical theories of category formation assume that one of infants' tasks during word learning is to discover the relevant features for each semantic category and so the process of matching their categories to adult ones might take some time. At the beginning, infant hypotheses are assumed not to be restrictive enough and therefore there exists an initial period of overextensions until baby discovers the correct features for the category in question and narrows down the set to which the label could correctly be applied (Clark, 1973) - a process similar to that of forming their perceptual categories.

This theory, however, has suffered some blows, on the basis of the fact that the semantic features advocated are characterised by vagueness and the process of discovering a word's extension cannot account for under-extension errors (Merriman, 1986). Two main theories currently aim to account for this aspect of language development: principles-and-constraints and exemplar-based accounts. These will be described below.

Principles-and-constraints Accounts

One line of thought assumes that in order to get around the *induction problem* (Gelman & Markman, 1986) and discover how to successfully extend labels to new items, infants rely on a number of (linguistic) constraints. Quine refers to the *induction problem* during word learning to emphasise the fact that, despite the infinite number of hypotheses infants can come up with when they hear a new word, they still manage to assign labels to the correct referents. Even with ostensive definitions - when for example mummy points to something and says "rabbit" - a baby can never be sure if the word refers to the whole animal, that specific animal, as a proper name would, a part of it, its texture, etc. Linguistic constraints undertake the task of limiting the hypotheses space of infants and guiding them to the correct interpretations of words. The most important ones are:

- *the whole-object constraint*: a bias to attach labels to whole objects, and not to parts or properties of them (Markman, Wasow, & Hansen, 2003)
- *the mutual exclusivity constraint*: a bias to attach a new label to an as yet unlabeled object(Merriman & Bowman, 1989)
- *the taxonomic constraint*: a bias to attach labels to whole categories of objects, and not just to individual instances (Markman & Hutchinson, 1984)

Of relevance to this discussion is the taxonomic constraint which is responsible for guiding children to generalising words to the appropriate objects. Some researchers assume that the taxonomic constraint is operant from very early on, when infants begin to learn their first words (Waxman & Markow, 1995; Behrend, Scofield, & Kleinknecht, 2001). Waxman (2003) suggests that infants enter the word learning task having an appreciation of the linkage between words and object categories; even preverbal infants are assumed to be able to use linguistic information in categorisation tasks, in a way that words highlight the commonalities among objects and facilitate

object categorisation (Waxman, 2003; Waxman & Braun, 2005). Others believe that this constraint is probably not present from so early on, but should be available at least at the beginning of the vocabulary spurt, that is sometime around the middle of the second year of life (Markman & Hutchinson, 1984; Hollich, Hirsh-Pasek, & Golinkoff, 2000).

The concepts-in-theories model suggests that infants enter the task of forming a lexicon with certain explanatory (high-level and abstract) biases which guide their first guesses about the properties which are most strongly causally linked to the central structure of a category. Thus word meaning and lexical categories are considered to be based not on perceptual regularities, but instead on "overarching distinct patterns of explanation associated with such domains as biological kinds, artifacts and sentient beings" (Keil, 1994). The assumption that several of these patterns exist points to a development of semantics which is based on existing cognitive structures and the switching among them, and not the acquisition of totally new forms of explanation (Carey, 1985; Keil, 1994). In other words, infants have already done most of the work in figuring out their environment by the time lexical development begins and they use their existing cognitive structures to build their developing semantic system, without the need to come up with new ones.

Nelson assumes that early words, present in the infant lexicon from ten to thirteen months of age, are matched to existing concepts and are likely to be *underextended*, as the child's concepts might be based on a single instance (Nelson, Rescorla, Gruenderl, & Benedict, 1978). Generalisation to new instances will be based on perceptual similarity of new items to the original item as well as event representations acquired through the infant's repetitive experience with instances of the particular type of event. A bit later on, a fundamental transition takes place, allowing infants to use their words in a *decontextualised* manner: event representations are replaced by specifications of the types of objects or action-schemes which are referred to by

particular words.

Nelson also suggests that a 12-month-old infant *sets out to assimilate names to well-formed concepts* (Nelson et al., 1978). Concepts based on functional relations - for example [things that move] - and static perceptual features are created first and infants are motivated to attach names to these concepts. Under- or over-extensions in this kind of lexical development are the result of generalisation on the basis of the "wrong" (in terms of adult categories) perceptual features.

All in all, what principles-and-constraints accounts assume is that infants will generalise a newly-learned word to whatever objects they consider as belonging to the same category as the object they have learned the word with; otherwise they will not. Whatever mistakes happen to occur along the way are due to a conceptual limitation, and not to a linguistic one. That is, infants understand that their words should refer to whole groups of objects, and not just to individual instances of them, and failure to apply them correctly is due to their failure to discover the correct corresponding category.

Exemplar-based Accounts

Exemplar-based theories on the other hand assume that there is no such thing as biases and constraints, and labels are extended on the basis of experience with different items (Barrett, 1986; Bowerman, 1978; Brooks, 1987). Tomasello's Usage-Based theory of language acquisition explicitly rejects principles and constraints and suggests instead that the infant brain is a powerful learning mechanism able to combine intention-reading (social aspect) and pattern-finding (statistical aspect) skills in the task of discovering the meaning and extension of a word. Infants are in position to follow the parent's attention during their word-learning task, and at the same time trace the statistical regularities in their environment in order to infer

which sound goes with which object (Tomasello, 2003).

Exemplar-based accounts assume that initially infants associate labels with specific items - the items they hear a label occurring with; only gradually and after exposure to a considerable number of naming instances do infants discover that labels not only refer to a whole range of items, but they also tend to occur with items that share certain properties. An initial period of *under-extension* is therefore assumed until baby learns to generalise the label to more and more members of a given category. Gradually infants come to appreciate that a newly-learned label should be generalised on the basis of similarity to a whole category of items. This approach predicts that initially cognitive categorisation might influence linguistic categorisation, whereas the latter should in its turn be able to influence the cognitive (re-) organisation of the input in later stages of development (Smith, Colunga, & Yoshida, 2003).

According to Wittgenstein (1978), we learn the extension of categories through experience with different exemplars; in order to learn the word "game", one is exposed to different instances of the category and gradually extends the word on the basis of similarity with the named instances (Wittgenstein, 1978).

Exemplar-based accounts assume that, at least initially, language does not interact with perceptual/conceptual categorisation. Lexical categories are instead created in pretty much the same way as perceptual categories, as described above. Bowerman's prototype model of word meaning is based on Rosch's work on categories and suggests that infants will initially attach a label to an abstracted underlying representation of the best exemplar and only subsequently will they extend the word to other exemplars which share the principal features with that representation. This assumed mental representation of the object corresponds to a prototypical referential exemplar of the word which functions as the specification of the type of object that can be labeled with the word (Bowerman, 1977, 1978). Initially, infants tend to *underextend* labels to items which resemble the original item in at least one feature. Gradually they

identify and abstract the principal perceptual or functional features characterising the prototype. The label is therefore extended to more and more referents, and at this stage it can also be *overextended* to items which are not part of the adult category but share a number of features with the prototype. People have assumed that these features are mainly perceptual in nature, and only a minority of the overextensions reported are based on functional properties of the objects (Clark, 1973; Bowerman, 1977; Rescorla, 1980). Exposure to multiple naming instances as well as similarity with other objects are, therefore, the important dimensions in generalising words to new objects (Brooks, 1987; Bowerman, 1978; Rosch, 1998).

Hybrid Accounts

Between the two extremes - principles-and-constraints and exemplar-based accounts, other models have been proposed in order to account for the diversity of findings on early words' extensions and the changes they undergo throughout language development. Barrett's multi-route model of early lexical development (Barrett, 1986) suggests that not all early words exhibit the same behaviour and follow the same route toward adult conventional meaning. *Context-bound and social-pragmatic* words are mapped onto holistic event representations, while *referential* words are mapped onto mental representations of categories organised around prototypes (Bowerman, 1978). Barrett further assumes that there are five stages in the acquisition of a word's meaning, but not all words need to go through all five stages:

- Level A: mapping of a word onto an event representation or the representation of a prototype of a category.
- Level B: modification of event representation, by adding actors, objects, etc.

- Level C: extraction of a single constituent in the event representation and use as the prototypical referent of the word.
- Level D: expansion of the referential scope of context-bound words and identification of principal features which characterise the prototype.
- Level E: refinement of underlying semantic representation of prototype - organisation of words into semantic fields, recession of overextension and contraction of referential scope.

This model satisfies Nelson's assumption that initially words are attached to well-formed concepts of events which occur in infants' daily lives (principles-and-constraints accounts). It also satisfies Bowerman's assumption that words are initially attached to individual exemplars, and only after a mental representation is abstracted from the best exemplar do words become generalised to other exemplars which share the principal features with it (exemplar-based accounts).

More recently, Hollich et al. (2000) have proposed the Emergentist Coalition Model (henceforth ECM) (Hollich et al., 2000). Word learning is considered as a complex multi-stage process, during which infants are able to make use of different cues ranging from statistical regularities in the input to language-specific principles; the focus of interest is shifted to the strategies available at each stage of language acquisition, their developmental trajectory and most crucially the driving force for abandoning one strategy and adopting another. The ECM suggests that principles and constraints are the outcome of attentional/associationistic factors in early development, which then become the "engines" of subsequent development, as each principle emerges from an immature to a mature state. Initially, infants use words in an associationistic fashion, and restrict them to the objects they have heard them occurring with; later, at around the second half of the second year of life, taxomonic

behaviour begins to emerge, as the product of experience with language and the use of words.

Prototypicality and Similarity. Exemplar-based accounts assume that generalisation is based on the similarity between the different members of a category; prototypicality-based accounts, such as Rosch's theory of categories described above, assume that prototypicality is the basis of generalisation. Both of these terms can be quite obscure on their own, unless clearly defined.

According to Tversky (1977), "prototypicality" as well as "similarity" are both proximity relations: the first holds between an item and a class (or category) and the second between two items. Similarity itself is a rather problematic term: items can be similar to each other in different dimensions, but what is it that defines which attributes are important for each category? These considerations are nicely captured by Tversky's (1977) "attribute model of similarity": prototypicality is computed as a function of the similarity between two items; this is expressed by the number of attributes shared by both, the number of attributes they do not have in common, as well as by the difference in attribute weighting. That is, some attributes could receive high weighting due to their perceptual salience (which remains relatively constant) while some others could be highly weighted because of their high diagnostic value (which changes considerably according to context). For example, in the category "bird", [has wings] or [has feathers] is a perceptually salient attribute and it should be weighted highly in deciding whether something is a member of the category, whereas [ability to fly] might not be an immediately salient attribute for deciding about category membership. At the same time, the attribute [has feathers] has high diagnostic value for category membership in an urban setting (i.e. if we are taking a walk in the parks), but might lose its value in the North Pole, if we are deciding

whether a penguin belongs to the "bird" category or not.

Perceptual salience then seems to play a crucial role in similarity judgments but it is only one dimension in computing similarity between items. Other dimensions of the relationship between two items could play a role, and these could change according to (cultural, environmental, social, functional) context. It is, therefore, very difficult to decide which perceptual features are the most salient for infants when investigating category structure and category formation. This issue will be discussed below.

Other researchers, however, have pointed out that similarity alone is not informative enough, in order to account for the formation of concepts and categories. That is, such theories provide no explanation as to why similarity on one aspect might be more important than similarity on a different aspect. Instead people use their background knowledge of *theories* about the world for including objects in a category and deciding which objects should be grouped together and which objects should not. *Theories* are used here to mean mental explanations about the world, and not any kind of scientific account. Infants are assumed to share such theories as well and early concepts are supposed to be based on them; however, the initial hypotheses derived by early theories may be wrong and will, therefore, have to be modified before infants can settle on the correct concepts (Murphy & Medin, 1985).

1.2 Categorisation Abilities in Infancy

Principles-and-constraints accounts assume that infants bring their perceptual categorisation abilities to bear in the task of word learning. Below is a description of categorisation abilities of infants in the first year of life, as well as any findings available about the interaction between perceptual categorisation and language.

1.2.1 Non-Linguistic Categorisation

The ability to respond to stimuli categorically is available to humans from very early in life, although the categories formed undergo considerable refinement throughout development (Oakes & Madole, 2003). Quinn and Eimas (1994) used a novelty-preference procedure to show that infants from as early as 3-4 months are capable of forming basic- (and global-)level representations of visual stimuli: infants were familiarised with multiple cats until their attention declined and then presented with pictures of other animals. A renewal of interest in response to the test items suggested that they saw something they considered different from what they were familiarised with. Infants were shown to be able to form a representation of "cat" which excludes "bird", "dog", "horse" and "tiger", but not "female lion" (while being able to discriminate between different members of their "cat" category). It is not until their sixth month that infants start to form more adult-like perceptual categories and exclude "female lions" from their representation of "cat" (Eimas & Quinn, 1994).

Younger's work on the development of perceptual categorisation gives rise to a developmental progression from processing features independently (at around 3 months) to processing feature relations within objects (around 7 months) to processing feature relations within categories (around 10 months) (see also (Gibson, 1967; Clark, 1973)). The ability to detect feature relations is supposed to be present from early infancy, but the types of features and feature relations infants are able to perceive changes with age. That is, younger infants respond to the presence of novel versus familiar featural information (independent features), than to feature relations, as the basis for their categorisation. One or more such salient features is the "glue that binds instances of categories" for younger infants, but shared feature relations are responsible for building categories for somewhat olds infants"' (Younger, 2003).

Toward the end of the first year of life (by 10 months), not only are infants able to attend to correlations among attributes to form perceptual categories, but they

also exhibit at least some characteristics of *explicit categorisation*. Younger and Cohen (1983) familiarised 10-month-old infants to (sketches of) novel "animals", which varied on five values (1-5) across four dimensions (length of neck, separation of ears, length of legs and thickness of tails), and then tested them with two novel items - one which contained the average values (3333) and one which contained all the extreme values (1111/5555) (see Figure 1.3). Depending on the distribution of values in the familiarisation set, infants responded differently to the novel items: if the values varied randomly across the four dimensions (Broad Condition), infants considered the average stimulus (3333) as more familiar, and spent more time looking at the modal stimulus (1111/5555). If the values were correlated such that 1s and 2s only occurred together and 4s and 5s only occurred together (Narrow Condition), infants spent more time looking at the average stimulus (3333), instead of the modal (1111/5555). This suggests that in the Broad Condition infants created a single category from the objects and, therefore, during test they exhibited a novelty preference by looking more at the sketches with the extreme values - the average stimulus which fell in the centre of the Broad category was considered more familiar and less attention was devoted to it. In the Narrow Condition, on the other hand, infants created two categories, and the modal stimulus was, as a result, more familiar than the average, which now fell in the borderline between the two categories. It, therefore, seems that infants at 10 months can spontaneously create two categories of items in a visual categorisation task, by attending to the correlations of features in the familiarisation set. These categories appear to be centered around an (abstract) prototype, as evidenced by the decreased attention to the average stimulus, which was equally novel for infants as the modal stimulus.

In a similar task, ten-month-old infants were presented with the Narrow Condition, but this time two non-members were included in each "category" in the familiarisation

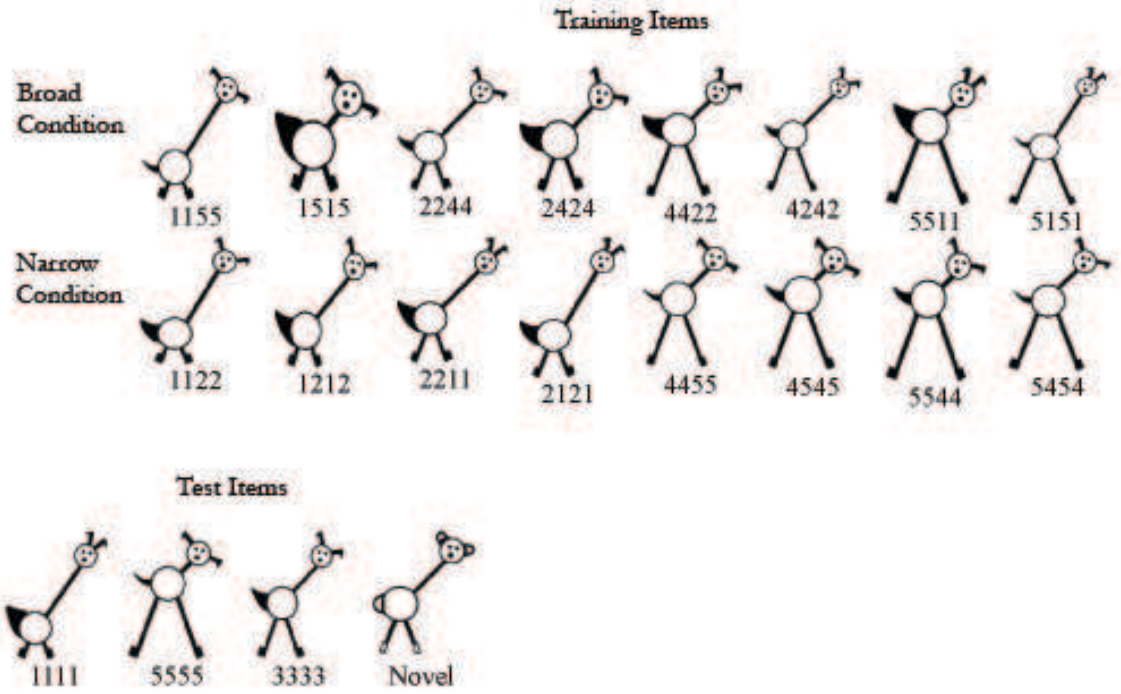


Figure 1.3: Made-up animal sketches used by Younger (1985) in their categorisation study with 10-month-old infants. In the Broad Condition values varied randomly in the training set, whereas in the Narrow Condition they were correlated so that two categories were spontaneously created by infants. Infants in the Broad Condition spent more time looking at the Modal stimuli - which suggests that a single broad category was formed, whereas infants in the Narrow Condition spent more time looking at the Average stimulus - which suggests the formation of two separate categories. Figure taken by Hu (2008) doctoral thesis.

set. Infants were then expected to attend to the average stimulus more, if they managed to ignore the non-members and still form their two categories. If, however, the non-members were affecting their category formation, then infants should pay more attention to the modal stimuli, since category formation would have been disrupted. Infants were shown to be able to compare items and disregard those that are discrepant from the rest as a basis for their categorisation response (Younger, 1993), thus exhibiting evidence for more *explicit categorisation* abilities.

1.2.2 The Impact of Labels on Category Formation

The relationship between language and concept development has been a matter of debate for a few centuries. Some people have argued that conceptual development proceeds independently from language and that the two do not interact with each other. Others have argued that the language one speaks shapes their conceptual system and decisively acts upon the concepts formed; this is known as the *Sapir-Whorf hypothesis* or *linguistic determinism or relativism.* According to this view there is nothing in the outside world to guide us in forming our concepts and we depend on the language we happen to hear in our environment in order to organise our world into meaningful concepts:

We dissect nature along lines laid down by our native languages. The categories and types that we isolate from the world of phenomena we do not find there because they stare every observer in the face; on the contrary, the world is presented in a kaleidoscopic flux of impressions which has to be organised by our minds − and this means largely by the linguistic systems in our minds. We cut nature up, organise it into concepts, and ascribe significances as we do, largely because we are parties to an agreement to organise it in this way − an agreement that holds throughout our speech community and is codified in the patterns of our language. The agreement is,

of course, an implicit and unstated one, but its terms are absolutely obligatory; we cannot talk at all except by subscribing to the organization and classification of data which the agreement decrees (Whorf, 1956).

Although language is assumed to crucially shape thought in certain ways, this form of extreme linguistic determination has been questioned through well-controlled, cross-linguistic, studies; Smith et al. (2003) argue that information relevant of category structure exists and is available to speakers of all languages. For example, prelinguistic infants have been shown to distinguish between solid and non-solid objects or between animate categories and objects which hold their shape and form (Smith et al., 2003; Yoshida & Smith, 2005).

Research on the developmental side of the issue of linguistic relativism has pointed to an early language impact on category formation; some people argue for facilitation of category formation by the presence of linguistic input (Xu, 2002; Waxman & Braun, 2005; Fulkerson & Waxman, 2007), while others consider linguistic input preventive of category formation due to its overshadowing effects in very young infants (Robinson & Sloutsky, 2004).

Robinson suggests that infants in their first year of life exhibit a preference for auditory stimuli and when both auditory and visual stimuli are presented simultaneously visual processing is disrupted. When 8-month-old infants were familiarised with pairs of non-linguistic sounds and optical patterns, they were able to notice a change in the stimulus only for the sounds, but not for the visual input. This suggests that in the first year of life, presentation of labels with visual categories could have a disruptive effect on category formation (Robinson & Sloutsky, 2004).

Researchers in favor of the first view, assume that language has an impact on conceptual development even in pre-linguistic infants: naming distinct objects with *distinct* names highlights their differences and promotes the process of object individuation (Xu & Tenenbaum, 2007). Additionally naming distinct objects with the *same*

name highlights their commonalities and promotes the formation of object categories (Balaban & Waxman, 1997). Waxman (2003) argues that *...Infants across the world begin the task of word learning equipped with a broad, universal expectation that directs them to link novel words to commonalities among objects. This initial link supports the rapid acquisition of the increasingly sophisticated language and conceptual systems that are the hallmark of human development...*(Waxman, 2003)

It has been suggested that in the second half of the first year of life categorisation of visual stimuli starts to be facilitated by the presence of linguistic sounds (words) (Balaban & Waxman, 1997; Fulkerson & Waxman, 2007; Fulkerson & Haaf, 2003). When infants are presented with a range of objects, they fail to exhibit evidence for categorisation if the items are presented with non-linguistic sounds (tones). However, infants are able to trace commonalities among familiarisation items and exhibit novelty preference in the test phase when the same visual stimuli are presented with labels. Waxman and Balaban (1997) suggest that at the early stages of lexical development (around the end of their first year of life) infants have broad universal expectations, which guide them to link novel words to commonalities among objects and words therefore function as *invitations to form object categories* - even when infants are not engaging in actual word learning.

Markow and Waxman (1995) used an object manipulation task to test categorisation in 12- and 13-month-olds: infants were familiarised with four toys from a given category (either basic or superordinate) and were then tested with two items, one from the familiar category and one from a novel contrasting category. Infants were divided in three Conditions, as in Waxman (1990): the *no-word* Condition involved an attention phrase but no labels; the *novel-noun* and *novel-adjective* Conditions involved the presentation of a novel noun and a novel adjective respectively. For the basic level, infants in both the novel-noun and novel-adjective Conditions exhibited a reliable novelty preference, while infants in the no-word Condition did not. For

the superordinate level, only infants in the novel-noun Condition showed a reliable novelty preference. This behaviour was interpreted as an expectation that labels will refer to categories of objects, even by infants at the very early stages of word learning (Waxman & Markow, 1995). This expectation does not depend on grammatical class (noun vs. adjective), although this has been shown to be the case for older babies (Waxman & Kosowski, 1990).

In the above studies, however, although there is a difference between the No Word and the Noun/Adjective Conditions, there is no baseline/Silent Condition to compare it with. In other words, it could be, as Robinson suggests, that auditory input is actually masking categorisation abilities of infants in the No Word Condition, in which there are no labels, but there still exists some kind of linguistic input. The study below aimed to overcome this confound and provide a more strict test of the impact of labels on category formation.

Recently Plunkett et al., 2008, replicated Younger's studies (see Section 1.2.1) with 10-month-old infants and also introduced new Conditions by adding labels to the task. Infants were presented with the same stimuli as those in the Narrow Conditions of Younger (1985) along with either two labels or one. The two labels in their turn were either randomly assigned to the visual stimuli or correlated with the two categories of stimuli (as defined by the occurrence of specific feature values). When labels correlated with the categories, infants exhibited the same behaviour as when there were no labels presented; that is, they formed two categories and attended to the average stimulus more during test. When labels were randomly assigned to the items, infants did not exhibit any preferences during test, indicating that labels disrupted category formation. Finally, when infants were presented with a single label, they exhibited a preference for the modal stimulus during test, suggesting that, instead of two categories, a single category was formed. These findings point to a powerful effect of labels on category formation even at these very early stages

of lexical development: labels can support the formation of categories (two labels-correlated items), disrupt categorisation when visual and linguistic information is in conflict (two labels-uncorrelated items) or even force the inclusion of all visual stimuli in a single category, in order to accommodate the linguistic information available.

1.2.3 Linguistic Categorisation

In the 70s a number of researchers tried to tackle the issue of early word reference by using diary studies and observations of children in their natural environment. Diary data have provided controversial information as far as behaviour in the early stages of vocabulary development is concerned highlighting the need for more strictly controlled methods. Experimental studies followed which aimed to overcome this limitation by providing a better control of the different factors affecting the use of words. These studies focused mainly on the comprehension of words, instead of production which seems to be more lax in its use of labels in different contexts. However, definite conclusions on the strategies for generalising labels to different objects have not been provided as yet.

Clark (1973) has used diary studies to show that infants' early words are characterised by a high degree of generality and exhibit overextensions from age 1.1 to 2.6. Gradually and with increased experience with different objects, infants are assumed to begin to discover what the criterial features are for each semantic category. New words also contribute in narrowing down the semantic domain for each of those very general first words. Infants are suggested to rely highly on perceptual characteristics of the items in discovering the most relevant features and forming their lexical categories: semantic features are considered to be universal because they rely on human perception and cognition in interpreting the outside world. The task of discovering the most appropriate features for meaning is, therefore, assumed to

proceed through the development of perception: just as infants first attend to single features and only later to combinations of features and feature relations (see Section 1.2.1), early words are characterised by limited semantic specificity and only later are infants able to trace more semantic features and relations with features in other semantic domains. Only after these tasks have been achieved, are infants able to include more types of information beyond the perceptual in their semantic domains, such as social and functional in a given cultural context (Clark, 1973).

This semantic feature approach, however, failed to gain any empirical evidence and proved neither appropriate nor usable in language development research. It was, therefore, abandoned in favour of more empirically-sound theories. Additionally, other researchers have assumed that overextensions of early words do not necessarily point to infants' semantically under-specified and overly general categories at these initial stages of word learning. Infants for example might use a word in a given context because they have no other label available or because of a perceived similarity between two things. This however does not mean that they do not know what the real meaning of the word is (Bloom, 2004).

Additionally it has been suggested that overextensions in the early lexicon are only manifested in production, but not comprehension. Infants might use a word in a given context to communicate about an object or action which does not fall within the word's extension, but deny to identify that same object or action as the referent of the word when asked to find one. This is what Labov, 1974, reports for their daughter's use of the word "Mommy": although she used the word to refer to both her mum and several other family members, she would seek for her mother only if asked "Where's Mommy?" (Kuczaj, 1986).

Following this logic, a number of researchers have assumed that early words are highly context-dependent (Bloom, 1973; Bates, Benigni, Camaioni, & Volterra, 1979; Barrett, 1983). Barrett (1993) has offered a more detailed account of the behaviour

of his son, Adam, who produced his first words in very limited contexts, and then learned to extend them to more and more situations. For example, the word *duck* was initially produced only in the context of hitting one of his ducks off the edge of the bath, and in no other case. This is taken to suggest that Adam had not associated the word with his own duck or real ducks, but instead had associated the label with the particular event in the context of which it was appropriate for him to produce the word. It was then concluded that Adam had acquired a mental representation of this event, by reference to which he was able to recognise the equivalence of these separate instances of the same event. This behaviour seems to accord with Nelson's hypothesis (described in Section 1.1.2), who assumes that early words are matched to existing concepts and are likely to be *underextended* until infants discover the correct features describing adult concepts.

According to Anglin (1986) diary studies and observational records of infants' speech might be misleading for drawing conclusions about early lexical development; firstly because such methods are inevitably biased towards overextensions, rather than underextensions or even overlap of reference[1]. In other words, overextensions are clear cases where an infant uses a word that an adult would not have used; but underextensions can only be inferred by the absence of that word in all contexts it could have been used by an adult. Also such studies focus on production when comprehension of words could be pointing to a different pattern of usage of those same labels. Additionally babies often engage in symbolic play and use words in a nonliteral pretentive kind of way - which can be misleading for people studying extension of labels on the basis of this behaviour. Finally Anglin suggests that these studies are not systematic, controlled or intensive enough to allow for a confident assessment of the denotative breadth of a child's individual words (Anglin, 1986).

[1]The term *overlap of reference* is used to refer to cases where the same word is both over- and under-extended; in some cases the use of the word agrees with the adult usage, but it is not used for all cases an adult would use the word, while in others it is used in cases where an adult would not have used the word.

In order to overcome the limitations of diary studies discussed above, Kay and Anglin (1982) tested 2-year-olds in a laboratory setting using strict experimental controls. They trained children with one of five words (*basket*, *candle*, *wheel*, *pin* and *card*) and subsequently tested the extension of those words with respect to several referents and non-referents of the adult category. They assessed extensions in both comprehension (by asking "Is this an X"?) and production (by asking "What's this?"), and counted over- and under-extensions only if they were made by a given child in both comprehension and production. The data suggested that infants tended to show a pattern of both over- and under-extensions (overlap of reference); less often they only underextended labels, and even less often they exhibited only overextensions (Kay & Anglin, 1982).

Markman and Hutchinson (1984) looked at young children's classification preferences using a sorting task: children were shown a picture (e.g., a poodle) and asked to choose another one which "goes with" the first. The two choices provided were a taxonomic match (e.g., a German Shepherd) and a thematic match (e.g., dog food). Two conditions were used; a *no-word* condition, where no label was provided for the target item, and the *novel-word* condition, where children were presented with a new word for the target item. Two-to-three year olds, as well as 5-year-olds chose the thematic and taxonomic match equally often in the no-word condition. However, when a novel word was presented children at both ages made significantly more taxonomic choices. The same result was replicated with 4-5-year-olds and artificial categories. These results suggest that even very young children can understand both taxonomic relations and that words refer to taxonomic relations and not thematic ones, even if the latter can be more salient in the absence of labels. In other words, young children should have in place "an abstract constraint on what words mean that is not mediated by known terms". This is what has been termed as the *taxonomic constraint*, a bias to attach labels to (taxonomic) categories of objects and not just

to individual objects (Markman & Hutchinson, 1984).

Waxman and Kosowski (1990) also showed similar behaviour by 2-year-olds in a picture categorisation task, where infants were presented with one image (e.g., a picture of a cow) and were asked to provide another one which was like the first. Infants were given two alternative types of items to choose from - taxonomic choices (e.g., two pictures of different cows) or thematic choices (e.g., a picture of milk and a picture of a barn). In a *no-word* Condition infants selected taxonomically less often than would be predicted by chance; in a *novel-word* Condition, as well as in a *novel-adjective* Condition, infants selected taxonomically more often than would have been predicted by chance. This behaviour was taken to suggest that by age two infants treat a new word as referring to a category of objects, and start to sort taxonomically upon being presented with it (Waxman & Kosowski, 1990).

The most direct evidence for the development of lexical categories comes from a study by Meints, Plunkett and Harris, 1999, who tested the extension of words which infants already understood. Using an Intermodal Preferential Looking (IPL) task they found that infants at the initial stages of word learning (12 months) were able to successfully identify pictures of typical, but not atypical, exemplars from an adult linguistic category as referents of a known label [2] (Meints, Plunkett, & Harris, 1999). By the end of the second year of life (18 and 24 months) infants were willing to accept both typical and atypical exemplars as referents for a label. These findings delineate a clear trajectory in the development of lexical extension, which coincides with previous data (Rosch & Mervis, 1975): labels are initially associated with typical items and only later, from about 18 months onwards, do labels follow the adult usage by including both typical and atypical exemplars as part of their extension. Therefore, infant lexical categories seem to start off with a more limited set of referents than that of the adult categories and only later do they extend to more objects until they

[2]Similarly Schafer, 2005, has used the IPL task to show that 12-month-old infants can recognise decontextualised and unfamiliar pictures as referents of known words (Schafer, 2005).

assimilate the adult ones.

Both principles-and-constraints, as well as exemplar-based accounts try to account for this behaviour, by providing alternative explanations: principles-and-constraints accounts assume that infants at all three ages share an understanding that the label refers to a category of objects; the failure of 12-month-olds to extend a word (such as "dog") to the atypical instances is due to the fact that their perceptual categories are not broad enough to also include the atypical members of the category (such as a "chihuahua"). By 18 months, infants have altered their perceptual/conceptual category of "dog" (as a result of experience) and consider a "chihuahua" as the "same sort of thing" as a "German Shepherd" or a "Labrador". Having achieved this cognitive hallmark, infants come to appreciate the fact that we should use the word "dog" to also refer to that peripheral member of the category. The inability of 12-month-olds to extend the label to the atypical instances is therefore a cognitive limitation regarding infant perceptual/conceptual categories, irrespective of language; infants do not learn anything linguistic when their behaviour changes over time. All they do is to discover the correct extension for their cognitive category and just apply that to their known label.

Exemplar-based accounts assume that these underextensions in the early stages of word learning are predicted by an associative model of language development. Infants initially have more experience with labels occurring with typical items (either because they are just more in terms of numbers, or because adults name those central items more often than atypical ones - see Section 1.1.1). Only after increased experience with a broader range of items do they come to appreciate that the label refers to a wider set - and they therefore extend their label to a corresponding conceptual category, if there is one at hand), or alter their perceptual category to "go with" their linguistic one "under construction".

1.3 Thesis Outline

The data above present a rather inconclusive picture of the development of lexical categories and its interaction with perceptual categories in the first year of life. Infants' ability for creating perceptual categories has been well-established, but the way these categories interact with the development of semantics is still a matter of debate.

Principles-and-constraints accounts suggest that infants should use those categories during word learning to generalise newly-learned labels to new objects; the time and manner by which the taxonomic constraint arises is still unknown, but it has been shown to be in place at least by 18 months of age. Exemplar-based accounts on the other hand suggest that infants build their lexical categories in the same manner as they build their perceptual categories; initially they limit a word to a specific object and only gradually do they include more objects in the reference of their word, on the basis of experience and similarity with different within-category objects.

In order to investigate the mechanisms by which infants generalise labels to new objects at the early stages of lexical development, this research project focuses on infants at the end of their first year of life (10 months). The experimental task involves teaching infants two object-label pairings and then testing their generalisation of labels to new exemplars from the same adult categories. On the basis of the theoretical approaches to categorisation (perceptual and linguistic) a decisive factor was manipulated in the design: amount of exposure to different naming instances. Do infants generalise newly-learned words to new objects when they have only encountered them with a single object, or do they need to be exposed to multiple word-object occurrences before they do so? Additionally, similarity and prototypicality of the objects were used to guide the selection of the exemplars, as defined by ratings from adult native

monolingual speakers of British English [3].

Chapter 2 describes two Experiments which constitute a first attempt to investigate the word-learning abilities of infants at the end of their first year of life. These experiments manipulate two experimental designs (one with and one without pre-test) and the factor of animacy (animates vs. inanimates) in this task.

Chapters 3 and 4 describe a further attempt to teach infants two word-object pairings, by applying a more infant-controlled procedure to the word-learning situation, and an attempt to test generalisation of newly-learned labels to other members of the same adult categories. The amount of experience with different category members during training is manipulated, in order to directly test the assumptions of principles-and-constraints and exemplar-based accounts as far as generalisation of labels is concerned: in Experiment Three infants are trained with a single object from each of two adult categories, whereas in Experiment Four multiple exemplars from each category are presented during training.

Chapter 5 describes a categorisation experiment aiming to test whether infants perceived the objects used in Experiments Three and Four as belonging to the same category of objects. Experiment Five replicates the training block of Experiment Four (with multiple objects from each adult category) and then uses a novelty-preference task without labels at test in order to investigate the perceptual categories infants form on the basis of the training and testing stimuli used in these series of studies.

Chapter 6 describes the adult typicality and similarity rating studies conducted in order to provide a basis for the selection of the objects in the infant studies. Finally Chapter 7 presents a discussion of the findings of the Experiments and their implications in the theoretical discussion presented in this Chapter.

[3]This research received ethics clearance by the Departmental Ethics Committee, CUREC Ethics protocol MSD/IDREC/2008/P11.1

Chapter 2

Experiments One and Two

2.1 Introduction

Experiments One and Two were a first attempt to investigate word learning abilities in infants before their first birthday. The challenge in this project was to achieve learning of words in the lab during brief experimental sessions and establish that this task is appropriate for infants at 10 months.

2.1.1 Infants' First Words

Lexical development has been thought to begin sometime around the second half of the first year of life; most researchers trace its start at around 10-12 months of age (Bloom, 1993; Nelson, 1973; Fenson et al., 1994; Gentner, 1982; Brown, 1957; Samuelson & Smith, 1998). This is a time which coincides with the emergence of crucial phonetic/phonological abilities, such as the fine-tuning to language-specific phonological contrasts[1] (Werker & Tees, 1984), the ability to segment words in fluent

[1]It is interesting that infants who fail to respond categorically to a non-native speech contrast were shown to also exhibit advanced categorisation abilities, contrary to infants who were still "universal listeners" (Lalonde & Werker, 1995), although recognition of invariant speech patterns has been exhibited by infants as young as 4,5 months for their own names (Mandel, Jusczyk, & Pisoni, 1995).

speech contexts (Jusczyk & Aslin, 1995) and sensitivity to prosodic and phonotactic constraints of ones native language (Mattys, Jusczyk, Luce, & Morgan, 1999). Some evidence of object - label associations has been provided for even younger infants, as early as 6 months of age, but only when using moving stimuli which display a high degree of social salience, such as "mommy" and "daddy" or "hand" and "foot" (Tincoff & Jusczyk, 1999). Parental reports about infant receptive vocabulary roughly agree with these findings, with the average 8-month-old estimated to understand about ten words and the average 12-month-old just above fifty (Fenson et al., 1994).

2.1.2 Word Learning in the Laboratory

Word learning studies in the laboratory have documented the ability to form word-object associations for two labels by infants as young as fourteen (Werker, Cohen, Lloyd, Casasola, & Stager, 1998) and fifteen months (Schafer & Plunkett, 1998) using an Inter-modal Preferential Looking (IPL) task. Word learning has also been exhibited by 13-month-olds in a multiple-choice task for only one label (Woodward, Markman, & Fitzsimmons, 1994), while in a series of habituation tasks 14-month-old infants have been shown to be able to learn object-label pairings for two labels only when the stimuli involved motion. However, in the same study eight to twelve-month-old infants failed to form object-label associations, although they were able to process word and category information separately (Werker et al., 1998). More recently, Pruden et al. (2006) have tried to teach two novel object-word pairings to infants at 10 months and claimed that infants were able to learn one of the words during the session (Pruden, Hirsh-Pasek, Golinkoff, & Hennon, 2006). However, their

On the basis of this finding, it was, therefore, suggested that some kind of mature form of behaviour starts to develop at around the end of the first year which is exhibited in different tasks; part of this process is a recategorisation ability which appears to emerge, whereby complex stimuli are explicitly categorised on the basis of correlations among stimulus features and new category boundaries are superimposed on earlier perceptual categories (Lalonde & Werker, 1995; Younger, 2003). This discussion, however, will not concern us here.

experimental design did not involve sufficient counter-balancing between presentation side and object, making any effects from the presentation of the target word on object preference difficult to interpret.

These studies offer the impression that early word learning is a very laborious process and many researchers have claimed that infants under the age of 18 months are not in position to engage in *true word learning* (Werker et al., 1998). However, there is growing literature on the subject providing indications to the contrary: infants at 15 months are able to learn two object-label pairings in a brief experimental session (Schafer & Plunkett, 1998), whereas at 12 months of age infants have been shown to be able to recognise de-contextualised, stationary pictures of socially non-salient items as referents of known labels without prior teaching (Meints et al., 1999) or after having received extensive training with different pictures of those items by parents (Schafer, 2005).

It is therefore possible that word learning studies might be underestimating infants' linguistic abilities at the early stages of lexical development. When addressing the issue of infant lexical development, most researchers focus on how quickly infants can form an association between an object and a label, which are both novel to the infants (Werker et al., 1998). This type of task requires infants to both process visual and acoustic stimuli individually and also to associate them to each other, considerably increasing the cognitive load imposed on them. Such high task demands could potentially mask very young infants linguistic abilities and create an impression of a language which in the very early stages of lexical development is qualitatively different from that of post-vocabulary-spurt infants.

2.1.3 Methods for Investigating Word Learning

When testing word comprehension (or word learning) in a laboratory setting, infants are often required to exhibit their understanding through means of attentional preferences to certain items. Infants before the end of their second year have a very limited productive vocabulary and even after that they might be reluctant to share their knowledge of words with a stranger in an unfamiliar environment. In order to investigate infant language, experimental techniques therefore exploit an ubiquitous behaviour of infants (and humans in general), that is present from very early in life: attention to different items through looking fixations.

The Preferential Looking Paradigm or Head-Turn procedure was pioneered by Fantz in the 50s. During his basic experimental paradigm, infants were presented with two visual stimuli simultaneously. Ceteris paribus[2], reliably more looking to one of the stimuli was considered as an instance of a spontaneous, pre-experimental attentional or perceptual preference which babies bring together in the task. This behaviour of infants, subsequently demonstrated in a number of studies, is what has been used in word-learning settings to exploit infant language comprehension, in the form of a spontaneous preference to look at the visual stimulus that matches an auditory stimulus (Golinkoff & Hirsh-Pasek, 1987b).

In an Inter-Modal Preferential Looking (henceforth IPL) Task, infants are presented with two visual stimuli for a fixed duration of time. Exactly half-way through the picture presentation time an auditory stimulus is presented (a word), which corresponds to one of the pictures that are on the screen. Comprehension of the word is evidenced by increased looking at the appropriate picture, over and above

[2]The assumption that two stimuli can be entirely matched for amount of interest or salience in a preferential looking task is in practice very difficult to maintain, especially for real objects. Novel stimuli can be easier to control, because their different perceptual dimensions (colour, shape, texture, etc.,) are liable to more strict control. However, even with novel categories one can never predict if an infant will exhibit a preference for red over blue colour or for a circle over a triangle for example. Because such inherent preferences are practically impossible to predict, a well-controlled preferential looking task always includes a pre-naming trial phase where baseline preferences are directly tested.

attention during the pre-naming phase of the trial.

An alternative way to investigate word learning in a laboratory setting is the "switch design". Infants are trained with two word-objects pairings until their attention declines; decreased looking has been widely used as an indicator of when infants have formed a relatively well-established representation of the association between a word and a (picture of an) object. When a decrease of attention to the repeated, and therefore familiar, stimulus is exhibited, the crucial test lies on orienting attention to a novel stimulus which mismatches the "habituated" stimulus. The novel stimulus/pair is achieved by "switching" around the words and the objects and presenting infants with an object and a word which were repeatedly presented with another word/picture, respectively, during training. Attention to the "switched" pair is considered to be indicative of the kinds of representations that have been formed during the task.

In the current research the IPL task was preferred over the switch design, because the former is considered to be easier for infants; during the IPL infants have both visual stimuli on display and they only need to use their recognition memory to activate the relevant association with the word. In the switch design only one of the pictures is available to infants and they, therefore, need to use their recall memory in order to be successful in the task. Because infants at the end of their first year are novice word learners, the IPL task was used in order to provide a more fair test for their word learning abilities.

2.1.4 Current Research

The current research, unlike other studies of its kind, aims to tap into infants word learning abilities by making the task as simple as possible for infants, while at the same time retaining a strict experimental procedure. A very simple design was, therefore, administered using items which are considered highly familiar for infants.

Real and familiar items were selected over novel ones for two reasons: firstly, familiar items are likely to reduce the cognitive load required by infants in order to complete the task. Word-learning abilities in brief experimental sessions have been shown for infants as young as 14 months when using novel items; however younger infants may be capable of learning new object-label associations, if the only task they are required to perform is that of learning the words - rather than also having to process new visual and auditory stimuli simultaneously. In other words, novel items could potentially distract infants from the word learning task, thus masking any abilities they might have from very early on (Werker et al., 1998). The second reason for the use of familiar items is that infants are expected to have some pre-established categories for these items (or at least that they will be able to quickly form new categories for them on-line); this would allow direct assessment of the impact of cognitive categories on generalising newly-learned labels to new objects.

In these experiments, infants were taught two word-object pairs and then tested on those same pairs as well as on new pairs with identical words but with new exemplars from the same category. A standard Inter-modal Preferential Looking (IPL) task was administered; two pictures were simultaneously presented to the infant along with a single word appearing exactly half way through the picture presentation time. This allows comparison of looking time to the pictures before and after the presentation of the word, and an increase in attention to the target after naming is typically considered as evidence for an established association between the label and the (picture of an) object.

In these experiments familiar objects were used, but words unlikely to be understood by the majority of 10-month-old infants, according to parental feedback. Also infants were either presented with animals or objects, in order to compare

possible effects of animacy during word learning. These two experiments are presented together because their design, materials and procedure are highly similar, apart from a crucial difference in the design: in Experiment One a pre-test block controlled for prior-knowledge of the to-be-learned pairings.

2.2 Experimental Design

2.2.1 Participants

The final sample consisted of 96 infants (50 boys) with a mean age of 9.9 months (R : $9.1 - 10.5$, SD : 0.29). Infants were recruited through the Oxford Babylab parental database, which contains personal information for people who have expressed their interest in our research, and are approached in local maternity wards (JR hospital, and others), local playgroups, NCT sales, through the Oxford Babylab website or publicity material regularly distributed to medical centres, doctors' surgeries and child-care centres. All infants participating in the study were healthy, full-term babies, from monolingual households in which the language used is British English.

Infants in each Experiment were further divided into two Conditions (24 infants each) on the basis of the animacy of the stimuli (animals/objects) presented to them. The two sexes were roughly equally represented in each condition. Ten additional infants were tested but excluded from the final set, eight due to fussiness, one due to experimenter error and one for parent interference during testing.

2.2.2 Stimuli

Six items were used across conditions: three animals ("bear", "cat" and "duck") and three objects ("cup", "phone" and "spoon"). Animate items were always paired with

animates, and likewise for inanimates. This resulted in six pairs of items: "bear-cat", "bear-duck", "cat-duck", "cup-phone", "cup-spoon" and "phone-spoon". In Condition A of each experiment infants were presented with objects, whereas in Condition B they were presented with animals. Animacy was introduced as a factor in the analysis, in order to test whether objects of increased salience for infants facilitate the forming of an association between the objects and the corresponding label, as has been recently assumed (Pruden et al., 2006). This proposed advantage is directly tested in this experiment.

According to the MacArthur Communicative Development Inventory (Fenson et al., 1994) less than 50% of 10-month-old infants are likely to understand the words chosen for this experiment. For exact percentages, see Table 2.1. These estimations were calculated from parental feedback provided for American infants; in order to be able to better estimate the comprehension abilities of infants tested in this experiment, parents were asked to fill in the Oxford Communicative Development Inventory (henceforth Oxford CDI) (Hamilton, Plunkett, & Schafer, 2000). However, the accuracy of parental feedback as far as individual words are concerned has recently been questioned, especially for non-verbal babies (Tomasello & Mervis, 1994; Houston-Price, Mather, & Sakkalou, 2007; Robinson & Sloutsky, 2004). In order to evaluate this concern, infants in Experiment One were tested for prior knowledge of these labels in a pre-test block before training.

Also according to the LEX2005 database, more than 25% of infants were reported to understand the selected words. This ensured that these objects were fairly familiar to infants, decreasing the cognitive load of the task and, therefore, allowing them to focus on word learning.

Word	Percentage of infants
bear	32.8
cat	44.8
duck	23.9
cup	41.8
phone	35.8
spoon	29.9

Table 2.1: Percentages of infants reported to understand the words used in the Experiment according to the LEX2005 database.

"Wake-up" trials were introduced at certain points during the experiment. Eight additional items were chosen to be presented in "wake-up" trials. None of these items was named. The items were: "clock", "house", "ball", "car", "turtle", "ladybug", "fish" and "octopus". Because wake-up trials aimed to regain attention to the task, the objects were chosen to be fairly familiar to infants and also fairly interesting to look at. The first four items were only presented to infants who were trained and tested with an animal pair, while the last four were only presented to infants who were trained and tested with an object pair. This was done in order to minimise interference in the word-learning and generalization tasks by items which exhibited high similarity to the test items.

Pairs of to-be-learned items were chosen to be highly dissimilar in terms of both their acoustic and visual form. Audio files were recorded in a single session by a female voice delivering them in high-affect child-directed speech (Fernald, 1992; Mills, Plunkett, Prat, & Schafer, 2005). The audio recordings were made with a digital audio tape recorder (DAT) in a sound-treated, recording booth. The audio stimuli were digitised at a sampling rate of 44.1 kHz and a resolution of 16 bits and spliced using Goldwave v. 5.10. Labels were produced by the speaker in citation form and the target words were not embedded in sentences, but rather presented in isolation, in order to make the task easier for infants (Jusczyk, 1999). An attention phrase always preceded the target word in order to attract and maintain infants' attention to the task. All

Experiment Phase	Attention Phrase
Pre-Test	*See*
Train	*What's this* and *There*
Test Block A	*Look*
Test Block B	*Look*
Wake-up Trials	*Look! What's this?* and *Look! Look!*

Table 2.2: Attention Phrases used in each phase of Experiments One and Two.

attention phrases can be found in Table 2.2.

Visual stimuli were digitised photographs of the items on a 10% gray background. All pictures measured 400 x 240 pixels and they were saved in 24-bit bitmap format. Photographs were all obtained from the digital photograph collection Hemera Photo Objects Volume I. Two pictures were selected from each test category and all pictures depicted typical exemplars from the given category. The typicality of the pictures was assessed by adult monolingual native speakers of British English. Typicality rating was considered to be crucial in this task for two reasons: firstly it has been shown that typicality, as judged by adult speakers of a language, is correlated with both the order and the rate at which adults learning a foreign language (Rosch et al., 1976), as well as infants acquiring their native tongue (Meints et al., 1999) include individual items in a linguistic category's extension. That is, both adults and infants learn to name typical items first and more easily than atypical items (See Sections 1.1.1 and 1.2.3). Second, although infants in Experiments One and Two are exposed to the items several times during training, typicality was considered to be an additional index of how familiar these items might be to infants and a possible indicator of the cognitive load that the task would pose for infants. All adult rating studies are described in 6.

For the "wake-up" trials, only one picture for each item was selected and processed

following the same procedure as that used for the test items. "Wake-up" pictures were very colourful and involved toy-objects as opposed to test pictures, which depicted real objects/animals. The pictures used in the "wake-up" trials were not included in the Adult Typicality Rating Study.

2.2.3 Design

Experiments One and Two involved a standard IPL task aiming to test both the word-learning abilities of infants at the end of their first year of life, as well as their generalization of newly-learned labels to new exemplars from the same adult category. Each infant was trained with two object-label pairings. When tested, infants were presented with the same labels and both with the pictures they were previously trained with (word-learning task), as well as with new typical exemplars from the same categories (generalization task). Each infant was trained and tested with only one pair of items, which differed across infants. Each pair of items (as described in section 2.2.2) was presented to eight infants; therefore each single word-object pairing was presented to sixteen infants, paired with two different distracters across infants. The two experiments differed in an initial pretest block, which was only presented in Experiment One; this yielded two different designs, Experiment One included four blocks (pretest/train/test/test) whereas Experiment Two only included three blocks (train/test/test). Each experiment was further divided in two Conditions: infants in Condition A were always presented with objects, whereas infants in Condition B were always presented with animals. The two Experiments therefore differed in only the pretest block; everything else was identical across these experiments.

During pretest (Experiment One), infants sat through a four-trial block; in each trial both pictures were simultaneously presented along with a word corresponding to one of the pictures on the screen. Pretest was used only in Experiment One in order to

test prior knowledge of the to-be-learned items.

During training (both Experiments) each infant was presented with the same two items multiple times. Each object-label pairing was presented individually and appeared randomly on the right or left side of the screen. Visual stimuli were identical to those used in the pretest block. The training block consisted of eight trials, four for each item.

The first test block after training (both Experiments) was identical to the pretest block of Experiment One. This block aimed to test the word learning abilities of infants and the effect of training on infant behaviour.

The second test block after training (both Experiments) differed from the two previous test blocks in only one aspect: the visual stimuli were new pictures of typical (as rated by adult native speakers of British English) items from the same categories. This block aimed to test generalization of newly-learned labels to new items from the same categories.

Between training and first test block as well as between the two post-training test blocks, a "wake-up" trial was presented. This was an IPL trial with two pictures presented simultaneously on the screen, including an attention-phrase, but *no labels*. "Wake-up" trials served to maintain infant attention to the task and prevent habituation effects caused by the repetitive presentations of the same items.

2.3 Procedure

After a brief familiarisation period in the Oxford Babylab, parent and child were seated in a dark booth. Throughout the experiment infants sat on their parent's lap approximately a meter away from a 1.3m (width) x .35m (height) projection screen. Pictures measured 32cm (width) x 19cm (height) and were presented on the left and

right of the screen, separated by a 17.7cm gap. Parents were asked not to interact with their child during the session. They were instructed to close their eyes and listen to music through headphones. Two cameras above each picture recorded infant eye movements, each sending a signal to a digital router to form a single twin split-image of the infant. A computer captured the whole session for later off-line scoring. The experimenter was able to manually control the experiment, which proceeded at the individual infant's own pace. A red light and buzzer just above the centre of the screen served to re-capture infant attention between trials.

Infants were randomly assigned to one of two Conditions in each Experiment. Infants in Experiment One received a four-trial pretest block before training. Each trial lasted 5000ms; at 1000ms the word "see" was presented and exactly half way through the trial, at 2500ms, the target word began. Each item was the target twice; order of presentation was randomised and side of presentation was counterbalanced within each block for each infant and for each object.

All infants sat through an eight-trial training block. Each trial lasted 5000ms but in this block only one picture was presented on the screen, either on the left or the right side. At 1000ms an attention phrase was presented (see Table 2.2, and at 2500ms the target word began. Each object was presented four times and the order of presentation was randomised. Side of presentation and attention phrase was counterbalanced across objects and randomised within this block.

The two post-training test blocks were identical in procedure to the pretest block and only differed in that the final test block included new pictures from each category of items.

Immediately after training, the first "wake-up" trial was presented: two pictures remained on the screen for a total of 5000ms ("clock-house" for infants who were tested with an animal pair and "turtle-ladybug" for infants tested with an object pair). At 1000ms infants heard "look" and at 2500ms half of the infants were presented with another "look" and half with "what's this?". The second "wake-up" trial was presented between the two post-training test blocks. Again, two pictures were presented for 5000ms ("ball-car" for infants who were tested with an animal pair and "fish-octopus" for infants tested with an object pair); audio files were identical to those used in the first "wake-up" trial. Side of presentation of the items was counterbalanced across infants.

2.4 Predictions

The two Experiments differed only in the presence/absence of the initial test block, which aimed to test prior knowledge of the to-be-learned labels. The post-training test blocks, common across Experiments, aimed to test word learning on the one hand and spontaneous generalisation of the to-be-learned items to new typical exemplars of the adult lexical categories on the other. Predictions are therefore common across both Experiments.

Constraints-and-principles accounts predict that infants have an understanding that words refer to categories of objects, rather than just individual instances of them. They should therefore be able to generalise a newly-learned label to the whole range of a corresponding, pre-established cognitive category. On the other hand exemplar-based (i.e., associative) accounts assume that babies need to discover, through experience, that words refer to categories of items and not just to the

individual instance with which they have heard the label occurring. These accounts therefore predict that initially infants should not be able to readily generalise the label to a new instance of the adult category due to their limited exposure to such object-label pairings.

2.5 Measures

During an Inter-Modal Preferential Looking Task two pictures are simultaneously presented on the screen, along with a single label which typically corresponds to one of the pictures. The label is presented exactly half way through the picture presentation time, dividing each trial in two equal parts. Looking time to the two pictures is compared in the pre- and the post-naming phases of the trial and an increase in target looking is considered as evidence for understanding. In the pre-naming phase infants are expected to look at the two pictures roughly equally long. However, occasionally infants exhibit an inherent interest in either visual stimulus and the pre-naming phase serves to control for such attentional preferences. In that case, an increase in target looking post-naming is considered as evidence for understanding only if it occurs *over and above* any systematic preferences in the pre-naming phase.

Three IPL measures were calculated for each infant, separately for the two Trial Phases (before and after the onset of the target word). An additional looking measure, latency of shift, was only calculated for the post-naming phase:

- Total Looking Time (TLT): the amount of time spent looking at both the target and the distracter. This measure of absolute looking time was used in order to test for the amount of attention infants exhibited in the task, and better control for habituation effects which might arise as a result of training. TLT

was calculated both for the two trial phases separately (pre- and post-naming), in order to look at infant attention in the task in relation to the presentation of the target word, and also for the total duration of the trial.

- Proportion of Target Looking (PTL): calculated by dividing the total looking time to target by the total looking time to both target and distracter. This proportional measure of infant attention allows the removal of background noise from the data (which is present for example in TLT) by adjusting for between- as well as within-subjects differences, which could be caused by uncontrolled factors interfering during an experimental session (e.g., habituation effects). Both PTL and LLD (described below) are considered as indicators of the time needed to match an incoming stimulus to an existing mental representation (Golinkoff & Hirsh-Pasek, 1987b).

- Longest Look Difference (LLD): calculated by subtracting the longest continuous look to the target from the longest continuous look to the distracter. The longest look measure exploits the idea that information processing takes place during fixations to visual stimuli (and not during saccades) (Bridgeman, Hendry, & Stark, 1975; Volkmann, 1986; Irwin, 1998) and an increase in longest look durations upon the presentation of a label is considered evidence for the activation of an association between a visual and an auditory stimulus.

- Latency of shift (LoS): the time between the onset of the target word presentation and the initialisation of infant's first shift away from the fixated picture. This measure also follows the logic behind PTL and LLD: a match between an auditory and a visual stimulus should attract increased looking time to the appropriate picture. If, therefore, an infant happens to fixate the target picture when the label is presented to her, it should take longer to disengage from that picture than if she happens to be looking at the distracter, because of the

absence of an association between the word and an urelated picture (Swingley & Aslin, 2000).

For this analysis, one "time-window" has been used: 300ms to 2000ms. These are both arbitrary limits, but decided upon on the basis of previous experimental findings. Infants at 18 months have been shown to need at least 200ms before they can initiate an eye movement in response to an auditory stimulus, with a mean considerably higher than this (Fernald, Pinto, Swingley, Weinberg, & McRoberts, 2001; Canfield, Smith, Brezsnyak, & Snow, 1997; Swingley & Aslin, 2000). Additionally, language has been shown to guide looking behaviour up until 2000ms from the onset of the target word (Swingley, Pinto, & Fernald, 1999). According to the above, shifts that were initiated less than 300ms or more than 2000ms after the onset of the target word were not considered to be affected by the naming event and were therefore excluded from the analysis. This measure has been shown to predict reliably comprehension with infants down to 14-15 months of age (Mani & Plunkett, 2007); it was not expected to be very informative regarding word understanding in 10-month-old infants, but it was used in order to evaluate directly this assumption.

Means for the above measures were calculated by aggregating data by participants. The picture is different as far as latency of shift is concerned: a number of infants happened to be fixating only the target or only the distracter at the onset of the target word for an entire test block, and it was therefore not possible to obtain two values for all of the infants; those missing values were filled in with the mean latency of the whole group to either target or distracter depending on which value was missing.

On the basis of the feedback provided by parents in the Oxford CDI (for the questionnaire see Appendix B), two additional measures of each infant's linguistic abilities were calculated and introduced as between-subjects factors in the analysis:

- Total Comprehension Score (TCS): the number of words reported to be understood by each individual infant.

- Comprehension Score for the testing items (CS). According to this latter, infants were divided in two groups: Comprehenders (C) (N = 13 for Experiment One and N = 12 for Experiment Two), reported to know at least one of the labels, and Non-Comprehenders (NC) (N = 31 for Experiment One and N = 35 for Experiment Two), reported not to know any of them. Infants who were reported to comprehend one or both of the testing words were included in the same group (*Comprehenders*), because knowledge of even one of the words could provide infants with cues as to which object the unknown word refers to. Although the mutual exclusivity (Markman, 1989; Merriman & Bowman, 1989) or N3C (novel-name-nameless-category; (Golinkoff & Hirsh-Pasek, 1987a)) principles have been shown to become available to infants much later in lexical development, certain cues could still be provided especially since infants are presented with the same two objects several times.

  Production scores were not included in this analysis because none of the infants participating in the study was reported to use these words in their productive vocabulary.

  Four infants were excluded from this analysis in Experiment One and one infant in Experiment Two because parents did not complete the Oxford CDI at the time of testing.

2.6 Scoring

Infant looking behaviour was recorded and scored off-line by a trained scorer. A digital-video scoring system (Observer 1.4.114 by Noldus Information Processing Systems) was used to assess visual events on a frame-by-frame basis (every 40ms).

Visual events received one of four tags: "right", "left", "mid" or "other"; "right" and "left" referred to fixations at the right and left pictures. "Mid" was used to describe shifts between the two sides, as well as blinks. "Other" was used for any other looking behaviour that involved infant attention shifting away from the task: looking away from the screen, crying, rubbing their eyes, etc.

Intra- and inter-reliability tests were carried out, with 15% of the trials scored again, both by the same scorer and by a different trained scorer. Pearson r correlations revealed high agreement in both reliability tests; all measures correlated above 0.95, intra-reliability: $r(112) = .97$, $p < .001$, and inter-reliability: $r(112) = .98$, $p < .001$.

An important methodological decision had to be taken concerning shifts *within* the same side of the screen: infants often fixate a picture and then move their eyes within the same picture to "scan" around. The crucial question is whether this "scanning" behaviour should be considered as an interruption to the fixation or as one continuous look; any decision on this issue would then have different implications both in methodological and theoretical terms. Deciding to consider a within-picture shift as a discontinuity in looking would affect the longest look measure (see Section 2.5) potentially introducing a bias towards shorter continuous looks when infants are actually involved in processing during that time. We know that new information is not being input during saccades; however old information could still cause neural activity and lexical processing has been shown to happen even during saccades (Irwin, 1998). In that case, considering shifts within the same picture as a discontinuation of the fixation to the stimulus could actually be underestimating infants' cognitive abilities. It was therefore decided to ignore shifts within a picture and consider this pattern as a single fixation, as long as the infant decided to maintain attention to the same visual stimulus.

2.7 Analysis

Each infant sat through twenty-two or eighteen trials in Experiment One and Two, respectively. The two experiments were analysed separately because the presence of the pretest block in Experiment One yielded two different designs.

All infants were presented with a training block, which consisted of eight trials. These trials were not included in the analysis; they were used to familiarise infants with the object-label pairings and there was only one picture on the screen for each trial. Wake-up trials were also excluded from the analysis, because they served as attention-getters and involved no labels. The remaining trials formed part of the statistical analysis; that is, three test blocks (a total of twelve trials) in Experiment One and two test blocks (a total of eight trials) in Experiment Two. Also individual trials in which infants failed to look at both pictures during the course of the trial were excluded, because their active involvement in the task was considered under question.

The initial data set (before excluding any trials) in Experiment One included a total of 547 trials whereas in Experiment Two it consisted of 382[3] trials. After the exclusion of trials, these data sets were reduced to a total of 529 and 359 trials - 3.30% and 6.12% reduction - for Experiment One and Experiment Two, respectively.

Normality tests were run for all measures and different factors included in the analysis using the Shapiro-Wilk test. The latter is used instead of the Kormikolov-Smith test when samples have fewer than 50 cases. Because data were aggregated either by participants or items, the data set for different factors and blocks was reduced below 50 cases and thus the Shapiro-Wilk test was used. Whenever a measure was not normally distributed, outliers were excluded in order to achieve normality. This resulted in the following exclusions:

[3]The difference in number of trials between the two Experiments is due to the pretest block.

- Experiment One: in the overall PTL and LLD one infant was excluded from the final set for the pretest block, two for the first test block and one for the second test block. The data excluded in each test block were obtained from different infants; however the data for an infant were only excluded from the test block in which they were shown to be outliers but kept in the other two test blocks. All of these data points belonged to infants in Condition A who were presented with Objects. After these exclusions normality was achieved for all measures in the analysis.

- Experiment Two: One infant was excluded from the pre-naming phase of the trial of both PTL and LLD measures in Test Block A, and two infants from the post-naming phase. In Test Block B, only one infant was excluded for the PTL measure.

As mentioned above the two Experiments were analysed separately, since the presence of the pretest block in Experiment One created two different designs. Three-way repeated measures Analyses of Variance (ANOVAs) were run for each measure with Trial Phase (pre- vs. post-naming) and Test Block (pretest (only in Experiment One), TestA, TestB) as within-subjects factors and Animacy (animals vs. objects) as a between-subjects factor, in order to test for any effects of Animacy on the understanding or learning of the labels.

Additionally, in order to evaluate the accuracy of parental feedback for their infant's comprehension abilities and whether vocabulary size was predictive of infant performance in the lab two analyses were conducted: a three-way repeated measures ANOVA with Trial Phase and Test Block as within-subjects factors and Comprehension Group ("C" vs. "NC" - as described above) as a between-subjects factor was conducted comparing performance of infants who were reported to know the words and those who were reported not to know the words prior to the study. In order

to evaluate the relationship between vocabulary size and infants' prior knowledge of target words as well as the ability to quickly learn new words, an Analysis of Covariance (ANCOVA) with trial phase as a within-subjects factors and TCS (Oxford CDI total comprehension score) as a covariate was conducted.

Baseline preferences for specific objects were tested by comparing the pre-naming phase for each of the measures to a chance level of 50% for PTL and of zero for LLD.

A latency analysis compared the mean time needed to shift to the target if fixating the distracter when the target word is presented to the mean time needed to shift to the distracter if fixating the target. A one-way ANOVA with Latency of Shift as a dependent variable and Look Direction at the onset of the target word was conducted in order to see if shifts to target differed from shifts to distracter. Language has been shown to guide looking behaviour up until 2000ms from the onset of the target word (Swingley et al., 1999). Shifts that were initiated less than 300ms or more than 2000ms after the onset of the target word were not considered to be affected by the naming event and were therefore excluded from the analysis. Also trials in which infants were not looking at either the target or the distracter (but instead somewhere else in or away from the screen) when the target word was being presented were excluded from this analysis. This meant that the dataset was considerably reduced compared to the rest of the analyses (a total of 99 unaggregated points for the pretest block (47.20% reduction) - 47 shifts to target and 52 shifts to distracter). Mean latencies of shift to target and to distracter were calculated for each infant separately.

Finally, as a result of the design, infants received repetitive presentations of the same picture pairs; it was therefore considered important to control for the possible

presence of habituation effects. TLT to screen (both target and distracter[4]) was used to compare differences between blocks. A 2 (Trial Phase) x 3 (Test Block) x 2 (Gender) ANOVA with Trial Phase and Test Block (pre- or post-naming) as within-subjects factors and Gender as a between-subjects factor was conducted for Experiment One. This analysis was also conducted in Experiment Two but the result should be considered with caution, because of the absence of the pre-training test block and the inability therefore to compare attention in that block with the post-training blocks.

2.8 Results

2.8.1 Experiment One

Experiment One included three test blocks and one training block (pretest/train/testA/testB). The pretest aimed to test prior knowledge of the to-be-learned labels. The first test block after training aimed to test word learning, while the second aimed to test generalisation of labels to new exemplars from the same category. Twenty-four infants participated in each Condition (22 female). Figure 2.1 shows data from the three test blocks separately for each Condition (Animals and Objects) for the LLD measure.

The figure indicates that infants show an increase in target looking in the post-naming phase of the trial *only* in the pretest block. This increase appears to be more robust for Objects than for Animals. In order to test these observations, a 2 (Trial Phase) x 3 (Test Block) x 2 (Condition) repeated-measures ANOVA with

[4]Time to switch between sides was excluded from TLT.

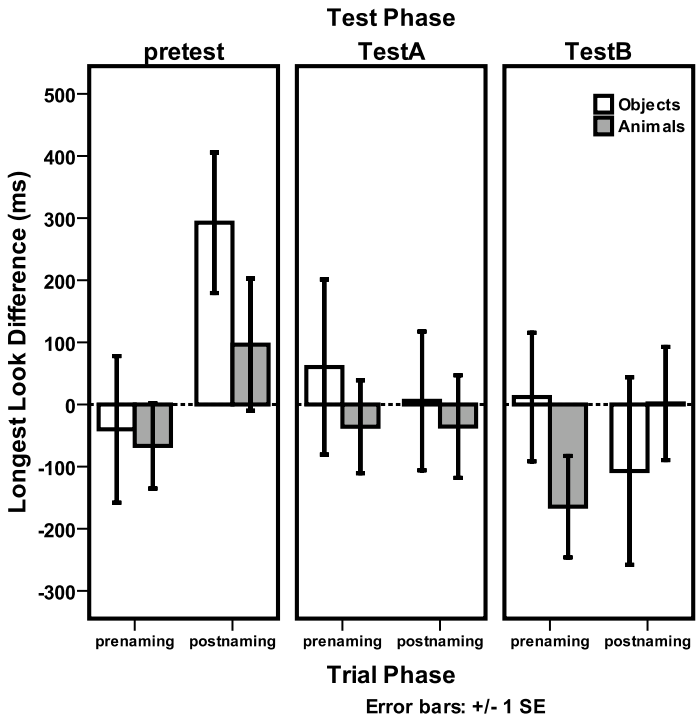


Figure 2.1: Longest Look Difference for the two Trial Phases (pre- and post-naming), in the three test blocks, separately for Conditions A (Objects) and B (Animals).

Trial Phase (before and after naming) and Test Block (pretest, testA, testB) as within-subjects factors and Condition (Animals and Objects) as a between-subjects factor was conducted for both PTL and LLD measures. PTL exhibited no main effects, Trial Phase: $F(1, 44) = .37$, $p = .55$; Test Block: $F(2, 43) = .84$, $p = .44$, Condition: $F(1, 44) = .06$, $p = .81$ or interactions. LLD showed no main effects, Trial Phase: $F(1, 41) = 2.83$, $p = .1$; Test Block: $F(2, 41) = .0$, $p = .97$; Condition: $F(1, 41) = 1.55$, $p = .22$. However, there was a non-significant trend for an interaction between Trial Phase and Test Block, $F(2, 40) = 3.13$, $p = .055$, ${\eta_p}^2 = .14$. Due to the absence of any significant main effects of Condition or interactions involving Condition, data were collapsed across Condition in subsequent analyses. On the basis of the interaction between Trial Phase and Test Block shown for the LLD measure, the three test blocks were analysed separately. Notice also that the individual test blocks are not strictly independent of each other since they are always conducted in

Measures	Trial Phase	Mean (SD)	N
PTL	pre-naming	.47 (.11)	47
	post-naming	.53 (.13)	47
LLD	pre-naming	-90.49 (380.73)	47
	post-naming	180.85 (534.66)	47

Table 2.3: Means and Standard Deviations for both PTL and LLD measures separately in the two phases of the trial for the pretest block.

the same order.

Pretest Block - Prior Knowledge

Figure 2.2 shows the proportion of infants looking at the target over the course of a trial, for 40ms-frames, in the pretest block. These data are calculated from aggregated means from all four trials in that block. The vertical dashed line indicates the onset of the target word. Before the presentation of the target word, approximately 50% of infants fixate the target picture. After the presentation of the target word, however, the majority of infants shift their fixation to the target picture. This increment in target looking suggests that the majority of infants associated the target word and picture.

Table 2.3 shows the means for PTL and LLD in the two Trial Phases of this first test block. In order to evaluate the statistical significance of the observed increase in target looking across the two trial phases, paired-samples t-tests between pre- and post-naming were run for PTL and LLD. Both measures exhibited a significant increase in target looking after the onset of the target word; PTL: $t(47) = -2.38$, $p < .05$; LLD: $t(45) = -2.99$, $p < .01$.

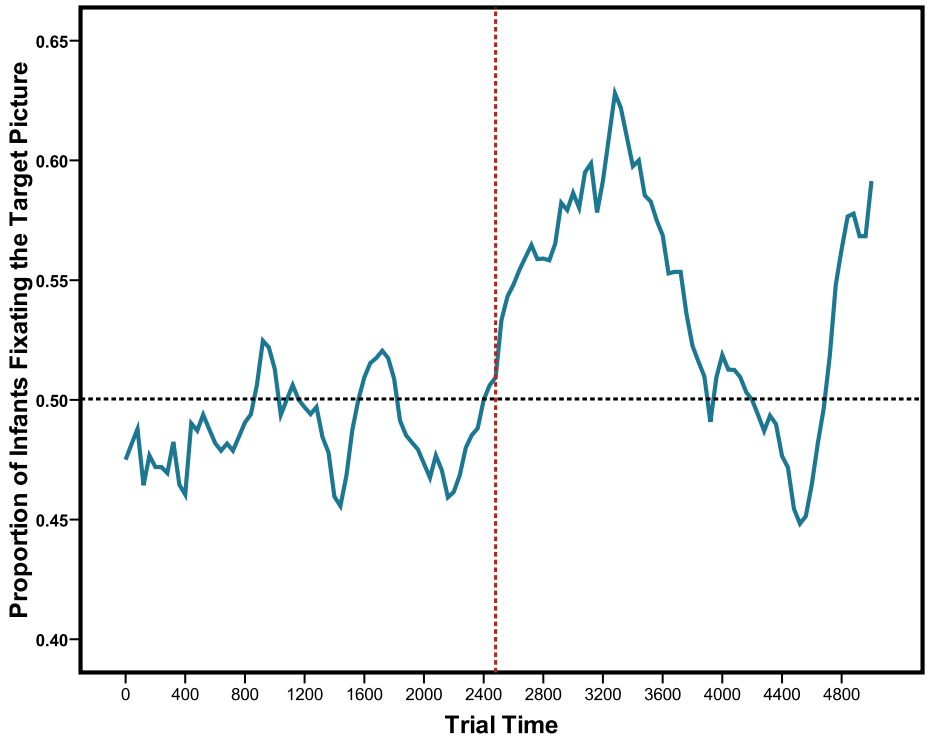


Figure 2.2: Proportion of infants looking at the target in each 40ms window of analysis. The vertical dashed line indicates the onset of the target word. The horizontal dashed line indicates the 50% chance level. Aggregated data from all four trials in this block are plotted.

One-sample t-tests were run for both PTL and LLD measures in the pre-naming phase, in order to evaluate whether infants showed any systematic biases to look at the objects before naming occurred. PTL was compared to a 50% chance level and LLD to a chance level of 0. Neither measure differed from chance before the presentation of the target word, confirming that infants looked roughly equally at the two pictures; PTL: $t(46) = -1.7$, $p = .10$; LLD: $t(46) = -1.6$, $p = .11$.

One-sample t-tests were also run for both measures in the post-naming phase. PTL did not differ from chance; $t(46) = 1.5$, $p = .14$, but LLD was significantly higher than chance; $t(46) = 2.32$, $p < .05$. See Table 2.3 for means and standard deviations. The sensitivity of the longest look measure has been found in previous studies using an IPL task (Meints et al., 1999) and will be discussed below.

Post Training Test Blocks - Word Learning and Generalisation of Labels

The first test block after training is identical to the pretest block, with an intervening training block separating them. It aims to test word learning of the training items. The second test block is identical to the other two in all respects, except that a new typical exemplar from each category is presented to the infants. Figures 2.3 and 2.4 show the proportion of infants fixating the target over the course of the trial, for aggregated means from all four trials of each block, for Test Block A and B respectively. It appears that the presentation of the label does not cause infant attention to shift to the target in either test block.

Tables 2.4 and 2.5 show the means for both PTL and LLD measures for the two post-training test blocks. Paired-samples t-tests comparing the two phases of the trial

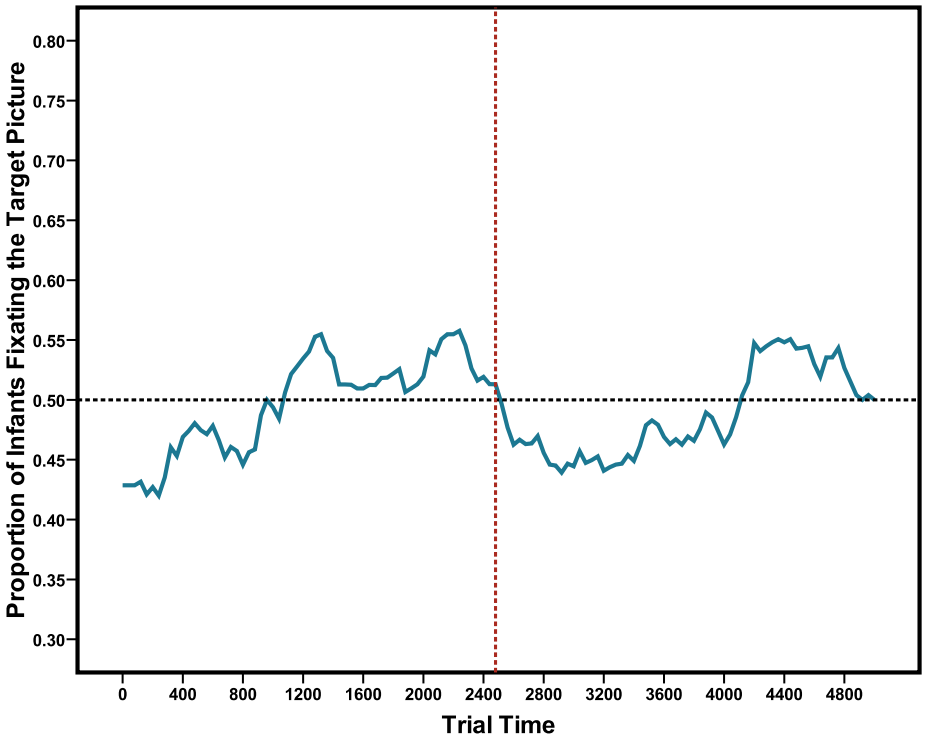


Figure 2.3: Proportion of infants fixating the target over the course of the trial in the first post-training test block. The vertical dashed line indicates the onset of the target word. The horizontal dashed line indicates the 50% chance level. Aggregated data from all four trials in this block are plotted.

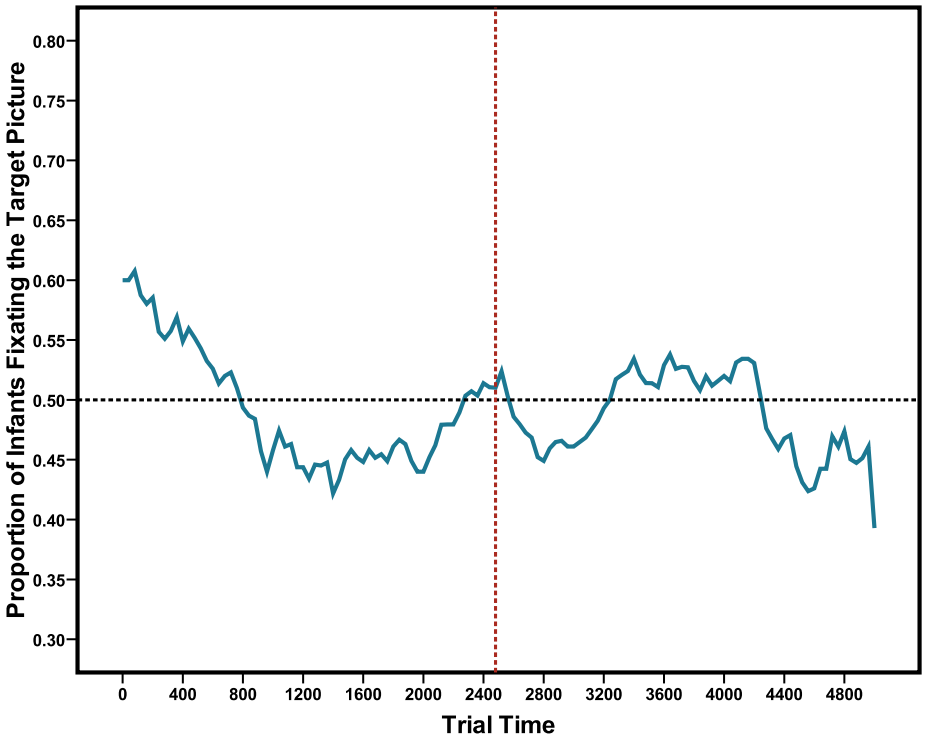


Figure 2.4: Proportion of infants fixating the target over the course of the trial in the second post-training test block. The vertical dashed line indicates the onset of the target word. The horizontal dashed line indicates the 50% chance level. Aggregated data from all four trials in this block are plotted.

Measures	Trial Phase	Mean	(SD)	N
PTL	pre-naming	.50	(.14)	48
	post-naming	.48	(.13)	48
LLD	pre-naming	12.24	(548.89)	48
	post-naming	-14.90	(476.52)	48

Table 2.4: Means and Standard Deviations for both PTL and LLD measures in the two phases of the trial for Test Block A.

Measures	Trial Phase	Mean	(SD)	N
PTL	pre-naming	.48	(.11)	47
	post-naming	.47	(.17)	47
LLD	pre-naming	-59.96	(459.57)	47
	post-naming	-76.63	(609.51)	47

Table 2.5: Means and Standard Deviations for both PTL and LLD measures in the two phases of the trial for Test Block B.

for both PTL and LLD measures confirmed the absence of any significant differences, Test Block A: PTL: $t(47) = .63$, $p = .53$; LLD: $t(47) = .27$, $p = .79$/ Test Block B: PTL: $t(46) = .3$, $p = .77$; LLD: $t(46) = .15$, $p = .89$.

The two measures were also compared to chance separately for the two phases of the trial using one-sample t-tests. For the pre-naming phase of the trial, attention to the target picture was shown to be at chance confirming the absence of any biases to look at any of the pictures before the presentation of the target word, Test Block A: PTL: $t(47) = -.12$, $p = .91$; LLD: $t(47) = .15$, $p = .88$/ Test Block B: PTL: $t(46) = -1.32$, $p = .19$; LLD: $t(46) = -.89$, $p = .38$. Target looking also was not shown to differ from chance in the post-naming phase of the trial, Test Block A: PTL: $t(47) = -1.13$, $p = .26$; LLD: $t(47) == .22$, $p = .83$/ Test Block B: PTL: $t(46) = -1.26$, $p = .21$; LLD: $t(46) = -.86$, $p = .39$.

Parental Feedback

Parental feedback on each infant's comprehension abilities was obtained through the Oxford CDI questionnaire. The accuracy of parental reports has been questioned (Tomasello & Barton, 1994; Houston-Price et al., 2007) especially with younger infants; in order to investigate the issue a separate analysis was conducted. Figure 2.5 shows the longest look measure separately for the two Comprehension groups in the three test blocks (see Section 2.5). Infants were included in the Comprehenders group if they were reported by parents to understand at least one of the testing items. Instead they were included in the Non-Comprehenders group if they were reported to understand neither of the words. The increase in target looking after the presentation of the word exhibited in the pretest block is also observed here; however, the two groups do not seem to behave differently.

In order to evaluate the accuracy of parental feedback regarding knowledge of the to-be-learned words, a 2 (Trial Phase) x 3 (Test Block) x 2 (Comprehension Group) repeated measures ANOVA was run for each measure. There were no significant main effects or interactions for either measure, apart from a significant interaction between Trial Phase and Test Block using the LLD measure, $F(2, 38) = 3.36$, $p < .05$, ${\eta_p}^2 = .15$. This analysis suggests that parents are not accurate at identifying whether their 10-month-old infants comprehend individual words.

The above analysis investigated whether parental reports on infants' understanding of the test items were a reliable predictor of their performance. However, in previous studies it has been suggested that parents are more accurate and consistent at estimating their infant's overall vocabulary size (Houston-Price et al., 2007). If this is the case, TCS (Total Comprehension Score) might be a better predictor of

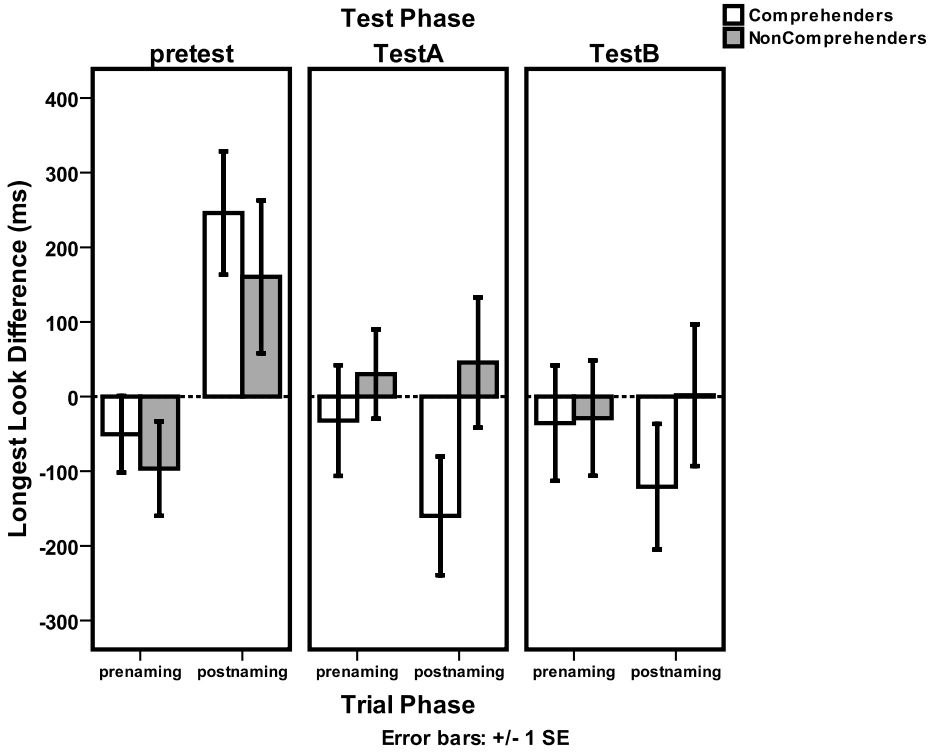


Figure 2.5: LLD in the two trial phases separately for each Comprehension group in the three test blocks.

infant performance in an IPL task (see Section 2.5 for a description of how TCS was calculated). In order to test this possible relationship between parental feedback and performance of each infant in the lab, a repeated-measures ANCOVA was conducted with Trial Phase as a within-subjects factor and TCS as covariate for both measures in the pre-test block where evidence for word understanding was exhibited. Also a Pearson r correlation was run comparing the change in target looking (post-naming - pre-naming) and TCS for the same test block for both measures. The main effect of Trial Phase was replicated for both measures, PTL: $F(1, 40) = 9.80$, $p < .01$, ${\eta_p}^2 = .2$/ LLD: $F(1, 39) = 5.73$, $p < .03$, ${\eta_p}^2 = .13$. There was no effect of TCS on Trial Phase for either measure, $F(1, 40) = 2.18$, $p = .15$/LLD: $F(1, 39) = .43$, $p = .51$ (see Figure 2.6). Also there was no significant correlation between the change in target looking for either measure ($r(40) = .10$, $p = .53$/ LLD: $r(39) = .09$, $p = .56$. This analysis suggests that parental feedback regarding infants overall vocabulary size

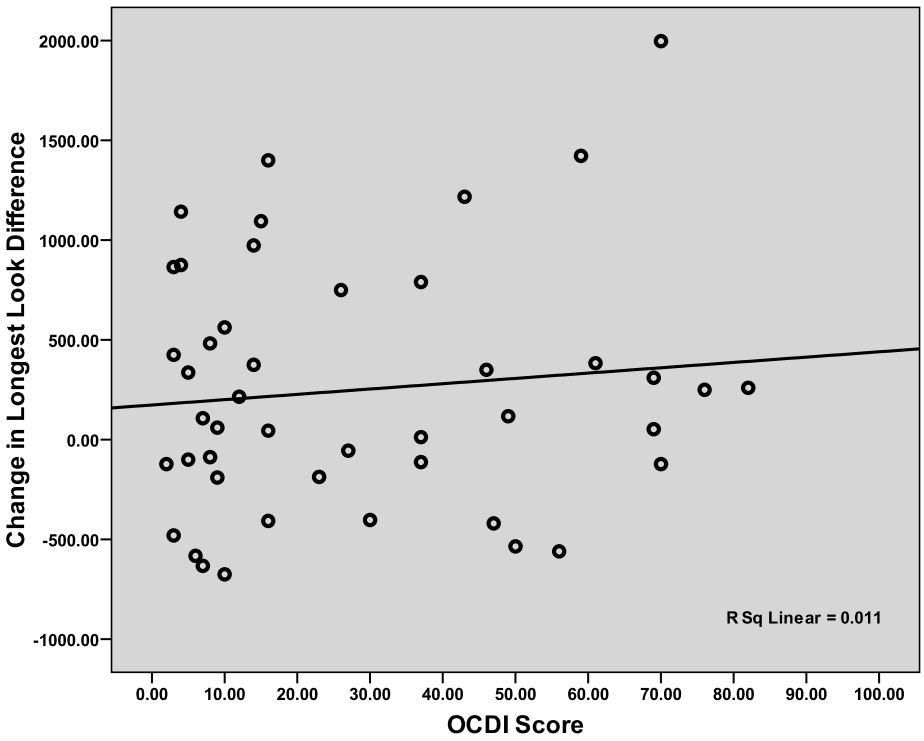


Figure 2.6: Scatter plot showing the relationship between each infant's Oxford CDI score and the change in the LLD measure (LLD post-naming - LLD pre-naming) in the pretest block.

does not predict their overall performance in an IPL task.

Latency Analysis

As mentioned in section 2.7, latency of shift is a measure most commonly used with older infants. A latency analysis was run, in order to test whether infants in the first year of life shift faster to the target if they are fixating the distracter at the onset of the target word than to the distracter if they are fixating the target at the onset of the target word.

In figure 2.7 the white bar shows the average time infants need to shift their gaze

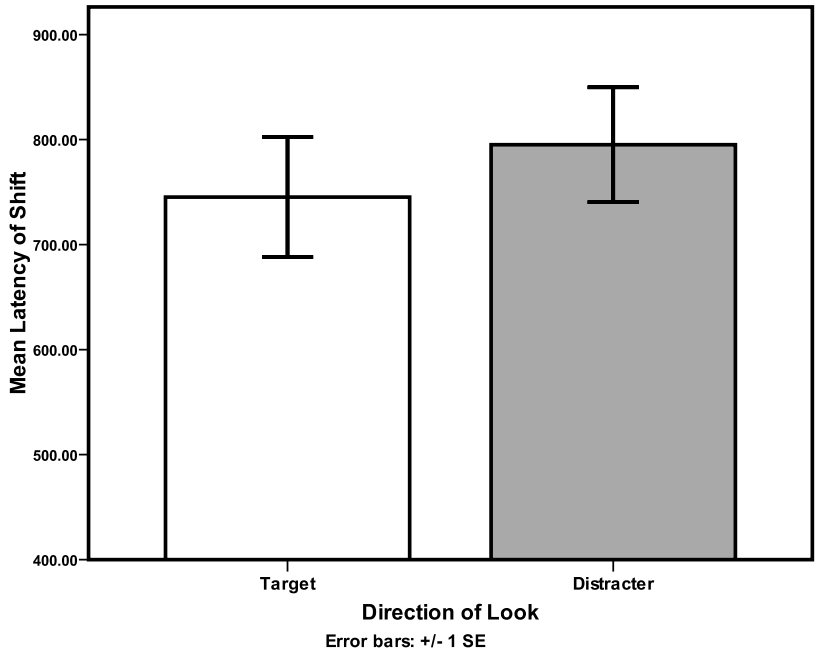


Figure 2.7: Mean time needed to shift away from the fixating picture when the target word is presented. The white bar shows shifts *to* target and the gray bar shifts *to* distracter. No significant differences were found between the two.

to the target picture if they are fixating the distracter at the onset of the target word, and the gray bar shows the average time to shift to the distracter picture when fixating the target at the onset of the target word. Figure 2.7 shows that there is no difference in the latencies to shift away from the target or the distracter. A one-way ANOVA compared the latency of first shift towards the target and latency of first shift toward the distracter. No significant differences were found between shifts to target and shifts to distracter ($F(1, 61) = 4$, $p = .53$) confirming the above observations.

It is difficult to interpret this null result. However, other studies with young infants also fail to find latency effects whilst reporting significant findings using other measures, such as PTL or LLD. It might be concluded that latency measures are not appropriate for use with such young infants.

Habituation Effects

The repeated presentations of the training items as well as the evidence for prior knowledge of the labels suggest that infants might be *habituated* over the course of the session causing them to lose interest in the task in the post-training test blocks. Figure 2.8 shows attention to the screen (in milliseconds) for the two phases of the trial separately in each test block. In the pretest block infants seem to pay similar attention to the screen before and after the presentation of the linguistic stimulus; also they seem to spend most of the picture presentation time focusing on the screen. Out of the 5000 milliseconds that each trial lasts infants spend on average approximately 4200ms fixating the two pictures. Of the remaining 800ms a fair amount of time is spent on shifts from one picture to the other, which are excluded from these calculations.

In the post-training test blocks, however, although attention remains relatively high in the pre-naming phase of the trial, it falls after the presentation of the word. It appears that infants exhibit effects of habituation after training, and not even the presentation of the linguistic stimulus can maintain their attention to the task.

In order to evaluate these observations, a 2 (Trial Phase) x 3 (Test Block) x 2 (Gender) repeated measures ANOVA with Trial Phase and Test Block as within-subjects factors and Gender as between-subjects factor was run for the TLT measure. Trial Phase was included as a factor in the analysis in order to test the effect of the auditory/linguistic stimulus on infant attention over the course of the experiment. Gender was added as a between-subjects factor in the analysis because gender effects have been observed in previous studies using training or habituation paradigms (Werker et al., 1998).

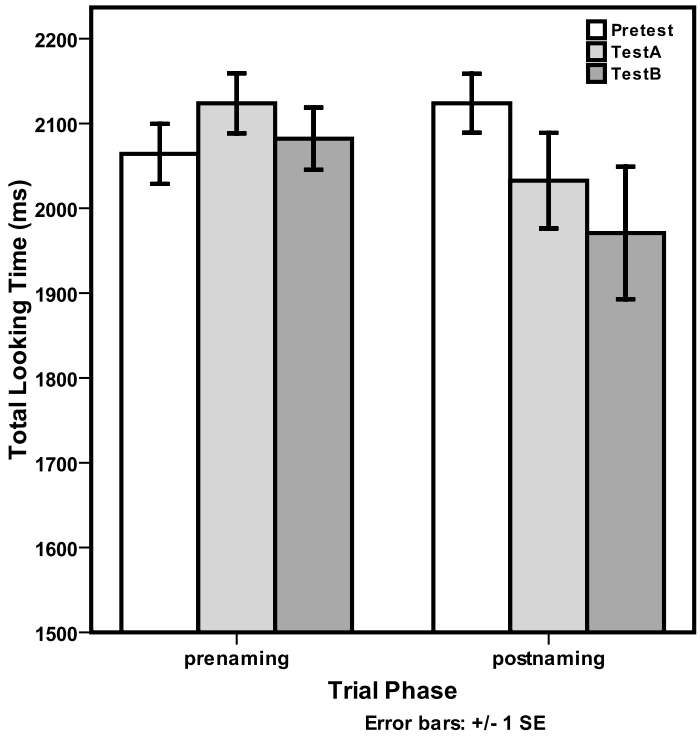


Figure 2.8: TLT in the two phases of the trial separately for each test block. Time for shifts between the two pictures is not included in the TLT measure.

This analysis revealed a marginally significant main effect of Trial Phase, $F(1, 31) = 3.81$, $p = .06$, ${\eta_p}^2 = .12$, a significant main effect of Test Block, $F(2, 30) = 5.93$, $p < .01$, ${\eta_p}^2 = .28$, and an interaction between Trial Phase and Test Block, $F(2, 30) = 8.56$, $p < .01$, ${\eta_p}^2 = .13$. No main effect of Gender, $F(1, 31) = .33$, $p = .57$, and no other interactions were exhibited. Post-hoc tests (paired samples t-tests) compared the two naming phases of TLT for each test block: Pretest: $t(46) = -.82$, $p = .42$/ Test A: $t(45) = 2.65$, $p < .02$/ Test B: $t(46) = 2.29$, $p < .03$. Table 2.6 shows descriptive statistics for each phase of the trial and each test block separately.

Test Block	Trial Phase	Mean (SD)	N
Pre-Test	pre-naming	2067.48 (226.35)	47
	post-naming	2113.71 (336.22)	47
Test Block A	pre-naming	2117.59 (244.91)	46
	post-naming	2000.04 (337.02)	46
Test Block B	pre-naming	2071.72 (241.17)	47
	post-naming	1942.69 (404.71)	47

Table 2.6: Means and Standard Deviations for the TLT (Total Looking Time) measure separately in the two phases of the trial for the three test blocks.

Discussion

In this first experiment infants were presented with a pretest-train-test-test paradigm: this design aimed to test prior knowledge of the to-be-learned object-label pairings (pretest), to train infants with these pairings (train) and to test them again for word learning with exactly the same pairings (testA) and also for label generalisation with the same labels but pictures of new typical members from the corresponding adult categories (test B). The purpose of this experiment is two-fold: to test whether infants at 10 months are able to form associations between (pictures of) objects and labels in brief experimental sessions; and most importantly to test generalisation abilities of infants at this very young age. In other words, when infants learn a new label for an object, do they assume that it refers to a group of objects, as principles-and-constraints approaches would suggest, or do they restrict this newly-learned piece of information to the object they initially heard it associated with, as assumed by exemplar-based accounts?

The analysis of infant attention to the pictures before and after the onset of the target word provided no evidence for comprehension of the test words in the two post-training test blocks. Infants spent similar times looking at the two pictures before and after they heard the target word. However, null results are not easy to interpret: is the absence of evidence for comprehension due to the failure of infants to form

an association between the target word and picture or is it due to some other factor which affected infants ability to exhibit this comprehension in the post-training test blocks?

At first glance, failure to find evidence for comprehension after training could be considered rather surprising, because such behaviour was found during pretest. It should, therefore, be expected that infants would also exhibit comprehension of the words at least in the first post-training test block, which was identical to the pretest.

Research on infant attentional preferences, however, suggests that this behaviour is not completely unexpected. If infants come to the task with some prior knowledge of the to-be-learned words, as shown in pretest, then the repetitive presentations of the same object-label pairings could cause infants to habituate and lose interest in the task. This explanation accords well with the Hunter and Ames (1988) Model of Infant Preferences (see Section 3.1 for a more detailed description of the model): initially infants are assumed to show a familiarity preference, before they move to a second phase when their preference is at chance, and then to a third phase when they prefer to attend to novelty. Such attentional preferences have been found using an IPL task with infants demonstrating their familiarity preference in the form of target looking after the presentation of the target word (attention to the familiar *pairing*) at the beginning of the experiment and then gradually switching to preference for novelty by shifting their attention to the distracter picture upon the presentation of the target word (Houston-Price & Nakai, 2004).

The assumption that infants habituated in Experiment One is supported by the TLT measure. Overall attention to the screen fell in the post-training test blocks, where infants were shown to turn away from the screen a lot quicker than in the

pretest block. Also the pre-naming phase of the trial was shown not to differ across test blocks, but post-naming infants exhibited an increase in attention only in the pretest block and a decrease in attention in the post-training test blocks.

It is interesting to note that attention does not seem to recover in test block B, where infants are presented with pictures of new exemplars from the same categories. One explanation for this behaviour is that infants do not treat these new pictures as something novel and do not therefore show renewed interest in the task. Their random looking as well as the decreasing TLT suggests that these new exemplars are treated as the "same sort of thing" as they had seen before (Younger & Cohen, 1983). This behaviour is well documented in habituation studies with infants at the same age when habituated with multiple objects from one category and tested with items from the same category and objects from a different category in the absence of labels (Younger & Cohen, 1986). However, if infants are habituated to a single item and tested with a new within-category item attention increases showing discrimination between the two items (Younger, 2003). This pattern appears inversed here: although infants are presented with a single exemplar from each category, attention does not recover when new exemplars are presented. Infants seem to understand something about the linguistic nature of the task and, even when presented with new exemplars, they fail to renew their attention[5]. A possibility is that even at such early stages of word learning, infants have discovered something important about the inter-modal nature of language, namely that words refer to groups of objects, and not just to individual instances of them. This overshadows the presentation of new visual stimuli and leads infants to treat both old and new stimuli as the "same sort of thing".

[5]An alternative explanation for infants' failure to increase attention to the new pictures is that they fail to discriminate between the old and new exemplars. This is unlikely, however, since the pictures differed in terms of colour and shape - two stimulus dimensions which have been shown to be highly salient from very early in life (Heider, 1970, 1971). See Appendix A for the pictures used in Experiments One and Two.

Word	LEX2005	Oxford CDI
bear	32.8	0.0
cat	44.8	8.3
duck	23.9	8.3
cup	41.8	0.0
phone	35.8	0.0
spoon	29.9	0.0

Table 2.7: Percentages of infants reported to understand the test labels in the LEX2005 database and the Oxford CDI word list.

Although infant-language databases, as well as parental feedback for individual infants in this experiment, indicated that the majority of 10-month-olds should not understand the words used in this study (for exact percentages see Table 2.7), evidence for prior knowledge of the labels was exhibited in the pretest block. Parents reported in the Oxford CDI check list that none of the infants comprehends any of the object-words, but the two Conditions (Animals and Objects) were shown to promote similar understanding of the test words. This suggests that the linguistic abilities of these infants might be more advanced than has previously been thought and that parental feedback might not be an accurate estimate for infant comprehension abilities, especially for young pre-verbal infants, as has been recently suggested (Houston-Price et al., 2007).

It is also worth noting that 10-month-old infants exhibit some preliminary categorisation ability during pretest, which allows them to recognise de-contextualised, stationary pictures of non-salient items as referents of a known label. This had been shown for infants as young as 12 months, but not earlier[6], under two conditions: after training with several exemplars from different categories over an extensive period of

[6]Some evidence for word comprehension has been exhibited for infants at 6 months, but only using socially salient stimuli, like "mummy" and "daddy" (Tincoff & Jusczyk, 1999)

time (Schafer, 2005) and without training with items that parents reported their infants already understand (Meints et al., 1999).

The two measures used for investigating word comprehension, PTL and LLD, showed similar results. However, in a number of IPL studies, the LLD measure has been shown to be a more sensitive indicator of language understanding (Meints et al., 1999; Schafer & Plunkett, 1998). In this experiment, although there was a significant increase for both measures in the post-naming phase of the trial during pretest, only LLD was shown to be significantly higher than chance after naming. This occasional "opacity" of the PTL and sensitivity of the LLD measure to naming effects could be explained on the basis of task characteristics as well as processing needs during the task: infants are presented with a new picture game, during which they are asked to find the picture that goes with the word they hear. If their object-label associations are not very strong, they might spend time looking at both pictures, but will need longer continuous looks to identify the correct picture than to reject the incorrect one. Increased looking to the target picture after the presentation of the target word appears robust enough to keep the difference between the two trial phases significant for both measures. However, if infant behaviour is compared to chance, naming effects are only traceable with LLD, but not with PTL. It appears that the LLD and PTL measures might be tapping different aspects of language processing: LLD might be an index of the time needed to match the incoming stimulus to a mental representation of the word-picture association. PTL on the other hand might capture, among other things, behaviour which is not guided by the linguistic stimulus, introducing noise in the data. However, the two are intimately connected since longer continuous looks to a picture will cause an increase to the PTL measure and shorter continuous looks will cause the PTL measure to decrease.

The latency of shift measure failed to trace any evidence for comprehension. This measure has been used quite extensively with older infants (18m and above) (Fernald

et al., 2001), but has also been used successfully with infants as young as 14-15 months of age (Mani & Plunkett, 2007). It appears that 10-month-old infants need to improve both their accuracy and speed in shifting to the target in an IPL task, before they are able to exhibit their understanding of words in this manner. Latency was, therefore, not used in any subsequent analyses.

2.8.2 Experiment Two

Experiment Two followed a train/test/test design with two test blocks after training. It only differed from Experiment One in that in the latter infants were presented with a pretest block aiming to test prior knowledge of the to-be-learned words. Although this was an important control, it was difficult to appreciate the impact of such a pretest block on the behaviour of such young infants. A pretest block could create a number of "adverse" outcomes, such as confusion as to what the task is, boredom due to the many repetitions or it could even provide infants with cues as to which picture goes with which word. It was, therefore, decided to also conduct Experiment Two without the pretest block and control for the role of the pretest block in this task.

Twenty four infants participated in each of the two Conditions (24 female). The first post-training test block aimed to test word learning whereas the second aimed to test generalisation of the labels to new typical members from the same adult categories. Figure 2.9 shows the PTL measure in the two phases of the trial separately for each test block.

In the pre-naming phase of the trial appears to remain at chance levels in both test blocks - that is, around 50%. In the post-naming phase of the trial, in the first

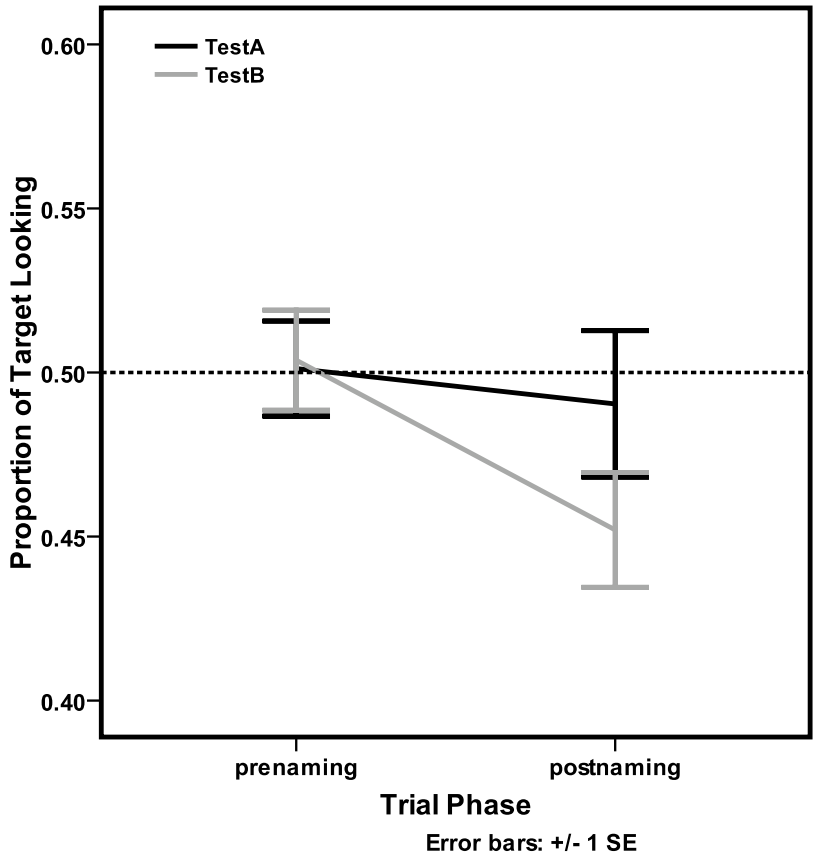


Figure 2.9: Proportion of Target Looking before and after naming separately for each post-training test block in Experiment Two.

Condition	Measure	Trial Phase	Mean	(SD)	N
Animals	PTL	pre-naming	.50	(.10)	24
		post-naming	.49	(.12)	24
	LLD	pre-naming	30.76	(420.90)	24
		post-naming	-11.88	(231.61)	24
Objects	PTL	pre-naming	.50	(.10)	24
		post-naming	.51	(.16)	24
	LLD	pre-naming	17.74	(407.30)	24
		pre-naming	19.55	(587.11)	24

Table 2.8: Means and Standard Deviations for both PTL and LLD measures separately in the two phases of the trial and the two Conditions (Animals and Objects) in Test Block A.

post-training test block, proportion of looking to the two pictures remains at chance level. In the second post-training test block, however, attention to the target picture appears to decrease after the presentation of the target word.

Table 2.8 and Table 2.9 show both PTL and LLD measures in Test Blocks A and B respectively, separated by Condition (Animals and Objects). It appears that the two Conditions do not differ in Experiment Two. No effects of Trial Phase seem to be present in Test Block A, but a decrease in target looking in Test Block B is apparent.

In order to evaluate these observations, a 2 (Trial Phase) x 2 (Test Block) x 2 (Condition) repeated measures ANOVA with Trial Phase (before and after the onset of the target word) and Test Block (A and B) as within-subjects factors and Condition (Animals and Objects) as a between-subjects factor was conducted for both PTL and LLD measures. No main effects or interactions were exhibited for either measure, except for a trend toward a main effect of Test Block for the PTL measure, PTL: $F(1, 44) = 3.44$, $p = .07$, ${\eta_p}^2 = .07$; LLD: $F(1, 45) = .37$, $p = .55$.

Condition	Measure	Trial Phase	Mean	(SD)	N
Animals	PTL	pre-naming	.51	(.10)	24
		post-naming	.45	(.12)	24
	LLD	pre-naming	86.88	(296.80)	24
		post-naming	-15.31	(444.83)	24
Objects	PTL	pre-naming	.48	(.10)	24
		post-naming	.45	(.12)	24
	LLD	pre-naming	-42.95	(469.23)	24
		pre-naming	-97.22	(608.69)	24

Table 2.9: Means and Standard Deviations for both PTL and LLD measures separately in the two phases of the trial and the two Conditions (Animals and Objects) in Test Block B.

Data were collapsed across Condition for subsequent analyses. Although there was no significant main effect of Test Block, the trend found in the previous analyses as well as Figure 2.9 suggests that infants could have behaved differently in these two test blocks; a separate analysis was, therefore, conducted in order to investigate this further.

Test Block A - Word Learning

Figure 2.9 suggests that target looking in Test Block A does not differ neither between the pre- and post-naming phases of the trial nor from chance. Figure 2.10 shows the proportion of infants fixating the target picture over the course of the trial, for aggregated means from all four trials of this block. It appears that infants exhibit similar behaviour before and after they hear the target word.

In order to evaluate these observations, paired-samples t-tests were conducted for both PTL and LLD measures; neither measure exhibited significant differences between the two trial phases, PTL: $t(46) = -.02$, $p = .98$; LLD: $t(46) = .17$, $p = .87$.

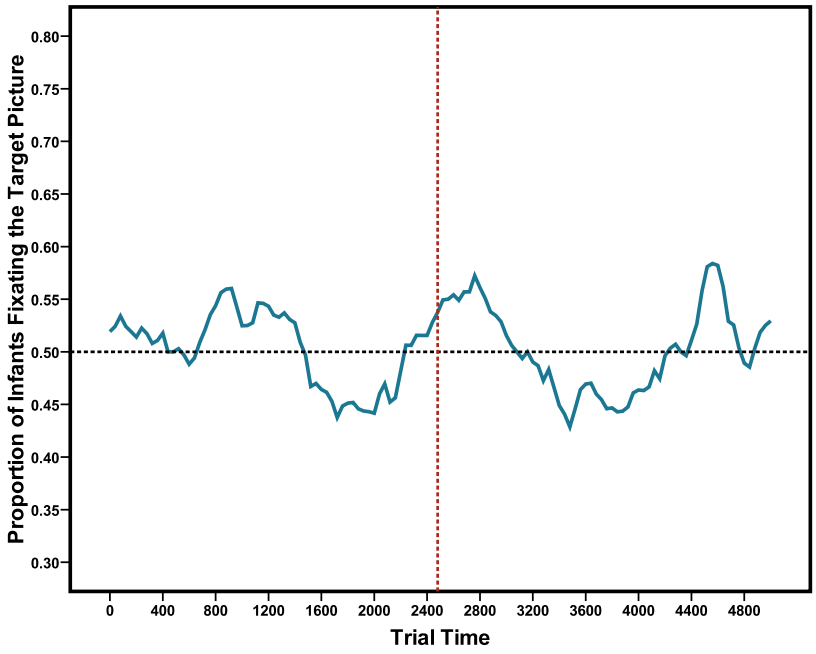


Figure 2.10: Proportion of infants fixating the target picture over the course of the trial, for aggregated means from all four trials in Test Block A. The vertical dashed line indicates the onset of the target word, whereas the horizontal dashed line indicates the chance level of 50%.

One-sample t-tests were also conducted comparing the two phases of the trial to chance for both measures. Neither of the measures differed from chance in either trial phase, pre-naming: PTL: $t(47) = .08$, $p = .94$; LLD: $t(47) = .41$, $p = .68$/ post-naming: PTL: $t(46) = .04$, $p = .97$; LLD: $t(46) = .06$, $p = .95$.

Test Block B - Generalisation of Labels

Figure 2.9 suggests that, in this second post-training test block, infants appear to behave more systematically than in the first test block, but showing a *decrease* in target looking in the post-naming trial phase. Figure 2.11 shows the proportion of infants fixating the target picture over the course of the trial, for aggregated means from all four trials in Test Block B. This figure suggests that more infants tend to orient to the distracter picture after the onset of the target word, although this does

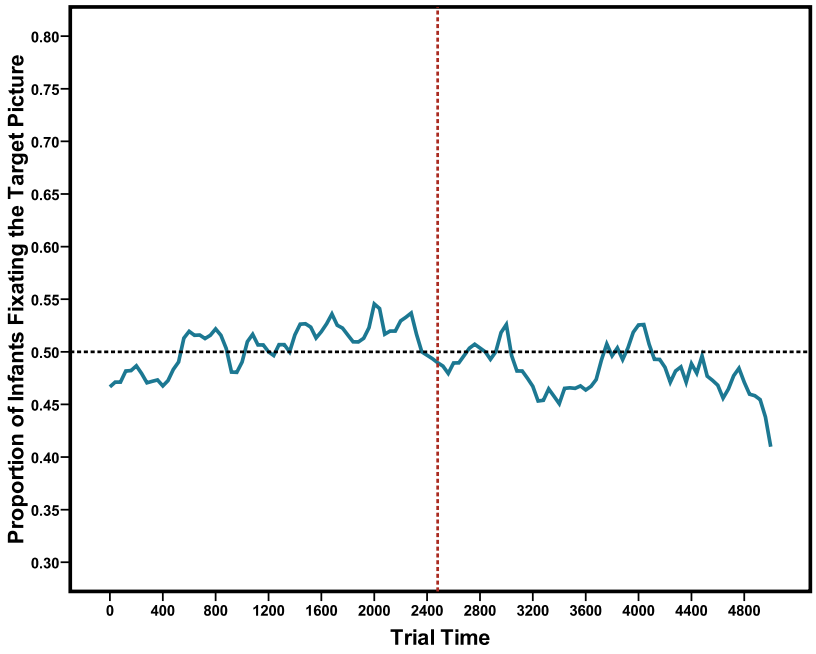


Figure 2.11: Proportion of infants fixating the target picture over the course of the trial, for aggregated means from all four trials in Test Block B. The vertical dashed line indicates the onset of the target word, whereas the horizontal dashed line indicates the chance level of 50%.

not appear to be a very strong effect.

In order to investigate these observations, paired-samples t-tests were conducted for both PTL and LLD measures. For the LLD measure, the two trial phases were not found to differ significantly from each other, $t(47) = .76$, $p = .45$. However, there was a weak trend for a difference for the PTL measure, $t(46) = 1.75$, $p = .09$. One-sample t-tests were also conducted for both measures comparing the two phases of the trial to chance. In the pre-naming phase of the trial, target looking was at chance for both measures, PTL: $t(46) = -.16$, $p = .87$; LLD: $t(47) = .39$, $p = .70$. In the post-naming phase, LLD remained at chance, $t(47) = -.74$, $p = .47$, but PTL showed a significant *decrease* in target looking, $t(47) = -2.74$, $p < .01$.

This analysis suggests that in Test Block B the presentation of the target word had a systematic effect on infant attentional preferences, but this effect was in the opposite direction of what is expected in an IPL task. Distracter looking has been found, however, in previous studies using a similar task (Houston-Price & Nakai, 2004) and its implications will be discussed below.

Parental Feedback

The accuracy of parental feedback regarding infant performance in an IPL task was also investigated in this Experiment. Figure 2.12 shows the PTL measure in the two test blocks separately for Comprehenders and Non-Comprehenders. It appears that infants in the two groups exhibit similar attention to the target in the pre-naming phase of the trial for both blocks. Post-naming, however, the decrease in target looking in Test B seems to be driven by the Non-Comprehenders.

In order to investigate this a 2 (Trial Phase) x 2 (Test Block) x 2 (Comprehension Group) repeated measures ANOVA with Trial Phase (before and after the onset of the target word) and Test Block (A and B) as within-subjects factors and Comprehension group (Comprehenders and Non-Comprehenders) as a between-subjects factor was conducted for both PTL and LLD measures. This analysis showed no main effects, Trial Phase: PTL: $F(1, 44) = .37$, $p = .55$; LLD: $F(1, 45) = .00$, $p = .99$/ Test Block: PTL: $F(1, 44) = 2.18$, $p = .15$; LLD: $F(1, 45) = .08$, $p = .78$/ Comprehension Group: PTL: $F(1, 44) = .04$, $p = .84$; LLD: $F(1, 45) = .41$, $p = .53$, and no interactions for either measure.

A second analysis investigated whether the total receptive vocabulary of individual

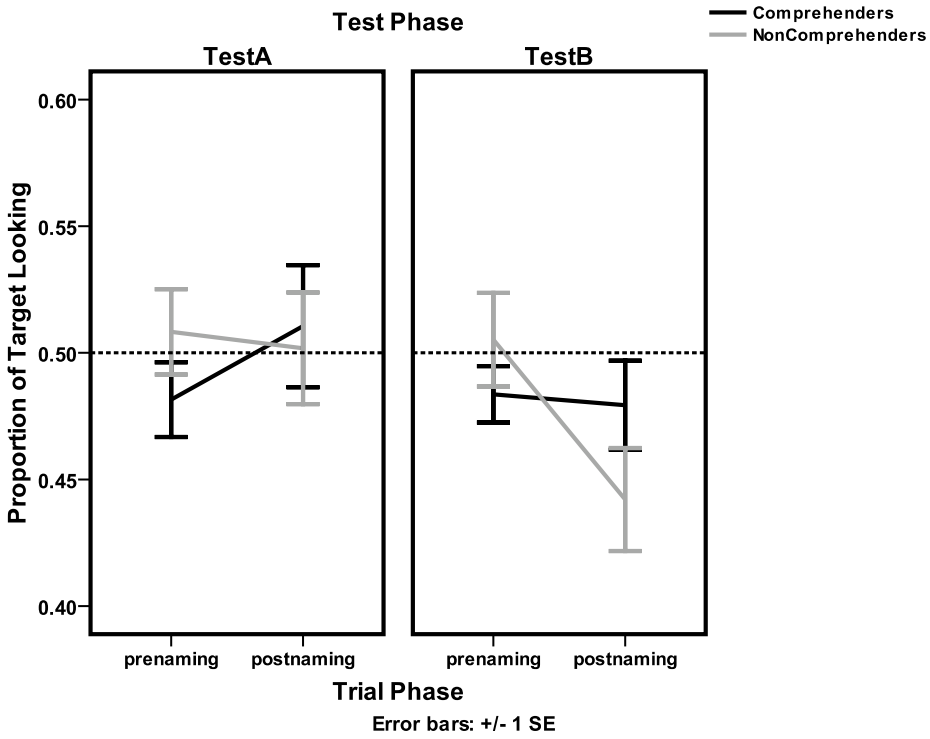


Figure 2.12: Proportion of target looking in the two trial phases separately for each test block and Comprehension group.

infants as reported in the Oxford CDI questionnaire predicted their performance in the task. A repeated measures ANCOVA with Trial Phase (pre- and post-naming) and Test Block (A and B) as within-subjects factors and TCS (total comprehension score) as a covariate was conducted for both measures. There were no main effects and no interactions for either of the measures, Trial Phase: PTL: $F(1, 43) = .05$, $p = .83$/LLD: $F(1, 43) = .02$, $p = .90$; Test Block, PTL: $F(1, 43) = .28$, $p = .59$/LLD: $F(1, 43) = .06$, $p = .81$; Comprehension group: PTL: $F(1, 43) = 1.56$, $p = .22$/LLD: $F(1, 43) = 1.77$, $p = .19$.

This analysis suggests that parental feedback on understanding of individual words and infants overall vocabulary size does not predict their performance in an IPL task. The interaction between Trial Phase and Test Block is not found in this analysis, due to the high variability within the two comprehension groups.

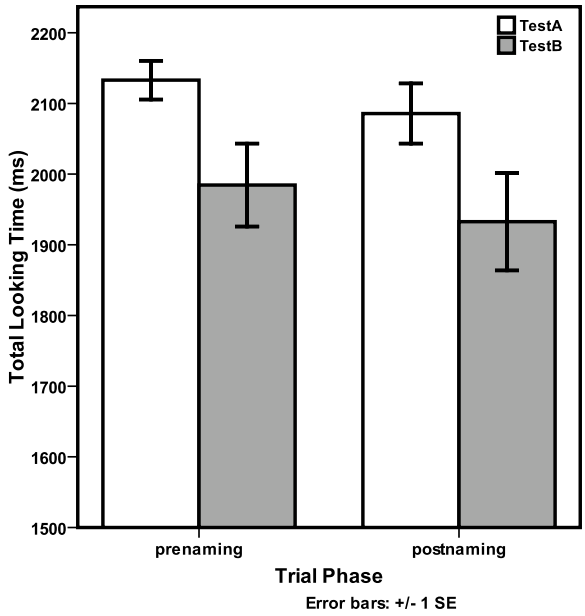


Figure 2.13: Total Looking Time to screen (TLT to target + TLT to distracter) for Experiment Two.

Habituation Effects

Although infants in Experiment Two were not presented with a pretest block, an analysis for habituation effects was run comparing attention to the task in the two post-training test blocks. Figure 2.13 shows the TLT measure in the two test blocks. It appears that, as in Experiment One (see Figure 2.8), infants show decreased attention in the post-naming phase of the trial.

In order to investigate the presence of any habituation effects, a 2 (Trial Phase) x 2 (Test Phase) x 2 (Gender) repeated measures ANOVA with Trial Phase (before and after the onset of the target word) and Test Block as within-subjects factors and Gender as a between-subjects factor was run for the TLT measure. A significant main effect of Trial Phase was found, $F(1, 46) = 8.88$, $p < .01$, ${\eta_p}^2 = .17$, but no main effect

of Test Phase, $F(1, 46) = 2.18$, $p = .15$, nor Gender, $F(1, 46) = .68$, $p = .41$. No interactions were found between Trial Phase and Test Phase, $F(1, 45) = .34$, $p = .56$, between Trial Phase and Gender, $F(1, 46) = .48$, $p = .49$, or between Test Phase and Gender, $F(1, 46) = .99$, $p = .33$.

This analysis provides results similar to those obtained in Experiment One. In the absence of a pretest block, it is difficult to interpret the main effect of Trial Phase, but it might indicate that, as in Experiment One, infants quickly lose attention to the screen in the post-training test blocks due to decreasing interest in the task.

Discussion

Experiment Two was identical to Experiment One, except for the pretest block. Infants were presented with a train-test-test task: this design aimed to train infants with two object-label pairings, to test for word learning of these pairings (Test A) and also to test for generalisation of the labels to new typical exemplars from the same adult categories (Test B). As with Experiment One, the purpose was twofold: to investigate the word learning abilities of infants toward the end of their first year of life as well as the mechanisms by which infants generalise their first words to different exemplars from the same categories. If infants are willing to generalise immediately a newly-learned label to different objects, that would indicate the existence of "taxonomic-like" behaviour. If on the other hand infants restrict a newly-learned label to the object they heard it occurring with, that would lend support to exemplar-based accounts of lexical development.

Infants in this Experiment showed no evidence of word learning in Test Block A. However, they showed a systematic preference for the *distracter* picture after the

presentation of the target word in Test Block B (see page 81). Distracter looking has been shown in previous studies using an IPL task and has been interpreted as an alternative form of evidence for word understanding (Houston-Price & Nakai, 2004). Infants are assumed to prefer to attend to stimuli of different nature at different points over the course of an experimental session. Initially a phase of familiarity preference is assumed, whereby infants try to match an existing mental representation to a familiar stimulus. When this process has been completed, a phase of novelty preference is assumed to follow, whereby infants shift their attention to a new stimulus[7] (Hunter & Ames, 1988). In an IPL task infants are assumed to demonstrate their familiarity preference with increased attention to the target picture after the presentation of the word. In designs which involve repetitive presentations of the same stimuli, infants might become habituated to these picture-word pairings and demonstrate their understanding of the labels through increased attention to the distracter picture after they are presented with a label. This is the kind of behaviour demonstrated in this Experiment, if we assume that infants come to the task with prior knowledge of the to-be-learned pairings. However, the absence of a pretest block makes these findings difficult to interpret.

As in Experiment One, Condition and Comprehension group were not shown to affect the understanding of the labels. Table 2.10 shows percentages of infants reported to understand the testing items in the LEX2005 database as well as in the Oxford CDI check list. All of the words were reported to be known by some of the infants who participated in Experiment Two, although no evidence for understanding was found in either Test Block.

[7]A phase of no preference in the transition between familiarity and novelty preference has also been assumed but it is a controversial issue in the literature (Roder, Bushnell, & Sasseville, 2000) and there is limited experimental evidence for it. This phase of no preference will, therefore, be omitted from this discussion.

Word	LEX2005	Oxford CDI
bear	32.8	8.33
cat	44.8	16.66
duck	23.9	25.00
cup	41.8	8.33
phone	35.8	12.50
spoon	29.9	16.66

Table 2.10: Percentages of infants reported to know the test labels in the LEX2005 database and the Oxford CDI word list.

Methodological Issues. Unlike Experiment One, in Experiment Two evidence for word understanding was detected primarily with the PTL measure and in the form of distracter looking, as discussed above. It appears that, as in Experiment One, infants might have come to the task with some prior knowledge of the to-be-learned word-object pairings. In the post-training test blocks infants seem to be involved in the task, since they exhibit some systematic behaviour. Their behaviour, however, is not necessarily, or exclusively, guided by the linguistic stimulus they are presented with, due to habituation effects caused by the repetitive presentations of the same word-object pairings.

As discussed in Experiment One, PTL and LLD might be tapping different aspects of information processing with PTL being better able to capture behaviour which is not strictly guided by the linguistic stimulus: in the post-training test blocks infants might be able to match the incoming stimulus with their mental representation very quickly, and, therefore, need shorter continuous looks to find the picture which goes with the word. This means that the LLD measure might not be a good index of infants' understanding of the labels. PTL on the other hand might be a better index in this case, because infants prefer to attend to novelty more, and even though they might turn to look at both visual stimuli, they focus more on the distracter.

2.9 General Discussion

Experiments One and Two aimed to teach infants two object-label pairings and to test generalisation of the labels to new typical exemplars from the same adult categories. Evidence for word learning in the post-training test blocks was limited (mainly in the form of distracter looking) and was only found in Experiment Two. In Experiment One prior knowledge of the labels was found during pretest. In Experiment Two the systematic preference for the distracter in the second post-training test block suggests that infants were habituated to the object-label pairings as a result of training. Although word learning using habituation has been shown in the past, older infants (14 months) were tested and they needed more trials to begin to decrease their attention (Werker et al., 1998). It would, therefore, be very surprising if infants at 10 months exhibited habituation effects after training with only eight trials, if they were engaging in word learning.

In both experiments parental feedback on their infant's overall vocabulary or concerning individual words was shown not to be predictive of infant performance in the task. This concern was recently raised by other researchers (Houston-Price et al., 2007), and it appears that there is plenty of evidence to suggest that parental feedback at the early stages of language development might be of limited value. This was taken into account in subsequent studies.

In terms of methodology, it appears that the two measures used to investigate word comprehension in an IPL task might be tapping different aspects of information processing. LLD was shown to be a better index of word comprehension in the pretest block of Experiment One, where infants possibly needed longer continuous looks to match the incoming stimulus to their mental representation. In the post-training blocks of Experiment Two the processing time needed for infants to identify the words decreases and, therefore, any effects from the presentation of the word become more evident through the PTL measure. The latency of shift measure did

not provide any evidence for word comprehension in Experiment One, contrary to the other two measures. It was, therefore, not used in subsequent analyses.

Attentional preferences have been shown to change over the course of the experiments. In the pretest phase of Experiment One infants preferred to attend to the target more after the presentation of the target word, while in the post-training blocks evidence of distracter looking was found for individual items. In the first post-training block of Experiment Two, infants exhibited understanding of the labels showing a target preference for some items and a distracter preference for others, whereas in the second post-training block they preferred overall to attend more to the distracter after they heard a word. Attentional preferences are thought to change on the basis of a number of factors which render the test stimuli more or less salient at different points within the same experimental session (Houston-Price & Nakai, 2004):

1. affective or physical properties of the stimuli: newborn infants for example prefer to listen to their mother's voice than to a stranger female's (DeCasper & Fifer, 1980)

2. level of familiarity: infants in their second year of life prefer to look at pictures of objects for which they already have a word instead of objects with unfamiliar labels, even in the absence of a label (Schafer, Plunkett, & Harris, 1999) and

3. match or mismatch in a cross-modal situation: infants in their first and second year of life orient their attention more to a picture which matches an auditory stimulus instead of one that does not (Golinkoff & Hirsh-Pasek, 1987b; Schafer, 2005; Ballem & Plunkett, 2005) - although under certain circumstances the reverse pattern is observed, whereby infants increase their looking to a visual display if it mismatches the auditory input (Werker et al., 1998) .

Both the level of familiarity and the match versus mismatch in a cross-modal situation, as well as the interaction between the two are of primary interest when interpreting

infant behaviour in preferential looking studies, especially in word-learning scenarios. Attentional preferences in the first years of life have been quite extensively investigated and have sometimes produced conflicting results (Houston-Price & Nakai, 2004; Cohen, 2004; Slater, 2004), such as those reported in (3) above. The pattern that often emerges from such discussions of infant attentional preferences is an early familiarity preference and a subsequent shift of attention to novelty (Hunter & Ames, 1988; Roder et al., 2000). Infant behaviour however is suggested to be intimately tied to the characteristics of the task, the age-group in discussion and the complexity of the stimuli (Hunter, Ames, & Koopman, 1983). It is worth noting that the first stage of familiarity preference is often not observed (Roder et al., 2000; Mather & Plunkett, July, 2007).

In an IPL task, an infant is typically required to exhibit the existence of a pre-established or newly-built association between two stimuli - typically a sound and (a picture of an object) - by *increased looking to the appropriate visual stimulus*. In other words, if an infant understands a word, she is expected to orient her attention to a picture of the appropriate object in the presence of the word - to a level over and above that which would be expected if she was allowed to look at the pictures freely (Schafer et al., 1999; Golinkoff & Hirsh-Pasek, 1987b). This is exactly the behaviour demonstrated in the pretest block of Experiment One. On the other hand, in habituation studies, infants are trained with two object-label pairings and then tested with a familiar pairing and a switched pairing. In this case infants are expected to show their understanding of an association between a sound and a picture by orienting their attention more to a picture which *mismatches* the auditory stimulus (Stager & Werker, 1997; Werker et al., 1998)! However contradictory this might sound, it is the direct outcome of the dynamic nature of attentional preferences during information processing: infant preferences have a temporal dimension which leads them to change over the course of an experiment according to various dimensions of

the experimental setting.

Attentional preferences and their use in experimental settings will be further discussed in the next chapter.

Although infants did not exhibit evidence for word learning, they did exhibit evidence for word comprehension using the IPL task. However, the main aim of the study was not merely to teach the words, but primarily to test generalisation of the labels to new objects. Prior knowledge of the words made this second aim impossible to investigate. Experiment Three was designed in order to overcome the limitations of the first two experiments and provide a better test for infants' word learning abilities.

Chapter 3

Experiment Three - Single Exemplar

Experiment Three is a word-learning task, during which 10-month-old infants are trained and tested with two object-label pairings for word learning and for word generalisation to new objects. In Experiments One and Two infants showed evidence for prior knowledge of the words, although parental feedback suggested the contrary. In order to overcome this problem in Experiment Three, two words were chosen which were reported to be understood by a small minority of infants in the LEX2005 database.

One consideration is that this might result in reduced familiarity with the words and objects used in the study, making the word learning task more demanding for infants. In order to overcome this concern, parental feedback on their infant's experience with different objects was obtained using the Category Experience Questionnaire (CEQ), which will be described below. Additionally, a number of task dimensions were manipulated, in order to make the task easier for infants, following previous word learning studies:

- a social interaction play-phase was introduced during which infants were allowed to actively manipulate two toy-objects, one at a time, while hearing the corresponding word (Mani & Plunkett, 2007).

- training trials became longer (Werker et al., 1998).

- a more infant-controlled (habituation) procedure was used during training (Werker et al., 1998; Fennell & Werker, 2004).

- the perceptual salience of the visual stimuli was increased by using moving pictures during training (Werker et al., 1998).

- Word learning and word generalisation were tested across infants, both in order to make the task shorter and easier for infants but also in order to avoid habituation or order effects due to the, unavoidably, consecutive presentation of the two post-training test blocks.

3.1 Habituation

3.1.1 Infant Attentional Preferences and Habituation

As described in the previous chapter, infant attentional preferences may vary during an experimental session on the basis of a number of factors (age of infants, task difficulty, complexity of the stimuli, etc.). The Hunter and Ames' model (1988) of Attentional Preferences (Hunter & Ames, 1988) as well as the Wagner and Sakovits' (1986) model (Wagner & Sakovitz, 1986) both assume that infants should exhibit a first stage of no preference to either familiarity or novelty, then attention should first shift to a familiar stimulus, and only subsequently should attention be focused to a novel stimulus. These models are rooted in optimal-level theories and conform to the assumption that familiarity preference is the consequence of infants' cognitive

preferences for an optimal level of information, which in some cases might be provided by a familiar rather than a novel stimulus (Berlyne, 1958). That is, attention is engaged for as long as an infant compares and/or updates the incoming information to an internal representation of the external stimulus. As soon as this process of comparison has been completed, attention begins to shift toward a novel and therefore informationally-rich stimulus. According to Hunt, 1970, *attentional preference may be expected to go to objects and patterns and places that have been perceived just often enough to develop central processes which permit tentative recognition, and, with more perceptual encounters, attentional preference should shift to what is unfamiliar or novel* (Hunt, 1970).

One interesting aspect of this switch between familiarity and novelty preference is the phenomenon of *habituation*, well-documented in both animals and humans. Habituation, at least for humans, has been associated to processing and learning mechanisms. It has been defined as a "response decrement as a result of repeated stimulation" (Harris, 1943) or a "relatively permanent waning of a response as a result of repeated stimulation" (Thorpe, 1956). According to Sokolov, a stimulus that is repeated produces habituation by means of a comparison between the incoming stimulus and an internal representation. This representation of the stimulus can be thought of as a pre-potentiated matrix of synapses built by the central representations of repeated stimulations; as the number of repetitions increases, the stimuli are processed quicker through the operation of the matrix, resulting in gradual habituation - which will persist until the representation of an incoming stimulus differs from the matrix (Sokolov, 1975; Sokolov, Spinks, Naatanen, & Lyytinen, 2002).

Although behavioural demonstrations of habituation have been widely documented, even with newborn infants (Nelson, 1995), the exact mechanisms underlying this behaviour have been a matter of extensive research (for a discussion, see Sokolov et al. (2002)). Recently there has been an effort to turn attention to the neural bases

of this behaviour (Mareschal et al., 2007) in order to provide a better understanding of habituation and attentional preferences during infancy, mainly drawing upon work from neuroscience, on both humans and animals. The central focus of this attention has been the area of the brain called *the hippocampus*, which has been shown to be responsible for the decrease in responding to specific features or feature combinations over repeated presentations of stimuli. This decrease in attention appears to be the outcome of *selective inhibition* of the reticular formation and brainstem areas involved in arousal and autonomic reflexes, and it appears to be short-lived and stimulus-specific (Mareschal et al., 2007; Sokolov et al., 2002; Nelson, 1995). In order to account for long-term habituation effects (Zelazo, Weiss, & Tarquinio, 1991), as well as for the role of prior experience (often manifested as age differences in habituation) (Hunter & Ames, 1988), hippocampal-related cortical areas are taken into consideration (Sirois & Mareschal, 2004).

3.1.2 Habituation In the Laboratory

In infant visual categorisation and word learning tasks, *habituation* has been widely used as an indicator of when infants have formed a relatively well-established representation of either the category (or categories) in question or the association between a word and a (picture of an) object. When a decrease of attention to the repeated and, therefore, familiar stimulus is exhibited, the crucial test lies on orienting attention to a novel stimulus which mismatches the "habituated" stimulus. This behaviour is considered to be indicative of the kinds of representations that have been formed during the task.

During discrimination tasks, infants are presented with a single item several times until looking has decreased to a pre-defined criterion. Presenting a new item at test is informative of whether infants have noticed the change that has occurred.

In categorisation studies, infants are habituated to multiple exemplars from a single category and tested with a new within-category item and with an out-of-category item. Persistence of habituation to the within-category item and increased looking to the out-of-category item - that is, novelty preference - is considered to be evidence that a category has been formed (Younger & Cohen, 1983).

Habituation has more recently been used in word learning studies, in order to maintain a more infant-controlled training regime (Werker et al., 1998; Fennell & Werker, 2004). Infants are trained with two object-label pairings until attention to the pictures decreases to a pre-defined criterion. Then infants are presented with a "same" trial and a "switch" trial. In "same" trials, infants are presented with a familiar word and a familiar object, which have been paired together during training. In "switch" trials, infants are presented with a familiar word and a familiar object, but this word-object pairing was not presented together during training. Renewed interested only to the "switch" pair is expected if infants have created an association between the word and the object during training.

In Experiment Three, a training procedure with habituation to a pre-defined criterion was used. However, at test an IPL task was administered. IPL is considered easier for infants because both pictures are presented on the screen and, therefore, infants only need to use their recognition memory in order to be successful in the task. Paradigms like the switch task require infants to use their recall memory in order to successfully solve the task, and are thus considered to be more difficult. Because 10-month-olds are just starting to acquire language, it was decided to use the IPL instead of the switch task.

3.2 Experimental Design

3.2.1 Participants

The final sample consisted of 60 (31 males) infants with a mean age of 10.04 months (R : $9.16 - 10.9$, SD : 0.37). Infants were recruited through the Oxford Babylab parental database, which contains personal information for people who have expressed their interest in our research, and were approached in local maternity wards (JR hospital, and others), local playgroups, NCT sales, through the Oxford Babylab website or publicity material regularly distributed to medical centres, doctors' surgeries and child-care centres. All infants participating in the study were healthy, full-term babies, from monolingual households in which the language used is British English.

Infants in each Experiment were further divided into two Conditions (30 infants each) on the basis of the testing stimuli (familiar/new exemplar) presented to them. The two sexes were roughly equally represented in each condition (Condition A - 16 males, Condition B - 15 males). Fifteen additional infants were tested but excluded from the final set due to fussiness.

3.2.2 Stimuli

As mentioned above, two words were used in this study; "cow" and "horse". Each infant was trained and tested with both words. "Cow" and "Horse" are reported in infant language databases (MacArthur CDI) to be understood by only 16.4 % and 10.4 % of 10-month-old infants, respectively. However, because in Experiment One infants were shown to understand the labels before coming to the laboratory, prior knowledge of the words was tested for each individual infant through parental feedback using the Oxford CDI, as well as in a pretest block.

As in the previous experiments, real objects were chosen in order to reduce the

cognitive load required by the infant in order to complete the task. However, because the two words chosen for Experiment Three were reported by parents to be understood only by a small minority of the 10-month-olds in the database, there was a concern about the familiarity of these objects to infants at that age. In order to control for familiarity effects, parents were asked to complete a second questionnaire in which they were asked to rate the frequency and the amount of exposure their infant has had with different categories of objects. The Category Experience Questionnaire (CEQ) was created on the basis of the Oxford CDI and included all object words in the list (for a description of the CEQ, see Section 3.2.6, and for the questionnaire see Appendix C).

The two animals, "cow" and "horse", were chosen over other animals for a number of reasons: firstly, they were reported by parents not to be understood by many infants in the LEX2005 database, so we expected infants in this experiment not to show any prior understanding of the labels. Second, these animals can be easily matched for level of interest, as defined by their perceptual salience. That is, their colours and shapes are fairly similar and, therefore, any systematic preferences to look at either animal should not be expected[1]. Third, for both animals it was possible to create categories with exemplars of varying typicality which are discriminable from each other. These were necessary because in subsequent studies (see Chapter Four) infants were presented with the same two animal categories but with multiple exemplars during training.

Two exemplars were chosen from each category, which were both judged as highly typical by adult native speakers of British English. Because this was primarily a generalisation task for infants, it was considered important to have a measure of the

[1]One concern could be that if the two objects are too similar, word learning can be difficult for infants. In order to control for this concern, adult native speakers of British English were asked to rate the visual similarity of the objects and a categorisation study without labels was run to test whether infants could form the corresponding animal categories. See Chapter Six

perceptual similarity of the objects used. The, within-category and between-category, visual similarity of these objects was, therefore, judged by adults as an index of how similar infants might perceive them to be. For both typicality and visual similarity rating studies see Chapter Six.

The two words, "cow" and "horse", are both monosyllabic and should, therefore, be easier for infants to learn than longer words. An effort was made to match them for phonological complexity: the word "cow" contains a diphthong and the word "horse" contains a fricative coda. This should create phonological forms of similar complexity, but an items analysis was also conducted in order to test whether there was an effect of item on infants' learning or understanding of words.

Infants were trained with both pictures of objects and real objects. The toy-objects were naturalistic animal figures of the German brand "Schleich". The toy-cows measured 13cm long and 7cm tall and the toy-horses 12cm long and 10cm tall. The pictures were digital photographs of the toy-objects on a white background. Photographs were digitally processed to remove any background using Adobe Photoshop CS2 and saved in 24-bit bitmap format. In test blocks pictures were presented stationary. For training blocks pictures were processed using Macromedia Flash 5 in order to add motion to the objects, which moved 45 pixels up and down the vertical axis in a smooth identical movement. The distance was covered over 25 frames, changing every 100ms. All items were saved looking in two directions, by flipping the canvas horizontally for both still and moving images; this ensured that items always faced the centre of the screen.

Audio files were recorded in a single session by a female voice delivering them in high-affect child-directed speech (Fernald, 1992; Mills et al., 2005). The audio

recordings were made with a digital audio tape recorder (DAT) in a sound-treated, recording booth. The audio stimuli were digitised at a sampling rate of 44.1 kHz and a resolution of 16 bits and spliced using Goldwave v. 5.10. All audio stimuli were digitally matched for volume to a typical -18.00dB. Labels were produced by the speaker both in citation form and embedded in sentences. In test blocks, words produced in citation form were used and they were presented in isolation, in order to make the task easier for infants (Jusczyk, 1999). In the training block, infants heard the words both in isolation and embedded in the following carrier phrases: "Look at the X!" and "Do you remember the X?".

3.2.3 Design

Experiment Three aimed to teach infants two object-label pairings and then test them for Word Learning (Condition A) as well as for Word Generalisation (Condition B) to new exemplars from the same adult categories . During test a standard IPL task was used. During training a habituation task was administered in order to achieve a more infant-controlled procedure.

Two exemplars from each category were used in this Experiment. The exemplars which were rated as the most typical by adult native speakers of British English (see Chapter 6 for a description of the Adult Rating Studies) were used at pretest and training in both Conditions. In Condition A the same exemplars were used for test, while in Condition B one new typical exemplar from each category was used.

During pretest, infants sat through a four-trial block; in each trial both pictures were simultaneously presented on the screen along with a word corresponding to one of the pictures. Pretest was used in order to test for prior knowledge of the to-be-

learned pairings.

During the interactive play-phase, infants were presented with the two toy-objects one at a time and heard the labels several times.

During the on-screen training phase, each infant was presented with the same two pairings multiple times. Each object-label pairing was presented individually on the right or left side of the screen. Visual stimuli were identical to those used in the pretest block. The training block consisted of twenty trials, ten for each item, divided in five subblocks of four trials - two for each pairing.

The test block in Condition A was identical to the pretest block and aimed to test the word learning abilities of infants and the effect of training on infant behaviour. In Condition B the test block differed from the pretest in only the visual stimuli and it aimed to test generalisation of the labels to new exemplars from the same adult categories.

3.2.4 Procedure

After a brief familiarisation period in the Oxford Babylab, parent and child were seated in a dark booth. Throughout the experiment infants sat on their parent's lap approximately a meter away from a 110(width) x 40(height)cm plasma flatscreen (except during the play-phase which took place in the Oxford Babylab seating area). Pictures measured 30cm(width) x 30cm(height) and were presented on the left and right of the screen, separated by a 28cm gap. Audio stimuli were presented through two speakers located just above the centre of the screen. Parents were asked not to interact with their child during the session. They were instructed to close their eyes

and listen to music through headphones. Two cameras above each picture recorded infant eye movements, each sending a signal to a digital router to form a single twin split-image of the infant. A computer captured the whole session for later off-line scoring. The experimenter was able to manually control the experiment, which proceeded at the individual infant's own pace. A tone delivered through the speakers served to re-capture infant attention between trials.
During the interactive play-phase, parent and infant were seated in the Oxford Baby-lab seating area and the experimenter sat opposite them at approximately 50cm distance so that the infant could reach the toy presented to her.

Conditions A and B were identical in the pretest and training blocks, and only differed in the test block, as described below.

Pretest

Infants were tested for prior knowledge of the words using a standard (IPL) task. Pretest consisted of four trials. In each trial, pictures of the two objects were presented simultaneously, along with a single label corresponding to one of the pictures. Each picture was the target twice, with order of presentation randomised and side of presentation counterbalanced within participants. Trials lasted 5000ms; at 1000ms infants were prompted to "look" and at exactly 2500ms the presentation of the target word began.

Interactive Play-Phase

After pretest, infants were taken out of the booth for the interactive play-phase. The researcher presented the infant with a toy-object and named it six times. The

infant remained seated on the caretaker's lap, and the latter was instructed not to interact with the infant at all. Each object was presented separately and the order of presentation was pseudo-randomised so that half of the infants saw "cow" first and half saw "horse" first.
The interactive play-phase lasted approximately 4-5 minutes. During that period infants were allowed to actively manipulate the objects (one at a time), while each object was labeled six times using the following sentences:

"Oh! Look! This one's a X! Look! It's a X! Do you want to play with the X? Do you like the X? Oh! Look! Where's the X? Look! Here's the X!"

On-Screen Training Phase

After the interactive play-phase, infants were taken back into the booth for an on-screen training phase. During training trials infants were presented with a single moving picture along with the appropriate word. Each trial lasted 12.5s; at 500ms infants were prompted to "Look". At 2000ms the target word was presented for the first time. At 4000ms the phrase "Look at the X" was presented. And finally at 6500ms infants were presented with the last phrase, "Do you remember the X?". This process was repeated for each item separately (see Figure 3.1).

The training block consisted of twenty trials, divided in five sub-blocks of four trials. Each sub-block consisted of two training trials for each word-object pairing. The order of presentation of the sub-blocks was randomised. Side of presentation for the pictures was counterbalanced within sub-blocks. Order of presentation was randomised within sub-blocks. This meant that there could be a maximum

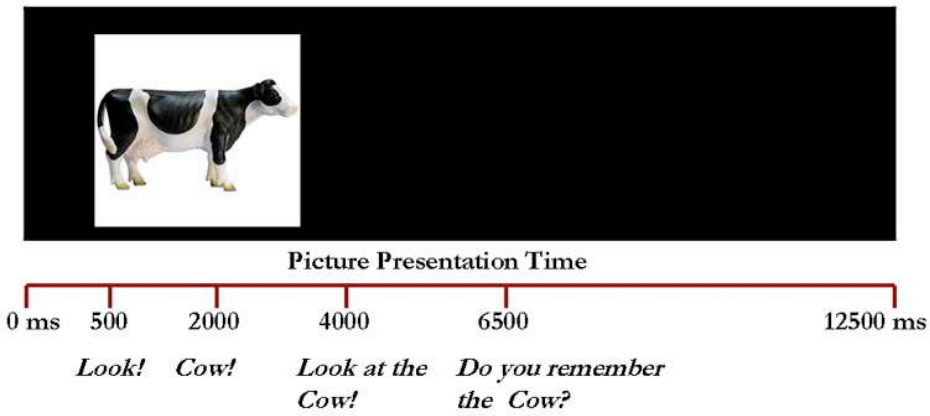


Figure 3.1: On-Screen Training: Time-course of a training trial, with a single picture on the screen and presentation of the corresponding label three times during the 12.5s of the picture presentation time.

of four consecutive same-picture trials presented, and only two consecutive same-picture/same-side trials (for the effect of alternating picture side on maintaining infant interest in the task see (Canfield et al., 1997)). As mentioned earlier, pictures in the training block moved smoothly up and down the vertical axis, in order to increase their perceptual salience and facilitate the word learning procedure (Werker et al.,94).

The trials were scored on-line by a trained researcher blind to the experimental conditions. The first sub-block was considered to be the baseline to which all subsequent sub-blocks were compared. Training ended when looking time had fallen below a pre-defined criterion (65% decrease from the baseline - first four trials) or when all twenty trials had been completed. The sub-blocks of four trials ensured equal exposure to both word-picture pairings before training stopped. Although decreasing attention to criterion for each individual pairing could not be ensured, a post-hoc analysis tested whether infants habituated to one or both of the pictures. The criterion of 65% has been used in previous word learning studies, and seems to be a good index of infants' decreasing attention, while ensuring that they are still relatively engaged in the task (Fennell & Werker, 2004).

Test Phase

Infants were finally tested in a post-training test block. In Condition A this test block was identical to pretest. In Condition B it only differed in that a new typical exemplar from each adult category was presented to infants.

3.2.5 Predictions

Condition A aimed to test whether infants at 10 months of age are able to learn two word-object pairings in a brief experimental session. According to previous findings (see Section 2.1.2) infants at this age are able to form such associations in their natural environment. By decreasing the processing load with the use of familiar items and the increase of the perceptual salience of the training items, it is expected that infants will be able also to form object-label associations in the laboratory.

For Condition B, constraints-and-principles accounts predict that infants will generalise the newly-learned labels to new within-category exemplars, as long as they consider the new exemplar and the training exemplar as members of the same category. This behaviour should be the result of a constraint, which guides infant linguistic behaviour and leads them to generalise a label to a group of objects. On the other hand, exemplar-based accounts predict that infants will fail to generalise the label to the new exemplars, because they need exposure to naming instances with different exemplars from the same category before they discover that the label refers to a whole category of objects and not to the individual object they have encountered the label with.

However, on the basis of previous findings (Werker et al., 1998), it is expected that only infants who show evidence for habituation during training might be able to show evidence for either word learning (Condition A) or word generalisation (Condition B)

at test.

3.2.6 Measures

Training Block

For the training phase, the TLT measure was used (see Section 2.5), calculated for each sub-block and object separately and for the whole duration of the trial.
Depending on whether infants had reached the pre-defined criterion during the training phase they were divided into two Habituation Groups - Habituators (HAB) (N = 33) and Non-Habituators (NH) (N = 27). This division was necessary because infant attentional preferences has been shown to change over the course of an experimental session, and, therefore, during test behaviour of infants who habituated could be different from those who did not. If this factor is not taken into account, any word learning taking place during training could be obscured at test because of the variability in infant preferences to different stimuli (for a description of infant attentional preferences see section 3.1).

Test Blocks

For the test blocks, two of the four IPL measures used in Chapter Two were also used here: PTL and LLD, both calculated for the pre- and post-naming phases of the trial.

1. Proportion of Target Looking (PTL): calculated by dividing the total looking time to target by the total looking time to both target and distracter. This proportional measure of infant attention allows the removal of background noise from the data (which is present for example in TLT) by adjusting for between- as well as within-subjects differences, which could be caused by uncontrolled

factors interfering during an experimental session (e.g., habituation effects). Both PTL and LLD (described below) are considered as indicators of the time needed to match an incoming stimulus to an existing mental representation (Golinkoff & Hirsh-Pasek, 1987b).

2. Longest Look Difference (LLD): calculated by subtracting the longest continuous look to the target from the longest continuous look to the distracter. The longest look measure exploits the idea that information processing takes place during fixations to visual stimuli (and not during saccades) (Bridgeman et al., 1975; Volkmann, 1986; Irwin, 1998) and an increase in longest look durations upon the presentation of a label is considered evidence for the activation of an association between a visual and an auditory stimulus.

On the basis of parental feedback three additional measures were used:

1. Comprehension Score: the total number of words infants were reported to understand in the Oxford CDI check list.

2. Comprehension Group: infants were divided in two groups, Comprehenders and Non-Comprehenders, on the basis of whether they were reported to understand the words used in the Experiment. Infants who were reported to understand at least one of the two words were tagged as Comprehenders, and those who were reported not to understand any of the words as Non-Comprehenders.

3. CEQ Score: familiarity of the testing objects was calculated using the Category Experience Questionnaire (CEQ). For a description of the CEQ see below. A detailed description of the CEQ along with a preliminary analysis from 50 infants can be found in Appendix III.

Category Experience Questionnaire. The Category Experience Questionnaire (CEQ) evaluated the experience of individual infants with different objects and ani-

mals. The labels used were selected from the Oxford CDI's sections: animals, vehicles, toys, food and drink, clothes, furniture, household items and outside. Only labels which could conform to a category structure were used in the CEQ (for example the words "sun" and "moon" were excluded), yielding a total of 117 categories.
Parents were asked to rate their infant's experience with the different categories by selecting one of the options on a pre-defined scale along three different category dimensions:

1. *frequency*: how often the infant sees an object from that category. Parents could choose between five options: never, once a month, once a week, several times a week, every day.

2. *number of exemplars*: how many exemplars from the category the infant might have seen in her life. Parents could choose between four options: 0-1, 2-3, 4-5, more than five.

3. *familiarity with the label*: whether the infant has ever heard the word for the category or not (without reporting on the understanding of the word). Parents could choose between yes and no.

Each category received a score on the basis of parental feedback on the dimensions above. The first option from each dimension (never, 0-1 and no respectively) received 0 points. Every subsequent option received an additional point. The three dimensions could, therefore, receive a score from 0-4, 0-3 and 0-1, respectively. In this manner an infant could receive a total score from 0-8 for each of the two testing items.

Both questionnaires (CDI and CEQ) were filled in before the study and parents were allowed to fill them in in whatever order they preferred.

3.2.7 Scoring

Infant looking behaviour at test was recorded and scored off-line by a trained scorer, as described in Section 2.6. Looking behaviour during training was scored on-line, but was also recorded for off-line reliability scoring. A digital-video scoring system was used to assess visual events on a frame-by-frame basis (every 40ms). Visual events received one of four tags: "right", "left", "mid" or "other"; "right" and "left" referred to fixations at the right and left pictures. "Mid" was used to describe shifts between the two sides, as well as blinks. "Other" was used for any other looking behaviour that involved infant attention shifting away from the task: looking away from the screen, crying, rubbing their eyes, etc.

Intra-reliability tests were carried out for the training block, with 25% of the trials scored off-line by the reseacher. On-line and off-line measurements were compared using Pearson r correlations. This off-line reliability scoring was compared to both the on-line looking time to screen as well as to a second "corrected" version of it calculated by subtracting 100ms from the on-line measurements. This is usually incorporated in on-line measurements made by habituation software (e.g. *Habit*) and it aims to correct for the time the scorer needs to initiate a motor movement from when the infant's gaze begins to move. Pearson r correlations revealed high agreement between on-line and off-line measurements of infant attention to the task $r(144) = .95$, $p < .001$.

3.2.8 Analysis

3.2.9 Training Block

In the training block, infants received repetitive presentations of two picture-label pairings, in sub-blocks of four trials. Within each sub-block each picture-label pair-

ing was presented separately twice. The unaggregated data for the training block consisted of 988 trials, 464 for Condition A and 524 for Condition B. In Condition A 16 infants reached the habituation criterion and in Condition B 17 infants did. On average infants completed 4.12 training sub-blocks ($R = 2 - 5$, $SD = 1.09$), 3.87 for Condition A ($R = 2-5$, $SD = 1.25$) and 4.37 for Condition B ($R = 3-5$, $SD = .85$). An Independent-samples t-test compared the mean number of blocks accumulated in each Condition and showed no significant differences between the two, $t(58) = -1.81$, $p = .08$. Data were further aggregated by sub-block and item. Mean TLT in each sub-block, as well as for the two pictures separately was calculated.

A univariate ANOVA with TLT as dependent variable and Condition and Item as fixed factors tested whether infants accumulated similar looking times to the two pictures in the two Conditions. Finally looking times in the different sub-blocks of the training block were calculated and compared for the two Habituation groups separately using a univariate ANOVA with TLT as dependent variable and Sub-block as a fixed factor.

Normality tests were conducted for the TLT measure, and all groups were found to be normally distributed.

3.2.10 Test Blocks

Infants were presented with four trials in each test block. The unaggregated data set for the pretest block consisted of 217 trials, 107 for Condition A and 110 for Condition B. For the test blocks the dataset consisted of 223 trials, 105 for Condition A and 118 for Condition B. The difference in the number of trials in each test block is due to the exclusion of trials in which the infant did not look at both pictures. Data were further aggregated by different factors depending on the analysis (item, habituation group, test block).

In order to test whether infants in the two Conditions and the pre- and post-training test blocks behaved in a similar way, a 2 (Trial Phase) x 2 (Test Block) x 2 (Condition) repeated measures ANOVA was run for both measures.

Paired-samples t-tests were run comparing the two trial phases in order to evaluate any significant changes in target looking before and after the onset of the target word. One-sample t-tests were also conducted comparing the two trial phases to chance, investigating any systematic biases to look at either picture before and after the onset of the target word.

In order to test the effect of Item and Habituation group on word comprehension, two-way repeated measures ANOVAs with Trial Phase and Item/Habituation group were conducted for both measures.

Finally, a repeated measures ANCOVA with Trial Phase as a within-subjects factor and Total Comprehension Score (TCS) or Familiarity Score (FS) as a covariate was conducted, as well as a Pearson correlation between TCS/FS and the change in target looking before and after naming for both measures, in order to test whether parental feedback was predictive of infant performance in the laboratory.

Normality tests were conducted for all measures. In order to achieve normality, in Condition A two infants were excluded from the final set in the pre-naming phase of the trial in the pretest block for PTL and two for LLD in the pre-naming phase of the trial in the pretest block. One infant was excluded in the pre-naming phase of the test block for PTL and two for LLD. No infants were excluded in Condition B. After the exclusions, all measures were normally distributed in both Conditions and test blocks.

3.3 Results

3.3.1 Training Phase

Infants' looking time to the visual stimuli was investigated using the TLT measure in order to control for level of attention at different stages during training. Figure 3.2 shows the total looking time accumulated for each target picture in the two Conditions. It appears that infants in each Condition behaved similarly toward the two different pictures. Overall they seem to have behaved similarly in the two Conditions as well, but accumulating slightly more looking time in Condition B. In order to evaluate these observations, data from the training block were submitted to a univariate ANOVA with TLT as a dependent variable and Condition and Target Picture as fixed factors. No main effects were found, Condition: $F(1, 120) = 2.86$, $p = .09$/ Target Picture: $F(1, 120) = .03$, $p = .86$, nor an interaction between Condition and Target Picture, $F(1, 120) = .00$, $p = .97$. Data were collapsed for Condition and Item in subsequent analyses in this block.

Figure 3.3 shows the five training sub-blocks separately for the two Habituation groups. It is apparent that Habituators' attention declines very quickly, whereas Non-Habituators exhibit a gradual decrease in attention, while still maintaining high interest in the task even in the final sub-block. This suggests that there was an overall decrease in attention over the course of training, but infants in the two Habituation groups exhibited different behaviours. In order to test whether infants in both groups showed a significant decrease in attention over the course of training - and, therefore, evidence for habituation, a Univariate ANOVA was run with TLT as a dependent variable and Sub-block as a fixed factor separately for each Habituation group. This analysis showed a significant main effect of Sub-block for both groups, Habituators:

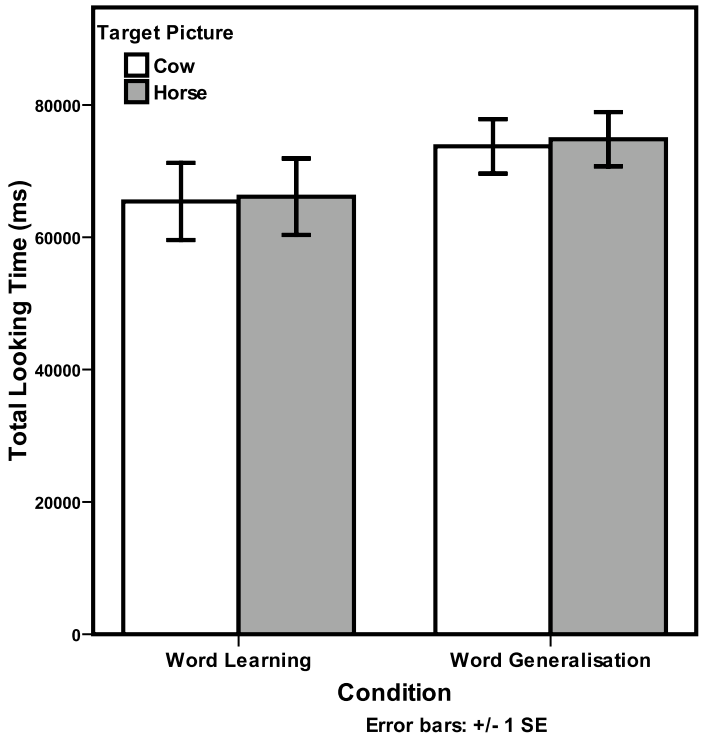


Figure 3.2: Mean Looking Time presented separately for each target picture in Condition A (Word Learning) and Condition B (Word Generalisation).

$F(4, 107) = 14.57$, $p < .001$, ${\eta_p}^2 = .35$, Non-Habituators: $F(4, 130) = 2.87$, $p < .03$, ${\eta_p}^2 = .08$.

It is, therefore, suggested that both groups showed a decrease in attention over the course of training, but in the Habituators' group (infants who reached criterion) this decrease was more pronounced. Non-Habituators decreased their looking time to the pictures in a gradual and slow manner, whereas Habituators exhibited a more sudden and dramatic decrease. It appears that both groups of infants were engaged in the task, but it could be that the two groups employ different learning styles. This should be taken into account during the analysis of the test blocks.

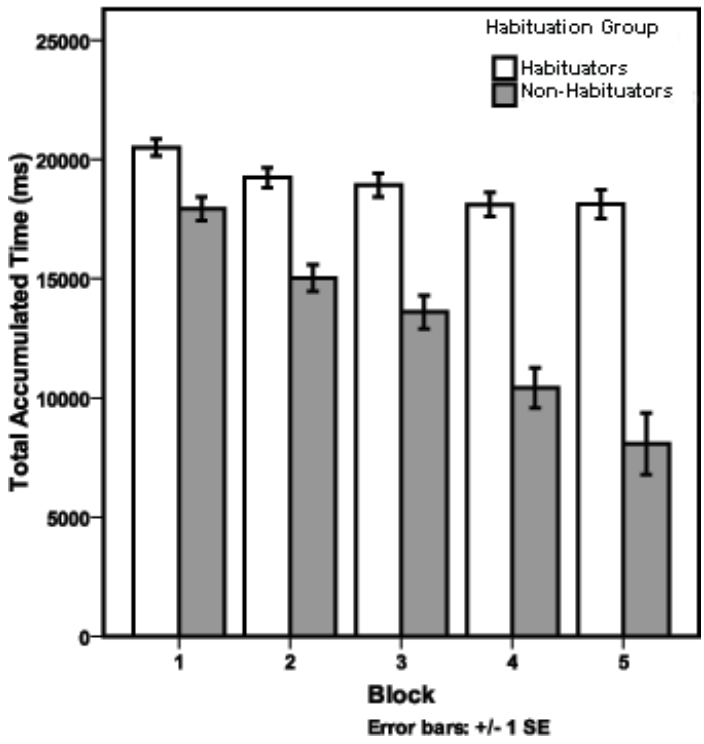


Figure 3.3: Mean Looking Time for each Habituation group separately for each training sub-block.

3.3.2 Test Blocks

Infants completed two test blocks - one before and one after training. Figure 3.4 shows the proportional measure before and after the onset of the target word for the two blocks and the two Conditions, separately for each Habituation group. It is evident that the two groups behaved in a different way, especially in the post-training test block. Infants who did not reach criterion during training - Non-Habituators (NH) - exhibit random looking behaviour at test; infants who did reach criterion - Habituators (HAB), however, seem to exhibit a systematic increase in target looking after the presentation of the target word, only in Condition A, but not Condition B. In the pretest block, NHs seem to exhibit a steep decline in target looking in the post-naming phase of the trial in Condition B. On the other hand, HABs in both Conditions showed similar attention to the target picture before and after the presentation of the target word, with a slight increase in target looking in the post-

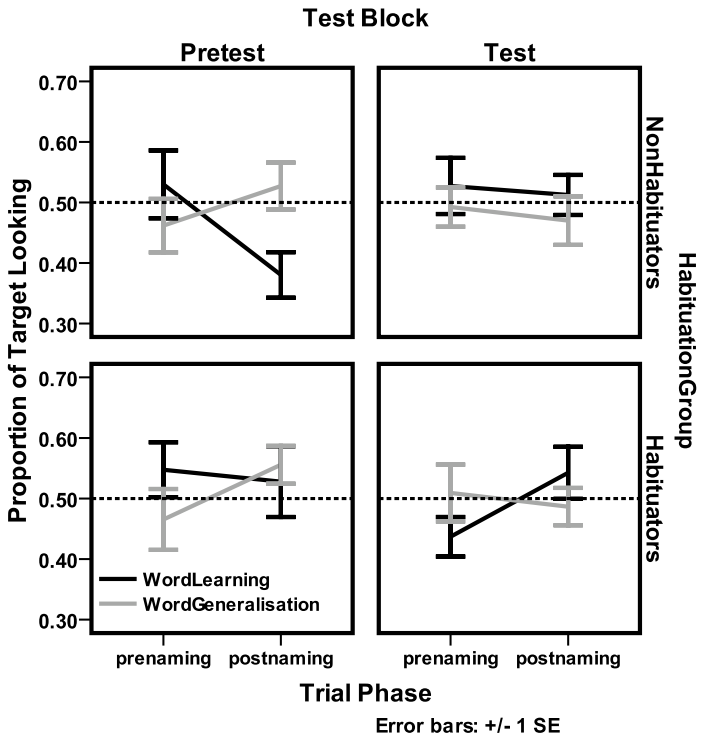


Figure 3.4: PTL for the two habituation groups in each test block and Condition separately.

naming phase of Condition B.

In order to test the observations above, a preliminary 2 (Trial Phase) x 2 (Test Block) x 2 (Condition) repeated measures ANOVA with Trial Phase and Test Block as within-subjects factors and Condition as a between-subjects factor was run for both PTL and LLD measures [2]. Neither measure exhibited any main effects, Trial Phase: PTL: $F(1, 58) = .08$, $p = .77$/ LLD: $F(1, 57) = 1.08$, $p = .30$; Test Block: PTL: $F(1, 58) = .08$, $p = .79$/ LLD: $F(1, 57) = .02$, $p = .90$; Condition: PTL: $F(1, 58) = .04$, $p = .85$/ LLD: $F(1, 57) = .27$, $p = .61$,, but did show a significant interaction between Trial Phase, Test Block and Condition, PTL: $F(1, 58) = 5.86$,

[2]Although in Figure 3.4 it seems that HABs and NHs behave differently, Habituation group was excluded from this analysis, because it was shown to reduce the power of the analysis. When added to a four-way ANOVA results considerably lost their statistical power and, therefore, became difficult to interpret.

$p < .02$, ${\eta_p}^2 = .10$; LLD: $F(1, 57) = 4.70$, $p < .05$, ${\eta_p}^2 = .08$.

On the basis of this analysis, the two Conditions were analysed separately.

Condition A - Word Learning

In order to investigate the interaction between Trial Phase, Test Block and Condition found in the previous analysis and to evaluate whether the two habituation groups behaved in a different way, a 2 (Trial Phase) x 2 (Test Block) x 2 (Habituation group) repeated measures ANOVA was conducted for both measures. No main effects were found, Trial Phase: PTL: $F(1, 28) = .19$, $p = .67$/ LLD: $F(1, 27) = .09$, $p = .77$; Test Block: PTL: $F(1, 28) = .01$, $p = .91$/ LLD: $F(1, 27) = .47$, $p = .50$; Habituation group: PTL: $F(1, 28) = .08$, $p = .77$/ LLD: $F(1, 27) = .20$, $p = .66$, but a non-significant trend for an interaction between Test Block and Trial Phase, $F(1, 28) = 3.25$, $p = .08$, ${\eta_p}^2 = .10$, as well as between Test Block and Habituation group was present for PTL, $F(1, 28) = 3.07$, $p = .09$, ${\eta_p}^2 = .10$.

This analysis suggests that infants behaved differently in the two test blocks of Condition A and that Habituators and Non-Habituators also exhibited different behaviour. A separate analysis was, therefore, conducted for each block (pre- and post-training) for only infants who reached criterion during training.

Pretest. In order the investigate the interaction between Trial Phase and Test Phase found in the previous analysis, paired-samples t-tests were conducted comparing the two trial phases for both PTL and LLD measures in only the pretest block of Condition A and for only those infants who reached criterion during training.

Both measures showed no differences between the two phases of the trial, PTL: $t(30) = 1.05$, $p = .30$/LLD: $t(30) = .44$, $p = .66$ (Pre-naming: PTL: $M = .51$, $SD = .23$; LLD: $M = 25.63$, $SD = 869.77$/ Post-naming: PTL: $M = .52$, $SD = .25$; LLD: $M = 150.00$, $SD = 879.55$).

One-sample t-tests were run comparing the two phases of the trial to chance. Neither measure was found to differ from chance in either trial phase, pre-naming: PTL: $t(30) = 1.38$, $p = .18$/ LLD: $t(30) = .97$, $p = .34$/ post-naming: PTL: $t(30) = -.33$, $p = .74$/ LLD: $t(30) = .27$, $p = .79$.

Paired-samples t-tests were also run for the Non-Habituators group in order to evaluate whether these infants had any prior knowledge of the to-be-learned words before training. Neither measure was shown to differ between the two trial phases, PTL: $t(26) = .97$, $p = .34$/LLD: $t(26) = 1.27$, $p = .22$. (Pre-naming: PTL: $M = .52$, $SD = .24$; LLD: $M = 79.44$, $SD = 828.35$/ Post-naming: PTL: $M = .44$, $SD = .25$; LLD: $M = -233.15$, $SD = 920.20$).

Figure 3.5 shows the LLD for the pretest block of Condition A for each target picture separately. It appears that the two pictures promote similar behaviour, in both trial phases: target looking appears to be at chance levels both before and after the presentation of the target word. In order to evaluate the observations above, a 2 (Trial Phase) x 2 (Target Picture) repeated measures ANOVA was conducted for both measures. No main effect of Trial Phase or Target Picture was found, Trial Phase: PTL: $F(1, 14) = .44$, $p = .52$; LLD: $F(1, 14) = .96$, $p = .81$/ Target Picture: PTL: $F(1, 14) = .01$, $p = .95$; LLD: $F(1, 14) = .01$, $p = .94$, and no interaction with Target Picture, PTL: $F(1, 14) = .02$, $p = .91$; LLD: $F(1, 14) = .02$, $p = .91$.

This analysis suggests that there was no evidence for word comprehension and,

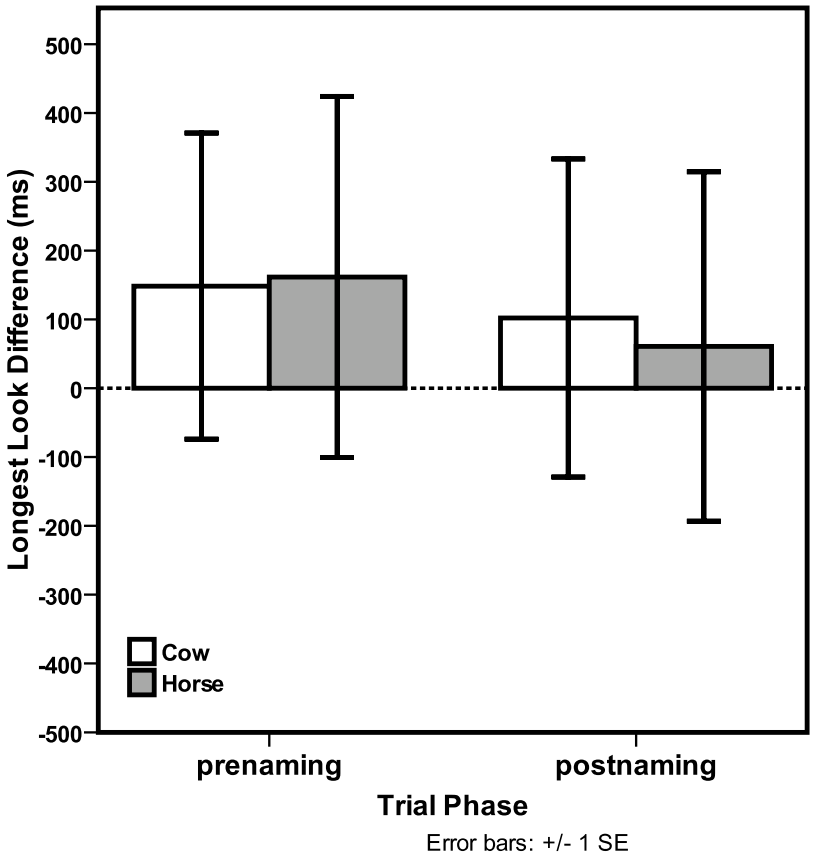


Figure 3.5: Longest Look Difference for the pretest block of Condition A for each target picture separately.

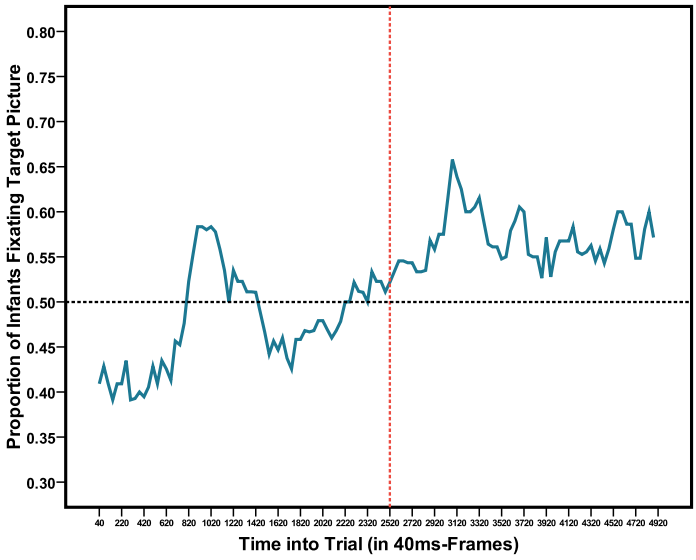


Figure 3.6: Proportion of infants in the HABs group fixating the target picture as the trial unfolds (for 40-ms frames) in the post-training test block in Condition A. The vertical dotted line denotes the onset of the target word, whereas the horizontal dotted line denotes the 50% chance level.

therefore, infants did not have any prior knowledge of any of the labels before training.

Test. Figure 3.6 shows the proportion of infants in the HABs group fixating the target picture at each point during the trial. It appears that after the presentation of the target word more infants shift to look at the target than the distracter.

Paired-samples t-tests were conducted in the post-training test block in order to test whether the two trial phases differed. The PTL measure showed no differences between the two phases of the trial, but the LLD measure showed a significant increase in target looking after the onset of the target word, PTL: $t(31) = -1.55$, $p = .13$/ LLD: $t(28) = -2.03$, $p = .05$ (Pre-naming: PTL: $M = .47$, $SD = .20$; LLD: $M = -28.57$, $SD = 597.22$/ Post-naming: PTL: $M = .50$, $SD = .20$; LLD: $M = 124.05$, $SD = 701.48$).

One-sample t-tests were run comparing the two phases of the trial to chance. In

the pre-naming phase of the trial the PTL measure showed a trend for a distracter preference, $t(31) = -1.87$, $p = .07$, while LLD did not, $t(28) = -.72$, $p = .48$. In the post-naming phase of the trial, neither measure differed from chance, PTL: $t(31) = .30$, $p = .76$/ LLD: $t(31) = 1.13$, $p = .27$.

Paired-samples t-tests were also run for the Non-Habituators group in order to evaluate whether these infants also showed evidence for word learning. Neither measure was shown to differ between the two trial phases, PTL: $t(26) = -.49$, $p = .63$/LLD: $t(25) = -.73$, $p = .47$. (Pre-naming: PTL: $M = .51$, $SD = .19$; LLD: $M = 30.39$, $SD = 636.10$/ Post-naming: PTL: $M = .54$, $SD = .27$; LLD: $M = 211.15$, $SD = 985.77$).

Figure 3.7 shows the LLD measure for the two pictures separately. It appears that both pictures exhibit an increase in target looking post-naming, with a slightly stronger increase for "horse". In order to evaluate whether the two pictures attracted similar looking times in the pre- and post-naming phases of the trial, a two-way repeated measures ANOVA was run with Trial Phase and Target Picture as within-subjects factors. No main effects were found, Trial Phase: PTL: $F(1, 14) = 2.27$, $p = .15$; LLD: $F(1, 11) = 2.78$, $p = .12$/ Target Picture: PTL: $F(1, 14) = .27$, $p = .61$; LLD: $F(1, 11) = .76$, $p = .40$, nor an interaction between Trial Phase and Target Picture, PTL: $F(1, 14) = .07$, $p = .77$; LLD: $F(1, 11) = .48$, $p = .50$.

This analysis suggests that infants in Condition A of Experiment Three exhibit some evidence for word comprehension only in the post-training test block by increasing their attention to the target picture after the presentation of the target word. This behaviour seemed to be unaffected by which picture was the target in each trial.

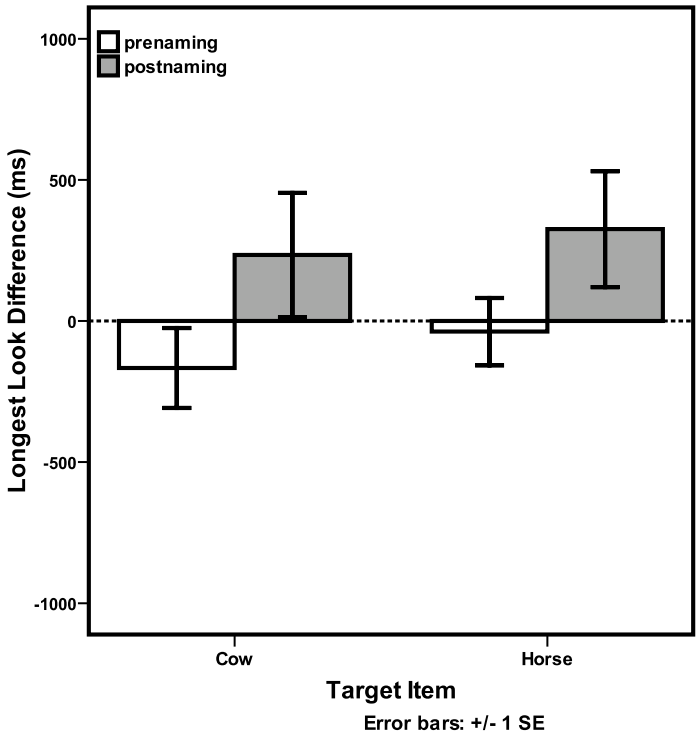


Figure 3.7: LLD for the post-training test block in Condition A showing the two pictures separately.

Condition B - Word Generalisation

In order to investigate the interaction between Trial Phase, Test Block and Condition, a separate analysis was also conducted for Condition B. Figure 3.4 suggests that in this Condition infants seems to show an increase in target looking after the presentation of the target word in the pretest block, but their attention to the target seems to remain at chance levels in both trial phases of the post-training test block. Also infants in the two habituation groups, HABs and NHs, did not exhibit different behaviours. In order to evaluate these observations a 2 (Trial Phase) x 2 (Test Block) x 2 (Habituation group) repeated measures ANOVA with Trial Phase and Test Block as within-subjects factors and Habituation group as a between subjects factor was conducted for both measures. No main effects were found for either measure, Trial

Phase: PTL: $F(1, 57) = 1.12$, $p = .29$/ LLD: $F(1, 57) = 1.47$, $p = .23$/ Test Block: PTL: $F(1, 57) = .24$, $p = .63$/ LLD: $F(1, 57) = .65$, $p = .42$; Habituation group: PTL: $F(1, 57) = .24$, $p = .62$/ LLD: $F(1, 57) = .10$, $p = .75$, but a significant interaction between Trial Phase and Test Block was present for the LLD measure, PTL: $F(1, 57) = 2.72$, $p = .11$/ LLD: $F(1, 57) = 3.86$, $p = .05$, ${\eta_p}^2 = .07$.

Although no interaction with Habituation group was found, on the basis of previous studies using a habituation training task (Werker et al., 1998) and of the literature concerning infant attentional preferences (Cohen, 2003), as well as due to the differences found between the two groups during the analysis of the training block, only infants who did reach criterion were included in the following analysis.

Pretest. Figure 3.8 shows the longest look measure for the two target pictures separately in the two phases of the trial in the pretest block of Condition B. It appears that attention to the target remains at chance levels for "Horse" both before and after the presentation of the target word. However, for "Cow" there appears to be a preference for the distracter before the presentation of the target word and an increase in target looking after naming.

In order to further investigate the interaction between Trial Phase and Test Block found for LLD in the previous analysis and evaluate whether behaviour toward the two pictures was different in the pretest block, a 2 (Trial Phase) x 2 (Target Picture) repeated measures ANOVA was conducted for both measures with both Trial Phase and Target Picture as within-subjects factors. No main effects were found for either measure, Trial Phase: PTL: $F(1, 16) = .94$, $p = .37$/ LLD: $F(1, 16) = 1.35$, $p = .26$/

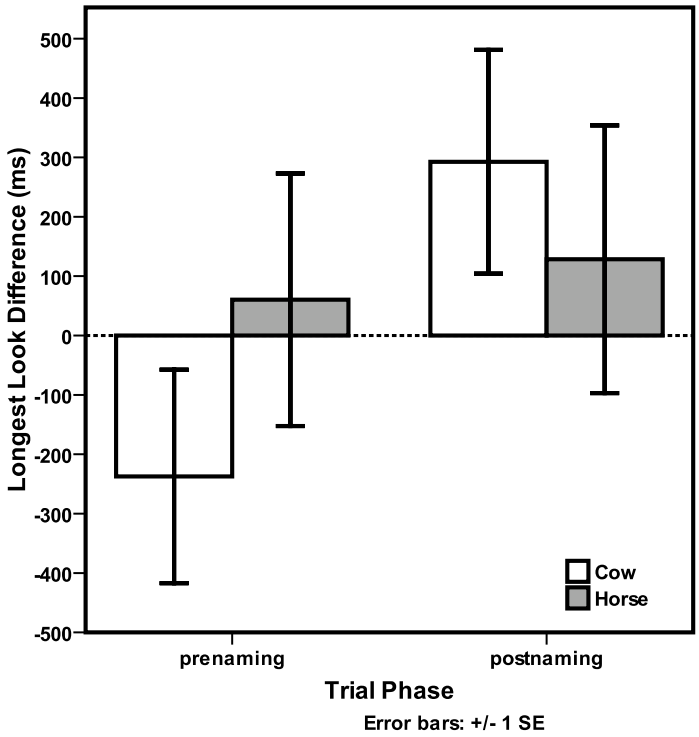


Figure 3.8: LLD for the two trial phases of the pretest block in Condition B showing the two target pictures separately.

Target Picture: PTL: $F(1, 16) = .26$, $p = .62$/ LLD: $F(1, 16) = .17$, $p = .69$, nor an interaction between Trial Phase and Target Picture, PTL: $F(1, 16) = .99$, $p = .33$/ LLD: $F(1, 16) = 1.04$, $p = .32$.

Although no main effect of Trial Phase was found and no interaction between Trial Phase and Target Picture, the increase in target looking for "Cow" appears quite pronounced in Figure 3.8. In order to confirm the absence of significant differences in target looking between the two trial phases, paired-samples t-tests were run for both measures comparing attention to each target picture separately, before and after the presentation of the target word. Neither measure showed any systematic behaviour for either picture, Cow: PTL: $t(16) = -1.55$, $p = .14$/ LLD: $t(16) = -.06$, $p = .95$; LLD: PTL: $t(16) = -1.70$, $p = .11$/ LLD: $t(16) = -.18$, $p = .86$.

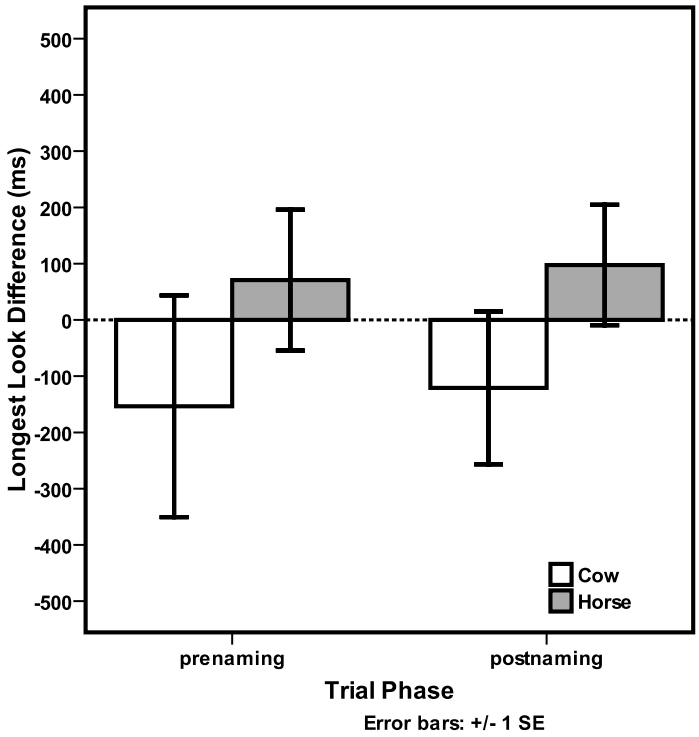


Figure 3.9: LLD for the two trial phases of the post-training test block in Condition B showing the two target pictures separately.

This analysis suggests that infants in the pretest block of Condition B did not show any evidence for prior word comprehension for either of the words used.

Test. Figure 3.9 shows the longest look measure for the two target pictures separately in the two phases of the trial in the post-training test block of Condition B. It appears that attention to the target remains at chance levels both before and after the presentation of the target word for both "Horse" and "Cow", with an overall slight preference for "Horse" in both phases of the trial.

In order to evaluate these observations, a 2 (Trial Phase) x 2 (Target Picture) repeated measures ANOVA was conducted for both measures with both Trial Phase

and Target Picture as within-subjects factors. No main effect of Trial Phase was found for either measure, PTL: $F(1, 16) = .00$, $p = .95$/ LLD: $F(1, 16) = .02$, $p = .88$, but a significant main effect of Target Picture was present for both measures, PTL: $F(1, 16) = 4.06$, $p = .06$, ${\eta_p}^2 = .20$/ LLD: $F(1, 16) = 4.54$, $p < .05$, ${\eta_p}^2 = .22$ ("Cow": PTL: $M = .47$, $SE = .03$; LLD: $M = -137.21$, $SE = 86.56$/ "Horse": PTL: $M = .52$, $SE = .03$; LLD: $M = 84.27$, $SE = 55.45$). Finally, no interaction between Trial Phase and Target Picture was found, PTL: $F(1, 16) = .18$, $p = .67$/ LLD: $F(1, 16) = .00$, $p = .99$.

This analysis suggests that infants in the post-training test block of Condition B showed no evidence of generalising the words to new typical objects from the same categories. The main effect of Target Picture indicates that during test infant attention was drawn more to one of the pictures, namely "Horse", and this did not differ in the two trial phases. This suggests that infants exhibited an inherent preference to one of the two objects and they did not perceive the task as a linguistic one even after the presentation of the word.

Parental Feedback

On the basis of parental feedback a total comprehension score (TCS) from the Oxford CDI and a familiarity score (FS) from the CEQ (see section 3.2.6) were calculated for each infant. Regarding prior knowledge of the to-be-learned words, in Condition A only one infant was reported to understand the word "horse" and in Condition B only one infant was reported to understand the word "cow". These infants were not excluded from the analysis, since no evidence for word comprehension and no outliers were found during pretest. Reports on individual words, therefore, were not analysed further.

Total Comprehension Score. In Experiments One and Two TCS was shown to be a poor predictor of infant performance in the laboratory. However, infants were shown to have prior knowledge of the to-be-learned words. It could be that TCS is a better predictor of word learning, rather than word comprehension. This analysis was, therefore, repeated in Experiment Three. Only the post-training test blocks of both Conditions were included in this analysis, in order to evaluate the effect of TCS on word learning. Also only infants who reached criterion during the training block were included.

Parents reported that infants in the final set could comprehend on average 24.35 words ($R = 0.00-61.00$, $SD = 16.95$) - in Condition A 23.61 words ($R = 0.00-61.00$, $SD = 17.89$) and in Condition B 25.10 words ($R = 0.00 - 54.00$, $SD = 16.22$). None of those infants was reported to understand more than 70 words, indicating that in general all infants were considered to be in the early stages of word learning.

Figure 3.10 shows the change in LLD before and after the presentation of the target word separately for the two test blocks of Conditions A and B. In order to test whether TCS can be predictive of performance in the IPL task and whether this differs between the two Conditions, a 2 (Trial Phase) x 2 (Condition) repeated measures ANCOVA with Trial Phase as a within-subjects factor, Condition as a between-subjects factor and TCS as a covariate was conducted for both measures.

No main effects of Trial Phase and Condition were found, Trial Phase: PTL: $F(1, 59) = .91$, $p = .34$/ LLD: $F(1, 56) = 1.32$, $p = .26$; Condition: PTL: $F(1, 59) =$

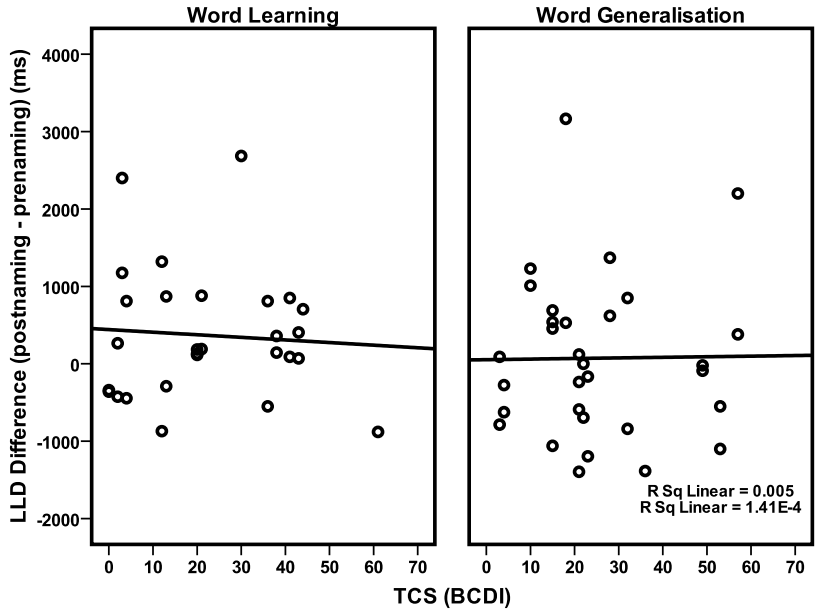


Figure 3.10: Relationship between the change in LLD before and after the presentation of the target word and TCS, in the post-training test blocks of Conditions A and B.

1.56, $p = .29$/ LLD: $F(1, 56) = 1.25$, $p = .27$, but a main effect of TCS was present, PTL: $F(1, 59) = 4.90$, $p < .05$/ LLD: $F(1, 56) = 4.73$, $p < .05$. TCS was not shown to affect Trial Phase in any way, PTL: $F(1, 59) = .83$, $p = .38$/ LLD: $F(1, 56) = .03$, $p = .86$.

This analysis suggests that parental feedback on infants' overall comprehension does not predict word learning in the laboratory. The main effect of TCS suggests that this covariate is doing an adequate job in adjusting the scores of the dependent variable - Trial Phase - and it, therefore, is a legitimate analysis to conduct. However, the introduction of TCS as a covariate eliminates the Trial Phase x Condition interaction, which suggests that this adjustment cannot account for the interaction presented above.

Familiarity Effects. A familiarity score for the test items was calculated using the CEQ. It could be that the more exposure infants have had to an object, the easier it is to learn a label for it in the laboratory.

Parents reported on their infant's exposure to each of the categories, across three dimensions: frequency of exposure, number of exemplars seen, exposure to the name of the object. For both categories, infants accumulated an average score of 7.63 ($R = 14.00$, $SD = 4.34$) - in Condition A 7.31 ($R = 14.00$, $SD = 4.25$) and in Condition B 7.97 ($R = 14.00$, $SD = 4.48$). For "cow", infants accumulated an average score of 3.86 ($R = 7.00$, $SD = 2.31$) - in Condition A 3.44 ($R = 7.00$, $SD = 2.33$) and in Condition B 4.29 ($R = 7.00$, $SD = 2.24$). Finally for "horse", infants accumulated an average score of 3.78 ($R = 8.00$, $SD = 2.72$) - in Condition A 3.88 ($R = 8.00$, $SD = 2.78$) and in Condition B 3.68 ($R = 8.00$, $SD = 2.70$).

Figure 3.11 shows the relationship between the change in the LLD measure between the two phases of the trial and the CEQ score for both categories, separately for each Condition. In order to investigate whether CEQ affects behaviour in any meaningful way, a two-way repeated measures ANCOVA was conducted for both measures with Trial Phase as a within-subjects factor, Condition as a between-subjects factor and CEQ Score as a covariate. Also Pearson r correlations were run for each measure comparing the change between the two trial phases and CEQ score for each Condition separately. No main effects of Trial Phase and Condition were found, Trial Phase: PTL: $F(1, 60) = .08$, $p = .79$/ LLD: $F(1, 57) = .07$, $p = .80$; Condition: PTL: $F(1, 60) = .15$, $p = .70$/ LLD: $F(1, 57) = 1.56$, $p = .22$, but a marginally significant main effect of CEQ score was present for the PTL measure, PTL: $F(1, 60) = 3.26$, $p = .08$/ LLD: $F(1, 57) = 1.03$, $p = .31$. CEQ was not shown to affect Trial Phase in any way, PTL: $F(1, 60) = .74$, $p = .40$/ LLD:$F(1, 57) = 1.18$,

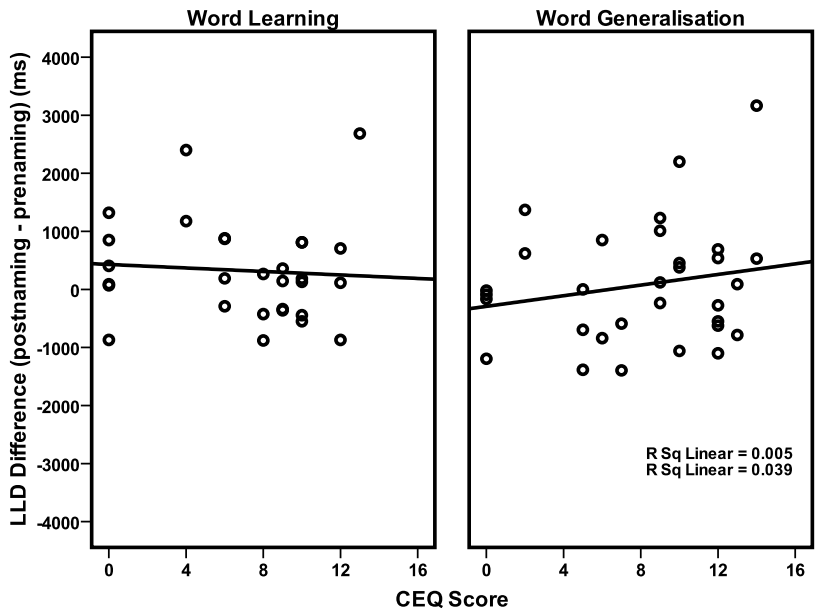


Figure 3.11: Relationship between the change in LLD before and after the presentation of the target word and Total CEQ Score, in the post-training test blocks of Conditions A and B.

$p = .28$. Also no significant correlation between change in target looking (either PTL or LLD) and CEQ score was found for either measure and Condition, PTL: Condition A: $r(32) = -.17$, $p = .35$/ Condition B: $r(31) = .20$, $p = .27$; LLD: Condition A: $r(29) = -.07$, $p = .70$ / Condition B: $r(31) = .20$, $p = .29$.

As in the previous analysis, for parental feedback on infants' comprehension abilities, this analysis suggests that parental feedback on infants' prior experience with the test items does not predict word learning in the laboratory. The interaction between Trial Phase and Condition is not present in this analysis, suggesting that CEQ score might not be an appropriate covariate for reducing the variance in the dataset and rendering the analysis more powerful.

3.4 Discussion

Experiment Three aimed to teach infants two word-object pairings, and then test them for word learning (Condition A) and for generalisation of the words to new typical exemplars from the corresponding adult categories (Condition B). Unlike Experiments One and Two, in Experiment Three there was no evidence for word comprehension in the pretest block of either Condition, suggesting that infants had no prior knowledge of the to-be-learned words.

Evidence for word comprehension was found in the post-training block of Condition A, but not Condition B. Although none of the measures differed from chance in either phase of the trial during the post-training test block of Condition A, there was a clear increase in target looking after the presentation of the target word in the LLD measure. The variance in the group was quite high, rendering this effect rather weak when other factors, such as target picture, were included in the analysis. The increase in target looking for LLD should, however, be considered as sufficient evidence for word learning, especially since these infants were shown to exhibit a systematic increase in attention to the appropriate picture after they heard the word, having been habituated to the very same pictures.

These results suggest that infants learned the labels during training, but they failed to generalise the newly-learned labels to new within-category exemplars. This pattern suggests that the taxonomic constraint is not in place for these infants, because they seem to lack the assumption that words refer to categories of objects, and not just to individual instances of them. However, it provides some support for exemplar-based accounts, which suggest that infants should not readily generalise newly-learned labels to new objects, at least before they encounter the word paired with multiple exemplars from a given category.

The evidence for word comprehension was exhibited for both target words, an ability which has not been found in any word learning studies before. Word learning in the laboratory has been shown for infants as young as 14 months for two novel labels (Werker et al., 1998) and for infants as young as 13 months for a single label (Woodward et al., 1994). There was also some evidence for word learning of a single novel label by 10 month olds (Pruden et al., 2006), but this result was confounded by the lack of proper counterbalancing of picture presentation side (for a more detailed discussion see Section 2.1.2).

As in previous studies using habituation (Werker et al., 1998; Fennell & Werker, 2004; Cohen, 2003), infants who reached criterion during training were shown to behave differently from infants who did not during both training and test. Habituators (HABs) were shown to accumulate significantly less attention during training than Non-Habituators (NHs) in both Conditions: HABs showed a steep and dramatic decrease in attention during training, whereas NHs showed a gradual drop in attention. In Condition A, Habituation group was shown to affect understanding of the labels, whereas in Condition B it did not. It was, therefore, decided to only include Habituators in the final set for both Conditions. It could be suggested that these infants belong to two different learning types and that infants who do exhibit evidence for a decrease in attention during training are those who are able to quickly and more efficiently solve the task. This is also evidenced in the absence of a significant increase in target looking in the post-naming phase of the test block of Condition A for infants who did not reach criterion during training.

As far as parental feedback is concerned, word comprehension was not shown to be affected by parental reports on either infants receptive vocabulary as reported in

the Oxford CDI or familiarity with the to-be-learned words and objects as reported in the CEQ. Also, when parental feedback was introduced as a covariate in the analysis, the post-training test blocks of Conditions A and B were not shown to be different; this suggests that parental feedback, either in the Oxford CDI or the CEQ, seems to increase the variance in the dataset and to render the analysis less powerful. It could be that parental feedback is of limited use at this early stage of lexical development and infants need more experience with language before these reports can be predictive of infant performance in the laboratory. The Oxford CDI was, therefore, not used in any of the subsequent Experiments; CEQ was used in both Experiment Four and Experiment Five, in order to test whether familiarity effects are more relevant when training infants with multiple objects from a given category and when testing for their ability to categorise, respectively.

In terms of methodology, the LLD measure was also shown to be more sensitive than PTL in detecting word comprehension, as was the case in the previous experiment. It could be that in designs which involve habituation of the infants to repetitive presentations of the same pictures/picture-label pairings, the LLD is better able to capture subtle behavioural changes, whereas the PTL is more prone to behaviour not guided by the linguistic input.

Although these conclusions lend support to exemplar-based accounts, there are two concerns which need to be tested before we can make safer assumptions about infants' first words:

- Exemplar-based accounts assume that infants need to experience several naming instances with the same word but different category members before they realise that words refer to categories of objects. Infants should, therefore, be more willing to generalise the newly-learned label to a new exemplar if they are

trained with multiple within-category items. This assumption is directly tested in Experiment Four. The latter is identical to Experiment Three, but during training infants are presented with multiple exemplars from each category, instead of just one.

- According to constraint-and-principles accounts, infants will generalise a newly-learned label to all members of a corresponding perceptual/conceptual category. We, therefore, need to establish that infants consider the training and testing objects used in Experiment Three as belonging to the same categories, namely "cow" and "horse", and that their failure to generalise the label is a linguistic and not a cognitive limitation. This concern is tested in Experiment Five. The latter is a categorisation study without any labels, aiming to test whether infants perceive the training and testing objects as belonging to the same categories.

Chapter 4

Experiment Four - Multiple Exemplars

Experiment Four is identical to Experiment Three, except for the on-screen training block during which infants are presented with multiple exemplars from each category, instead of just one. As in the previous Experiments, it aims to test word learning and label generalisation. However, the task is slightly different, because infants need to associate labels to groups of objects, and not to single objects, and to generalise the labels to new objects from the adult categories. The latter can happen under two conditions: either because infants are able to create on-line categories (or at least activate their existing categories) and associate the labels to those categories; or because one of the training objects is highly similar to the testing object and infants, therefore, generalise the label on the basis of similarity between the two objects. In the former case, infants are simultaneously involved in category formation/activation and word learning, while in the latter they use a perceptual dimension of the objects (in this case, similarity) to successfully solve the task.

This experiment was conducted in order to test directly for the predictions of exemplar-based accounts in relation to label generalisation: infants should not be

willing to generalise a newly-learned label to new exemplars from the corresponding category before they experience a number of naming instances with multiple within-category[1] exemplars. The first part of this prediction was confirmed in Experiment Three, where infants failed to generalise the newly-learned label to a new typical exemplar from the corresponding adult category. Experiment Four was designed to test the second part of this prediction.

4.1 Creating Categories in the Laboratory

The word-learning task has been shown to be successful in Experiment Three with a single exemplar from each category. However, in Experiment Four infants are asked to associate multiple objects to the labels, and not just one object to each label. Infants could, therefore, be involved in categorisation and word learning at the same time. Visual categorisation tasks have been used extensively in infant research during the first year of life and they typically involve the creation of a single category of objects (Eimas & Quinn, 1994; Younger & Cohen, 1983; Fulkerson & Waxman, 2007). That is, infants are presented with multiple within-category exemplars (for example, "cats") until their attention begins to drop and then they are tested with a new within-category exemplar ("cat") and a new out-of-category object ("horse"). Renewed interest in the out-of-category object and low attention to the within-category exemplar suggests that infants have created a category (in this instance, a "cat" category) which excludes the out-of-category object.

However, being able to create a single category does not imply that infants can also create two categories on-line, even if these objects are fairly familiar to them. Evidence for such an ability has been shown in the past with infants at 10 months,

[1] As far as the adult categories are concerned.

with and without labels associated with the categories (Younger & Cohen, 1983; Plunkett, Hu, & Cohen, 2008). The stimuli were sketches of animals which varied systematically along specific dimensions and, therefore, should have been fairly easy for infants to process. On the other hand, in Experiment Four infants were presented with pictures of real objects, "cows" and "horses", which should be more familiar to them, than the made-up animals. In terms of task difficulty in this Experiment as opposed to the Younger studies, the increased complexity of the objects (real animals instead of sketches) and the simultaneous increased familiarity (real, and not made-up) should be expected to make the task similarly difficult for infants in the two studies and they should be expected to categorise the stimuli successfully.

4.2 Experimental Design

4.2.1 Participants

The final sample consisted of 60 (27 males) infants with a mean age of 9.97 months ($R : 9.23 - 10.67$, $SD : 0.30$). Infants were recruited in the same way as in previous experiments.

Infants in this Experiment were divided into two Conditions (30 infants each) on the basis of the testing stimuli (familiar/new exemplar) presented to them. The two sexes were roughly equally represented in each condition (Condition A: 15 males/ Condition B: 12 males). Ten additional infants were tested but excluded from the final set; five due to fussiness, three due to equipment failure and two due to experimenter error.

4.2.2 Stimuli

As in Experiment Three, two words were used in this study; "cow" and "horse". Each infant was trained and tested with both words. Prior knowledge of the words was tested for each individual infant through parental feedback using the Oxford CDI, as well as in a pretest block.

Five exemplars from each category, "cow" and "horse", of varying typicality were used in this study. Because this was primarily a generalisation task for infants, it was considered important to have a measure of the perceptual similarity of the objects used as well. The, within-category and between-category, visual similarity of these objects was, therefore, judged by adults as an index of how similar infants might perceive them to be. For both typicality and visual similarity rating studies see Chapter Six.
All the objects used were rated by adult native speakers of British English for typicality, as well as visual similarity. No objects judged as atypical were included in the final set.

Infants were trained with both pictures of objects and real objects. The toy-objects were naturalistic animal figures of the German brand "Schleich". The same objects as those used in Experiment Three were used for training. The toy-cows measured 13cm long and 7cm tall and the toy-horses 12cm long and 10cm tall. The visual stimuli were processed and presented in the same way as in Experiment Three.

Audio files were recorded in a single session by a female voice delivering them in high-affect child-directed speech (Fernald, 1992; Mills et al., 2005). Audio stimuli were processed and presented in the same way as in Experiment Three.

4.2.3 Design

Experiment Four aimed to teach infants two category-label pairings and then test them for Word Learning (Condition A), as well as for Word Generalisation (Condition B) to new exemplars from the same adult categories. During test a standard IPL task was used. During training a habituation task was administered in order to achieve a more infant-controlled procedure.

Five exemplars from each category were used in this Experiment. The exemplars which were rated as the most typical (see Section 6.2.1 for a description of the Typicality Rating Study) were used at pretest and test in both Conditions. These were the same objects as those used in Experiment Three. For the training block, four exemplars of varying typicality were used - training stimuli were chosen from a set of seven objects from each category and were those rated by adult native speakers as being typical and borderline objects. No atypical objects were used.

All other things remained identical between Experiment Three and Experiment Four. During pretest, infants sat through a four-trial block; in each trial both pictures were simultaneously presented on the screen along with a word corresponding to one of the pictures. Pretest was used in order to test for prior knowledge of the to-be-learned pairings.

During the interactive play-phase, infants were presented with the two toy-objects one at a time and heard the labels several times.

During the on-screen training phase, each infant was presented with the multiple object-label pairings. Each object-label pairing was presented individually on the right or left side of the screen. Four images from each category were used in this

block. The training block consisted of twenty trials, ten for each category, divided in five sub-blocks of four trials - two for each category. This meant that one object from each category had to be presented twice. Although ideally, five objects from each category would have been included in the training set, the selection was limited on the basis of what was available in the market, since real toy-objects were used. Each object was presented equal times across infants. Pairs of between-category objects were exhaustive, in order to avoid any between-category similarity effects.

The test block in Condition A was identical to the pretest block and aimed to test the word learning abilities of infants and the effect of training on infant behaviour. In Condition B the test block differed from the pretest in only the visual stimuli and it aimed to test generalisation of the labels to new exemplars from the same adult categories.

4.2.4 Procedure

After a brief familiarisation period in the Oxford Babylab, parent and child were seated in a dark booth. Throughout the experiment infants sat on their parent's lap approximately a meter away from a 110(width) x 40(height)cm plasma flatscreen (except during the play-phase which took place in the Oxford Babylab seating area). Pictures measured 30cm(width) x 30cm(height) and were presented on the left and right of the screen, separated by a 28cm gap. Audio stimuli were presented through two speakers located just above the centre of the screen. Parents were asked not to interact with their child during the session. They were instructed to close their eyes and listen to music through headphones. Two cameras above each picture recorded infant eye movements, each sending a signal to a digital router to form a single twin split-image of the infant. A computer captured the whole session for later off-

line scoring. The experimenter was able to manually control the experiment, which proceeded at the individual infant's own pace. A tone delivered through the speakers served to re-capture infant attention between trials.
During the interactive play-phase, parent and infant were seated in the Oxford Baby-lab reception area and the experimenter sat opposite them at approximately 50cm distance so that the infant could reach the toy presented to her.

The procedure used was identical to that of Experiment Three, except from the on-screen training block.

Pretest

Identical to Experiment Three.

Interactive Play-Phase

Identical to Experiment Three.

On-Screen Training Phase

After the interactive play-phase, infants were taken back into the booth for an on-screen training phase. During training trials infants were presented with a single moving picture along with the appropriate word. Each trial lasted 12.5s; at 500ms infants were prompted to "Look". At 2000ms the target word was presented for the first time. At 4000ms the phrase "Look at the X" was presented. And finally at 6500ms infants were presented with the last phrase, "Do you remember the X?". This process was repeated for each item separately.

The training block consisted of twenty trials, divided in five sub-blocks of four trials. Each sub-block consisted of two training trials for each object-label pairing.

The on-screen training block differed from Experiment Three in that four exemplars from each category were used, instead of a single object. Within each sub-block of the training block, infants were presented with one "cow" and one "horse". The four "cows" and four "horses" in the training block were paired with each other exhaustively across infants and order of presentation of the sub-blocks varied randomly within-participants.

The trials were scored on-line by a trained researcher blind to the experimental conditions. The first sub-block was considered to be the baseline to which all subsequent sub-blocks were compared. Training ended when looking time had fallen below a pre-defined criterion (65% decrease from the baseline - first four trials) or when all twenty trials had been completed. The sub-blocks of four trials ensured equal exposure to both word-category pairings before training stopped. Although decreasing attention to criterion for each individual category could not be ensured, a post-hoc analysis tested whether infants habituated to one or both of the categories. The criterion of 65% has been used in previous word learning studies, and seems to be a good index of infants' decreasing attention, while ensuring that they are still relatively engaged in the task (Fennell & Werker, 2004).

Test Phase

Identical to Experiment Three.

The exemplar used to test Word Generalisation was only presented in the test block of Condition B. For the interactive play-phase, the objects used in Experiment Three were also used here and they remained constant across infants. Although this rendered those two objects more familiar to infants than the rest in the training set, it was allowed because typical exemplars are considered to be more frequent in the

environment (Rosch & Mervis, 1975; Rosch et al., 1976) and it, therefore, made the word-learning scenario more naturalistic.

4.2.5 Predictions

Condition A aimed to test whether infants at 10 months of age are able to learn two word-category pairings in a brief experimental session when multiple objects from each category are presented to them. Although Experiment Three established that word learning is possible for such young infants, the increased complexity in this study makes word learning more difficult for them.

For Condition B, principles-and-constraints accounts make the same predictions as in Experiment Three. Infants should generalise the newly-learned labels to new within-category exemplars, as long as they consider the new exemplar and the training exemplar as members of the same category. On the other hand, exemplar-based accounts predict that infants should generalise the label to the new exemplars, because during training they have received increased exposure to the category-label pairing, through the presentation of multiple exemplars from each category along with the appropriate word.

4.2.6 Measures

The same measures used in Experiment Three were also used here.

Training Block

For the training phase, the TLT measure was used (see Section 2.5), calculated for the different sub-blocks separately. In Experiment Four, infants were also divided in

two Habituation Groups on the basis of whether they reached criterion or not. As in Experiment Three, two habituation groups were created: "Habituators" and "Non-Habituators". However, the multiple objects presented during training in Experiment Four have caused attention to the pictures to increase in comparison to Experiment Three, and more infants reached criterion only in the last sub-block, completing the whole training block.

Test Blocks

For the test blocks, two of the four IPL measures described in Section 2.5 were also used here: PTL and LLD, both calculated for the pre- and post-naming phases of the trial.

On the basis of parental feedback one additional measure was used: Familiarity Score (CEQ): identical to Experiment Three. Although word comprehension was shown not to be affected by infants' familiarity with different objects, the CEQ was used in this study, because it could be a better predictor of performance when infants are trained with multiple objects from each adult category instead of individual objects. The CEQ was filled in before the study and returned when parent and infant visited the laboratory. In Condition B all parents returned the CEQ; in Condition A three parents failed to provide it. Data for those infants were excluded from the parental feedback analysis, but not from the rest of the analysis.

4.2.7 Scoring

Infant looking behaviour was scored in the same way as Experiments Two and Three.

Intra-reliability tests were carried out for the training block in this Experiment, with

25% of the trials scored off-line by the reseacher and on-line and off-line measurements were compared using Pearson r correlations. This off-line reliability scoring was compared to both the on-line looking time to screen as well as to a second "corrected" version of it calculated by subtracting 100ms from the on-line measurements. This is usually incorporated in on-line measurements made by habituation software (e.g. *Habit*) and it aims to correct for the time the scorer needs to initiate a motor movement from when the infant's gaze begins to move (Cohen, 2003). Pearson r correlations revealed high agreement between on-line and off-line measurements of infant attention to the task, $r(160) = .95$, $p < .01$, for both measures.

4.2.8 Analysis

4.2.9 Training Block

In the training block, infants received multiple presentations of two word-category pairings, presented in sub-blocks of four trials. Within each sub-block, two picture-word pairings - one from each category - were presented separately twice. The unaggregated data for the training block consisted of 1087 trials, 551 for Condition A and 536 for Condition B. On average infants went through 4.52 training sub-blocks ($R = 2 - 5$, $SD = 0.78$), 4.60 for Condition A ($R = 3 - 5$, $SD = 0.62$) and 4.43 for Condition B ($R = 2 - 5$, $SD = 0.86$). An Independent-samples t-test compared the mean number of blocks accumulated in each Condition and showed no significant differences between the two, $t(58) = .86$, $p = .39$. In Condition A 13 infants reached the habituation criterion and in Condition B 15 infants did (overall Habituators = 28 and Non-Habituators = 32). Mean TLT in each sub-block, as well as for the two categories was calculated.

	Category			Habituation Group		
	Cow	Horse	Total	Non-Habituators	Habituators	Total
Condition A	79.82	79.47	79.14	183.40	125.45	158.29
Condition B	77.36	75.83	76.60	178.67	127.71	153.19
Total	78.59	77.15	77.87	181.19	126.66	155.74

Table 4.1: Total Looking Time (in seconds) for each Category and Habituation group in the two Conditions.

	Cow			Horse		
Item	ConditionA	ConditionB	Total	ConditionA	ConditionB	Total
1	66	64	130	68	71	139
2	67	69	136	72	67	139
3	67	67	134	70	64	134
4	75	68	143	66	66	132
Total	275	268	543	276	268	544

Table 4.2: Number of presentations for each exemplar of Cow and Horse in the two Conditions.

Tables 4.1 shows the total looking time to the pictures, accumulated in the two Conditions for the group total and for each Category and Habituation Group separately.

Because infants were presented with multiple exemplars from each category during training, the number of times each picture was presented across infants was calculated, in order to ensure that infants received similar exposure to the whole training set. Table 4.2 shows the number of times each picture was presented across infants in each Condition separately. Pictures appeared roughly equally across infants suggesting that infants have had equal exposure to them overall.

A univariate ANOVA with TLT as dependent variable and Condition and Item as fixed factors tested whether infants accumulated similar looking times to the two

pictures and in the two Conditions. Finally looking times in the different sub-blocks of the training block were calculated and compared for the two Habituation groups separately using a univariate ANOVA with TLT as dependent variable and Sub-block as a fixed factor.

Normality tests were conducted for the TLT measure, and all groups were found to be normally distributed.

4.2.10 Test Blocks

Infants received four trials at test, two for each word. The unaggregated data set for the pretest block consisted of 206 trials, 105 for Condition A and 101 for Condition B. For the test blocks the dataset consisted of 219 trials, 107 for Condition A and 112 for Condition B. The difference in the number of trials in each test block is due to the exclusion of trials in which the infant did not look at both pictures. Data were further aggregated by test block, word and habituation group.

In order to test whether infants in the two Conditions and the pre- and post-training test blocks behaved in a similar way, a 2 (Trial Phase) x 2 (Test Block) x 2 (Condition) repeated measures ANOVA was run for both measures.

Analyses of Variance were also conducted for both measures with Trial Phase as within-subject factors and Condition and Habituation group as between-subjects factors. Normality tests were conducted for all measures and outliers were excluded in order to achieve normality. In Condition A, one infant was excluded from the pretest block for both the PTL and LLD measures. In Condition B, one infant was excluded from the post-training test block for both measures.

4.3 Results

4.3.1 Training Phase

During the on-screen training block, infants were presented with multiple word-object pairings, using four exemplars from each adult category. Figure 4.1 shows the total looking time accumulated during training for each of the two categories, "cow" and "horse", in the two Conditions separately. It appears that the two categories accumulated similar attention both overall and in the two Conditions. In order to test this observation, a Univariate ANOVA with TLT as a dependent variable and Condition and Category as fixed factors was conducted. No main effects were found, Condition: $F(1, 120) = .43$, $p = .52$/Category: $F(1, 120) = .14$, $p = .71$, and no interaction between Condition and Category, $F(1, 120) = .00$, $p = .98$. Data were, therefore, collapsed across Condition and Category for subsequent analyses in the training block.

Figure 4.2 shows the TLT measure for the two Habituation groups separately, from aggregated means from the two Conditions. The two groups appear to have accumulated different attention to the pictures during training: Habituators looked at the pictures far less than Non-Habituators. Figure 4.3 shows the different sub-blocks separately for the two Habituation Groups. As in Experiment Three, infants in the Habituators group show a more sudden decrease in attention, whereas the Non-Habituators retain high levels of attention throughout the training block. It is worth noting that some infants in the Habituators group in Experiment Four go through the whole training block, but behave a lot differently from infants in the Non-habituators group.

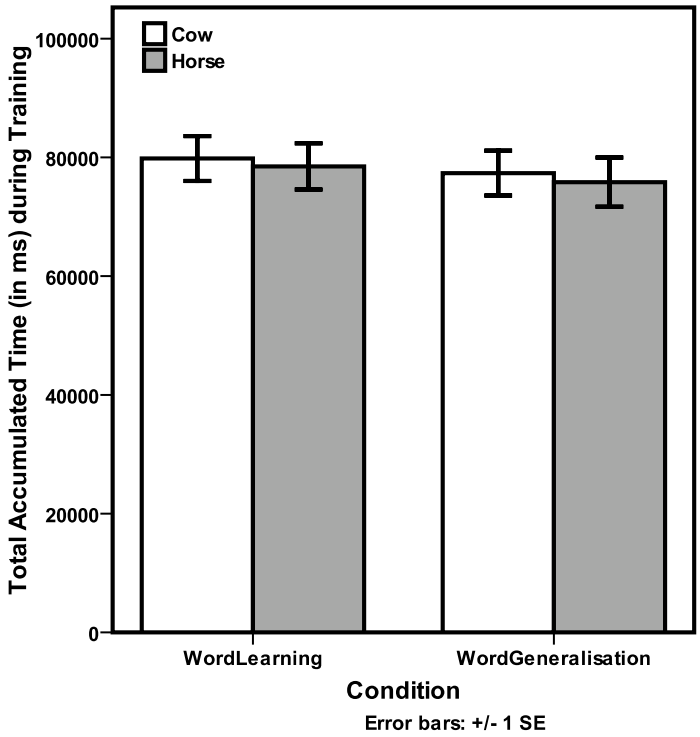


Figure 4.1: Mean Looking Time separately for each Category in the two Conditions of Experiment Four.

A Univariate ANOVA with TLT as a dependent variable and Sub-block as a fixed factor was conducted for Habituators and Non-Habituators separately. For the first group, a highly significant main effect of Sub-block was found, $F(4, 222) = 29.65$, $p < .001$, ${\eta_p}^2 = .35$, with a decrease in attention as training progressed. Bonferroni post-hoc comparisons showed that all sub-blocks differed from each other ($p < .001$) except for the pair Sub-block 1 and 2 ($p = .17$). For the Non-Habituators, there was no main effect of Sub-block, $F(4, 222) = .40$, $p = .81$, suggesting that infant attention during training did not fall significantly. Bonferroni post-hoc tests confirmed the absence of a main effect, by showing no differences between any of the sub-blocks.

This analysis suggests that during training infants spontaneously divided into two groups, exhibiting different behaviours; this was taken into account during the

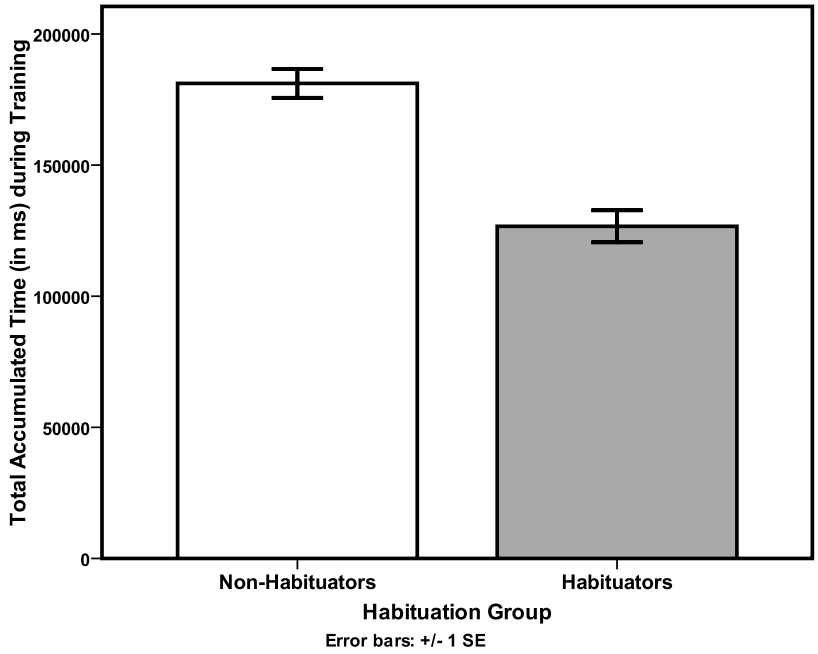


Figure 4.2: Mean Looking Time presented separately for each Target Picture in the two Conditions of Experiment Four.

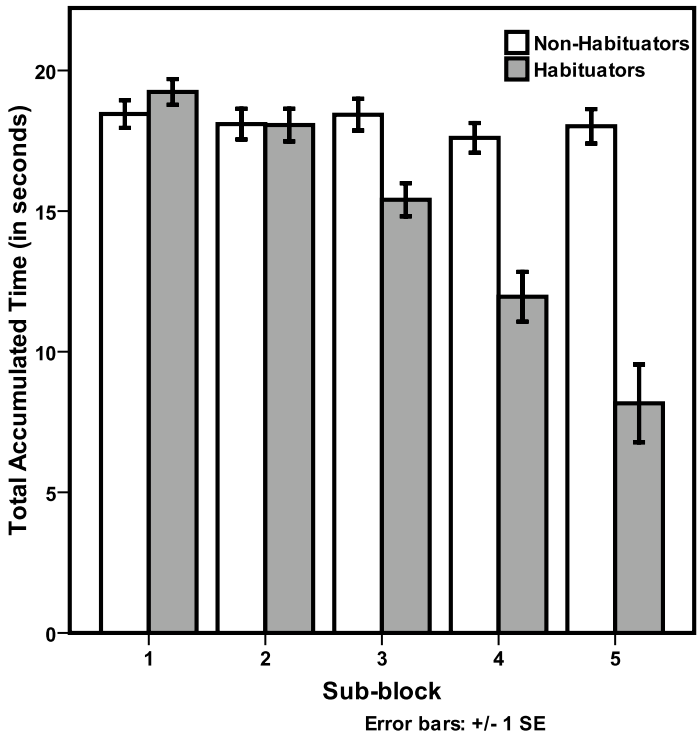


Figure 4.3: Mean Looking Time for each training sub-block presented separately for the two Habituation groups.

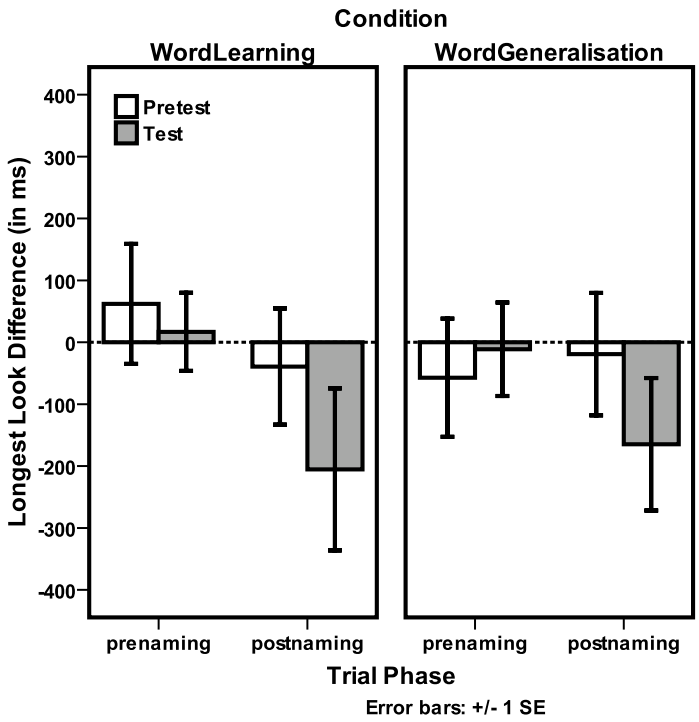


Figure 4.4: LLD for each test block separately for the two Conditions.

analysis of the post-training test blocks.

4.3.2 Test Blocks

Infants were presented with two tests blocks of four trials in this Experiment, one before and one after training. Figure 4.4 shows the LLD measure for the two test blocks in each Condition separately. It appears that in the pre-naming phase of the trial infants do not show any systematic behaviour in any of the test blocks. In the post-naming phase, however, there appears to be a preference for the distracter picture in the post-training test block of both Condition A (Word Learning) and Condition B (Word Generalisation), but more pronounced in Condition A.

In order to test these observations a three-way repeated measures ANOVA with

Trial Phase (pre- and post-naming) and Test Block (Pretest and Test) as within-subjects factors and Condition (Word Learning and Word Generalisation) as a between-subjects factor was conducted for both measures. No main effects were exhibited for either measure, Trial Phase: PTL: $F(1, 57) = 1.65$, $p = .20$/ LLD: $F(1, 56) = 2.59$, $p = .11$; Test Phase: PTL: $F(1, 57) = 1.02$, $p = .32$/ LLD: $F(1, 56) = .52$, $p = .48$; Condition: PTL: $F(1, 57) = .40$, $p = .53$/ LLD: $F(1, 56) = .13$, $p = .72$. Also no interactions were found for either measure.

Although no interactions between Trial Phase and Test Block were found, paired-samples t-tests were run for each Condition separately comparing the two phases of the Trial for the pretest blocks, in order to ensure that infants had no prior knowledge of the labels. The two trial phases were shown to be similar in both Conditions, Condition A: PTL: $t(29) = .24$, $p = .82$ (pre-naming: $M = .51$, $SD = .13$ and post-naming: $M = .49$, $SD = .18$)/ LLD: $t(29) = .77$, $p = .45$ (pre-naming: $M = 62.11$, $SD = 530.88$ and post-naming: $M = -.39$, $SD = 513.61$); Condition B: PTL: $t(28) = -.46$, $p = .65$ (pre-naming: $M = .47$, $SD = .12$ and post-naming: $M = .49$, $SD = .17$) / LLD: $t(28) = -.06$, $p = .94$ (pre-naming: $M = -.57.21$, $SD = 514.08$ and post-naming: $M = -45.89$, $SD = 530.62$).

This analysis suggests that in Experiment Four infants as a group exhibited no systematic looking in any of the test blocks and Conditions, indicating a failure to exhibit any evidence for word learning or for word generalisation. In Experiment Three, however, there was evidence that Habituators and Non-Habituators behaved differently, not only during training, but also during test, with stronger evidence for learning of the object-label pairings from the Habituators group. Although no evidence for word comprehension was found in this analysis, the post-training test blocks were further analysed in relation to Habituation group, in order to test whether the

two groups behaved differently after having been exposed to repetitive presentations of several object-label pairings.

Habituation Group

Figure 4.5 shows the Habituators and Non-Habituators in the post-training test blocks of the two Conditions. In the pre-naming phase of the trial, both groups appear to behave at chance levels. In the post-naming phase, in the Word Generalisation condition there seems to be no systematic behaviour from the two groups. In the Word Learning condition, however, a systematic preference for the distracter picture is evident for the Habituators group, but no systematic behaviour for Non-Habituators.

In order to evaluate these observations a three-way repeated measures ANOVA was conducted for both measures with Trial Phase as a within-subjects factor and Condition and Habituation group as between-subjects factors. A main effect of Trial Phase was found for the LLD measure, Trial Phase: PTL: $F(1, 55) = 2.91$, $p = .09$/ LLD: $F(1, 56) = 4.17$, $p < .05$, ${\eta_p}^2 = .07$, with a decrease in target looking after the onset of the target word, Pre-naming: $M = .50$, $SE = .02$/ Post-naming: $M = .44$, $SD = .02$. No other main effects were found, Habituation Group: PTL: $F(1, 55) = .74$, $p = .40$/ LLD: $F(1, 56) = 2.33$, $p = .13$; Condition: PTL: $F(1, 55) = .00$, $p = .97$/ LLD: $F(1, 56) = .20$, $p = .66$. An interaction between Condition and Habituation Group was also present for the LLD measure, $F(1, 56) = 3.98$, $p = .05$, ${\eta_p}^2 = .07$.

On the basis of the interaction between Condition and Habituation group, a repeated-measures ANOVA with Trial Phase as a within-subjects factor and Habit-

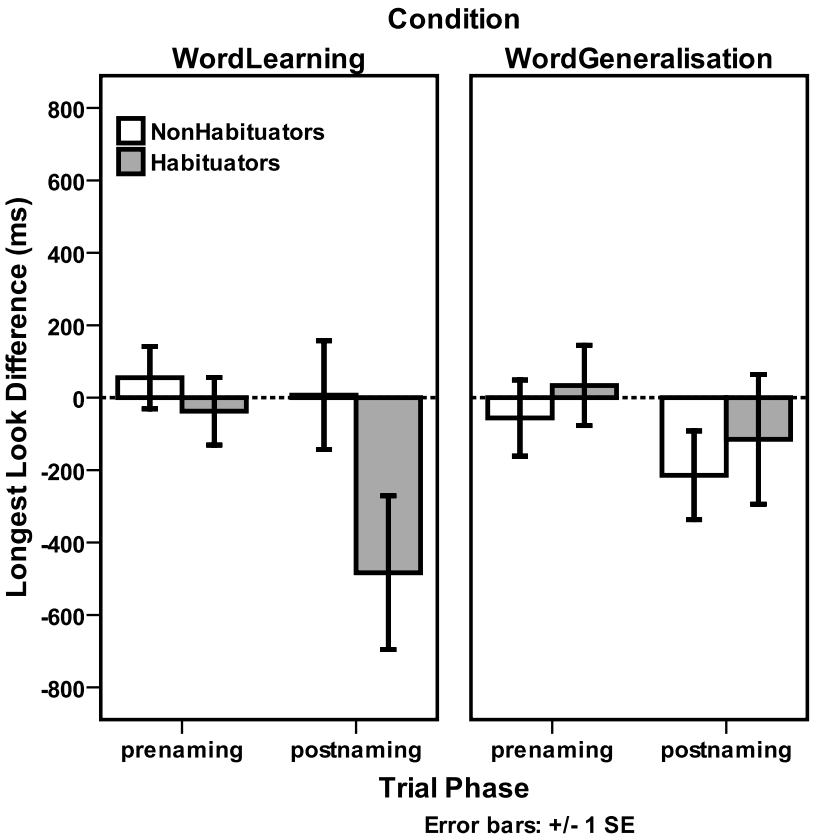


Figure 4.5: LLD for each Habituation group in each test block and Condition separately.

uation group as a between-subjects factor was run for the LLD measure for each Condition separately. No main effects and no interactions were found in either Condition, except for a non-significant trend for a main effect of Habituation group in the Word Learning Condition, $F(1, 27) = 3.35$, $p = .08$, ${\eta_p}^2 = .11$. On the basis of this trend and Figure 4.5, paired-samples t-tests were run comparing the pre- and post-naming phases of the trial to each other only for Habituators in Condition A. Also one sample t-tests were run comparing the two phases of the trial to chance for Habituators and only for the Word Learning Condition. These tests showed no significant differences between the two trial phases, PTL: $t(11) = 1.34$, $p = .21$ (pre-naming: $M = .47$, $SD = .14$ and post-naming: $M = .38$, $SD = .21$)/ LLD: $t(10) = 1.30$, $p = .22$ (pre-naming: $M = -81.44$, $SD = 298.52$ and post-naming: $M = -388.03$, $SD = 751.59$), but a significant preference for the distracter picture after the onset of the target word was present for the LLD measure and a non-significant trend for an interaction for the PTL measure, PTL: $t(11) = -1.87$, $p = .08$/ LLD: $t(10) = -2.13$, $p = .05$.

The main effect of Trial Phase found in this analysis suggests that infants exhibited an overall preference for looking at the distracter picture after hearing the target word. The interaction between Condition and Habituation group points to the fact that infants who have reached criterion and infants who have not behaved differently in the two Conditions. Habituators in the Word Learning Condition seem to be driving this result, by showing decreased target looking post-naming, although in a weak form, possibly due to the small number of infants in that group. Distracter looking has been interpreted as evidence for word understanding in the past and in Chapter Two of this thesis, and will be discussed below.

Parental Feedback - Familiarity Effects

On the basis of parental feedback a familiarity score (FS) from the CEQ (see section 3.2.6) was calculated for each infant. Although no evidence for familiarity effects using the CEQ was found in the previous Experiment, it could be that parental feedback on infants' experience with different objects is more relevant when multiple exemplars from an adult category are used during training.

As in Experiment Three, parents reported on their infant's exposure to each of the categories, across three dimensions: frequency of exposure, number of exemplars encountered, exposure to the name of the object. For both categories, infants accumulated an average score of 6.33 ($R = 14.00 - 0.00$, $SD = 4.51$) - in Condition A 6.56 ($R = 14.00 - 0.00$, $SD = 4.30$) and in Condition B 6.13 ($R = 14.00 - 0.00$, $SD = 4.76$). For "cow", infants accumulated an average score of 3.14 ($R = 8.00 - 0.00$, $SD = 2.54$) - in Condition A 3.37 ($R = 7.00 - 0.00$, $SD = 2.59$) and in Condition B 2.93 ($R = 7.00 - 0.00$, $SD = 2.52$). Finally for "horse", infants accumulated an average score of 3.19 ($R = 7.00 - 0.00$, $SD = 2.45$) - in Condition A 3.19 ($R = 7.14 - 0.00$, $SD = 2.37$) and in Condition B 3.20 ($R = 7.00 - 0.00$, $SD = 2.57$).

Figure 4.6 shows the relationship between the change in the LLD measure between the two phases of the trial and the CEQ score for both categories, separately for each Condition. In order to investigate whether CEQ affects behaviour in any meaningful way, a two-way repeated measures ANCOVA was conducted for both measures with Trial Phase as a within-subjects factor, Condition as a between-subjects factor and CEQ Score as a covariate. No main effects were found, Trial Phase: PTL: $F(1, 54) = .84$, $p = .36$/ LLD: $F(1, 53) = 2.00$, $p = .16$; Condition: PTL: $F(1, 54) = .04$, $p = .84$/ LLD: $F(1, 53) = .01$, $p = .95$; CEQ: PTL: $F(1, 54) = .15$, $p = .70$/ LLD: $F(1, 53) = .09$, $p = .77$. CEQ was not shown to affect Trial Phase in any way, PTL:

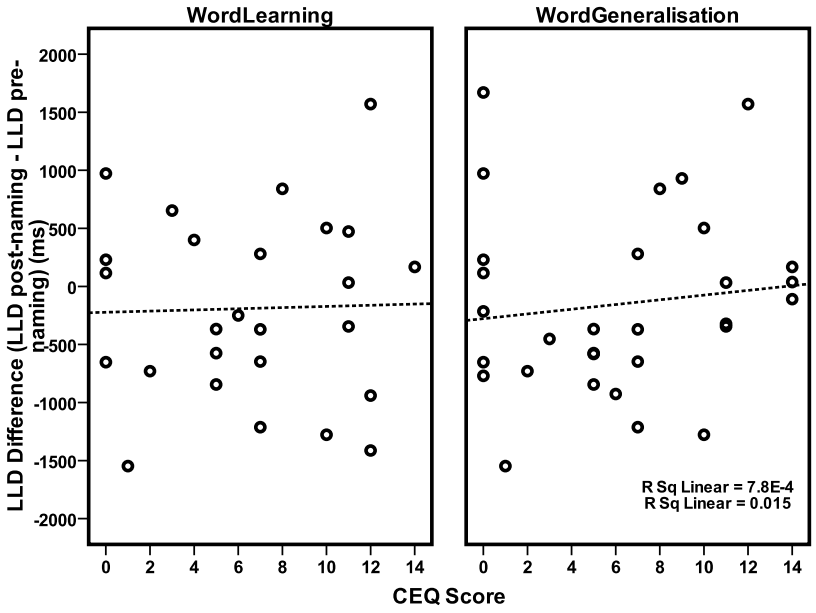


Figure 4.6: Relationship between the change in LLD before and after the presentation of the target word and CEQ score, in the post-training test blocks of Conditions A and B.

$F(1, 54) = .05$, $p = .83$/ LLD:$F(1, 53) = .35$, $p = .56$.

Additionally Pearson r correlations were run for each measure comparing the change between the two trial phases and CEQ score for each Condition separately. No significant correlation was found between the two for either measure and either Condition, PTL: Condition A: $r(30) = -.11$, $p = .59$/ Condition B: $r(30) = .03$, $p = .86$; LLD: Condition A: $r(30) = .03$, $p = .89$ / Condition B: $r(30) = .12$, $p = .52$.

Although no interaction between Trial Phase and Condition was found, a separate analysis was run only for infants in the Habituators group of Condition A, for which a systematic behaviour related to the presentation of the label was present in previous analyses. A repeated measures ANCOVA was conducted for both measures with Trial Phase as a within-subjects factor and CEQ Score as a covariate. No main effect of

Trial Phase or CEQ was found, Trial Phase: PTL: $F(1, 25) = .07$, $p = .79$/ LLD: $F(1, 24) = .60$, $p = .45$; CEQ: PTL: $F(1, 25) = .02$, $p = .88$/ LLD: $F(1, 24) = .01$, $p = .94$; CEQ was not shown to affect Trial Phase in any way, PTL: $F(1, 25) = .29$, $p = .60$/ LLD:$F(1, 24) = .02$, $p = .89$. A Pearson r correlation was, also, run for both measures, comparing the change between the two trial phases and CEQ score. Again no significant correlation was found between the two for either measure, PTL: $r(30) = -.11$, $p = .59$; LLD: $r(30) = .12$, $p = .52$.

This analysis suggests that parental feedback on infants' prior experience with the test items does not have any meaningful relationship with word comprehension.

4.4 Discussion

In Experiment Four, as in Experiment Three, infants received a test-train-test design: a pretest block served to test for prior knowledge of the to-be-learned labels, training followed with multiple presentations of the word-object pairings and then post-training test served to test for word-learning (Condition A) and for generalisation of labels to new objects (Condition B). The only difference between Experiments Three and Four was that in the latter Experiment infants were trained with multiple exemplars from each of the two categories, "cow" and "horse", instead of just a single exemplar.

In this Experiment no evidence for a target preference after naming was found in the post-training test blocks of either Condition; however, some infants showed a distracter preference during the post-naming phase of the trial in the test block of Condition A. Distracter looking has been interpreted as evidence for word learning in the past, and it has been discussed in Chapter Two of this thesis as well. It ap-

pears that, unlike Experiment Three, in this Experiment infants exhibit a systematic stronger preference for the distracter picture after the presentation of the label and, therefore, prefer to attend more to novelty, rather than familiarity, during test.

Ceteris paribus, it appears that this behaviour and the contrasting results of Experiment Three can only be due to the design of the training block: in Experiment Three infants are presented only with a single exemplar from each adult category, namely one cow and one horse. In Experiment Four infants are presented with multiple exemplars from each adult category. This causes infants in Experiment Four to take longer to habituate than infants in Experiment Three; this difference in the amount of attention during training could be a possible source of infants' different attentional preferences during test in each Experiment. That is, during the training block in Experiment Three infants' only task can be that of learning which label goes with which object. However, infants in Experiment Four can be engaging in an additional task: in parallel to attaching the label to (one of) the familiar objects, they could be creating perceptual categories on the basis of the multiple exemplars from each category (and, thus, taking longer to habituate due to the task complexity). Their longer attention might, therefore, be an artifact of the increased complexity of the visual stimuli, and not the word-learning task itself. In this case, they might still be getting habituated to the word learning task, but retain attention because they are also creating perceptual categories on the basis of the visual stimuli, and during test they exhibit their task-specific habituation in the form of distracter looking.

Infants also spontaneously divided into two groups on the basis of their attention pattern during training: Habituators and Non-Habituators. These two groups behaved very differently during training, as in Experiment Three; Habituators showed a steep and dramatic fall of attention to the pictures as training progressed, whereas Non-Habituators showed high interest in the task throughout the training block. In addition to their different behaviours during training, these two groups were shown to

be rather different during test, although this difference was not found to be significant in some analyses due to the small numbers in each group. Habituators in Condition A were shown to prefer to orient more at the distracter picture during the post-naming phase of the trial in the post-training test block, exhibiting evidence for learning the labels.

As was the case in previous Experiments, parental feedback was shown not to have any meaningful relationship with word comprehension in this Experiment. It appears that prior experience with the words and the objects might not have an effect on infants' linguistic abilities at this stage in development or that parents are not reliable in reporting on their infant's experience. More research on this is needed, in order to establish the reliability of parental reports for infants in their first year of life.

This experiment aimed to test the assumption of exemplar-based accounts that infants "discover" the extension of words through exposure to naming instances with multiple within-category exemplars. Evidence for word learning was found in Condition A, but infants failed to generalise the labels once again. Infants might or might not have associated the label pairings with all exemplars from a given category, but only with the most typical exemplars, which are used as test objects in this Experiment. In either case, it could be that infants need more exposure to naming instances of within-category exemplars before they figure out the extension of a word, or that they need more experience with word learning altogether before they realise that words refer to classes of objects, and not just to individual objects.

An alternative explanation could be that infants did not consider the training exemplars paired with each label as the "same kind of thing", and this hindered their ability to also learn the words. In order to test this possibility, Experiment Five was designed aiming to test infants perceptual categories, in a setting identical to that of Experiments Three and Four.

Chapter 5

Experiment Five - Categorisation

Experiment Five was a categorisation study, aiming to test whether infants perceive the training and test stimuli presented in Experiments Three and Four as belonging to two categories of objects, namely "cows" and "horses". This study was necessary, in order to evaluate whether

1. Infants in Experiment Three refused to generalise the newly-learned labels to new objects from the same categories because of a linguistic "limitation" or because they did not consider the training and testing exemplars as being the "same kind of thing".

2. Infants in Experiment Four spontaneously divided the eight training exemplars into two categories - with the help of the labels provided, and whether their failure to both learn and generalise the labels was due to their inability to successfully categorise the training set or to some other cognitive limitation.

As described in Section 4.1, categorisation studies have been extensively used to study the formation of a single category in the laboratory (Eimas & Quinn, 1994; Younger & Cohen, 1983; Fulkerson & Waxman, 2007). The only studies which found evidence for the spontaneous formation of two categories were Younger's series of studies on categorisation in the first year of life and, more recently, Plunkett et

al., 2008, who replicated and further developed Younger's research (for a detailed description of those studies, see Section 1.2.1). In both cases, the visual stimuli were sketches of made-up animals which varied on five values (1-5) across four dimensions (length of neck, separation of ears, length of legs and thickness of tails). When the set of objects presented to infants during familiarisation contained animals which varied randomly across the four dimensions (Broad Condition), evidence for the creation of a single category was found. When the set of objects contained animals which were made up of either values 1 and 2 or 4 and 5 (Narrow Condition), evidence for the creation of two categories was found. In Plunkett et al., 2008, labels were presented along with the animals in the Narrow Condition; in the presence of two labels which correlated with the two perceptual categories, evidence for the creation of two categories was found, but in the presence of a single label evidence for a single category was found (Plunkett et al., 2008).

In Experiment Four, the two adult categories, "cows" and "horses", were presented with two labels systematically correlated with them. On the basis of previous findings, hearing a label with the objects should give infants an extra cue in the categorisation task (Plunkett et al., 2008; Balaban & Waxman, 1997; Fulkerson & Waxman, 2007), whether infants are learning the labels for the objects or not. Alternatively, the presentation of the auditory stimulus along with the pictures could be disrupting visual categorisation (Robinson & Sloutsky, 2004). If the auditory-dominance hypothesis is correct, then in Experiment Four infants could just be failing to do any learning altogether, either of the word-object pairings or of the visual categories they were presented with.

5.1 One or Two Categories: A Crucial Test

In order to test this, Experiment Five was designed. The latter was identical to Experiment Four, except for the final test block. During test infants were presented with three novelty-preference trials without any labels. In each of these trials, a new exemplar from the familiar categories, either a new "cow" or a new "horses", and a completely new animal was presented. Experiment Five aimed to test both whether infants created any categories at all, and also whether they created a single category, which contained both "cows" and "horses", or two categories, a "cow" category and a "horse" category. The former was evaluated by looking at a novelty preference to the new animals. The latter was more challenging, because even if a novelty preference was found during test, it would not be straightforward to evaluate whether all familiar animals belong to the same category or to two different ones.

The only available studies testing the formation of two categories in the laboratory are the Younger et al. series of studies, as well as Plunkett et al., 2008. In both cases, category membership was tested using artificial animals: average and modal stimuli were created by systematically varying the values across the four dimensions described above. Modal stimuli, that is stimuli which contained the extreme values from each category, were those preferred when infants created a single category, making the extremes more novel. Average stimuli, that is stimuli which contained the mean values from the familiarisation set, were those preferred when infants created two categories, making the average stimulus more novel. The challenge for Experiment Five was, therefore, the task of finding exemplars that would represent the Average and Modal Stimuli for the two real categories used, namely "cow" and "horse".

Modal stimuli were rather straightforward: atypical exemplars seem to function as modal stimuli in the sense that they contain features which are not very common

or frequent in the category and are, thus, considered peripheral. Average stimuli, however, were harder to obtain, since the objects in the current series of experiments were real and, therefore, difficult to manipulate both in terms of feature values, but also in respect to the dimensions which are considered important for a given category.

In order to obtain such stimuli a morphing software was used to create a picture which looked half way between a "cow" and a "horse". Morphing softwares work by shifting pixels in a relevant area according to a vector field. That is, each picture is marked with points which then become distorted as one picture fades away into the other. This ensures that the two visual images "blend" without the experimenter having to select one dimension over another as more relevant in creating an average stimulus for the two categories.

5.2 Experimental Design

5.2.1 Participants

The final sample consisted of 16 (12 males) infants with a mean age of 10.09 months ($R : 9.66 - 10.54$, $SD : 0.26$). Five additional infants were tested but excluded from the final set, one due to equipment failure and four due to fussiness. Infants were recruited through the Oxford Babylab parental database in the same way as in the rest of the Experiments.

5.2.2 Stimuli

As in Experiment Four, infants were trained and tested with two real categories of objects; "cow" and "horse". Prior knowledge of the words was tested for each individ-

ual infant through parental feedback using the Oxford CDI; in this study, however, infants did not receive a pretest block. This block was omitted because infants in Experiment Five were not tested for word-learning (or knowledge), but instead for visual categorisation of the training and testing stimuli. Additionally, infants in Experiment Three and Four were not shown to have had any prior knowledge of the to-be-learned words, so the pretest block was not considered to be a crucial test for infants' comprehension abilities.

The training block of Experiment Five was identical to Experiment Four. However, in order to test category membership in Experiment Five additional test stimuli were used: two out-of-category exemplars ("bear" and "tiger") were paired with the typical "cow" and typical "horse" used in the test block of Condition B in Experiments Three and Four, in order to test whether infants were successful in identifying the familiar animals over the new ones. A "cow-horse" combination (Average) was presented against an atypical "cow" and an atypical "horse" (Modal), in order to test whether infants created a single "cow-horse" category or two different ones, a "cow" and a "horse" category.

In Experiment Five, six exemplars from each of the familiar adult categories ("cows" and "horses") were used during familiarisation and test, as well as three additional test stimuli: a "bear", a "tiger" and a "cow-horse" combination. Infants were familiarised only with pictures in this Experiment, without the interactive play-phase of Experiments Three and Four during which toy animals were used. All the exemplars used in Experiment Four were also used here, but in each category an atypical exemplar was also added in the test phase. A single picture of a "bear" and a single picture of a "tiger" were used during test.

All pictures were digital photographs of toy-objects, produced by the German brand "Schleich", against a white background. The toy-bear measured 9.5cm long and 5.5cm tall and the toy-tiger 11cm long and 5.5cm tall. Photographs were digitally

processed to remove any background using Adobe Photoshop CS2 and saved in 24-bit bitmap format. During familiarisation, pictures moved up and down the y-axis as in Experiments Three and Four, whereas during test blocks pictures were presented stationary. All pictures were saved looking in two directions, by flipping the canvas horizontally for both still and moving images; this ensured that objects always faced the centre of the screen.

For the "cow-horse" combination (Average stimulus), morphing software (Morpheus Photo Morpher v3.10) was used: the "cow" and "horse" pictures used during test of Condition B in Experiments Three and Four were morphed together (for the original pictures see Appendix A; for the Average stimulus see Figure 5.1). The vector field contained points all over the two images: along the periphery of the animals, the face, the tail and the body, giving a total of 55 points. The morphing took place over 31 frames, and frame number 16 - which fell exactly half way through the morphing process - was chosen as the Average stimulus. The image was saved in 24-bit .bmp format and saved looking in two directions, as the rest of the pictures.

Auditory stimuli were only used in the familiarisation block and were identical to those used in Experiments Three and Four; during test pictures were presented in silence.

5.2.3 Design

Experiment Five aimed to evaluate whether infants could form visual/perceptual categories for the objects used in Experiments Three and Four. During test a novelty-preference procedure was used, without any auditory stimuli. During familiarisation a habituation task was administered in order to achieve a more infant-controlled



Figure 5.1: Average stimulus obtained by the morphing of the two typical test objects, a cow and a horse, used in Experiment Four in order to test generalisation of labels to new objects (for the original objects see A).

procedure.

Experiment Five did not have a pretest block or an interactive play-phase. The on-screen familiarisation phase was identical to the habituation procedure used in Experiment Four. Two test blocks followed, with a total of three novelty-preference trials presented to infants *in silence*. Test block A consisted of two trials, whereas Test Block B consisted of a single trial. In the first test block, the new typical within-category exemplars, "cow" and "horse", were paired with a new out-of-category exemplar, either a "bear" or a "tiger". In the second test block, the "cow-horse" combination was paired with either the atypical "cow" or the atypical "horse".

5.2.4 Procedure

Parent and child followed an identical procedure to Experiments Three and Four during the session.

On-Screen Familiarisation Phase

The training block of Experiment Five was identical to the habituation procedure used in Experiment Four.

Test Phase

Infants were tested for visual categorisation of the familiarisation stimuli in two test blocks, of two trials each. Three 5-second trials were presented to infants with two pictures simultaneously presented on the screen, but no labels. In the first test block, the typical within-category exemplars, "cow" and "horse", were paired with an out-of-category exemplar, either a "bear" or a "tiger". The within- and out-of-category test objects were paired with each other exhaustively across infants. The order of presentation of each trial was randomised for each infant. Presentation side for the within- and out-of-category stimuli was counterbalanced within infants and presentation side for each individual animal was counterbalanced across infants.
In the second test block, the "cow-horse" combination (Average) was paired with either the atypical "cow" or the atypical "horse" (Modal). The side of presentation for the average and modal stimuli, as well as the presentation side for each individual stimulus was counterbalanced across infants.

5.2.5 Predictions

Experiment Five aimed to test whether infants were able to form visual categories on the basis of the familiarisation set, and whether they formed one category, containing both "cows" and "horses", or two categories, a "cow" and a "horse" category. Younger (1985) as well as Plunkett et al. (2007) suggest that infants should be in position to form two categories in the laboratory, especially since the categories in question are perfectly correlated with the presentation of two labels.

5.2.6 Measures

For the familiarisation block the measures used in Experiments Three and Four were also used here. However, in the test phase of Experiment Five no labels were presented, so the trial could not be divided in a pre- and post-naming phase. Only one of the measures used in the previous Experiments was used, namely PTL, and was calculated for the whole duration of the trial.

Training Block

For the training phase, the TLT measure was used (see Section 2.5), calculated for the different sub-blocks separately. In Experiment Five, infants were also divided in two Habituation Groups on the basis of whether they reached criterion or not (a decrease in attention at least 65% of the first four trials). As in Experiments Three and Four, two habituation groups were created: "Habituators" and "Non-Habituators". As in Experiment Four, the multiple objects presented during training have caused increased attention to the pictures overall and some infants in the habituation group reached criterion only in the last sub-block, completing the whole training block.

Test Blocks

For the test blocks, two of the four IPL measures described in Section 2.6 were also used here: PTL and LLD, both calculated for the whole duration of the trial. The LLD measure in this Experiment is a measure of novelty preference. It is calculated by subtracting the longest continuous look to the familiar stimulus - new within-category animals ("cow" and "horse") - from the longest continuous look to the novel stimulus (out-of-category animals in Test Block A or Average stimulus in Test Block B). Positive values indicate a preference for the novel objects, whereas negative values indicate a preference for the within-category animals.

On the basis of parental feedback, one additional measure was used: Familiarity Score (CEQ): identical to Experiments Three and Four. The CEQ was used in order to test whether prior experience with the test categories could have an effect on infant's ability to successfully categorise them in the laboratory.

The CEQ was filled in before the study and returned when parent and infant visited the laboratory. Two parents failed to return the CEQ; data for those infants were excluded from the parental feedback analysis, but not from the rest of the analysis.

5.2.7 Scoring

Infant looking behaviour at test was recorded and scored off-line by a trained scorer, as described in Section 2.6. Looking behaviour during training was scored on-line, but was also recorded for off-line reliability scoring, as described in Section 3.2.7.

Intra-reliability tests were carried out for the training block in this Experiment, with 25% of the trials scored off-line by the reseacher and on-line and off-line measurements were compared using Pearson r correlations. This off-line reliability scoring was compared to both the on-line looking time to screen as well as to a second "corrected"

Category			Habituation Group		
Cow	Horse	Total	Habituators	Non-Habituators	Total
77.18	79.96	78.57	131.14	177.36	157.14

Table 5.1: Total Looking Time (in seconds) for each Category and Habituation group.

version of it calculated by subtracting 100ms from the on-line measurements[1]. Pearson r correlations revealed high agreement between on-line and off-line measurements of infant attention to the task, $r(60) = .96$, $p < .001$, for both measures.

5.2.8 Analysis

5.2.9 Training Block

In the training block, infants received several presentations of multiple word-objects pairings, in sub-blocks of four trials. Within each sub-block, two picture-word pairings - one from each adult category, "cow" and "horse" - were presented separately twice. The unaggregated data for the training block consisted of 288 trials across all infants. On average infants went through 4.50 training sub-blocks ($R = 2 - 5$, $SD = 0.97$). Seven infants reached the habituation criterion (Habituators) and 9 infants did not (Non-Habituators). Mean TLT in each sub-block, as well as for the two categories was calculated.

Table 5.1 shows the total looking time to the pictures, for each Category and Habituation Group separately and for the group totals.

Because infants were presented with multiple exemplars from each category during

[1]The corrected version was used in order to take into account the time lapsed between infants' reaction and the initiation of a motor movement on the experimenter's behalf. For a more detailed description see 2.6

Exemplar	Cow	Horse	Total
1	34	36	70
2	38	36	74
3	35	36	71
4	37	36	73
Total	144	144	288

Table 5.2: Number of presentations for each exemplar of Cow and Horse.

training, the number of times each picture was presented across infants was calculated, in order to ensure that infants received similar exposure to the whole training set. Table 5.2 shows the number of times each picture was presented across infants. Pictures appeared roughly equally across infants suggesting that infants have had equal exposure to them overall.

A univariate ANOVA with TLT as dependent variable and Item as a fixed factor tested whether infants accumulated similar looking times to the two categories. Finally looking times in the different sub-blocks of the training block were calculated and compared for the two Habituation groups using a univariate ANOVA with TLT as dependent variable and Sub-block and Habituation group as fixed factors.
Normality tests were conducted for the TLT measure, and all groups were found to be normally distributed.

5.2.10 Test Blocks

Infants were presented with three test trials, in two post-familiarisation blocks. The unaggregated data set for the test blocks consisted of 48 trials - 32 for test block A and 16 for test block B. Data were further aggregated by test block and familiarity (within- and out-of-category exemplars for Test A/ Average vs Modal stimuli for Test

B).

The two blocks were analysed separately. The following analyses were run for each block: A univariate ANOVA with PTL as a within-subjects factor and habituation group and familiarity as fixed factors was conducted in order to test whether the two habituation groups behaved similarly at test and whether infants discriminated between the familiar and new objects (Test Block A) or whether they formed one or two categories during familiarisation (Test Block B). A univariate ANCOVA with PTL as a within-subjects factor, familiarity as a fixed factor and Familiarity Score (FS) as a covariate was conducted for each test block separately in order to test whether parental feedback was predictive of infant performance in the laboratory. Data from two infants were excluded from this analysis, because parents failed to provide the CEQ.

Normality tests were conducted for all measures and outliers were excluded in order to achieve normality. Two trials were excluded for the PTL measure and three for the LLD measure in Test Block A and two trials from both measures in Test Block B.

5.3 Results

5.3.1 Training Phase

During the on-screen familiarisation block, infants were presented with multiple word-object pairings, using four exemplars from each adult category. Figure 5.2 shows the total looking time accumulated during training for each of the two categories, "cow" and "horse". It appears that the two categories accumulated similar attention. In order to test this observation, a Univariate ANOVA with TLT as a dependent variable and Category as a fixed factor was conducted. No main effect of Category

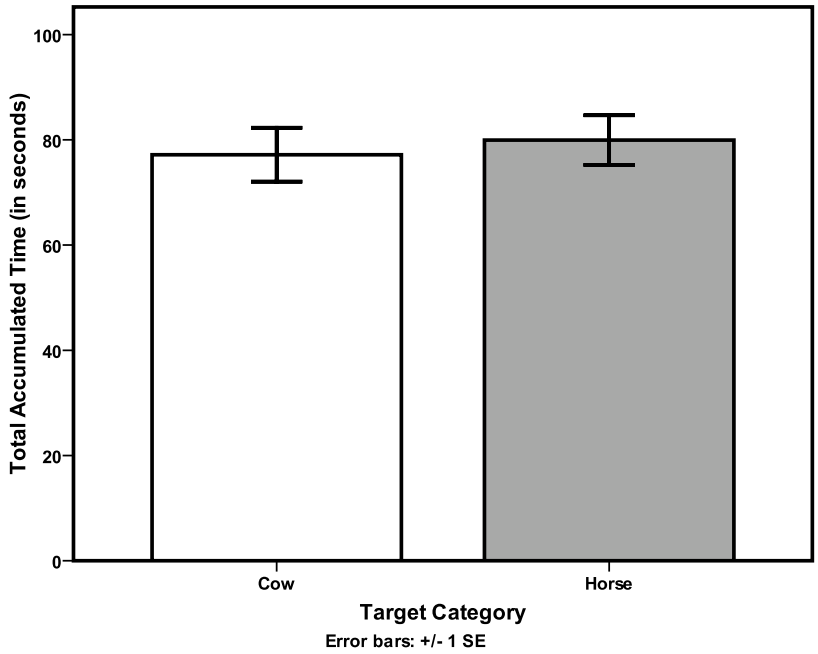


Figure 5.2: Mean Looking Time separately for each Target Category.

was found, $F(1, 32) = .16$, $p = .69$. Data were, therefore, collapsed across Category for subsequent analyses in the familiarisation block.

Figure 5.3 shows the TLT measure for the two Habituation groups separately. The two groups appear to have accumulated different looking times to the pictures during familiarisation: Habituators looked at the pictures far less than Non-Habituators. Figure 5.4 shows the different sub-blocks separately for the two Habituation Groups. As in the two previous experiments, infants in the Habituators group show a more sudden decrease in attention, whereas the Non-Habituators retain high levels of attention throughout the familiarisation block. Also as in Experiment Four, some infants in the Habituation group go through the whole block, but behave a lot differently from infants in the Non-Habituators group.

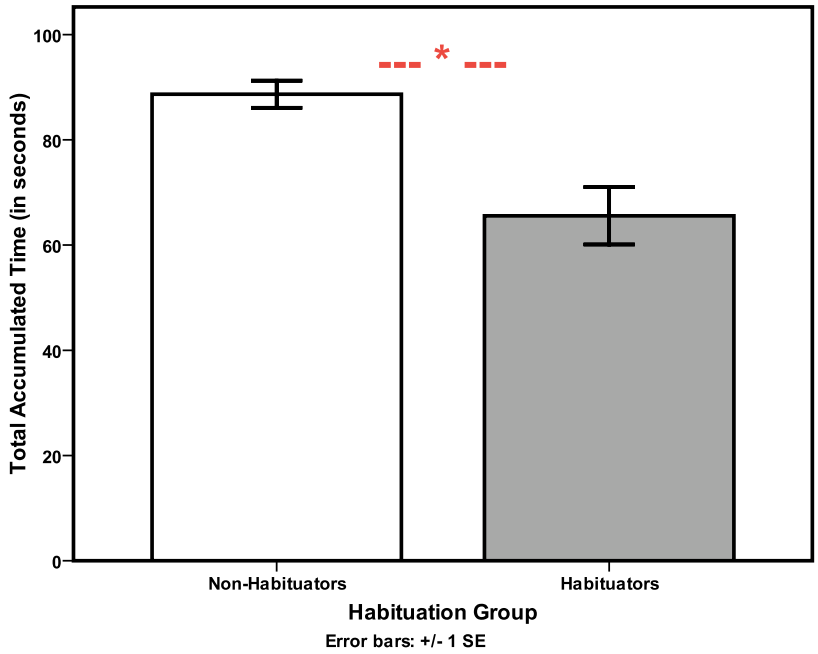


Figure 5.3: Mean Looking Time presented separately for each Habituation group in the two Conditions.

On the basis of Figure 5.4, a Univariate ANOVA with TLT as a dependent variable and Sub-block as a fixed factor was conducted for Habituators and Non-Habituators separately. For the Habituators, a highly significant main effect of Sub-block was found, $F(4, 52) = 8.51$, $p < .001$, ${\eta_p}^2 = .42$, with a decrease in attention as training progressed. Bonferroni post-hoc comparisons were conducted comparing all sub-blocks to each other. The following pairs were shown to differ from each other: Sub-block 1 from 3 ($p < .02$), 4 ($p < .01$) and 5 ($p < .001$), and Sub-blocks 2 and 5 ($p < .01$). For the Non-Habituators, there was no main effect of Sub-block, $F(4, 89) = 1.81$, $p = .13$, suggesting that infant attention during training did not fall significantly. Bonferroni post-hoc tests confirmed the absence of a main effect, by showing no differences between any of the sub-blocks.

This analysis suggests that, as in previous Experiments, during familiarisation

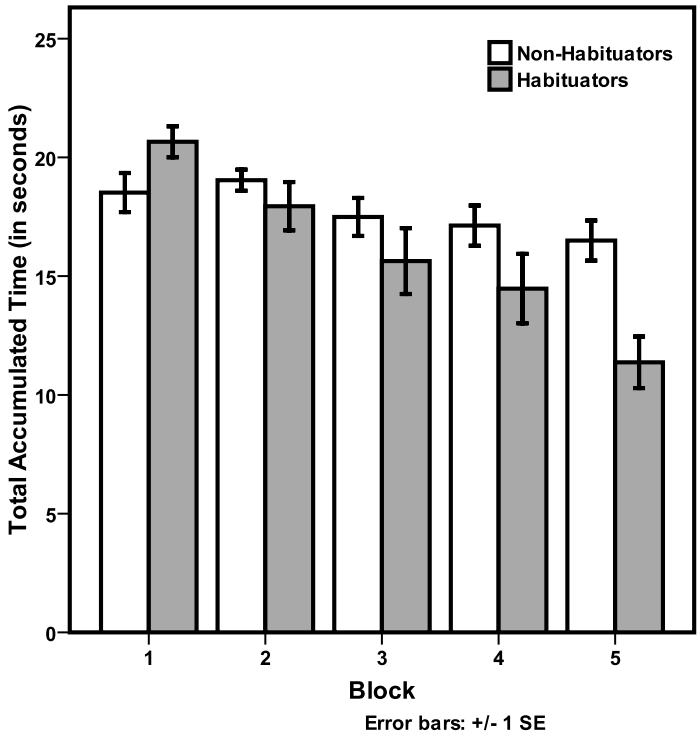


Figure 5.4: Mean Looking Time for each familiarisation sub-block presented separately for the two Habituation groups.

infants spontaneously divided into two groups, exhibiting different behaviours; this was taken into account during the analysis of the post-training test blocks.

5.3.2 Test Blocks

Infants were presented with two tests blocks in this Experiment, both after familiarisation. Test Block A consisted of two trials, while Test Block B consisted of a single trial. Figure 5.5 shows the LLD measure for the two test blocks separately. LLD measures novelty preference in this Experiment; it appears that in Test Block A infants exhibit a preference for novelty, especially infants who belong to the Non-Habituators group. In Test Block B, however, this pattern is inversed with infants exhibiting a familiarity preference, appearing to be driven by Habituators this time.

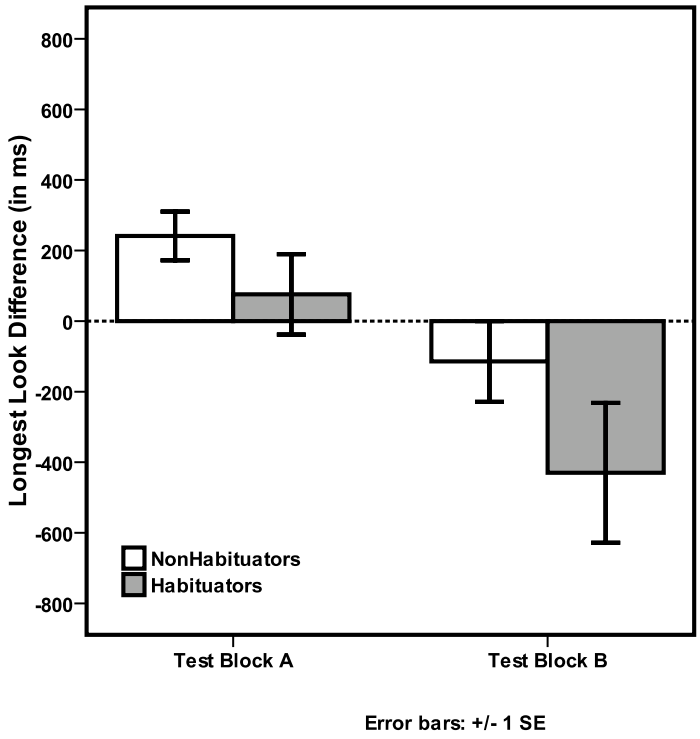


Figure 5.5: LLD for each test block and Habituation group separately. The LLD measure here is a measure of novelty preference calculated by subtracting the longest continuous look to the new within-category animals (cow and horse) from the longest continuous look to the new out-of-category animals (bear and tiger in Test Block A or Average stimulus in Test Block B).

The two test blocks were analysed separately in this Experiment, because they were different in their design and tested different assumptions. One-sample t-tests were run for both measures comparing novelty preference to chance in both blocks. A significant preference for the new out-of category exemplars was found in Test Block A for the LLD measure, $t(14) = 2.49$, $p < .03$, $M = 164.00$, $SD = 255.44$ and a significant preference for the Modal stimuli was found in Test Block B for the same measure, $t(15) = -2.26$, $p < .05$, $M = -252.00$, $SD = 446.16$. PTL showed no significant differences from chance in either test block, although a trend for a preference for the out-of-category exemplars seems to be present in Test Block A, Test Block A: $t(15) = 1.63$, $p = .13$, $M = .54$, $SD = .09$/ Test Block B: $t(15) = -.29$, $p = .78$, $M = .49$, $SD = .12$.

One-sample t-tests for the LLD measure were also run comparing attentional preferences in each habituation group separately, for the two Test Blocks. Non-habituators were found to exhibit a preference for the out-of-category exemplars in Test Block A, $t(7) = 3.49$, $p < .02$, $M = 241.25$, $SD = 195.39$, but no systematic behaviour was exhibited by Habituators, $t(6) = .67$, $p = .53$, $M = 75.71$, $SD = 301.10$. Habituators, on the other hand, exhibited a trend for a preference for the Modal stimuli in Test Block B, $t(6) = -2.17$, $p = .07$, $M = -430.00$, $SD = 524.60$, but behaviour which did not differ from chance was found for Non-Habituators, $t(8) = -1.00$, $p = .35$, $M = -144.44$, $SD = 343.26$.

This analysis suggests that infants in the two habituation groups behaved differently in this Experiment; non-Habituators were able to discriminate between familiar and novel animals, by looking significantly more at the out-of-category exemplars in Test Block A, and, therefore, showed evidence for their ability to create perceptual categories on the basis of the familiarisation set. During Test Block B, they failed to show any systematic behaviour altogether. Habituators, on the other hand and rather surprisingly, showed no evidence for discriminating between the within- and out-of-category exemplars in Test Block A. However, they showed systematic behaviour in Test Block B, by attending more at the Modal, atypical, objects during Test Block B, instead of the Average stimulus. This behaviour will be discussed further below.

Habituation group

In order to test whether the two habituation groups behaved differently in the two test blocks, a Univariate ANOVA with LLD/PTL as a dependent variable and Habituation group as a fixed factor was conducted for both measures in each test block separately.

No main effect of Habituation group was found for either measure in neither test block, Test Block A: PTL: $F(1, 15) = 1.55$, $p = .22$/ LLD: $F(1, 15) = 11.28$, $p = .17$; Test Block B: PTL: $F(1, 15) = .49$, $p = .50$/ LLD: $F(1, 15) = 2.12$, $p = .92$.

Although Habituators and Non-Habituators were shown to have different behaviour in the previous analysis, this analysis suggests that the two groups did not show different behaviour. However, it should be emphasised that the numbers of infants in each group might not be enough to give power to this analysis, but the effects from individual groups arise in the t-tests. Because of the small sample numbers any effects should, therefore, be interpreted with caution.

Parental Feedback - Familiarity Effects

On the basis of parental feedback a familiarity score (FS) from the CEQ (see Section 3.2.6) was calculated for each infant. Although no evidence for familiarity effects using the CEQ were found in previous Experiments, it could be that parental feedback on infants' experience with different objects is more relevant when testing for the creation of perceptual categories, instead of word learning.

In the CEQ parents reported on their infant's exposure to each of the categories, across three dimensions: frequency of exposure, number of exemplars encountered, exposure to the name of the object. For both categories, infants accumulated an average score of 4.79 ($R = 14.00 - 0.00$, $SD = 4.81$) - for "cow" 2.43 ($R = 8.00 - 0.00$, $SD = 2.57$) and for "horse" 2.36 ($R = 14.00 - 0.00$, $SD = 2.42$).

Figure 5.6 shows the relationship between the LLD measure and the CEQ score for both categories, separately for each Test Block. In order to investigate whether

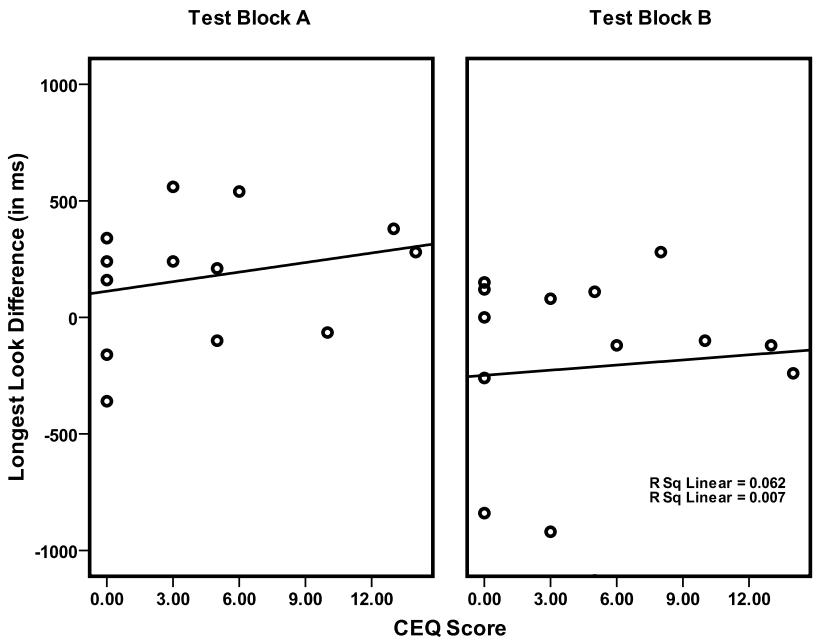


Figure 5.6: Relationship between the LLD measure and CEQ score, in Test Blocks A and B.

CEQ can be predictive of infant categorisation abilities, a Univariate ANCOVA was conducted with LLD as a dependent variable and CEQ Score as a covariate separately for the two test blocks. No effect from CEQ on infant preferences was found for either test block, Test A: $F(1, 13) = .73$, $p = .41$; Test B: $F(1, 13) = .08$, $p = .78$.

Also a Pearson r correlation was conducted for each block, in order to test directly whether LLD and CEQ score shared any meaningful relationship. No significant correlation was found in either block, Test Block A: $r(14) = .25$, $p = .41$, Test Block B: $r(14) = .08$, $p = .78$.

This analysis suggests that parental feedback on infants' prior experience with the test items does not have any meaningful relationship with category creation in this Experiment.

5.4 Discussion

Experiment Five was a categorisation study aiming to test whether infants at the end of their first year are able to create perceptual categories on the basis of the training stimuli of Experiment Four. In Test Block A, evidence for novelty preference was found indicating that infants discriminated between the familiar animals, "cow" and "horse", and the novel ones, "bear" and "tiger", and preferred to look at the latter after being familiarised to the former. This behaviour has been documented with infants before and Experiment Five serves as a replication of previous findings and a control for whether infants can successfully categorise the specific objects used in these studies (Younger, 2003; Plunkett et al., 2008).

Experiment Five also aimed to test whether during familiarisation infants created a single category including both "cows and horses" or two separate categories, a "cow" and a "horse" category. In order to test this, infants were presented with objects which contained the Modal (atypical objects) and Average ("cow - horse" made-up animal) values from the categories in the familiarisation set. In Test Block B evidence for a familiarity preference was found, in the form of increased looking at the Modal stimuli, the atypical cows and horses. One would expect, however, that infants would show a novelty preference in this test block as well, by attending to the Average stimulus. Two explanations are at hand: either infants failed to create two separate categories and instead included both cows and horses in a single "cow - horse" category, or the way familiarity and novelty are defined in this test block might be problematic. That is, infants might not perceive atypical objects as part of their perceptual categories and, therefore, they attend to those during test still exhibiting an out-of-category preference. It could also be that the "cow - horse" made-up animal is not perceived as realistic enough from infants, causing them to become confused and orient to the other picture more. Finally, another possibility is that infants prefer

to attend to familiarity instead of novelty in this final test trial. This possibility will be discussed below in relation to the behaviour of the two habituation groups.

The two habituation groups in this Experiment appear to behave differently in the two Test Blocks: in Test Block A non-Habituators are those showing a novelty preference, while Habituators behave at chance. In Test Block B, Habituators prefer to look at the Modal stimuli more, whereas Non-Habituators behave at chance. Although no interaction was found between Test Block and Habituation group - possibly due to the small numbers of infants in each group, there was still evidence that those groups differed in their behaviour.

An obvious first question is why in Test Block A Habituators seem to behave at chance, failing to categorise, whereas Non-Habituators seem to have been successful in creating visual categories during familiarisation. And the obvious next question is why this pattern seems to be inversed in the second Test Block? One possibility is that Habituators are just not participating in this Experiment and should the numbers be higher the result would be different. However, there is another possibility for this behaviour: it could be that this task - word learning and visual categorisation - is too much for infants at the end of their first year. Infants who have begun to discover that there is something special between words and objects attend to the word learning task, and exhibit evidence of word comprehension in Experiment Four - these infants belong to the Habituators group. Some other infants are not yet highly interested in the relationship between words and objects and find the relationship between perceptually similar objects more appealing - these are the infants who fail to exhibit naming in Experiment Four, the Non-Habituators. Infants who are successful in word learning, fail in visual categorisation, and infants who are not successful in learning the labels exhibit evidence for perceptual categorisation.

If then this is the case, what happens in Test Block B? Habituators exhibit a familiarity preference attending more at the stimulus which bears some relationship with the ones they have been looking at during familiarisation. Non-Habituators, who are already at the stage of a novelty preference, might just be getting confused by the non-realistic Average stimulus and they fail to show any systematic preferences.

What are the implications of these interpretations for the theories aiming to account for label generalisation in infancy? If infants who are involved in word learning cannot visually categorise the stimuli simultaneously, then it appears that the taxonomic constraint cannot be in place so early, due to processing limitations. Exemplar-based accounts seem to be able to better account for this behaviour by emphasising on the need for more experience with language before infants can understand how words correspond to categories of objects and not just to individual objects.

Chapter 6

Adult Rating Studies

The objects used in Experiments One to Five were rated by adult native speakers of British English for their typicality and similarity. Typicality ratings were considered as an index of the familiarity of the objects and, as a result, of the cognitive load that the task would impose on infants. In previous research on categories and category formation (see Section 1.1.1), typical exemplars were found to be more frequent in the environment, those named more often and also those learned easier and quicker by both adults and infants (Rosch & Mervis, 1975; Rosch et al., 1976; Meints et al., 1999; Meints, Plunkett, Harris, & Dimmock, 2002). Also they have been shown to be predictive of performance in different tasks, such as categorisation and retrieval of instances of a concept from memory (Rosch & Mervis, 1975; Rosch, 1998; Smith, Sera, & Gattuso, 1988).

Visual similarity ratings were considered necessary for Experiments Three and Four, because of the need to choose the training and testing exemplars in a strictly-controlled manner. Two considerations led to this:

- Exemplar-based accounts of category formation assume that generalisation of a label might depend on the similarity distance between items both within- and across-categories and it is, therefore, important to have a measurement of the

perceived distance of the training and testing stimuli.

- The similarity distances both within- and across- categories could also be an index of the cognitive load imposed on infants by the task of categorising and generalising labels to new objects. That is, categories that are more diverse might facilitate generalisation of labels to new (typical or atypical) items, whereas categories with high similarity among their members might make it difficult for atypical items to be incorporated in their extension. Similarly, if the distance between two different categories is large, generalisation to new exemplars might be facilitated, whereas close similarity between two categories might make the task of discovering each category's extension difficult for infants (see for example, (Quinn, Eimas, & Rosenkrantz, 1993; French, Mareschal, Mermillod, & Quinn, 2004).

In Experiments One and Two, six categories were used; three object ("cup", "phone" and "spoon") and three animal ("bear", "cat" and "duck") categories. The typicality of these items was rated in Typicality Rating Study One. In Experiments Three and Four two animal categories were used, "cow" and "horse". For each category seven exemplars were rated by adult native speakers of British English for their perceived typicality and similarity in Typicality Rating Study Two and Visual Similarity Rating Study, respectively.

6.1 Experiments One and Two: Typicality Ratings

The visual stimuli used in infant Experiments One and Two were rated for typicality in this study.

6.1.1 Participants

Eighteen (13 female) undergraduate and graduate students (mean age: 25, range: 18-39) at the University of Oxford took part in the study. Four participants were excluded; one for equipment failure, one for reversing the typicality scale, one for experimenter error and one for being bilingual. All of the participants were native speakers of British English and none of them had familiarity to any other language beyond the point of second language learning, i.e. none of them was bilingual.

6.1.2 Stimuli

In total 36 pictures were included in the study belonging to six categories: three animal categories ("bear", "cat" and "duck") and three object categories ("cup", "phone" and "spoon"). Six exemplars were presented from each category. Pictures were pre-assessed for typicality by the experimenter and divided into three typicality levels: typical, borderline and atypical. Only pictures from the first level (*typical*) were used in the infant study. Items from the other two levels (*borderline* and *atypical*) were included in order to ensure that participants judged the typicality of the pictures as part of the whole range of a given linguistic category: if only typical items were presented, participants could have been misled to rating them as more atypical because the only available comparison would have been other typical exemplars.

Pictures were presented on a computer monitor using Presentation software. Pictures were 24-bit digitised photographs in .bmp format (450 x 450 pixels). The pictures were presented against a 10% grey background. The names of the objects were presented above the centre of the pictures, in Arial, bold, sized 14 letters. Both the picture and the label were simultaneously presented against a black background, vertically centred on the screen.

6.1.3 Procedure

Participants were given written instructions immediately before the study and informed consent was obtained. The instructions were as follows:

Instructions

You are asked to judge the typicality of the pictures in relation to the words you will be presented with; that is to evaluate how good an example of a word each picture is. In order to clarify this, think of a "true red" as compared with an "orangish red" or a "purple red"; or that a robin and a crow are good examples of birds whereas an ostrich or a penguin are not necessarily so. Make your decisions as quickly as possible by pressing the buttons from 1 (very typical) to 7 (very atypical) on the keyboard. Once a rating is made, you will not be able to go back to previous pictures and that the displayed picture will automatically change only when you have made a decision for it.

Participants were asked to read the instructions carefully, make sure they understood them and ask if something was not clear. They were then reminded that the scale would be visible on the screen throughout the experiment, and were asked to begin whenever they felt ready.

Before the study, they were asked to fill in a questionnaire, indicating their age, level of education and familiarity with languages other than English. Their familiarity to other languages was assessed through their answers to the following three questions:

1. whether a member of their family was a native speaker of a language other than English

2. which languages they speak and at what proficiency level and

3. whether they had spent more than a month in a foreign country and for what reason

The Experiment started as soon as the participant pressed "enter". The picture along with the label remained on the screen for as long as the participant needed for making their decision. As soon as the participant pressed one of the buttons from 1-7 on the keyboard the next picture was automatically presented. The order of presentation of the 36 pictures was randomised and each picture was presented only once.

6.1.4 Measures

Two measures were used to assess the typicality of the items:

1. Typicality Ratings (TR): participants rated the pictures on a scale from 1 (very typical) to 7 (very atypical).

2. Reaction Times (RT): the time lapsed between the appearance of the picture on the screen and the rating of the participant (by pressing a button on the keyboard). Reaction Times were used in order to test whether the two measures agreed and whether there was a meaningful relationship between typicality and the speed with which participants rated the items.

6.1.5 Analysis

The final dataset consisted of 859 unaggregated points. Mean TR and mean RT were calculated for each item. Sapiro-Wilk normality tests were run for each picture and each typicality level separately for both measures. Both measures were normally

distributed for all groups after the exclusion of individual trials which behaved as outliers - 4 trials for the TR measure and two for the RT measure.

Univariate ANOVAs were run for each different measure with Typicality level (Typical, Atypical and Borderline) and Animacy (Animals and Objects) as within-subjects factors. Bonferroni post-hoc tests were run in order to test for differences between the pre-defined typicality groups.

In order to check the validity of the pre-defined Typicality levels, a cluster analysis was run for the TR measure, in which the data were treated as if there were no initial expectations for the behaviour of individual items. A K-means cluster analysis was run for the TR measure and the output was tested against the pre-defined Typicality levels using a Pearson r correlation.

6.1.6 Results

Pre-defined Typicality Levels

Table 6.1 shows the TR and RT for each pre-defined Typicality level separately. Typical objects receive the lower scores on the typicality scale (more typical), whereas atypical objects receive the highest scores (more atypical). Borderline objects have a mean between the two other groups. Also typical objects seem to be rated much quicker than both borderline and atypical objects. Figure 6.1 shows the mean TR for animals and objects separately; animals seem to be considered more typical overall than objects.

In order to test these observations a univariate ANOVA was run with Typicality level (Typical, Atypical and Borderline) and Animacy (Animals and Objects) as fixed factors for both measures. A main effect of Typicality level was found for both

Typicality Level	Mean TR (SE)	Mean RT (SE)
Typical	2.03 (0.21)	2.18 (0.08)
Borderline	3.59 (0.25)	2.51 (0.10)
Atypical	4.51 (0.25)	2.56 (0.10)

Table 6.1: Mean Typicality Ratings and Reaction Times for the different pre-defined typicality groups (aggregated data from the six categories).

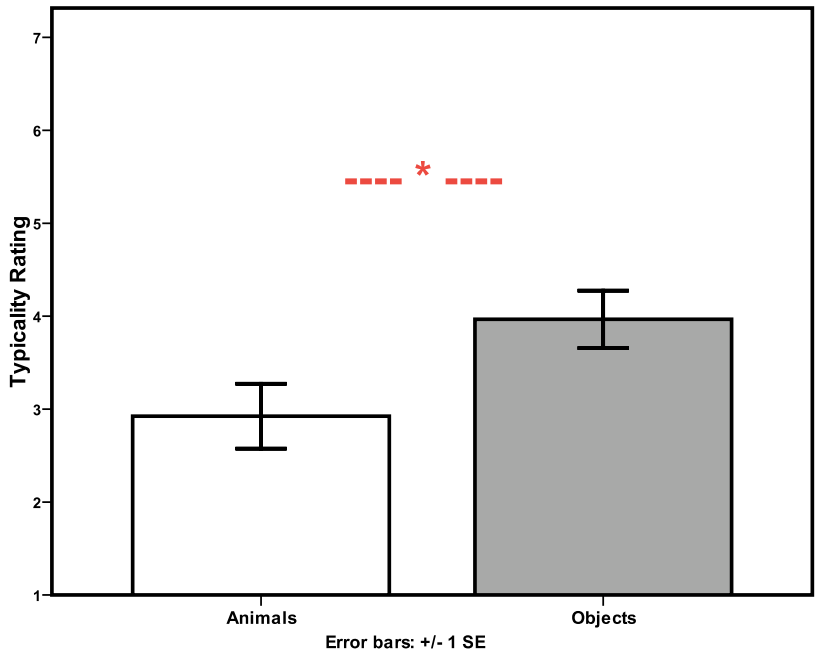


Figure 6.1: Mean Typicality Ratings for the animal and object categories separately.

measures, TR: $F(2, 36) = 22.53$, $p < .001$, ${\eta_p}^2 = .68$; RT: $F(2, 36) = 4.70$, $p < .02$, ${\eta_p}^2 = .24$, as well as a main effect of Animacy, TR: $F(1, 36) = 9.77$, $p < .01$, ${\eta_p}^2 = .32$; RT: $F(1, 36) = 10.24$, $p < .01$, ${\eta_p}^2 = .25$. No interaction between Typicality level and Animacy was found for either measure, TR: $F(2, 36) = .35$, $p = .78$; RT: $F(2, 36) = .93$, $p = .41$.

On the basis of the main effect of Typicality level, Bonferroni tests were run comparing the different typicality levels for both measures. For the TR measure, all groups were shown to significantly differ from each other, Typical-Borderline: $p < .01$, Typical-Atypical: $p < .001$, Borderline-Atypical: $p < 05$. For the RT measure, only typical and atypical differed, $p < .05$, but no other differences were found, Typical-Borderline: $p = .13$, Borderline-Atypical: $p = 1.00$.

This analysis suggests that the objects used in the infant studies were considered by adult native speakers of British English as typical exemplars from the corresponding categories. On the basis of the main effect of Animacy, data were analysed separately for animals and objects in subsequent analyses.

Cluster Analysis

In the previous analysis a main effect of Typicality level was found using the pre-defined typicality groups. However, a cluster analysis was also conducted in order to confirm that individual typical exemplars (used in the infant studies) belonged to the group which was rated as most typical. A K-means cluster analysis was run with three clusters, for each animacy group separately. The clustering was shown to be significant in a univariate ANOVA for both animals and objects; Animals: $F(2, 15) = 103.63$, $p < .001$; Objects: $F(2, 15) = 66.23$, $p < .001$.

Table 6.2 shows the means for each cluster, separately for animals and objects, while Figure 6.2 shows the clusters produced by the K-means analysis in relation to the pre-defined typicality groups. A Pearson correlation analysis tested the relationship

Animals			Objects		
Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3
1.56	3.14	4.90	2.62	4.29	5.75

Table 6.2: Final Centres for the three-cluster grouping produced by the K-means analysis. Animals and Objects are presented separately.

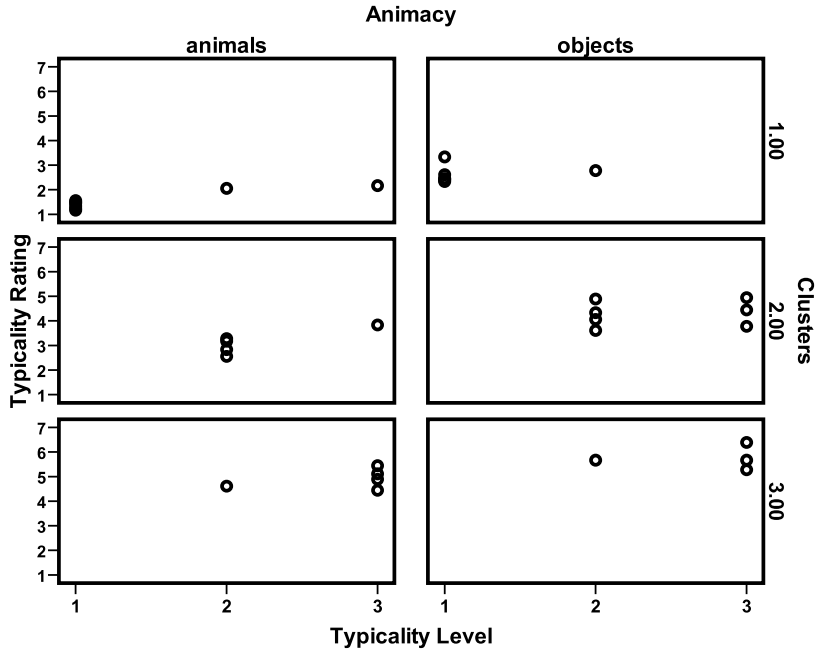


Figure 6.2: Scatter-plot showing the relationship between the pre-defined typicality groups and the clusters produced by the K-means cluster analysis. The Y-axis shows the mean typicality rating for each of the pictures. Animals and Objects are presented separately. All of the items used in the infant studies belong to the first cluster with the lowest mean TR, reported to be the most typical exemplars from the corresponding categories.

between the pre-defined typicality groups and the outcome of the K-means analysis. This revealed a high correlation between the two, $r(36) = .77$, $p < .01$.

This analysis confirms that the pre-defined typicality groups are a legitimate division of the objects.

6.1.7 Discussion

Typicality Study One has shown that adult native speakers of British English consider categories as having an internal structure; some exemplars are perceived as more central, more typical, while others fall in the periphery of the category and are considered as more atypical. As expected from previous studies on the structure of categories (Rosch & Mervis, 1975), this is true for both object categories, as well as for natural kinds, in this case animal categories. Although members of the animal categories are considered as more typical than objects overall, the same internal structure seems to exist in both kinds of categories.

As far as the specific objects used in the infants studies are concerned, they were shown to be considered as the most typical exemplars by adult speakers of British English. The K-means cluster analysis confirmed that the pre-defined Typicality Levels statistically correlated with the way participants rated the objects.

Finally, in terms of methodology, the two measures used, Typicality Ratings (TR) and Reaction Times (RT), provided similar results. The RT measure makes plausible the hypothesis that there might be an underlying psychological basis to the internal structure of categories, and the advantage of some exemplars over others in category membership. An equivalent of this from the developmental side can be found in Meints et al., 1999, where infants are better at recognizing typical exemplars of a category as referents of a known label at 12 months, but at 18 months they are good at identifying both typical and atypical exemplars. This suggests that the process of creating an internal structure for (perceptual or linguistic) categories might start very early in development.

6.2 Experiments Three and Four: Typicality and Similarity Ratings

6.2.1 Typicality Rating Study

The visual stimuli used in infant Experiments Three and Four were rated for typicality in this study.

Participants

20 (10 female) native speakers of British English (mean age: 19, range: 18-25) took part in the study. Participants were recruited in the Department of Experimental Psychology, University of Oxford, from undergraduate and graduate students. No participants were excluded from the study, but individual trials were excluded if found to be outliers during the analysis. All participants were monolingual speakers of British English and none of them had familiarity with any language other than English beyond the point of second language learning.

Stimuli

Two animal categories were tested: "cow" and "horse". Seven exemplars from each category were included in the test set. The visual stimuli were photographs of toy animals produced by the German brand "Schleich", and, therefore, the choice of the exemplars was limited to those objects which could be found in the market. The objects were pre-assessed by the experimenter and divided into two typicality levels: typical (N = 5 per category) and atypical (N = 2 per category). This resulted in a total of 10 typical and 4 atypical exemplars. To avoid any biases due to the difference in the number of typical and atypical items, the stimuli were divided in ten groups

which consisted of only two typical and two atypical items from each category. Each typical exemplar was combined once with all other typical exemplars from the same category to yield the ten groups. Thus each typical exemplar was included in four groups, whereas atypical exemplars were included in all ten groups. Participants were randomly assigned to one of the ten groups.

The visual stimuli were photographs of the toy animals, presented on a white background. Objects in the pictures were canonically oriented, that is in side view. Pictures were processed using Adobe Photoshop CS2, in order to remove any background. Digitised photographs in .jpeg format were used. The objects, their labels ("cow" or "horse") and a typicality scale (1-7) were simultaneously presented against a black background. The position of the objects, the labels and the scale remained constant across exemplars. Above the object the label was presented in white Arial fond, size 26. The Typicality Scale (1-7) was presented just above the label in red Garamond fond, size 24 for letters and 28 for numbers, and remained on the screen for as long as the pictures and labels were presented (for an example see Figure 6.3). The study was presented on a 30 cm (height) x 38.5 cm (width) computer monitor using Presentation software, and participants sat approximately one metre from the screen. Pictures measured 16 cm (height) x 15.5 cm (width) on the screen.

Procedure

The procedure and instructions were identical to Typicality Rating Study One, except from one difference: the typicality scale was inversed. The atypical end of the scale was now rated as 1 and the typical end as 7, in order to give a more intuitive rating to items. This is also the way typicality scales in Rosch (1975, 1976) and Meints (1999) were defined. Each participant rated the set of pictures twice; the whole set

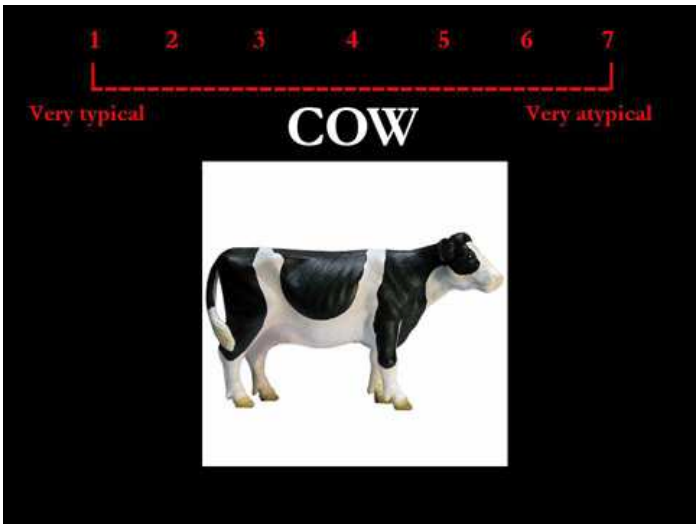


Figure 6.3: An example of a trial from the Typicality Rating Study.

was rated once before the pictures were presented again. This was done in order to test whether participants gave different rating having been exposed to exemplars of varying typicality. The order of presentation of the eight items was randomised and there was a different randomised order for the two blocks.

Measures

In order to investigate the typicality of the objects used in the study, two measures were used:

1. Typicality Ratings (TR): participants were asked to rate the objects on a scale from 1 - very atypical to 7 - very typical.

2. Reaction Times (RT)

Also, as in Typicality Rating Study One and on the basis of the pre-defined typicality groups and those provided by the Cluster analysis, objects were divided in

a Typicality Group (Typical or Atypical) and Cluster Group, respectively.

Analysis

The unaggregated set resulted in 316 trials, after the exclusion of 4 trials which were found to be outliers. Mean TR and RT were calculated for each individual item.

A one-way ANOVA with Presentation Order as a Grouping factor was run for TR and RT in order to test whether first-round ratings differed from second-round ratings.

A univariate ANOVA with Typicality Group (typical and atypical) and Category ("cow" and "horse") as fixed factors was run in order to test whether pre-defined typicality groups were rated differently and whether the two categories, "cow" and "horse", promoted similar behaviour.

Finally a cluster analysis was run for the TR measure in order to investigate whether the objects naturally fell into different groups on the basis of adult ratings. A K-means cluster analysis was run, because it both allows to test for different numbers of clusters and also it assigns each item to the cluster whose center is nearest, giving a rough approximation of how cognitive processing of natural categories might take place. Two and three clusters were used in the analysis, because there were two already defined typicality groups and because there could be an additional level of peripheral items between the two extremes. Finally Pearson r correlations were conducted between the Typicality and Cluster Groups in order to test whether these groups were related in any meaningful manner.

Normality tests were run for both measures and normality was confirmed.

Results

Presentation Order. A preliminary one-way ANOVA with Presentation Order as a Grouping factor was conducted in order to test whether first- and second-round ratings differed from each other. Presentation order was not shown to affect TR, $F(1, 26) = .06$, $p = .81$, and, therefore, data were collapsed across presentation order for subsequent analyses. On the other hand, Presentation Order was shown to affect the RT measure, $F(1, 26) = 4.38$, $p < .05$, with the mean time needed to make the ratings the second round significantly lower than the first, M1: 2427.38 (SD: 498.49), M2: 2106.49 (SD: 283.53). Only first-round reaction times were, therefore, used in subsequent analyses.

Relationship between Typicality Ratings and Reaction Times. Figure 6.4 shows the relationship between Typicality Ratings and Reaction Times. It appears that the two measures are inversely related: the more typical an object is considered, the less time participants need for making a decision about its typicality. In order to investigate the relationship between Typicality Ratings and Reaction Times a Pearson r correlation was run. A significant correlation was found between TR and RT, $r(28) = -.43$, $p < .05$.

Pre-defined Typicality Levels. Table 6.3 shows the means for all the objects, separately for the two categories. Numbers 1-5 are items which were pre-defined as typical, whereas 6-7 are items which were pre-defined as atypical. Items were rated on a scale from 1-very atypical to 7-very typical. Therefore, items with high scores indicate higher degree of typicality. The lower a mean is the more atypical an item was considered by adult native speakers of British English. Pre-defined atypical

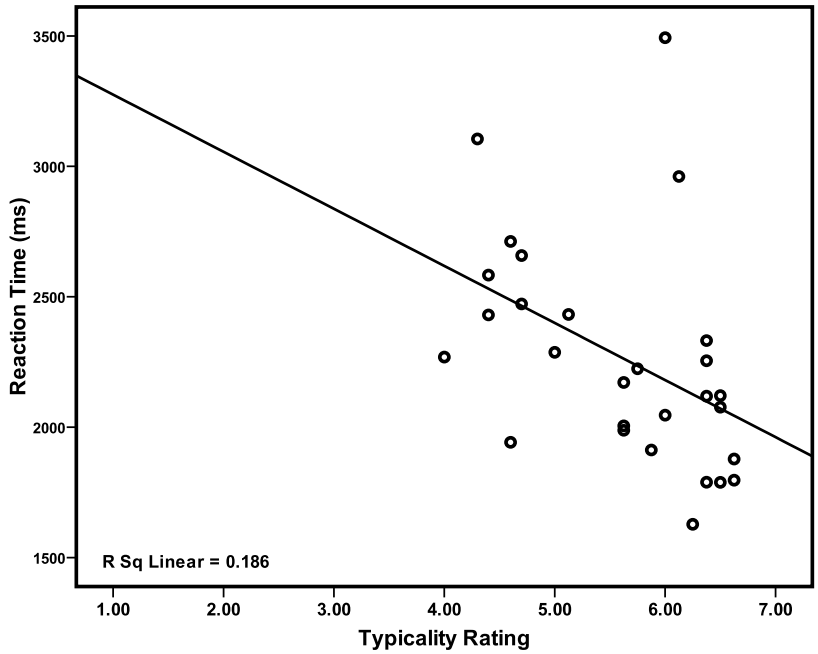


Figure 6.4: Relationship between Typicality Ratings and Reaction Times. The higher the typicality, the shorter times are needed for participants to rate the pictures.

Item	Cow	Horse
1	6.71 (0.61)	6.44 (0.73)
2	6.25 (0.77)	6.50 (0.73)
3	5.93 (1.16)	5.75 (1.06)
4	6.44 (0.81)	6.73 (0.46)
5	5.31 (1.82)	5.44 (1.63)
6	4.55 (1.50)	4.50 (1.80)
7	4.65 (1.70)	4.15 (1.66)

Table 6.3: Mean Typicality Ratings for the seven exemplars in the two categories.

items have all received mean scores between 4.15 and 4.65, whereas typical items have received scores between 5.31 and 6.73. It is noteworthy that as typicality is considered higher the standard deviation is reduced as well. Table 6.3 shows the TR measure for the two categories; it appears that the two categories, "cow" and "horse", promoted similar behaviour to typical and atypical items.

In order to test these observations, a univariate ANOVA was run for both measures

with Typicality group (typical vs. atypical) and Category as fixed factors. A main effect of Typicality group was found, TR: $F(1, 80) = 54.04$, $p < .001$, ${\eta_p}^2 = .35$/ RT: $F(1, 80) = 4.67$, $p < .05$, ${\eta_p}^2 = .06$, with typical items receiving significantly higher and quicker ratings than atypicals, Typical Items, TR: $M = 6.0$, $SE = .18$/ RT: $M = 2170.56$, $SE = 116.44$; Atypical Items, TR: $M = 4.45$, $SE = .18$/ RT: $M = 2526.46$, $SE = 116.44$. No main effect of Category was found for either measure, TR: $F(1, 80) = .08$, $p = .80$/ RT: $F(1, 80) = 1.03$, $p = .31$; "cow": TR: $M = 5.31$, $SE = .18$/ RT: $M = 2431.92$, $SE = 116.44$; "horse": TR: $M = 5.24$, $SE = .18$/ RT: $M = 2265.10$, $SE = 116.44$, nor an interaction between Typicality group and Category, TR: $F(1, 80) = 1.30$, $p = .45$/ RT: $F(1, 80) = .15$, $p = .70$.

Cluster Analysis. A cluster analysis was run in order to test whether the pre-defined grouping on the basis of typicality was a valid division of the objects into typical and atypical exemplars. A K-means cluster analysis was run for the TR measure with two and three clusters. In both cases the clusters were shown to be significantly different in a one-way ANOVA, Two Clusters: $F(1, 12) = 48.3$, $p < .001$; Three Clusters: $F(2, 11) = 100.28$, $p < .001$. In the two-cluster grouping all atypical items belonged to one group, with the lowest mean, along with one typical item from each category. In the three-cluster grouping all atypical items belong to one group, with the lowest mean, then two typical items from each category belong to the group with the second lowest mean and the rest of the items (all typical - three from each category) belonged to the group with the highest mean (see Table 6.4. This behaviour indicates that there is a typicality grading within both categories, which appears to be similar and should be taken into account when choosing items for the infant study and when interpreting the results.

Item	Clusters (Mean TR)		
	A (4.46)	B (5.61)	C (6.51)
Cow	6 (AT)	3 (T)	1 (T)
	7 (AT)	5 (T)	2 (T)
			4 (T)
Horse	6 (AT)	3 (T)	1 (T)
	7 (AT)	5 (T)	2 (T)
			4 (T)

Table 6.4: Three Clusters provided from the K-means analysis, separately for each category. The brackets next to each item denote the pre-defined typicality group: AT for atypical and T for typical.

A correlational analysis was run in order to test whether the items in each cluster coincide with those in the pre-defined typicality groups; Pearson r correlations were run for Cluster and Typicality Group, for both clusters, and were significant at the .01 level (Two Clusters: $r(14) = -.83$ $p < .01$; Three Clusters: $r(14) = -.87$ $p < .001$).

Discussion

The adult Typicality Rating study Two showed that the objects naturally fell into different groups on the basis of their perceived typicality, and that the pre-defined typicality levels agreed with the participants ratings. Some objects were considered to be more typical than others, and this could affect infants' ability to categorise them. Atypical exemplars, that is objects 6 and 7 from each category, were, therefore, not included in the infant studies.

As in Typicality Rating study One, the two measures, TR and RT, were shown to provide similar results; participants are quicker in rating typical objects, and slower in rating atypical ones.

For Experiment Three, two of the three most typical instances from each category, that is 1, 2 and 4, were used. For Experiment Four all five objects belonging to the clusters with the highest typicality were used. However, before making the final

decision on which objects to be used during pretest and test in both Experiments, a visual similarity study was conducted, in order to have an indication of the perceived distance between objects within and across categories.

6.2.2 Similarity Rating Study

The visual stimuli used in Experiments Three and Four were rated for their perceived similarity in this study.

Participants

20 (16 female) native speakers of British English (mean age: 19.15, range: 18-31) took part in the study. Participants were recruited in the Department of Experimental Psychology, University of Oxford, from undergraduate students and parents visiting the Oxford Babylab. Three participants were excluded from the study because they were bilinguals. All participants in the final set were monolingual speakers of British English and none of them had familiarity with other languages beyond the level of second language learning.

Stimuli

Two animal categories, "cow" and "horse", which were used in the previous study, were tested in this study. Seven exemplars (of varying typicality) from each category were included in the test set. Participants were asked to rate the similarity of each object to the rest of the objects from the same category as well as from the other category. Combinations within and across categories were exhaustive and yielded 42 within-category pairs and 49 across-category pairs.

The visual stimuli were photographs of toy animals produced by the German brand "Sleich", presented on a white background. All objects were presented in side view. Pictures were processed using the Adobe Photoshop CS2, in order to remove any background. Digitised photographs in .jpeg format were used. The photographs of the objects were those used in the previous study. However, in this study pairs of pictures instead of individual pictures were presented in each trial. Each pair

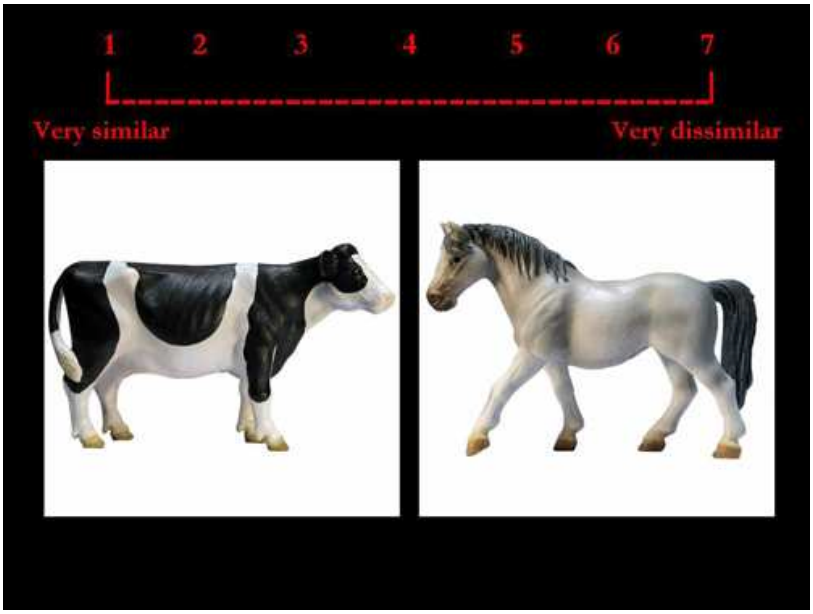


Figure 6.5: An example of a trial in the Similarity Study.

of objects and a Similarity Scale, from 1 (very dissimilar) to 7 (very similar), were simultaneously presented against a black background. The position of the objects and the scale remained constant across exemplars. Above the object the Typicality Scale (1-7) was presented in red Garamond fond, size 24 for letters and 28 for numbers, and remained on the screen for as long as the pictures were presented (for an example see Figure 6.5). The study was presented on a 30 cm (height) x 38.5 cm (width) computer monitor using Presentation software, and participants sat approximately one metre from the screen. Pictures measured 11 cm (height) x 10.5 cm (width) on the screen.

Procedure

Participants were given written instructions immediately before the study and informed consent was obtained. The instructions were as follows:

Instructions. *In this study, you will be presented with pairs of pictures; all you are asked to do is to tell us how similar, in your opinion, the two pictures are. Pictures in a single pair could be from the same category, or from a different category (for example, two cats or a cat and a dog). Please give us your ratings irrespective of the category they belong to, depending only on how much you think the pictures look alike. To make this clearer, imagine that you are presented with pairs of cats and dogs: a Golden Retriever and a German Shepherd can look very similar, but both look very different to a European cat. However, both the Retriever and the German Shepherd look very different to a Chihuahua - perhaps even more different than the cat does! This is the kind of judgment you are being asked to make.*
You need to make your decisions as quickly as possible by pressing the buttons from 1 (very dissimilar items) to 7 (very similar items) on the keyboard (please use the number buttons on the row above the letters, not the keypad on the right-hand side). Each pair will be presented individually. The pair will automatically disappear as soon as you make your decision, and the next pair will appear on the screen (please be patient if this takes some time and do not try to rate any pair twice); this will continue until all pairs have been rated. Please note that you will not be able to go back to previous pairs, so you will not have the chance to change the ratings you have made. Finally, when you have read and understood the instructions, you may begin the experiment by pressing "enter".

Participants were asked to read the instructions carefully and ask the experimenter for any clarifications. They were then reminded that the Similarity Scale would be visible on the screen throughout the experiment, and were asked to begin whenever they were ready.

Before the study participants were asked to fill in a questionnaire, identical to the

one in Typicality Rating Studies One and Two.

Each participant rated all pairs once. The pairs were included in two groups depending on the side of each item, so that if item A of pair A appeared on the right side in group 1, it appeared on the left in group 2; the opposite holds for item B of pair A, and so on and so forth. Pairs were pseudo-randomly assigned in each group; each group contained 91 pairs to be rated. The order of presentation of the 91 pairs was randomised and participants were randomly assigned to either Group 1 or Group 2.

Measures

Two measures were used to assess the visual similarity of the pairs:

1. Similarity Ratings (SR): Participants were asked to rate the similarity of each picture pair on a scale from 1 - very dissimilar to 7 - very similar

2. Reaction Times (RT): the lapsed between the presentation of the picture pair and the rating through the button press.

An additional division of the testing items into two groups was done:

1. Typicality group (TG): depending on the combination of objects in each pair (both within- and across-categories), three typicality groups were created - typical-typical, typical-atypical or atypical-atypical. The pre-defined typicality assessments were used in this analysis, since they were found to correlate significantly with the cluster analysis in the previous study.

2. Category group (CG): depending on which category each picture of a pair belonged to, two different category groups were created: "cow-cow" and "horse-

horse". The intermediate "cow-horse" category was not included in this grouping because the pictures belonged to different categories. In order to compare pairs from the same and from different categories the next grouping was created.

3. Category Membership group (CMG): depending on whether both pictures of a pair belonged to the same or a different category two groups were created, Within or Across Category pairings.

Analysis

The unaggregated set consisted of 1820 trials. Mean SR and RT were calculated for individual pairs. Normality tests were run for each item and outliers were excluded. After the exclusion of individual trials which behaved as outliers (six trials for SR and five for RT), all measures were normally distributed.

A univariate ANOVA with SR as a dependent variable and Category Membership group was conducted, in order to test whether within- and across-category pairings promoted similar behaviour.

A univariate ANOVA with SR as a dependent variable and Category was conducted only for Same(-category) pairs in order to test whether the two categories behaved differently.

A Univariate ANOVA with SR as a dependent variable and TG (Typicality group) as a fixed factor was run for both within- and across-category pairs in order to test whether typicality level had any systematic effect on similarity ratings.

A Univariate ANCOVA with RT as a dependent variable and SR as a covariate was conducted in order to test whether the two measures shared any systematic relationship.

Results

Category Analysis. Figure 6.6 shows the mean similarity ratings for picture pairs from the same category and from different categories for different combinations of typicality. It appears that overall objects which come from the same category are considered to be more similar than objects from different categories. Also, pairs of objects from the same level of typicality seem to be perceived as more similar than pairs from different levels of typicality. As far as atypical items are concerned, according to Rosch (1975), one would expect them to be the ones maximally different from each other due to the fact that they share the least attributes with other members of the category.

In order to test these observations, a univariate ANOVA with SR as a dependent variable and Category Membership group (CMG) (within and across categories) and Typicality Group (TG) as fixed factors was run. A highly significant main effect of CMG was found, $F(1, 90) = 122.28$, $p < .001$, $eta_p{}^2 = .58$; Within: $M = 4.97$, $StdError = .17$/ Across: $M = 3.06$, $StdError = .13$, as well as an interaction between CMG and TG, $F(2, 90) = 11.82$, $p < .001$, $eta_p{}^2 = .22$), but no main effect of TG, $F(2, 90) = 2.03$, $p = .14$.

Paired-samples t-tests were run comparing within- and across-category pairings for each typicality group separately. All groups were found to be highly different from each other, Typical-Typical: $t(44) = 145.68$, $p < .001$; Typical-Atypical: $t(39) = 18.89$, $p < .001$; Atypical-Atypical: $t(5) = 214.39$, $p < .001$.

The interaction between Typicality Group and Category Membership Group appears to be related to the different behaviours toward the typical-atypical pairs; when within-category they are perceived as the least similar, whereas when across-categories they are perceived as the most similar pairings. This pattern will be further discussed

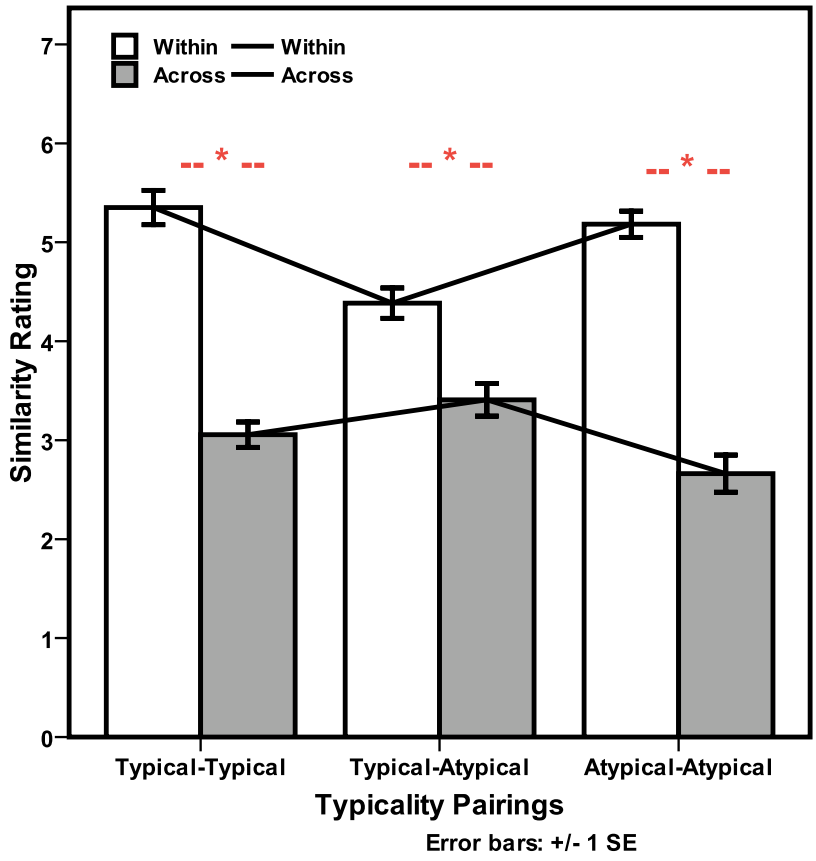


Figure 6.6: Similarity Ratings for pairs of pictures from the same category (Within) and from different categories (Across).

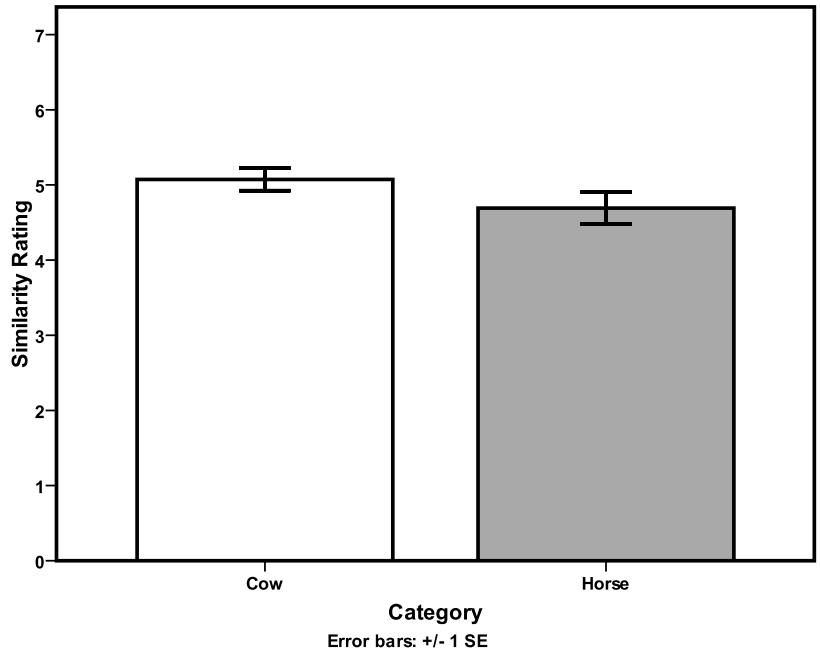


Figure 6.7: Similarity Ratings for the two categories, "cow" and "horse". Only data from same-category pairs were included.

below.

Figure 6.7 shows the mean similarity ratings for the two categories, "cow" and "horse", from data from only same-category pairs. It appears that the two categories did not differ overall. In order to test this observation, a univariate ANOVA with SR as a dependent variable and Category as a fixed factor was conducted. No main effect of Category was found, $F(1, 40) = 2.15$, $p = .15$; Cow: $M = 5.07$, $StdError = .18$/ Horse: $M = 4.69$, $StdError = .18$.

Reaction Times Analysis. Figure 6.8 shows the relationship between Similarity Ratings and Reaction Times. It appears that the more similar objects are perceived to be, the quicker participants are to rate them. In order to confirm this observation, a Pearson r correlation was conducted, showing that the two measures were highly correlated, $r(91) =$, $p < .01$.

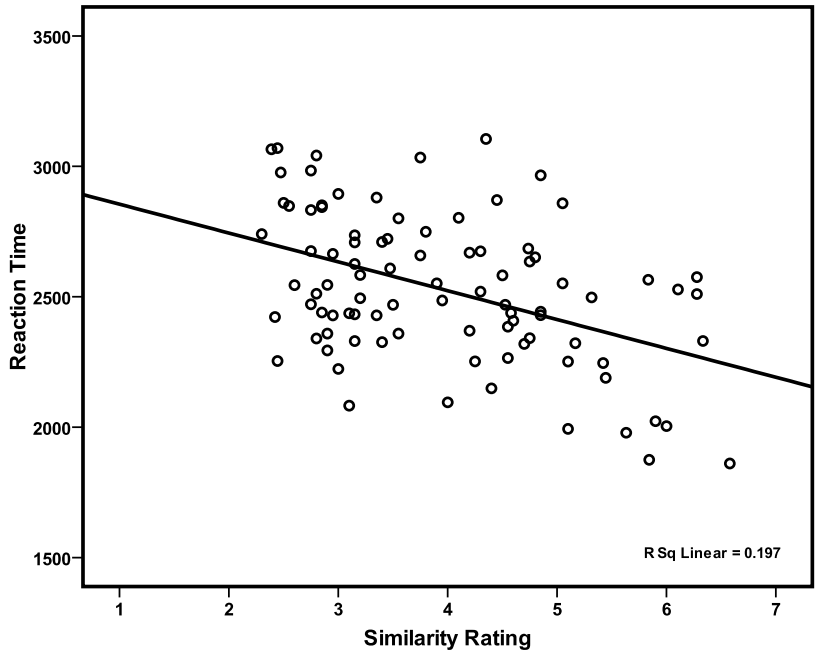


Figure 6.8: Relationship between Similarity Ratings and Reaction Times: the more similar two objects are perceived to be, the quicker participants are to rate them.

Discussion

The Similarity Rating study showed that overall adults perceive same-category objects as more similar than different-category objects, in accordance with Rosch (1975). Additionally categories are shown to have an internal structure and participants perceive objects as more or less similar depending on how typical the objects of a pair are considered. When both objects are typical, similarity is high; this is also the case when both objects are atypical. Pairings between typical and atypical objects, however, are those considered as the least similar.

The high similarity between typical-typical pairings and the low similarity between typical-atypical pairings are in accordance with Rosch's findings that typical objects of a category are those sharing the most common attributes, whereas atypical ones

are those sharing the least common attributes. However, this should make atypical-atypical pairings less similar as well. It appears that this result is idiosyncratic of the specific objects used in this study. The atypical objects were quite similar to each other and this was confounded by the objects available in the market.

When same-category and across-category pairings are compared, an interesting cross-over is evident in Figure 6.6: when across-category both typical-typical and atypical-atypical pairs receive lower ratings than typical-atypical pairs. In other words, typical items are rated as very similar to other typical items from the same category and very dissimilar to typical items from a different category. This holds true for atypical items as well; if they are paired with atypicals from the same category they are rated as very similar, whereas if they are paired with atypicals from a different category they are rated as very dissimilar.

The intermediate pairings, typical items paired with atypical, show an interesting pattern: when within-category they are rated as the least similar pairs; when across categories they are rated as the most similar pairs. It, therefore, appears that there is something about the perceived typicality of an object than affects how similar or dissimilar to other objects from the same or a different category it is perceived to be. It also seems to accord with Rosch's findings about category structure: typical exemplars are those which share the most common attributes with other same-category objects and the least attributes with out-of-category objects. Atypical exemplars, on the other hand, spread toward the periphery of a category; these may share very few common attributes with other same-category objects and they may share more attributes with out-of-category objects. This should be what is driving the interaction found between Category Membership and Typicality Group in the above analysis.

Finally, as in the typicality studies, the two measures, Similarity Ratings and Reaction Times were shown to be significantly correlated; the more similar two objects are, the quicker participants are to rate them.

6.2.3 Similarity Ratings and Typicality Ratings

Tables 6.5 and 6.6 show the Similarity Ratings for within- and across-category pairs, respectively. It appears that some pairs are perceived as more similar than others. Similarly, in the Typicality Rating study, the two categories were shown to have a (similar) internal structure with different exemplars perceived as more or less typical than others. Three objects from each category, 1, 2 and 4 were rated as the most typical, with Cow 1 and Horse 4 rated as the most typical, followed by Cow 4 and Horse 2.

When choosing which objects to include in Experiments Three and Four, it was important to take into account the similarity distance of the objects. For example, when training infants with two word-object pairings, the perceived similarity of the objects should not be very small because this could confuse infants. On the other hand, when testing infants for word generalisation, the test objects should not be very dissimilar from the training objects from the same category, but also not too similar so that infants can discriminate between the two.

On the basis of the typicality and similarity ratings, Cow 1 and Horse 4 were used during pretest, the interactive play-phase, on-screen training and test for word learning, because they were considered to have an advantage over other items due to their higher typicality. For the label generalisation task, Horse 2 was used, but not Cow 4 which was rated as the second most typical object from that category. The reason for not using that object was that Cow 1 and Cow 4 had the same colour,

	C2	C3	C4	C5	C6	C7		H2	H3	H4	H5	H6	H7
C1	**5.83**	6.33	**4.80**	4.45	3.95	4.60	H1	**4.85**	6.28	**6.00**	4.30	4.20	3.10
C2		4.75	**6.28**	5.63	4.40	5.44	H2		5.42	**4.85**	6.58	4.55	3.55
C3			5.84	4.20	5.10	4.58	H3			6.11	5.05	4.74	3.47
C4				5.17	4.70	4.53	H4				4.30	4.50	3.50
C5					4.75	5.90	H5					4.35	3.80
C6						5.32	H6						5.05

Table 6.5: Similarity Ratings for each same-category pair of objects - C denotes Cow and H denotes Horse. The numbers next to the letters denote the exemplar from each category.

	H1	**H2**	H3	**H4**	H5	H6	H7
C1	**2.60**	**3.35**	3.40	**2.90**	3.10	5.10	3.00
C2	**3.15**	**2.39**	3.75	**2.90**	2.85	3.20	3.55
C3	2.75	2.75	2.75	2.80	2.85	2.95	4.85
C4	**2.55**	**2.55**	3.15	**2.95**	2.75	4.00	3.75
C5	2.44	2.44	3.45	4.55	2.80	3.15	2.50
C6	3.00	3.90	2.30	2.42	4.25	2.47	2.80
C7	3.40	3.15	3.20	3.35	3.15	2.85	2.90

Table 6.6: Similarity Ratings for each across-category pair of objects - C denotes Cow and H denotes Horse. The numbers next to the letters denote the exemplar from each category.

and this could cause confusion to infants or it could make it harder to discriminate between the two objects. It was, therefore, decided to use Cow 2 instead. The within- and across-category similarity ratings can be found in Table 6.5 and 6.6 respectively. Across-category pairings (C1-H4 and C2-H2) received similar ratings. Within-category pairings (C1-C2 and H2-H4) seem to differ more with the cow pairing being more similar than the horse pairing. The higher degree of similarity for the cow pairing could cause infants to generalise more easily for that category, and this had to be taken into account in Experiments Three and Four.

Finally, the most atypical exemplar from each category, namely "Cow 6" and "Horse 7", was used in Experiment Five as a modal stimulus; this constitutes the most extreme case of category membership and is also highly dissimilar from the typical instances of the other category.

Chapter 7

General Discussion and Conclusions

7.1 Introduction

Infants at the end of their first year have been reported to begin to associate words to objects; within a couple of years they determine how words are used, what each word means and which objects each word refers to, successfully integrating with the linguistic community of which they form a part. In order to solve the "induction problem" in language, some researchers have proposed the existence of word learning constraints, certain biases which guide infants to discover the correct meaning of words (Markman & Hutchinson, 1984; Markman et al., 2003; Merriman & Bowman, 1989). A different line of thought assumes that such constraint-like behaviour is the outcome of accumulated experience of a powerful learning machine which functions on the basis of similarity and analogy (Smith et al., 2003; Barrett, 1986).

This thesis investigated the manner in which infants at the early stages of word learning extend labels to different objects, and the role of perceptual (or conceptual) categorisation abilities in this process. The lack of information on word learning

abilities of infants at the end of their first year of life made the establishment of a successful experimental paradigm a necessary prerequisite before the investigation of the main aim of this thesis. The final design was based on previous findings and related research with older infants. We expect that this thesis will contribute to a more thorough understanding of lexical development and will offer an additional step toward disambiguating between different theories which aim to account for how infants discover the extension of words.

This chapter discusses the implications from the principle findings of the empirical studies in this thesis. The first part of this chapter describes on overview of the outcome of Experiments 1-5, investigating both word learning and word generalisation, as well as perceptual categorisation; the second part discusses the major issues arising from the experimental results, that is word learning and word generalisation in the first year of life, as well as those arising from the experimental design, that is attentional preferences and measures applied to the IPL task. Finally the third part of the chapter presents the limitations of this research and suggests future studies to extend the investigation to relevant issues.

7.2 Overview of Experiments

7.2.1 Experiments One and Two

Investigating Word Learning in the first year of life

Experiments One and Two in Chapter 2 of this thesis were a first attempt to examine the word learning and word generalisation abilities of infants at the end of their first year of life. Ninety-six 10-month-old infants participated in these two Experiments (48 infants in each); each Experiment was divided in two Conditions on the basis of

whether infants were presented with animate or inanimate objects. All infants were trained and tested with two object-label pairings; these were selected from six possible labels - three animals and three objects, so that infants received different labels. The only difference between the two Experiments was the inclusion of a pretest block in Experiment One, which tested prior knowledge of the to-be-learned object-label pairings.

In Experiment One evidence for word comprehension was found in the pretest block, indicating that infants had some prior knowledge of at least one of the to-be-learned pairings. In Experiment Two evidence for word comprehension was found in the post-training test blocks in the form of distracter looking. Animacy was not shown to play a role in the understanding of the labels, and neither was parental feedback on both infants' overall comprehension and comprehension of the test items.

Conclusions

The strong naming effects exhibited in this study provide the first experimental evidence for the ability of 10-month-old infants to both form and hold a long-term association between an object and a label; theorists of word learning have argued that infants begin to comprehend words at the end of their first year, but no study to date has shown this ability in the lab without explicit training and when using socially non-salient stimuli. These items were reported by parents to be understood by considerably less than 50% of ten-month-olds in the LEX2005 database, which suggests that the linguistic abilities of these infants might be more advanced than has previously been thought.

However, the main aim of these studies, the investigation of label generalisation, could not be directly realised, because of infants' prior knowledge of the to-be-learned object-label pairings. Experiments Three and Four aimed to directly test the issues

of word learning and word generalisation using a more infant-controlled design.

7.2.2 Experiments Three and Four

Experiments Three and Four applied a more infant-controlled procedure for testing word learning in the lab, on the basis of the Werker et al. (1998) experimental paradigm. Sixty 10-month-old infants participated in each Experiment; they were habituated to criterion to two object-label pairings, and then tested in an IPL task for Word Learning (Condition A) and Word Generalisation (Condition B). The two Experiments differed only in their on-screen training block: in Experiment Three infants were trained with a single exemplar from each adult category, whereas in Experiment Four infants were trained with multiple exemplars from each category. These two experiments aimed both to test whether 10-month-old infants can learn two object-label pairings in a brief experimental session and to test directly the assumptions made by principles-and-constraints and exemplar-based accounts of lexical development; the former assume that infants' extension of labels is guided by the taxonomic constraint, which leads them to generalise a newly-learned label to a category of objects. The latter assume that infants initially restrict a label to the object they hear it occurring with, and only after experience with multiple naming instances of different within-category exemplars do they realise that labels refer to whole classes of objects, and not to individual instances of them.

Word Learning - Condition A

In both Experiment Three and Four, infants showed evidence for learning the object-label pairings at test; in Experiment Three infants showed a significant increase in target looking after the presentation of the target word, whereas in Experiment Four

infants showed a systematic preference for the distracter picture after naming. The different attentional preferences in the two Experiments can be attributed to the different training blocks and are discussed below.

Word Generalisation - Condition B

Besides the evidence for word learning in both Experiments, there was no evidence for label generalisation in either Experiment; infants failed to generalise the newly-learned words irrespective of the amount of exposure they had to labeling instances of within-category exemplars.

Conclusions

Experiments Three and Four offered the first experimental evidence of young infants' ability to learn two object-label pairings in a brief experimental session. To date word learning abilities in the laboratory have been shown for infants as young as 14 months for learning two words (Werker et al., 1998) and some limited evidence has been provided for 10-month-olds learning a single word (Pruden et al., 2006). Additionally, infants failed to generalise those labels to new typical objects from the same adult categories, contrary to what principles-and-constraints accounts assume. Exemplar-based accounts predict this outcome for Experiment Three, in which infants were trained with a single object from each category, but predict that generalisation should occur in Experiment Four, in which infants are trained with multiple object-label pairings.

7.2.3 Experiment Five

Perceptual Categorisation

Experiment Five was conducted in order to test whether infants could form perceptual categories on the basis of the training and testing set of Experiments Three and Four. Sixteen 10-month-olds were presented with multiple exemplars from each category used in Experiments Three and Four, and tested with a new exemplar from each category along with completely new objects in the absence of any linguistic input. In Test Block A infants were presented with new within-category objects and new out-of-category objects, in order to test whether infants formed any categories during familiarisation; in Test Block B infants were presented with a Modal (atypical object) and an Average (combination of features from two categories) stimulus, in order to test whether infants formed one or two categories on the basis of the familiarisation set. This experiment was conducted in order to test whether the failure of infants to generalise the labels in Experiments Three and Four was due to a linguistic or a more general cognitive limitation.

In the first test block of Experiment Five some infants preferred to attend to novelty, by spending more time looking at out-of-category objects, indicating that they were able to form visual categories on the basis of the familiarisation set. Some other infants, however, failed to show such an ability and showed no systematic behaviour during test. Infants who showed a novelty preference belonged to the group who failed to habituate, whereas infants who habituated failed to show a novelty preference. For the second test block interpretation is more problematic, because only Habituators showed some systematic behaviour, with increased looking at the Modal stimuli. It appears that this difference in behaviour might be due to the fact that infants are presented with two tasks in these Experiments - namely, word learning and visual categorisation. As a result infants might differentially attend to either task, but

possibly fail to be successful in both simultaneously during this brief experimental session. This difference in the behaviour of the two groups will be further discussed in Section 7.6.2 below.

Conclusions

Overall, results seem to suggest that infants have formed the categories in question. However, this needs to be treated with caution, because of the small number of infants tested and due to the different behaviours of the two habituation groups. The group expected to have categorised the labels, namely Habituators, failed to do it, whereas infants who did not habituate were successful in the categorisation task. If infants are in position to perceptually categorise these stimuli, then it could be that the inability of infants in Experiments Three and Four to extend the newly-learned labels to new within-category objects is due to processing limitations, which prevent infants at this stage in development from exhibiting taxonomic behaviour in language tasks.

7.3 (True) Word Learning

In Section 1.2.3 of Chapter One, we have seen that some researchers argue that early words are highly context-dependent and often under-extended as well (Bloom, 1973; Bates et al., 1979; Barrett, 1983). In Experiments One and Two, however, strong evidence for word comprehension was found showing that ten-month-old infants exhibit some preliminary categorisation ability, which allows them to recognise decontextualised, stationary pictures of non-salient items as referents of a known label. Although this has been shown for 12-month-olds (Schafer, 2005) after explicit training with several exemplars from different categories over an extensive period of time,

it has not been exhibited for younger infants in naturalistic word learning settings (i.e. without any intervention to the parent-infant interactions on the experimenter's behalf).

The main naming effect exhibited by infants was not shown to be affected by characteristics of the stimuli, such as animacy (animals vs. objects), suggesting that the importance of perceptual salience (in this case in the form of animacy) might have been exaggerated in the literature (Pruden et al., 2006). Infants at the end of their first year exhibit word comprehension for animate items, which are assumed to bear a high-degree of perceptual salience, as well as for inanimate, stationary, every-day objects. This strongly suggests that assumptions about the decisive role of perceptual salience during early word learning need to be reconsidered; perceptual salience might be an important, and facilitative, but not a necessary factor for efficient word learning and comprehension. These findings indicate that attention should probably be directed to both the role of increased experience with objects and exposure to a high number of labeling instances in these early stages of lexical development.

Although both developmentalists and parents trace the beginning of lexical development to somewhere around the end of the first year of life, the first evidence for word comprehension by infants well before their first birthday and without any explicit training has been presented in this thesis. One of the basic issues in the study of lexical development concerns the nature of these first words. Some developmentalists have argued that true word learning is not possible in such early stages of language acquisition and they contrast it to mere "associative learning", the formation of an association between temporally contiguous visual and acoustic stimuli (Mervis & Bertrand, 1995; Gopnik & Meltzoff, 1986; Markman, 1991). This approach to word learning suggests important maturational changes during the course of language acquisition and considers lexical development as involving qualitatively different kinds of mechanisms before and after the vocabulary spurt.

In order to identify different types of word learning, a number of criteria have been proposed: Werker et al. (1998) suggest that "true word learning" occurs only when infants can

1. detach a word from an object
2. use the word when the object is absent
3. extend it to other category members

Oviatt differentiates between recognitory and symbolic comprehension, on the basis of whether infants can extend labels to new within-category items and on whether they are in position to use the label in the absence of the original object (Oviatt, 1980).

In Experiment One infants well before the vocabulary spurt have been shown to fulfill at least two of the three criteria considered necessary for symbolic comprehension - namely detachment from original object and extension to new members from the same adult linguistic category (Werker et al., 1998; Schafer, 2005). Although we are not in position to appreciate the actual symbolic status of a word, these infants do exhibit abilities that have not yet been previously documented for such "immature" word learners. These naive word learners are able to exhibit word comprehension without any explicit training and for pictures of items they have never seen before. Even if their mental representations are not as robust as those of post-vocabulary spurt infants, their behaviour appears to be clearly guided by linguistic input. "Full referential understanding" might be a more elaborate form of learning, but the linguistic abilities of infants less than one-year-old might also be more elaborate than previously documented.

7.4 Word Generalisation

7.4.1 Disambiguating between Different Accounts

Experiments Three and Four aimed to test the assumptions of principles-and-constraints and exemplar-based accounts for the way infants extend labels to objects. As described in Section 1.2.3, principles-and-constraints accounts assume that infants are guided by the taxonomic constraint in order to extend a newly-learned label to a whole class of objects; any failure to apply a label correctly is due to a perceptual/conceptual and not a linguistic limitation. Nelson (1978) suggests that early words are matched to existing concepts and that infants at the end of their first year *set out to assimilate words to well-formed concepts*. On the other hand exemplar-based accounts assume that infants are initially restricted to applying a label to the object they heard it occurring with; after experience with multiple naming instances with different within-category exemplars they discover that words refer to classes of objects, and not to individual instances of them, and begin to extend the label on the basis of similarity to the familiar exemplars (Barrett, 1986; Smith et al., 2003).

Infants in both Experiment Three and Four exhibit understanding of the words in Condition A; as Rosch (1978) and Bowerman's (1978) prototype model of word meaning would predict, infants attach the label to the most representative exemplar of the category - the objects that adults have rated as most typical of the categories in question. However, in both experiments infants fail to generalise the labels to the new exemplars, suggesting that taxonomic-like behaviour is not readily available to infants during word learning. Infants inability to recognise other typical exemplars as referents of newly-learned words goes contrary to what principles-and-constraints accounts predict: infants fail to generalise a label to a category of objects, although they might have a pre-established horse or cow category. Instead, they seem to be attaching a label to an object in an associative-like way, at least in this short

experimental session.

It appears that 10-month-old infants in this Experiment engage in the word learning task *ignoring* whatever concepts they have already formed and focusing on the association between the object and the label. This is in contrast to what principles-and-constraints approaches assume: infants do not seem neither to set out to match labels to existing concepts (Nelson et al., 1978) neither to use their conceptual knowledge during their word learning task (Waxman, 2003; Waxman & Braun, 2005; Markman & Hutchinson, 1984).

Instead they underextend the label and restrict it to the most representative exemplar of the category (Rosch, 1998; Bowerman, 1978; Brooks, 1987). Initially, they appear as statistical learning machines tracing the regularities in their environment in order to infer which sound goes with which object (Tomasello, 2003). It appears, however, that they require more exposure naming instances of a given word-object pairing before they realise that a word refers to a category of objects and before identifying which are the important dimensions in generalising the labels to new exemplars for each category (Smith et al., 2003; Wittgenstein, 1978; Bowerman, 1978).

This behaviour of infants is also in accordance with hybrid models of word learning; Barrett's model (1986) assumes that initially newly-learned labels should be restricted to a single instance, in a prototype-like way. Infants need to identify what the principal features which characterise the prototype are before they abstract them and are able to generalise the label to new exemplars (Barrett, 1986). On the other hand, the Emergentist Coalition Model (ECM) (2000) suggests that infants begin the word learning task with associative-like behaviour, before the taxonomic constraint becomes available around the second half of the second year of life, through maturational changes. Although the ECM is able to account for the behaviour of infants in Experiments Three and Four, it falls into difficulty in accounting for the behaviour of infants in Experiment Two, who do exhibit word comprehension for

previously unseen objects and taxonomic-like behaviour. If infants younger that 18-months only engage in associative learning, then generalisations of this sort should not have been demonstrated for infants as young as 10 months!

Another explanation for infants failure to generalise is because of processing limitations; 10-month-olds might find it too difficult to attend to the word learning and the visual categorisation task simultaneously. However, infants at 10 months are in position to recognise de-contextualised objects as the referents for familiar labels. Infants do not, therefore, lack taxonomic behaviour altogether, but they fail to exhibit it in brief experimental sessions, under conditions of limited exposure to the object-label pairings. If infants can show generalisation of labels with known words, then failure to generalise newly-learned words offers support to exemplar-based accounts which emphasise on the role of experience before infants discover the extension of a word. It also puts the assumptions of principles-and-constraints accounts under question because this failure to generalise the newly-learned words is not a global characteristic of early words, but only of those with which infants had limited experience. Biases and constraints apply universally, in an "all or nothing" fashion; if such a thing as a taxonomic constraint existed, it should have applied to both familiar and newly-learned words.

7.4.2 Conclusions

This thesis has provided the first evidence that one of the proposed constraints on word learning, the taxonomic constraint, does not appear to be functional in the early stages of word learning, although infants do exhibit taxonomic behaviour when tested with known words. Infants fail to generalise words with which they have had limited experience, but are willing to generalise those which they are very familiar with. Evidence for taxonomic generalisations for newly-learned words in later stages

of lexical development could, therefore, be due to the experience of infants with words and their usage, as proposed by exemplar-based accounts, and could be based on quantitative, instead of qualitative, changes in infant learning mechanisms.

7.5 The Impact of Labels on Category Formation

In Section 1.2.2 of Chapter One, we have seen that some researchers argue for an early impact of language on category formation (Waxman & Braun, 2005; Fulkerson & Waxman, 2007; Plunkett et al., 2008), while some others argue for an auditory dominance, which disrupts visual processing (Robinson & Sloutsky, 2004). This thesis does not aim to disambiguate between these views and the results are not conclusive either; however, there seems to be an interesting interplay between language and category formation favouring the second view: infants who habituated in Experiment Five showed no evidence for category formation, whereas infants who did not habituate showed a novelty preference suggesting that they perceptually categorised the objects during familiarisation. In Experiment Four, this pattern appears inversed, with Habituators showing word learning and Non-Habituators failing to show any evidence for learning the object-label pairings. This picture could be suggestive of a "language dominance", instead of just an auditory dominance, for infants who habituated, resulting in the disruption of category formation for those infants. On the other hand, infants who did not habituate, and also failed in the word-learning task, do not seem to have been affected in their perceptual categorisation abilities.
This language dominance hypothesis is a little bit different from the auditory dominance hypothesis proposed by Robinson (2004). The latter habituated 8-month-old infants to visual-auditory pairings, using visual displays with geometrical patterns and non-linguistic sounds; at test infants showed a novelty preference only when the

auditory stimulus changed but not when the visual stimulus did. In that study infants showed an auditory dominance by selectively attending to the auditory stimulus and failing to process the visual stimulus. One explanation is that they could have been "put off" by the auditory stimuli, which were not word-like, and, therefore, did not consider the task a linguistic one, failing to attend to the pairing. Alternatively infants at 8 months could be at a stage in development where they are attracted by sounds and focus attention on that dimension of the task. However, infants at 10 months could be a step further in development where they attend to the relationship between the sound and the visual display which goes with it. Therefore, we could be having two different stages here: an initial phase of auditory dominance, exhibited by 8-month-olds, and a subsequent phase of language dominance, exhibited by 10-month-olds in this study.

In previous studies, however, infants who habituated also showed a novelty preference at test, suggesting that they were able to perceptually categorise the objects (Balaban & Waxman, 1997; Fulkerson & Waxman, 2007; Fulkerson & Haaf, 2003). Why do those infants benefit from the labels and infants in the current research do not? The simple answer is that infants in previous research *might not* really benefit from linguistic input during their category formation. That is, in the absence of a silent control condition during which infants habituate in the absence of any auditory input whatsoever, we cannot claim that such thing as "facilitation" really exists. All we can say is that labels do not prevent category formation, but not that they facilitate it.

But why do labels prevent category formation in Experiments Four and Five? The answer might lie in the fact that the task might be too easy for infants in previous studies because they are exposed to only one category and to a single label and their categorisation abilities do not get interrupted. Another explanation might be that in the current research the labels are familiar to infants and this familiarity leads infants

to attend to the word-object pairing more and engage in word learning, thus failing to also perceptually categorise the pictures.

Evidence for perceptual categorisation in the presence of labels and using two categories has been found in Plunkett (2008); however, unlike Experiments Four and Five, in those studies infants were not divided in Habituators and Non-Habituators during the analysis. It could be that overall infants in Plunkett (2008) exhibited a novelty preference, but some infants failed to create perceptual categories. As in Experiment Five, it could be that infants showed a novelty preference overall, but this preference could have been driven by Non-Habituators.

7.5.1 Conclusions

Assumptions about an early impact of labels on category formation facilitating the categorisation process were not confirmed in this thesis. Instead evidence for an early "language dominance" has been found, with category formation being interrupted for infants who were successful in the word-learning task. Further research is needed in order to evaluate this hypothesis.

7.6 Attentional Preferences and IPL

7.6.1 Target versus Distracter Looking

The repetitive presentations of the same stimuli/object-label pairings in the basic design of this research caused some infants to habituate and show signs of decreasing attention to the task. This was evident during training in Experiments 3-5 with infants spontaneously dividing into two groups on the basis of whether they reached criterion or not - Habituators and Non-Habituators, respectively. However, this decrease in attention was also evident at test; some infants demonstrated their word

comprehension through a systematic target preference and some through a systematic distracter preference. In Experiment One, infants showed a significant preference for the target during *pretest* after the presentation of the word; in Experiment Two, infants showed significant distracter looking in the *post-training* test blocks. In Experiment Three, where infants were trained with a *single exemplar* from each category, infants demonstrated their learning of the words through increased attention to the target after naming; in Experiment Four, where infants were trained with *multiple exemplars* from each adult category, learning was demonstrated through distracter looking.

Traditionally, in an IPL task infants exhibit their word comprehension by preferentially fixating the corresponding picture upon hearing the word. However, distracter looking has been found in such tasks when they involved repetition of the same stimuli (Houston-Price & Nakai, 2004) and has been interpreted using the Hunter and Ames' model of infant attentional preference: infants initially attend to familiarity and after repeated exposure to the same stimuli they begin to orient their attention more to novelty (Hunter & Ames, 1988). Experiments One and Two follow this pattern: in the pretest block of Experiment One infants show a familiarity preference and they increase their attention to the target when they hear the word; in the post-training block of Experiment Two, and after repetitive presentations of the same object-label pairings, infants switch to a novelty preference and orient their attention to the distracter upon hearing the word.

Experiment Four is an even more interesting demonstration of the dynamic nature of infant preferences: infant exhibit a task-specific habituation at test, which had been obscured by an additional process during training. That is, during training infants are presented with multiple word-object pairings: they are, therefore, spending time to both learn the words and to visually process (and maybe also to categorise) the multiple exemplars from each category. This means that during training they take

longer than infants in Experiment Three to habituate; although they take longer, due to an additional task dimension, they might have developed larger habituation to the specific word-object pairings. This means that, unlike in Experiment Three, they do not show their comprehension of the words by attending to the target, but instead they orient more to the distracter picture upon hearing the word in search of novelty. It is, therefore, suggested that both target and distracter looking can be considered as forms of evidence for word learning/comprehension. Predicting when infants are going to use one or the other is rather difficult, but shall depend on a number of task dimensions (see (Hunter & Ames, 1988)):

- age of infants: older infants are more likely to exhibit a novelty preference that younger infants due to their more advanced processing abilities.
- complexity of stimuli: complex stimuli should be expected to take longer to process than simpler stimuli.
- complexity of task: multiple task dimensions should be expected to make the task more difficult and infants might take longer to habituate. However, as in Experiment Four, infants might habituate to one aspect of the task, while still maintaining relatively high interest. This should be taken into account when interpreting distracter looking behaviour.
- time available for infants to process the task: longer times are more likely to produce habituation and drive infants to look for novelty.

Additionally, when comprehension is demonstrated via distracter looking, evidence for habituation should be sought. Evidence for habituation during training/familiarisation should be considered necessary, along with the task dimensions described above, in deciding whether distracter looking is a legitimate form of evidence for word comperehension.

7.6.2 Habituation and its Implications

As described in Section 3.1, habituation is a "response decrement as a result of repeated stimulation" (Harris, 1943) or a "relatively permanent waning of a response as a result of repeated stimulation" (Thorpe, 1956). For Experiment One to Three, the task is that of word learning with a single exemplar from each category being presented, and, therefore, the attentional preferences of infants evolve around those stimuli (pairs) which are presented repeatedly. On the other hand, in Experiments Four and Five the task is not equally straightforward for infants; two words are still presented here but multiple pictures are paired with each word. This means that besides the word learning task available in the rest of the experiments, there is an additional experimental dimension: infants could possibly be processing the visual stimuli and creating perceptual categories on the basis of the training/familiarisation set. So what do attentional preferences tell us about mental processing during the experimental session? Do infants both learn the words and visually categorise the stimuli? Or do they only focus on one of those tasks and ignore the other?

I suggest that the answer comes from the behaviour of the two habituation groups in Experiments Four and Five; in Experiment Four infants who habituated were those who exhibited evidence for word comprehension at test, albeit in the form of distracter looking. In Experiment Five infants who habituated did not exhibit a novelty preference at test, and those who did not habituate showed a preference for novelty, providing evidence that they could discriminate between the within- and out-of-category novel objects. This pattern could be pointing to a period in development when infants are just beginning to discover the relationship between words and objects. Those infants who have already discovered this relationship could be attending to the word learning task during training, and possibly ignoring the categorisation part, which might be confusing for them at this stage. These infants are those who have reached criterion in Experiments Four and Five. Infants who

have not reached criterion might be those who have not yet discovered that there is something special about words and objects and are, therefore, finding the visual categorisation task more appealing, showing a novelty preference in Experiment Five and no word learning in Experiment Four.

This interpretation only constitutes a hypothesis which could explain the behaviour of infants in these Experiments. The small number of infants in Experiment Five makes it less robust, because it could be that infants who reached criterion are just those who do not wish to participate. More research is needed, in order to draw safer conclusions.

7.6.3 Conclusions

Infant attentional preferences have been shown to be dynamic in nature and change over the course of an experimental session; comprehension in an IPL task might be demonstrated by increased looking at either the target or the distracter. Novelty preference, in the form of distracter looking, can be the result of either a stimulus-specific habituation, as in Experiment Two, or a task-specific habituation, as in Experiment Four. It is suggested that a number of task dimensions should be taken into account and evidence for habituation should be sought before it is legitimate to use distracter looking as a good index for word comprehension.

Also attentional preferences and habituation criteria are only a conventional means of studying behaviour and they could be the result of different underlying mental processes. Infants in the two habituation groups of Experiments Four and Five could be exhibiting entirely different types of learning. The results of such studies need to be interpreted with caution.

7.7 Measuring Infant Attention in an IPL Task

Two measures were used to test word comprehension in this research: Proportion of Target Looking and Longest Look Difference. The two measures generally agreed, but in certain cases one was more sensitive than the other. LLD was shown to give stronger evidence for word comprehension in Experiment One and Experiment Three, in which the least habituation effects were present and a target preference was found. On the other hand, in Experiments Two and Four PTL was shown to show stronger evidence for comprehension, in the form of distracter looking and as a result of increased levels of habituation.

It appears that the two measures used to investigate word comprehension in an IPL task might be tapping different aspects of information processing. LLD appears to mirror the "uncertainty" in matching the mental representation of infants and the incoming stimulation; PTL appears to also mirror non-linguistic behaviour reflecting habituation effects as a result of multiple repetitions in the design. These observations should be taken into account when using these measures in different designs.

7.8 Evaluation

The main contributions of the studies in this thesis include the investigation of comprehension abilities of infants at the end of their first year of life, the improved training paradigm for teaching object-label pairings in the laboratory and the examination of the manner in which infants extend labels to objects, disambiguating between opposing theories and controlling for the impact of perceptual categorisation in this process. This thesis has shown that infants at the end of their first year, and the early stages of word learning, are in position to show linguistic abilities not previously documented for such young infants:

1. they have shown evidence for recognising de-contextualised and newly-encountered pictures as the referents of known words, satisfying at least two of the three criteria set by researchers for "true word learning". Also this provides evidence that taxonomic generalisations in lexical development are in place at least for *some words* even at this very early stage.

2. they have shown evidence for their ability to learn two word-object pairings in brief experimental sessions - an ability previously documented for infants as young as 14 months of age.

3. infants failed to generalise those words to new typical objects from the corresponding adult categories, although as a group they showed evidence for their ability to perceptually categorise those words.

4. infants who were successful in the word learning task appear to be those who failed to categorise, and vice versa. This suggests a period of "language dominance" during early lexical development, which prevents word learners from also processing other tasks.

These findings suggest that we should be very cautious when referring to biases and constraints during word learning: among the very first words of infants one can find both taxonomic generalisations and also failure to generalise, possibly depending on the amount of experience with the objects and their labels. This offers support to exemplar-based accounts of lexical development and points to language mechanisms which do not undergo qualitative changes during development but offer the impression of constraint-like behaviour on the basis of accumulated experience.

The findings of the experiments described above are, of course, only a part of the lexical development jigsaws and provide implications for further research into the extension of labels and the interaction between cognitive and linguistic categorisation. A similar series of studies with *novel* - made-up objects might offer a more strict

control and more insights into both the features of interest when building categories, both perceptual and linguistic, as well as into the role of frequency and similarity when extending labels to objects.

The use of novel objects would also allow the investigation of the extension of labels with older infants, using the same stimuli. The use of real objects poses limitations in the age range which can be tested, since older infants are likely to know the labels before coming to the laboratory. Research with infants in their second year of life can provide a better understanding of the developmental trajectory of lexical categories, and can help disambiguate between the two theories available for accounting for this phenomenon, namely principles-and-constraints and exemplar-based accounts.

Novel objects would also allow the investigation of the number of categories created on the basis of the training set in Experiment Four; in Experiment Five the investigation of the number of perceptual categories formed during familiarisation was restricted by the choice of modal and average stimuli. Modal stimuli were peripheral members of the categories in question, whereas an average stimulus was created by fusing the two categories together using morphing software. The end product, however, was not as naturalistic as the rest of the exemplars and this could have contributed to the increased attention of infants to the modal stimuli. Novel categories could allow more control in the creation of those test items and help investigate whether infants are creating one or two categories during training.

An additional question arising from Experiment Four is whether infants attach the label to all same-category exemplars or just to the most typical exemplar; in the latter case, failure to generalise could be due to the fact that infants are not learning that the label "goes with" the whole class of objects, but only with the most typical one, and are, therefore, neglecting to extend it during test. If the hypothesis about the source of different behaviours in the two Habituation groups is correct, then it could be that infants at the end of their first year of life are limited in their processing capabilities

and cannot both learn an object-label pairing and perform visual categorisation at the same time. This would mean that the result of Experiments Three and Four is not contrastive to either account of lexical development, due to infants' processing limitations. Instead in order to draw safer conclusions about label generalisation, we would need to test an age group, in which infants both learn the words and can visually categorise the stimuli. Any failure to generalise in that case would constitute strong support for exemplar-based accounts, and would provide evidence for the lack of any learning biases, at least at that point in development.

In this kind of research, both with familiar and novel objects, methods such as eye-tracking can prove valuable in detecting the relevant features during (linguistic and non-linguistic) category formation and in the creation of effects, such as prototypicality gradings. In this thesis, prototypicality and similarity ratings were obtained from adults, but these can be quite different from how infants perceive objects to be. Eye-tracking can follow the creation of categories in real time and closely monitor the rise of prototypicality effects, in relation with frequency and similarity.

7.9 Final Conclusions

This thesis has shown that infants at the end of their first year, and the early stages of lexical development, have started associating words to objects and are in position to identify newly-encountered, de-contextualised, (pictures of) objects as referents of familiar labels, indicating their ability to detach from the original word learning setting and exhibit evidence for "true" word learning. Additionally, these infants exhibited the ability to quickly form word-object associations in the laboratory over the course of a brief experimental session. This ability, however, was not evident in all infants' behaviour: it appears that we might be witnessing a

transitional period between unimodal and cross-modal processing, with some infants having just discovered the special relationship between words and objects and some others still focusing on the relationship between perceptually similar objects. During this period a form of interference may arise for infants who have discovered the special relationship of sounds and objects, which we termed as "language dominance". Although evidence for taxonomic generalisations has not been found in the word learning task it was present in the word comprehension task; it appears that (at least some) infants at the end of their first year have more sophisticated linguistic abilities than has previously been thought and that lexical development might not proceed through qualitative shifts but instead through the accumulation of experience with language and its use.

Appendix A

Visual Stimuli in Infant and Adult Studies

A.1 Infant Studies

A.1.1 Experiments One and Two

Training Stimuli

Animals.



Objects.



Testing Stimuli

Animals.



Objects.



A.1.2 Experiment Three

Training Stimuli



Testing Stimuli

Condition A.

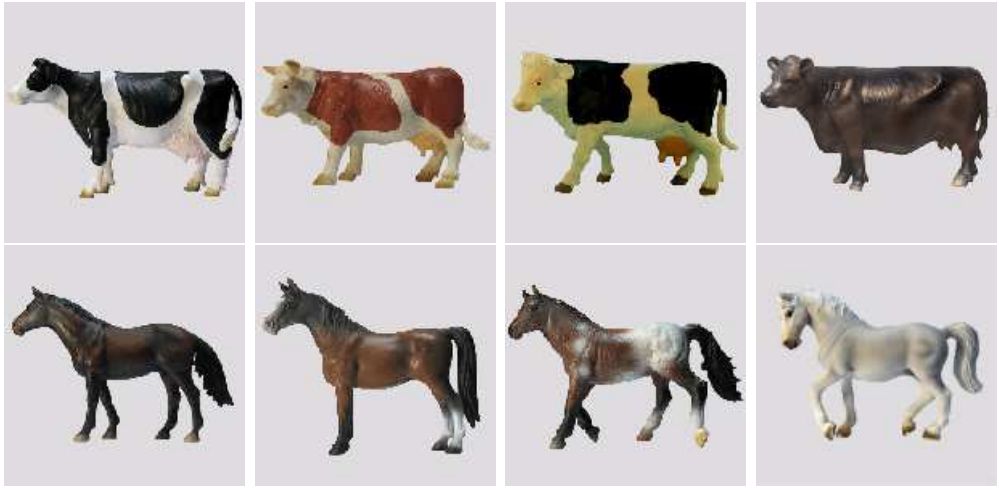


Condition B.



A.1.3 Experiments Four and Five

Training Stimuli



Testing Stimuli - Experiment Four

Identical to Experiment Three.

Testing Stimuli - Experiment Five

Within-Category Typical Exemplars.



Within-Category Atypical Exemplars.



Out-of-Category Exemplars.



Cow-Horse Combination.



A.2 Adult Studies

A.2.1 Typicality Rating Study One

Typical Items

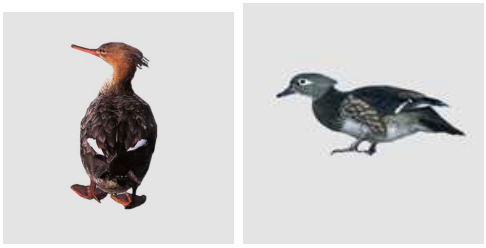
All pictures used in Experiments One and Two.

Borderline Items

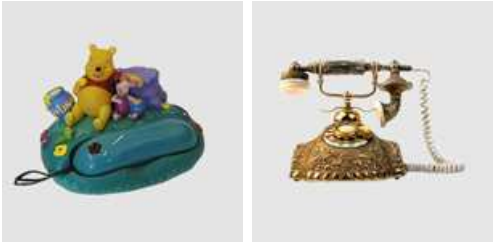














Atypical Items














A.2.2 Typicality Rating Study Two

The Training and Testing Items used in Experiment Four were also included in this study. An additional atypical exemplar from each category was also used. These are presented below:



A.2.3 Similarity Rating Study

The pictures used in Typicality Rating Study Two were used in this study. All pictures were paired with each other exhaustively.

Appendix B

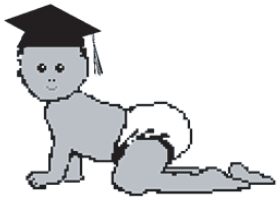
Communicative Development Inventory

The Oxford Communicative Development Inventory (BCDI) is a British adaptation of the MacArthur CDI (Fenson et al., 1994) as prepared by Hamilton et al. in 2000. More information on the Oxford CDI can be found in Chapter Two, Section 2.2.2. The questionnaire can be found below:

OXFORD UNIVERSITY BABYLAB

Communicative Development Inventory

- A UK adaptation of the MacArthur CDI [*] -



Dear parent,

The following is a list of words that are typical in children's vocabularies.

For words that your child **understands but does not yet say**, place a mark in the first column, labelled "**U**".

	U	U/S
crocodile	●	○

For words that your child **understands and also says,** place a mark in the second column, labelled "**U/S**".

	U	U/S
crocodile	○	●

If your child uses a different pronunciation of a word (e.g., 'bickie' for biscuit, or 'telly' for television) - mark the word anyway.

Occasionally we list two alternative forms - please **underline** the one your child understands and/or produces.

	U	U/S
pool/pond	○	●

Please fill in the whole circle exactly as shown above, do not just tick or partly fill the circle.

correct marking - ● incorrect markings - Ⓥ or ⊙

This inventory is a comprehensive "catalogue" of words that are used by many different children across a wide age range, so do not worry if your child knows only a few of them at the moment!

If you have any additional comments or information that you think we should consider, please add these at the end of this inventory.

Thank you very much!

---

[*] For information and original copies of the MacArthur CDI, please contact the Developmental Psychology Lab, San Diego State University, San Diego, CA 92182, USA.

OXFORD UNIVERSITY BABYLAB

Communicative Development Inventory

Subject code

……

Your name: ……

Child's name: ……. Male/female: …….

Birth date of child: …../……/…… Today's date: …../……./…….

Animal sounds	U	U/S		U	U/S
baa baa	O	O	ouch	O	O
choo choo	O	O	quack	O	O
cockadoodledoo	O	O	uh oh	O	O
grr	O	O	vroom	O	O
meow	O	O	woof	O	O
moo	O	O	yum	O	O

Animals	U	U/S		U	U/S
animal	O	O	horse	O	O
bear	O	O	kitten	O	O
bee	O	O	lamb	O	O
bird	O	O	lion	O	O
bunny / rabbit	O	O	monkey	O	O
butterfly	O	O	mouse	O	O
cat	O	O	owl	O	O
chicken	O	O	penguin	O	O
cow	O	O	pig	O	O
deer	O	O	pony	O	O
dog	O	O	puppy	O	O
donkey	O	O	sheep	O	O
duck	O	O	spider	O	O
elephant	O	O	squirrel	O	O
fish	O	O	tiger	O	O
frog	O	O	turkey	O	O
giraffe	O	O	turtle	O	O
goose	O	O			

Vehicles	U	U/S		U	U/S
aeroplane / plane	O	O	bus	O	O
bicycle / bike	O	O	car	O	O
boat	O	O	fire engine	O	O
lorry / truck	O	O	pushchair/buggy	O	O
motor-bike	O	O	train	O	O

Toys	U	U/S		U	U/S
ball	O	O	doll	O	O
balloon	O	O	pen	O	O
block / brick	O	O	teddy bear	O	O
book	O	O	toy	O	O
bubble	O	O			

Food and Drink	U	U/S		U	U/S
apple	O	O	food	O	O
banana	O	O	ice cream	O	O
biscuit	O	O	jam	O	O
bread	O	O	juice	O	O
butter	O	O	meat	O	O
cake	O	O	milk	O	O
carrot	O	O	orange	O	O
cereal	O	O	pasta / spaghetti	O	O
cheese	O	O	peas	O	O
chicken	O	O	pizza	O	O
chips	O	O	sweets	O	O
coffee	O	O	tea	O	O
drink	O	O	toast	O	O
egg	O	O	water	O	O
fish	O	O			

Body Parts	U	U/S		U	U/S
arm	O	O	hair	O	O
belly button / tummy button	O	O	hand	O	O
cheek	O	O	head	O	O
ear	O	O	knee	O	O
eye	O	O	leg	O	O
face	O	O	nail	O	O
finger	O	O	nose	O	O
foot	O	O	toe	O	O
tongue	O	O	tummy	O	O
tooth	O	O	mouth	O	O

Clothes	U	U/S		U	U/S
bib	O	O	dress	O	O
boot(s)	O	O	glasses / specs	O	O
button	O	O	hat	O	O
coat	O	O	jacket	O	O

Clothes	U	U/S		U	U/S
jeans	O	O	shoe	O	O
jumper / sweater	O	O	shorts	O	O
nappy	O	O	sock	O	O
necklace	O	O	trousers	O	O
pyjamas	O	O	zip	O	O
shirt	O	O			

Furniture and Rooms	U	U/S		U	U/S
bath / bathtub	O	O	living room	O	O
bathroom	O	O	play pen	O	O
bed	O	O	potty	O	O
bedroom	O	O	refrigerator / fridge	O	O
chair	O	O	rocking chair	O	O
cooker / stove / oven	O	O	settee / sofa	O	O
cot	O	O	sink	O	O
door	O	O	stairs	O	O
drawer	O	O	table	O	O
garage	O	O	TV / television	O	O
high chair	O	O	window	O	O
kitchen	O	O			

Outside	U	U/S		U	U/S
beach	O	O	outside	O	O
bucket	O	O	park	O	O
church	O	O	party	O	O
flower	O	O	pool	O	O
garden	O	O	rain	O	O
house	O	O	school	O	O
moon	O	O	shop	O	O
sky	O	O	swing	O	O
slide	O	O	tree	O	O
snow	O	O	wall	O	O
spade	O	O	water	O	O
star	O	O	work	O	O
stone	O	O	zoo	O	O
sun	O	O			

Household items	U	U/S		U	U/S
bin	O	O	bowl	O	O
blanket	O	O	box	O	O
bottle	O	O	broom	O	O

Household items	U	U/S		U	U/S
brush	O	O	paper	O	O
clock	O	O	penny	O	O
comb	O	O	picture	O	O
cup	O	O	pillow	O	O
dish	O	O	plant	O	O
dummy	O	O	plate	O	O
fork	O	O	purse	O	O
glass	O	O	radio	O	O
hammer	O	O	rubbish	O	O
hoover / vacuum	O	O	scissors	O	O
jug	O	O	soap	O	O
key	O	O	spoon	O	O
lamp	O	O	telephone	O	O
light	O	O	toothbrush	O	O
medicine	O	O	towel	O	O
money	O	O	watch	O	O
mug	O	O			

People	U	U/S		U	U/S
aunt	O	O	girl	O	O
baby	O	O	grandma	O	O
boy	O	O	grandpa	O	O
brother	O	O	lady	O	O
child	O	O	man	O	O
daddy	O	O	mummy	O	O
doctor	O	O	nanny	O	O
friend	O	O	people	O	O
person	O	O	teacher	O	O
policeman	O	O	uncle	O	O
sister	O	O			

Games and Routines	U	U/S		U	U/S
bath	O	O	no	O	O
breakfast	O	O	pat-a-cake	O	O
bye bye	O	O	peekaboo	O	O
dinner	O	O	please	O	O
don't	O	O	shh / hush / shush	O	O
hello	O	O	tea	O	O
hi	O	O	thank you	O	O
lunch	O	O	wait	O	O
nap	O	O	want to	O	O
night night	O	O	yes	O	O

Action Words	U	U/S		U	U/S
bite	O	O	know	O	O
blow	O	O	like	O	O
break	O	O	look	O	O
bring	O	O	love	O	O
bump	O	O	make	O	O
call	O	O	open	O	O
carry	O	O	play	O	O
catch	O	O	pull	O	O
clean	O	O	push	O	O
cry	O	O	put	O	O
cuddle	O	O	read	O	O
cut	O	O	ride	O	O
dance	O	O	run	O	O
draw	O	O	say	O	O
drink	O	O	scratch	O	O
drive	O	O	see	O	O
drop	O	O	show	O	O
eat	O	O	shut / close	O	O
fall	O	O	sing	O	O
feed	O	O	sleep	O	O
find	O	O	smile	O	O
finish	O	O	splash	O	O
get	O	O	stop	O	O
give	O	O	swim	O	O
go	O	O	swing	O	O
have	O	O	take	O	O
hear	O	O	tell	O	O
help	O	O	throw	O	O
hit	O	O	tickle	O	O
hug	O	O	walk	O	O
hurry	O	O	wash	O	O
jump	O	O	watch	O	O
kick	O	O	wipe	O	O
kiss	O	O	write	O	O

Descriptive Words	U	U/S		U	U/S
all gone	O	O	clean	O	O
asleep	O	O	cold	O	O
bad	O	O	dark	O	O
big	O	O	dirty	O	O
blue	O	O	dry	O	O
broken	O	O	empty	O	O
careful	O	O	fast	O	O

Descriptive Words	U	U/S		U	U/S
fine	O	O	old	O	O
gentle	O	O	pretty	O	O
good	O	O	red	O	O
green	O	O	sad	O	O
happy	O	O	scared	O	O
hard	O	O	sick	O	O
hot	O	O	sleepy	O	O
hungry	O	O	soft	O	O
hurt	O	O	thirsty	O	O
little	O	O	tired	O	O
nasty	O	O	wet	O	O
naughty	O	O	yellow	O	O
nice	O	O			

Question words	U	U/S		U	U/S
how	O	O	where	O	O
what	O	O	who	O	O
when	O	O	why	O	O

Time	U	U/S		U	U/S
day	O	O	now	O	O
later	O	O	today	O	O
morning	O	O	tomorrow	O	O
night	O	O	tonight	O	O

Pronouns	U	U/S		U	U/S
her	O	O	my	O	O
his	O	O	that	O	O
I	O	O	this	O	O
it	O	O	you	O	O
me	O	O	your	O	O
mine	O	O			

Prepositions	U	U/S		U	U/S
away	O	O	on	O	O
back	O	O	out	O	O
down	O	O	there	O	O
in	O	O	under	O	O
inside	O	O	up	O	O
off	O	O			

Quantifiers	U	U/S		U	U/S
all	O	O	not	O	O
again	O	O	other	O	O
another	O	O	same	O	O
more	O	O	some	O	O
none	O	O			

Extra words	U	U/S		U	U/S
chase (action)	O	O		O	O
smell (action)	O	O		O	O
	O	O		O	O
	O	O		O	O
	O	O		O	O
	O	O		O	O

Additional Questions:

Does anyone speak to your child in a language other than English (if so, which language)?

Has your child ever had any hearing problems, including glue ear?

Was your child born more than six weeks premature?

Thank you for your help.
If you have any further comments, please write them below.

Appendix C

Category Experience Questionnaire

C.1 Introduction

In order to choose the (categories of) objects used in Experiments Three, Four and Five, and to also control for the amount of exposure infants have had with those (categories of) objects, the Category Experience Questionnaire (CEQ) was created. Although infants in these Experiments might or might not have had any pre-established categories with the to-be-learned objects (due to their prior experience with these objects), familiarity was considered as an important factor for successful performance in the task. That is, infants who have been exposed to a category repeatedly in the past might find it easier to process both the visual and auditory stimuli, to associate them to each other and to generalise the labels to new objects.

The categories evaluated in the CEQ were taken from the Oxford CDI word-list; only words which corresponded to categories of objects were used, and no verbs. This resulted in a total of 117 categories, split into 8 groups: Animals ($N = 34$), Vehicles ($N = 10$), Toys ($N = 7$), Food ($N = 7$), Clothes ($N = 16$), Furniture ($N = 10$), Outside ($N = 5$) and Household Items ($N = 28$).

According to various accounts of category formation, different dimensions could affect

the structure of categories, both in adulthood and infancy - see Section 1.1.1. Three of those dimensions were evaluated in this questionnaire:

1. *Frequency of occurrence*: the more familiar an object/a group of objects is, the easier might be to categorise it. In other words, the more exposure one has to objects from a given category, the more likely she is to create a category of those objects (Rosch & Mervis, 1975; Rosch et al., 1976).

2. *Number of Exemplars*: One can only speak for a category in a meaningful way only when multiple objects from that category are available to the participant (Rosch & Mervis, 1975; Rosch et al., 1976). Thus the more objects an infant sees from a given category, the more likely she is to create a category which contains those objects - perceptual similarity should come into play when multiple objects from a category are viewed and typicality gradings are expected to arise in the process (Meints et al., 1999). Both the latter are evaluated in Chapter Six for the object categories used in Experiments 3-5.

3. *Exposure to the label*: Infants have been shown to benefit during category formation from the presentation of a label which is correlated with a category (Waxman, 2003; Waxman & Braun, 2005; Fulkerson & Waxman, 2007; Plunkett et al., 2008).

For each dimension a rating scale was provided:

1. *Frequency*: "never", "once a month", "once a week", "several times a week", "every day"

2. *Number of Exemplars*: "0-1", "2-3", "4-5", "more than 5"

3. *Exposure to the label*: "yes" or "no"

During the analysis, "never", "0-1" and "no" from the three dimensions respectively received 0 (zero) points. The rest of the answers were given 1, 2, 3 and 4

points as the scale raised. The three answers which received 0 points were rated as such because they were considered to contribute a minimum to the formation of *categories*. Each category could, therefore, receive a score from 0 to 8 points. The three dimensions varied in their ratings: for Frequency 0-4, Number of Exemplars 0-3 and Exposure to Label 0 or 1.

C.2 The Questionnaire

The CEQ can be seen below:

Category Experience

In this questionnaire we aim to discover the amount of experience your child has had with different categories of objects and animals. We are interested in experience with a category in all different sorts of contexts: from real objects and toys to pictures and television programs; and from cases when you have pointed out an object to those cases when your child was left to observe an object on his/her own. Please note that we don't want you to judge whether your child knows a word (for example "cat"), but the actual things that word is used to refer to.

For each item on the list we would like you to answer three questions:

1. How often your child sees a member of a given category, selecting a frequency from *never* to *every day*.

2. How many examples of that category your child has seen in his/her lifetime, selecting a number from *0* to *more than 5*.

3. Whether your child has ever heard a word for that category, by only selecting *yes* or *no* depending on whether the child has actually ever heard an item being named or not.

Make your selection by filling in the circle at the left of the word.

An example is given in order to make this clearer; let's take the item "cat":

For question one, if your child sees granny's cat once every couple of weeks and a neighbour's cat once every couple of weeks, you would choose *once a week* as an average frequency. For question two, following the same logic, you would say that your child has seen *2* cats in his/her lifetime.

If you have a book with different pictures of cats that you child enjoys looking at quite often but not every single day, you would choose *several times a week* for frequency and *more than five* for number of examples.

It is often the case that a child has daily experience with an object but has never actually heard a word for it. For example, you might have a cat at home and never refer to it as "cat" but instead use its name, say "tobbie". In that case, your child would have only heard the cat's name, but not the actual word "cat". So you will answer *no* to question three.

1

OXFORD UNIVERSITY BABYLAB
CATEGORY EXPERIENCE QUESTIONNAIRE

Subject Code
……..

Your name: …...…… Today's date: …../…../…..

Child's name: …… Birth date of child: …./…./….

Male/ Female: ……...

Animals						
bear	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
bee	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
bird	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
bunny/rabbit	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
butterfly	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
cat	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
chicken	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
cow	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
deer	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
dog	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			

donkey	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
duck	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
elephant	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
fish	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
frog	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
giraffe	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
goose	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
horse	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
kitten	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
lamb	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
lion	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
monkey	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
mouse	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
owl	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			

penguin	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
pig	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
pony	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
puppy	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
sheep	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
spider	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
squirrel	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
tiger	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
turkey	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
turtle	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			

Vehicles

aeroplane	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
bicycle	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
boat	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			

lorry / truck	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
motor-bike	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
bus	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
car	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
fire engine	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
pushchair	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
train	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			

Toys

ball	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
balloon	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
block / brick	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
book	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
doll	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
pen	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			

teddy bear	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			

Food

apple	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
banana	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
biscuit	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
bread	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
cake	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
carrot	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
orange	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			

Clothes

bib	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
boot(s)	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
coat	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
jeans	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
sweater	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			

necklace	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
pyjamas	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
shirt	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
dress	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
glasses	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
hat	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
jacket	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
shoe	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
shorts	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
sock	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
trousers	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
Furniture						
bed	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
chair	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			

cot	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
high chair	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
potty	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
fridge	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
rocking chair	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
settee / sofa	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
table	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
TV	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			

Outside

bucket	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
flower	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
	Word?	**O** no	**O** yes			
stone	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
swing	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
tree	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			

Household items

bin	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
bottle	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
brush	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
clock	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
comb	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
cup	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
dish	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
dummy	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
fork	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
glass	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
hammer	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
vacuum	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
jug	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
key	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			

lamp	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
mug	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
bowl	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
box	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
broom	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
pillow	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
plate	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
radio	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
scissors	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
soap	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
spoon	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
telephone	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
toothbrush	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			
watch	*How often?*	**O** never	**O** once a month	**O** once a week	**O** several times a week	**O** every day
	How many?	**O** 0-1	**O** 2-3	**O** 4-5	**O** More than 5	
	Word?	**O** no	**O** yes			

C.3 Infant Data

Before Experiments Three to Five were conducted, data from 50 infants were obtained using this questionnaire. These data were useful in choosing the test objects. The questionnaire can, however, also be a useful tool in future studies with babies at this age, in order to better control for level of familiarity with different objects/categories of objects.

C.3.1 Participants

Parental reports for fifty 10-month-old infants were obtained; the two sexes were roughly equally represented in the sample (23 males - 27 females). The mean age of those infants was 10.00 months ($SD = 0.35$, $R = 8.83 - 10.64$). Infants were recruited through the Oxford Babylab parental database, in the same way as in Experiments 1-5.

C.3.2 Scoring

As described above, for each category the three dimensions, Frequency - Number of exemplars - Exposure to label, were rated on a scale from 0-4, 0-3 and 0-1, respectively. Then the three dimensions were summed up to provide a single CEQ score for each object category (from 0 to 8).

C.3.3 Analysis

The data from the fifty infants were aggregated by object category, super-ordinate category (the eight groups of object categories in the CEQ - animals, vehicles, toys,

food, clothes, furniture, outside and household items) and gender.

A list of the categories in the CEQ was created on the basis of the CEQ Score in descending order. Percentages from the CDI scores in the LEX2005 database were also provided and a correlation analysis was run in order to test whether the CEQ data and the CDI data share any meaningful relationship.

In order to test exposure to different super-ordinate categories and whether the two genders differed in their experience with different objects, a Univariate ANOVA was conducted with CEQ Score as a dependent variable and super-ordinate category and gender as fixed factors.

Normality tests were run and all groups were normally distributed.

C.3.4 Results

Super-Ordinate Category and Gender Effects

Figure C.1 shows the percentage contribution of each super-ordinate category in the total CEQ mean. In other words, if the whole questionnaire is considered to be 100%, then each super-ordinate category has a contribution in this 100 depending on how high its mean score is. For example, if infants happen to be exposed to lots of animals, then the mean CEQ score of the "Animal" category will be quite high and it will contribute a high percentage to the total.

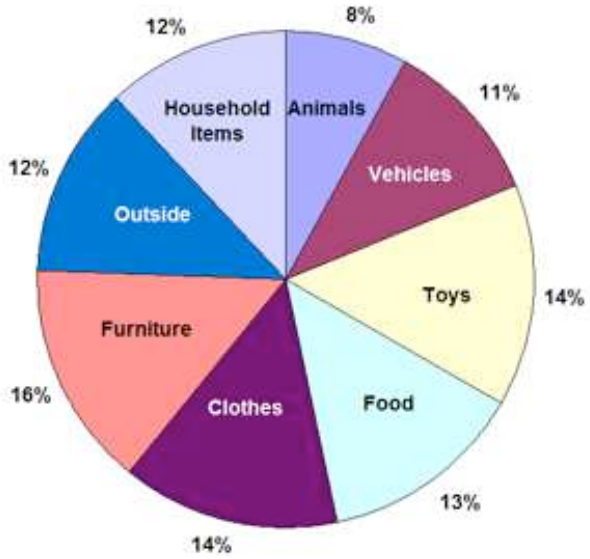


Figure C.1: Percentage contribution of each super-ordinate category to the total CEQ score.

It appears that infants have increased exposure to four of those super-ordinate categories; Furniture (16%), Clothes and Toys (14%) and Food (13%). Quite surprisingly, animals have not contributed highly to the total, although some of the first words of infants are animal words (e.g. "cat" and "dog"). However, the animal super-ordinate category in the CEQ consists of a large number of basic-level categories (34) which results in increased variance in the dataset and a rather low mean CEQ score. The super-ordinate categories with the highest mean consist of fewer object categories which are those most familiar to infants, rendering their mean score quite high.

Table C.1 shows the mean CEQ Score for each super-ordinate category and each gender separately. It appears that the two genders are exposed to rather similar object categories, while infants as a group seem to also have similar experience with the different super-ordinate categories, except from the animal category. In order to test these observations, a Univariate ANOVA with CEQ Score as a dependent variable and Super-ordinate Category and Gender as fixed factors was conducted.

Super-Ordinate Category	Females		Males	
	Mean	SD	Mean	SD
Animals	2.73	1.49	2.45	1.45
Vehicles	3.63	1.55	3.31	1.78
Toys	4.76	1.99	4.45	2.29
Food	4.45	1.37	4.19	1.65
Furniture	4.77	2.23	4.54	2.09
Clothes	4.77	1.44	4.29	2.12
Outside	3.83	2.15	4.06	2.25
Household Items	3.84	1.81	3.83	1.91

Table C.1: Mean CEQ score for each Super-Ordinate Category in the questionnaire presented for each gender separately.

A main effect of Super-Ordinate Category was found, $F(7, 218) = 20.54$, $p < .001$, ${\eta_p}^2 = .18$, but no main effect of Gender, $F(1, 218) = 1.67$, $p = .46$, and no interaction between the two, $F(7, 218) = .27$, $p = .99$.

On the basis of the main effect of Super-Ordinate Category, Bonferroni comparisons were run for all pairs of categories. "Animals" differed from all other categories of objects at the .05 level. The only other pairs which were found to be different were "vehicles" and "clothes" ($p < .05$) and "vehicles" and "furniture" ($p < .05$).

Items Analysis

Although infants were reported to be exposed more to some super-ordinate categories than others, individual object categories might not necessarily follow this pattern. Table C.2 shows all the object categories of the CEQ ranked on the basis of their mean score - also the MacArthur CDI Score is presented alongside as reported in the LEX2005 database.

Figure C.2 shows how CDI score changes as CEQ score increases. It appears that the two reports seem to be positively related, although data from these two questionnaires are taken from infants in different countries and possibly populations. In order to evaluate whether there was any meaningful relationship between CEQ and CDI score a Pearson r correlation analysis was conducted. This showed a highly significant correlation between CEQ and CDI score, $r(117) = .58$, $p < .001$.

Object Category	Total CEQ Score	CDI Score
Book	7.48	50.7
Sock	7.24	32.8
Bib	7.22	26.9
Spoon	7.14	29.9
Bowl	6.78	9.0
Trousers (Pants)	6.72	16.4
Car	6.66	46.3
Tree	6.52	9.0
Chair	6.44	17.9
Teddy Bear	6.30	32.8
Telephone	6.18	35.8
Highchair	6.16	29.9

Cup	6.16	41.8
Bed	6.08	31.3
Table	6.08	11.9
Bottle	6.02	68.7
Banana	6.00	46.3
Pushchair	5.96	
Ball	5.94	70.1
Shoe	5.74	40.3
TV	5.74	20.9
Flower	5.74	13.4
Bird	5.52	34.3
Bread	5.52	16.4
Settee/Sofa (Couch)	5.50	11.9
Cot (Crib)	5.46	32.8
Coat	5.44	7.5
Cat	5.38	44.8
Duck	5.30	23.9
Pajamas	5.30	22.4
Plate	5.30	10.4
Key	5.26	40.3
Apple	5.04	23.9
Dog	4.98	65.7
Glasses	4.94	13.4
Bin	4.9	
Biscuit	4.88	43.3
Hat	4.84	20.9

Toothbrush	4.82	17.9
Mug	4.56	
Block/Brick	4.54	17.9
Box	4.48	9.0
Fish	4.36	16.4
Watch	4.22	10.4
Bus	4.18	6.0
Fridge (Refrigerator)	4.16	16.4
Carrot	4.10	20.9
Bunny/Rabbit	3.98	26.9
Dish	3.98	9.0
Jeans	3.96	0
Jacket	3.94	10.4
Cow	3.80	16.4
Pillow	3.78	13.4
Pen	3.76	16.4
Glass	3.76	13.4
Swing	3.72	17.9
Brush	3.68	19.4
Shirt	3.56	14.9
Elephant	3.54	9.0
Monkey	3.52	4.5
Lamp	3.42	6.0
Pig	3.40	11.9
Sheep	3.40	7.5
Clock	3.36	11.9

Fork	3.32	7.5
Radio	3.30	11.9
Dress	3.28	4.5
Train	3.14	7.5
Sweater	3.14	7.5
Lorry/Truck	3.10	17.9
Bicycle	3.02	10.4
Bear	2.98	32.8
Horse	2.98	10.4
Necklace	2.92	4.5
Soap	2.92	10.4
Giraffe	2.82	3.0
Vacuum	2.82	16.9
Lion	2.78	7.5
Boat	2.78	
Dummy	2.76	
Chicken	2.74	9.0
Shorts	2.68	6
Stone	2.58	
Airplane	2.54	11.9
Orange	2.52	1.5
Butterfly	2.44	3.0
Balloon	2.34	35.8
Bee	2.24	6.0
Mouse	2.20	4.5
Cake	2.20	7.5

Motorbike (Motorcycle)	2.16	1.5
Tiger	2.12	9.0
Frog	2.10	6
Boots	1.96	1.5
Doll	1.90	19.4
Spider	1.88	
Lamb	1.86	9.0
Penguin	1.62	3.0
Scissors	1.60	0
Owl	1.58	4.5
Squirrel	1.50	3.0
Broom	1.50	7.5
Fire Engine (Firetruck)	1.46	4.5
Donkey	1.42	4.5
Goose	1.38	7.5
Bucket	1.34	
Puppy	1.10	22.4
Comb	1.08	14.9
Rocking Chair	1.02	6.0
Kitten	0.98	38.8
Jug	0.86	
Turtle	0.80	7.5
Hammer	0.80	0
Pony	0.62	6.0
Deer	0.50	3.0
Potty	0.46	5.0

Turkey	0.18	1.5

Table C.2: The object categories in the CEQ listed on the basis of their mean in descending order. The CDI score as reported in the LEX2005 database is also provided alongside. CEQ and CDI score were shown to be highly correlated; however, the two questionnaires are not identical: some categories in the CEQ (and the Oxford CDI) do not exist in the CDI (and vice versa), while some other words are present but using the American English word. In the former case, the value for the CDI is left blank; in the latter case, the words which are different in the CDI are provided next to the CEQ words in brackets.

Additionally, although the two genders were shown not to differ in the overall means of the eight super-ordinate categories, the top 10 object categories for each gender were calculated and are presented in Table C.3. It appears that the two genders do not differ much neither as far as individual object categories are concerned; eight out of ten categories with the highest mean are the same, while one other is a different version of the category, "chair" vs. "high chair".

C.3.5 Discussion

The Category Experience Questionnaire was created in order to evaluate the prior exposure of infants to different object categories and have some control over prior knowledge of the visual and auditory stimuli in Experiments 3-5.

Although in Experiments 3-5 parental reports in the CEQ were shown not to have any meaningful relationship with performance in the task (word learning, word

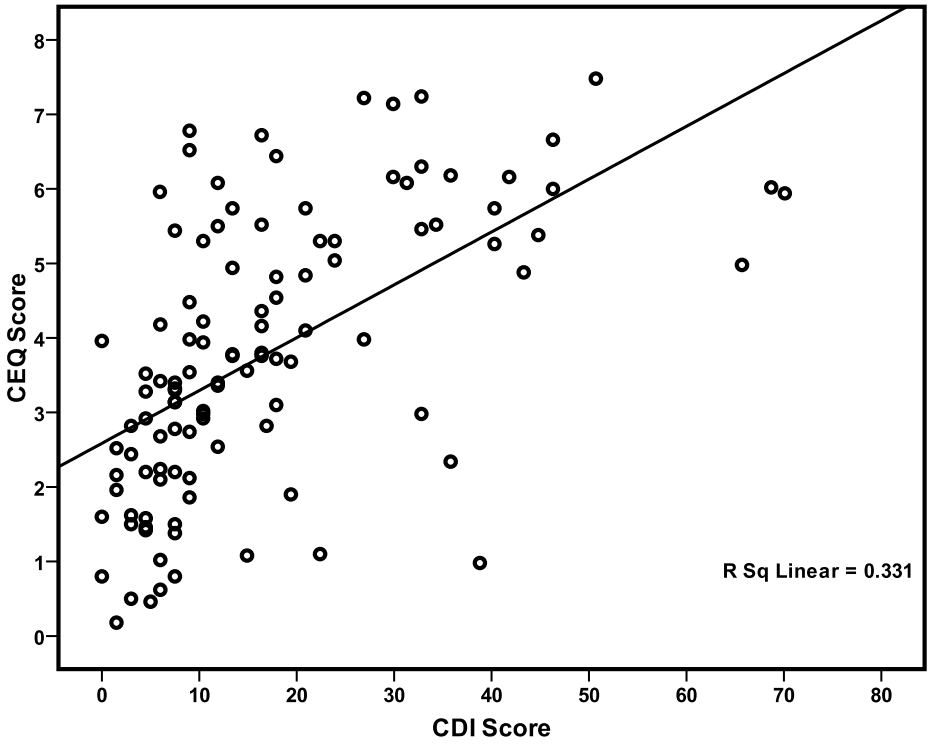


Figure C.2: Relationship between CEQ Score and CDI Score (LEX2005 Database)

Females		Males	
Object Category	Total CEQ Score	Object Category	Total CEQ Score
Book	7.65	Bib	7.73
Bib	7.04	Sock	7.73
Bowl	7.04	Book	7.32
Spoon	7.00	Spoon	7.23
Sock	6.85	Trousers	6.96
Chair	6.81	Car	6.91
Trousers	6.50	Tree	6.82
Teddy Bear	6.46	Bowl	6.46
Car	6.39	High Chair	6.27
Tree	6.27	Ball	6.14

Table C.3: Top 10 object categories with the highest CEQ Score for the two genders separately.

generalisation and categorisation, respectively), a highly significant correlation was found here between CEQ score and CDI score - even when these reports were from parents in different countries. This result suggests that the CEQ could be useful and informative about infants' experience with different categories, but the age of infants in these studies possibly obscures this. As the Oxford CDI in this thesis shows no relationship with infant performance but has been repeatedly shown to be a good predictor of word comprehension in the lab for older children (Styles & Plunkett, Under Review), it could be that, if the CEQ is used with older infants, it can be a useful tool in controlling infants' prior experience with different categories. This remains to be tested.

The super-ordinate categories in this questionnaire seem to differ between them, with some appearing more familiar than others. This, however, can be slightly misleading, since some categories, such as "animals", include many more exemplars than others, such as "food". The latter only include the most common categories infants may encounter and, therefore, end up having a higher mean than other categories which are characterised by higher variance in their mean CEQ score. An items analysis might, therefore, provide a more appropriate use of this questionnaire, in parallel to how the CDI is mainly used.

The two genders seem to be exposed to very similar categories overall, and it seems that we could only expect to find small differences for individual object categories, like for example "doll" and "dress" which are very female-oriented.

References

Anglin, J. M. (1986). Semantic and conceptual knowledge underlying the child's words. In S. Kuczaj & M. D. Barrett (Eds.), *The development of word meaning.* New York: Springer-Verlag.

Balaban, M. T., & Waxman, S. R. (1997). Do words facilitate object categorization in 9-month-old infants? *Journal of experimental child psychology, 64*(1), 3-26.

Ballem, K. D., & Plunkett, K. (2005). Phonological specificity in children at 1;2. *Journal of Child Language, 32*(1), 159-173.

Barrett, M. D. (1983). The early acquisition and development of the meanings of action-related words. In T. Seiler & W. Wannenmacher (Eds.), *Concept development and the development of word mearning.* Berlin: Springer-Verlag.

Barrett, M. D. (1986). Early semantic representations and early word usage. In S. Kuczaj & M. D. Barrett (Eds.), *The development of word meaning.* New York: Springer-Verlag.

Barsalou, L. W. (1987). The instability of graded structure: implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (p. 101-140). Cambridge: Cambridge University Press.

Bates, E., Benigni, I., L. Betherton, Camaioni, L., & Volterra, V. (1979). *The emergence of symbols: Cognition and communication in infancy.* New York: Academic.

Behrend, D. A., Scofield, J., & Kleinknecht, E. E. (2001). Beyond fast mapping: Young children's extensions of novel words and novel facts. *Developmental Psychology*, *37*(5), 698-705.

Berlin, B., & Kay, P. (1969). *Basic color terms* (U. of California Press, Ed.). Berkeley and Los Angeles.

Berlyne, D. E. (1958). The influence of complexity and novelty in visual figures on orienting responses. *Journal of Experimental Psychology*, *55*(3), 289-296.

Bloom, L. (1973). *One word at a time: The use of single word utterances before syntax.* The Hague: Mouton.

Bloom, L. (1993). *The transition from infancy to language: Acquiring the power of expression.* New York, NY: Cambridge University Press.

Bloom, P. (2004). Myths of word learning. In G. D. Hall & S. R. Waxman (Eds.), *Weaving a lexicon.* Cambridge, MA: MIT Press.

Bowerman, M. (1977). The acquisition of rules governing "possible lexical items": Evidence from spontaneous speech errors. *Papers and Reports on Child Language Development*, *13*, 148-156.

Bowerman, M. (1978). The acquisition of word meaning: An investigation of some current conflicts. In N. Waterson & C. Snow (Eds.), *Development of communication.* New York: John Wiley.

Bridgeman, G., Hendry, D., & Stark, L. (1975). Failure to detect displacement of visual world during saccadic eye movements. *Vision Research*, *15*, 719-722.

Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (p. 169-211). New York: Wiley.

Brooks, L. R. (1987). Nonanalytic cognition. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (p. 141174). Cambridge: Cambridge University Press.

Brown, R. (1957). Linguistic determinism and the part of speech. *Journal of Abnormal and Social Psychology*, *55*, 1-5.

Canfield, R. L., Smith, E. G., Brezsnyak, M. P., & Snow, K. L. (1997). Information processing through the first year of life: A longitudinal study using the visual expectation paradigm. *Monographs of the Society for Research in Child Development*, *62, No2*.

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

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121-152.

Clark, E. V. (1973). Non-linguistic strategies and the acquisition of word meanings. *Cognition*, *2*(2), 161-182.

Cohen, L. (2003). Comment on part i: Unresolved issues in infant categorization. In D. Rakison & L. Oakes (Eds.), *Early category and concept development: Making sense of the blooming, buzzing confusion.* New York: Oxford University Press.

Cohen, L. (2004). Uses and misuses of habituation and related preference paradigms. *Infant and Child Development*, *13*, 349-352.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: newborns prefer their mothers' voices. *Science*, *208*(4448), 1174-1176.

Eimas, P. D., & Quinn, P. C. (1994). Studies on the formation of perceptually based basic-level categories in young infants. *Child Development*, *65*(3), 903-917.

Fennell, C. T., & Werker, J. F. (2004). Infant attention to phonetic detail: Knowledge and familiarity effects. *Proceedings of the 28th Annual Boston University Conference on Language Development*.

Fenson, L., Dale, P., Reznik, S., Bates, E., Thal, D., & Pethick, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59*(242).

Fernald, A. (1992). Human maternal vocalizations to infants as biologically relevant

signals: An evolutionary perspective. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (p. 391-428). New York, NY: Oxford University Press.

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (2001). Rapid gains in speed of verbal processing by infants in the 2nd year. In M. Tomasello & E. Bates (Eds.), *Language development: The essential readings* (p. 49-56). Malden, MA: Blackwell Publishing.

Fillmore, C. (1982). Frame semantics. *Linguistics in the Morning Calm*, *ed. by The Linguistic Society of Korea*(Soeul: Hanshin), 111-137.

French, R. M., Mareschal, D., Mermillod, M., & Quinn, P. C. (2004). The role of bottom-up processing in perceptual categorization by 3- to 4-month-old infants: Simulations and data. *Journal of Experimental Psychology: General*, *133*(3), 382-397.

Fulkerson, A. L., & Haaf, R. A. (2003). The influence of labels, non-labeling sounds, and source of auditory input on 9- and 15-month-olds' object categorization. *Infancy*, *4*, 349-369.

Fulkerson, A. L., & Waxman, S. R. (2007). Words (but not tones) facilitate object categorization: Evidence from 6- and 12-month-olds. *Cognition*, *105*, 218-228.

Gelman, S. A., & Markman, E. M. (1986). Young children's inductions from natural kinds: The role of categories and appearances. *Cognition*, *23*, 183209.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj (Ed.), *Language development: Vol. 2. language, thought, and culture* (p. 301-334). Hillsdale, NJ: Erlbaum.

Gibson, E. J. (1967). *Principles of perceptual learning and development.* New York: Appleton Century Crofts.

Golinkoff, R. M., & Hirsh-Pasek, K. (1987a). Early object labels: the case for a developmental lexical principles framework. *Journal of Child Language*, *21*,

125-155.

Golinkoff, R. M., & Hirsh-Pasek, K. (1987b). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, *14*, 23-45.

Gopnik, A., & Meltzoff, A. (1986). Words, plans, things and locations: Interaction between semantic and cognitive development in the one-word stage. In S. Kuczaj & M. Barrett (Eds.), *The development of word meaning.* New York: Springer-Verlag.

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

Hampton, J. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *18*, 441-461.

Hampton, J. (1987). Inheritance of attributes in natural concept conjunctions. *Memory and Cognition*, *15*, 55-71.

Harris, J. (1943). Habituation response decrement in the intact organism. *Psychological Bulletin*, *40*, 385-422.

Heider, E. R. (1971). "focal" color areas and the development of color names. *Journal of Experimental Psychology*, *4*, 447-455.

Heider, E. R. (1972). Universals in color naming and memory. *Journal of Experimental Psychology*, *93*, 10-20.

Heider, K. G. (1970). *The dugum dani: a papuan culture in the highlands of west new guinea*. Chicago: Aldine.

Hollich, G., Hirsh-Pasek, K., & Golinkoff, R. M. (2000). Breaking the language barrier: an emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development*, *65*(3, Serial No. 262).

Houston-Price, C., Mather, E., & Sakkalou, E. (2007). Discrepancy between parental

reports of infants' receptive vocabulary and infants' behaviour in a preferential looking task. *Journal of Child Language*, *34*, 701-724.

Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development*, *13*(4), 341-348.

Hunt, J. (1970). Attentional preference and experience: I. introduction. *Journal of Genetic Psychology*, *117*, 99-107.

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in infancy research*, *5*, 69-95.

Hunter, M. A., Ames, E. W., & Koopman, R. (1983). Effects of stimulus complexity and familiarization time on infant preferences for novel and familiar stimuli. *Developmental Psychology*, *19*, 338-352.

Irwin, D. E. (1998). Lexical processing during saccadic eye movements. *Cognitive Psychology*, *36*, 1-27.

Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends in Cognitive Sciences*, *3*(9), 323-328.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*(1), 1-23.

Kay, D., & Anglin, J. (1982). Overextension and underextension in the child's expressive and receptive speech. *Journal of Child Language*, *9*, 8398.

Keil, F. C. (1994). Explanation, association, and the acquisition of word meaning. In L. R. Gleitman & B. Landau (Eds.), *Lexical acquisition* (p. 169-197). Cambridge, MA: MIT Press.

Kuczaj, S. A. (1986). Thoughts on the intensional basis of early object word extension: Evidence from comprehension and production. In S. Kuczaj & M. D. Barrett (Eds.), *The development of word meaning.* New York: Springer-Verlag.

Lalonde, C. E., & Werker, J. F. (1995). Cognitive influences on cross-language speech perception in infancy. *Infant Behaviour and Development*, *18*, 459-75.

Lynch, E. B., Coley, J. D., & Medin, D. L. (2000). Tall is typical: Central tendency, ideal dimensions and graded category structure among tree experts and novices. *Memory and Cognition*, *28*(1), 41-50.

Madole, K. L., & Oakes, L. M. (1999). Making sense of infant categorization: Stable processes and changing representations. *Developmental Review*, *19*, 263296.

Mandel, D. R., Jusczyk, P. W., & Pisoni, D. B. (1995). Infants' recognition of the sound patterns of their own names. *Psychological Science*, *6*, 314-317.

Mandler, J. M. (2000). Perceptual and conceptual processes in infancy. *Journal of Cognition and Development*, *1*, 3-36.

Mandler, J. M., & McDonough, L. (1993). Concept formation in infancy. *Cognitive Development*, *8*, 291318.

Mani, N., & Plunkett, K. (2007). Phonological specificity of consonants and vowels in early lexical representations. *Journal of Memory and Language*, *57*, 252-272.

Mareschal, D., Johnson, M., Sirois, S., Spratling, M., Thomas, M., & Westermann, G. (2007). *Neuroconstructivism* (Vol. I: How the brain constructs cognition). Oxford, UK: Oxford University Press.

Markman, A. B. (1999). *Knowledge representation.* Mahwah, NJ: Erlbaum.

Markman, E. M. (1989). *Categorisation and naming in children: problems of induction.* Cambridge, MA: MIT Press.

Markman, E. M. (1991). The whole-object, taxonomic, and mutual exclusivity assumptions as initial constraints on word meanings. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on language and thought: Interrelations in development* (p. 72-106). New York, NY: Cambridge University Press.

Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology*, *16*(1), 1-27.

Markman, E. M., Wasow, J. L., & Hansen, M. B. (2003). Use of the mutual exclusivity

assumption by young word learners. *Cognitive Psychology*, *47*(3), 241-275.

Mather, E., & Plunkett, K. (July, 2007). Mutual exclusivity as an intermodal preferential looking task. *Poster presented to the 30th Child Language Seminar, University of Reading, UK.*

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, *38*(4), 465-494.

Medin, D. L., & Barsalou, L. W. (1987). Categorization processes and categorization perception. In S. Harnad (Ed.), *Categorical perception* (p. 455-490). Cambridge: Cambridge University Press.

Medin, D. L., & Schaffer, M. M. (1978). A context theory of classification learning. *Psychological Review*, *85*, 207-238.

Medin, D. L., & Smith, E. E. (1984). Concepts and concept formation. *Annual Review of Psychology*, *35*, 113-138.

Meints, K., Plunkett, K., & Harris, P. (1999). When does an ostrich become a bird? the role of typicality in early word comprehension. *Developmental Psychology*, *35*(4), 1072-1078.

Meints, K., Plunkett, K., Harris, P. L., & Dimmock, D. (2002). What is on and under to 15-, 18- and 24-month-olds? typicality effects in early comprehension of spatial prepositions. *British Journal of Developmental Psychology*, *20*, 113-130.

Merriman, W. (1986). How children learn the reference of concrete nouns: A critique of current hypotheses. In S. A. Kuczaj & M. D. Barrett (Eds.), *The development of word meaning.* New York: Springer-Verlag.

Merriman, W., & Bowman, L. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development*, *54*(3-4), 1-132.

Mervis, C. B., & Bertrand, J. (1995). Early lexical acquisition and the vocabulary spurt: A response to goldfield and reznick. *Journal of Child Language*, *22*, 461-68.

Mervis, C. B., Johnson, K. E., & Scott, P. (1993). Perceptual knowledge, conceptual knowledge and expertise: Comment on jones and smith. *Cognitive Development*, *8*, 149-155.

Mills, D. L., Plunkett, K., Prat, C., & Schafer, G. (2005). Watching the infant brain learn words: Effects of vocabulary size and experience. *Cognitive Development*, *20*(1), 19-31.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

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

Nelson, C. A. (1995). The ontogeny of human memory: A cognitive neuroscience perspective. *Developmental Psychology*, *31*, 723-738.

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development*, *38*.

Nelson, K., Rescorla, L., Gruenderl, J., & Benedict, H. (1978). Early lexicons: What do they mean? *Child Development*, *49*, 960-968.

Oakes, L. M., & Madole, K. L. (2003). *Principles of developmental change in infants' category formation* (L. M. Oakes & D. H. Rakison, Eds.). New York: Oxford University Press.

Oviatt, S. L. (1980). The emerging ability to comprehend language: An experimental approach. *Child Development*, *51*(1), 97-106.

Plunkett, K., Hu, J.-F., & Cohen, L. (2008). Labels can override perceptual categories in early infancy. *Cognition*, *106*(2), 665-81.

Pruden, S. M., Hirsh-Pasek, K., Golinkoff, R. M., & Hennon, E. A. (2006). The birth of words: ten-month-olds learn words through perceptual salience. *Child*

Development, *77*(2), 266-280.

Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.

Quinn, P. C., & Eimas, P. D. (1997). A reexamination of the perceptual-to-conceptual shift in mental representations. *Review of General Psychology*, *1*, 271287.

Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception*, *22*(4), 463-475.

Quinn, P. C., & Oates, J. (2004). Early category representation and concepts. In J. Oates & A. Grayson (Eds.), *Cognitive and language development in children* (p. 21-60). Oxford: Blackwell Publishing Ltd.

Quinn, P. C., Westerlund, A., & Nelson, C. A. (2006). Neural markers of categorization in 6-month-old infants. *Psychological Science*, *17*(1), 59-66.

Rakison, D. H. (2003). Parts, motion and the development of the animate/inanimate distinction in infancy. In D. H. Rakison & L. M. Oakes (Eds.), *Early category and concept development: Making sense of the blooming, buzzing confusion* (p. 159-192). New York, NY, US: Oxford University Press.

Rescorla, L. (1980). Overextensions in early language development. *Journal of Child Language*, *7*, 321-335.

Rips, L., & Medin, D. (2005). Concepts, categories, and semantic memory. In K. Holyoak & R. Morrison (Eds.), *Cambridge handbook of thinking and reasoning.* Cambridge: Cambridge University Press.

Rips, L., Shoben, E., , & Smith, E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, *12*, 1-20.

R.M., N., & Zaki, S. (2002). Exemplar and prototype models revisited: Response strategies, selective attention and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *28*, 924-940.

Robinson, C. W., & Sloutsky, V. M. (2004). Auditory dominance and its change in

the course of development. *Child Development*, *75*(5), 1387-1401.

Roder, B., Bushnell, E., & Sasseville, A. (2000). Infants' preferences for familiarity and novelty during the course of visual processing. *Infancy*, *1*(4), 491-507.

Rosch, E. (1973). Natural categories. *Cognitive Psychology*, *4*, 328-50.

Rosch, E. (1998). Principles of categorization. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (p. 27-48). Hillsdale, N.J.: Erlbaum Associates.

Rosch, E., Mervis, C., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382-439.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573-605.

Samuelson, L. K., & Smith, L. B. (1998). Memory and attention make smart word learning: An alternative account of akhtar, carpenter, and tomasello: Erratum. *Child Development*, *69*(3).

Schafer, G. (2005). Infants can learn decontextualized words before their first birthday. *Child Development*, *76*(1), 87-96.

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

Schafer, G., Plunkett, K., & Harris, P. (1999). What's in a name? lexical knowledge drives infants' visual preferences in the absence of referential input. *Developmental Science*, *2*(2), 187-194.

Sirois, S., & Mareschal, D. (2004). An interacting systems model of infant habituation. *Journal of Cognitive Neuroscience*, *16*, 1352-1362.

Slater, A. (2004). Novelty, familiarity and infant reasoning. *Infant and Child Development*, *13*, 353-354.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge,MA: Harvard University Press.

Smith, L. B., Colunga, E., & Yoshida, H. (2003). Making an ontology: Cross-linguistic

evidence. In D. H. Rakison & L. M. Oakes (Eds.), *Early category and concept development: Making sense of the blooming, buzzing confusion* (p. 275-302). New York, NY, US: Oxford University Press.

Smith, L. B., Sera, M., & Gattuso, B. (1988). *The development of thinking* (R. J. Sternberg & E. E. Smith, Eds.). New York, NY: Cambridge University Press.

Sokolov, E. N. (1975). The neuronal mechanisms of the orienting reflex. In E. N. Sokolov & O. S. Vinogradova (Eds.), *Neuronal mechanisms of the orienting reflex* (p. 217-235). Hillsdale, NJ: Erlbaum.

Sokolov, E. N., Spinks, J. A., Naatanen, R., & Lyytinen, H. (2002). *The orienting response in information processing*. New Jersey: Lawrence Erlbaum Associates.

Stager, C., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word learning tasks. *Nature*, *388*, 381-382.

Styles, S., & Plunkett, K. (Under Review). What is word understanding for the parent of a one-year-old? matching the difficulty of a lexical comprehension task to parental cdi report. *Journal of Child Language*.

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, *76*(2), 147-166.

Swingley, D., Pinto, J., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, *71*, 73-108.

Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder. *Cognitive Psychology*, *23*, 457-482.

Taylor, J. R. (1995). *Linguistic categorization: Prototypes in linguistic theory*. New York: Oxford University Press.

Taylor, M., & Gelman, S. A. (1989). Incorporating new words into the lexicon: Preliminary evidence for language hierarchies in two-year-old children. *Child Development*, *60*, 625-636.

Thorpe, W. H. (1956). *Learning and instinct in animals*. London: Methuen.

Tincoff, R., & Jusczyk, A. M. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science*, *10*, 172-175.

Tomasello, M. (2003). *Constructing a language: a usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tomasello, M., & Barton, M. (1994). Learning words in non-ostensive contexts. *Developmental Psychology*, *30*, 639-650.

Tomasello, M., & Mervis, C. B. (1994). The instrument is great, but measuring comprehension is still a problem. commentary on fenson, dale, reznick, bates, thal and pethick, 1994. *Monographs of the Society for Research in Child Development*, *59*, 1749.

Volkmann, F. C. (1986). Human visual suppression. *Vision Research*, *26*(9), 1401-1416.

Wagner, S., & Sakovitz, L. (1986). A processing analysis of infant visual and cross-modal recognition: Implications of an amodal code. In L. P. Lipsitt & C. Rovee-Collier (Eds.), *Advances in infancy research* (Vol. 4, p. 195217). Norwood, NJ: Ablex.

Waxman, S. R. (2003). Links between object categorization and naming: Origins and emergence in human infants. In D. H. Rakison & L. M. Oakes (Eds.), *Early category and concept development: Making sense of the blooming, buzzing confusion* (p. 213-241). New York, NY, US: Oxford University Press.

Waxman, S. R., & Braun, I. (2005). Consistent (but not variable) names as invitations to form object categories: New evidence from 12-month-old infants. *Cognition*, *95*(3), B59-B68.

Waxman, S. R., & Kosowski, T. D. (1990). Nouns mark category relations: Toddlers' and preschoolers' word-learning biases. *Child Development*, *61*, 1461-1473.

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

Werker, J. F., Cohen, L., Lloyd, V., Casasola, M., & Stager, C. (1998). Acquisition of word-object associations by 14-month-old infants. *Developmental Psychology*, *34*(6), 1289-1310.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49-63.

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

Wierzbicka, A. (1985). *Lexicography and conceptual analysis*. Karoma: Ann Arbor.

Wittgenstein, L. (1978). *Philosophical investigations. translated by g.e.m. anscombe.* Oxford: Basil Blackwell.

Woodward, A. L., Markman, E. M., & Fitzsimmons, C. M. (1994). Rapid word learning in 13- and 18-month-olds. *Developmental Psychology*, *30*(4), 553-566.

Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition*, *85*, 223-250.

Xu, F., & Tenenbaum, J. B. (2007). Word learning as bayesian inference. *Psychological Review*, *114*(2), 245-272.

Yoshida, H., & Smith, L. B. (2005). Linguistic cues enhance the learning of perceptual cues. *Psychological Science*, *16*(2), 90-95.

Younger, B. A. (1993). Understanding category members as the same sort of thing: explicit categorisation in ten-month infants. *Child Development*, *64*, 309-320.

Younger, B. A. (2003). Parsing objects into categories: Infants' perception and use of correlated attributes. In D. H. Rakison & L. M. Oakes (Eds.), *Early category and concept development: Making sense of the blooming, buzzing confusion* (p. 77-102). New York, NY, US: Oxford University Press.

Younger, B. A., & Cohen, L. (1983). Infant perception of correlations among attributes. *Child Development*, *54*, 858-867.

Younger, B. A., & Cohen, L. B. (1986). Developmental change in infants' perception of correlations among attributes. *Child Development*, *57*, 803-815.

Zelazo, P. R., Weiss, M. J. S., & Tarquinio, N. (1991). *Habituation and recovery of neonatal orienting to auditory stimuli* (M. J. S. Weiss & P. R. Zelazo, Eds.). Norwood, NJ: Ablex.