Children's use of teleological explanations about the natural world

Jonathan Grant Halls

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy
February 2018

Abstract

There is a significant body of research on children's preconceptions concerning scientific concepts and the impact this can have upon their science education. One active issue concerns the extent to which young children's explanations for the existence of natural kinds rely on teleological explanations. These explanations either propose a purpose (e.g., muddy puddles are for jumping in) or suggest goal-driven behaviour (e.g., tigers decided to grow stripes for better camouflage). It has been argued that this teleological tendency is a major barrier to children's understanding of causality in the natural world, particularly when learning about evolution. This thesis details four studies which explore and seek to limit children's use of teleological explanations for natural kinds.

Study 1 investigated two factors (question wording and questioned topic) that may have led to an overestimation of children's use of teleological explanations for natural phenomena (i.e., time-constrained, natural events or processes such as snow, clouds or night). The findings indicated that children's (aged 5- to 8-years-old, n = 66) responses to open questions involved significantly fewer teleological answers than to the typical leading form used in prior research. Furthermore, the concept of teleology is more nuanced than often suggested, as levels of teleological explanation varied considerably within the category of natural phenomena. Consequently, young children may be more able to learn about causal explanations than the literature implies.

Study 1b used methods from corpus linguistics to explore the naturally occurring discourse surrounding the examples of natural phenomena investigated in Study 1. Teleological and scientific patterns of discussion were identified in a specially created corpus of children's discourse. The results of this analysis indicate that the patterns of language that children experience can influence their use and the form of teleological explanations.

Studies 2 and 3 consisted of efficacy tests of an intervention designed to limit children's endorsement of teleological explanations to account for the existence of a) natural phenomena (e.g., snow or rainbows) and b) organisms' traits (e.g., giraffes' necks or zebras' stripes). The interventions consisted of short discussion-based activities which sought to develop children's understanding of why teleological explanations are inappropriate to use in scientific contexts. (Study 2, 5- to 7-year-olds, n = 54; Study 3, 9- to 10-year-olds, n = 24). These studies suggest teleology need not be a major barrier to teaching and learning about causality in Early Years education. After only limited discussion about styles of explanation in science, the results revealed that children's teleological preconceptions were malleable to change, a finding which runs counter to current thinking on children and teleology.

In terms of a contribution to knowledge, it is proposed that children's use of teleological explanations is more nuanced than currently suggested by the literature. Furthermore, children's use of these explanations can be limited with a series of simple, short discursive lessons. Lessons that do not require additional pedagogical knowledge on the part of primary school teachers.

Acknowledgements

Firstly, I want to express my immense gratitude for the advice and supervision of Shaaron Ainsworth. Shaaron was involved with this thesis from the very beginning, even helping me to write the proposal despite me being 4500 miles away and without any regular internet access. It goes without saying that this thesis would not have existed without her irreplaceable input. While her guidance was much valued at every stage, both in relation to my doctoral work and broader career development. I am perhaps most indebted to Shaaron for her encouragement to ruminate on the nature of night as a type of weather.

Secondly, I am especially thankful for the involvement of Mary Oliver. Throughout the doctoral process Mary has provided unparalleled support and advice about working in educational settings and understanding scientific ideas. She has showed endless patience with my interpretation of and writing about evolution and trait function. I'm sure at times it was not easy to show such restraint.

Thirdly, I am forever beholden for the practical and emotional support provided by my wife, Katherine Crossley. It cannot have been easy for Kate when I was regularly late for our plans because "I've just got to finish this little bit before I leave", whilst neglecting to mention that a "little bit" could mean over 2 hours of analysis.

Fourthly, I'd like to thank my friends and colleagues at the Learning Sciences Research Institute and the wider School of Education. Special mentions should go to Jessica Mason for her advice on corpus linguistics and to Rafael De Albuquerque and Dimitrios Panagiotopoulos for their help with second coding of data.

Finally, I would like to acknowledge the schools, teachers and, most importantly, the children that were involved in these studies. The studies described here could not have taken place without their participation.

Table of Contents

1. Introduction 1

1.1 Conception of thesis 1

1.2 Overarching themes 1

1.3 Overview of chapters 2

2. Literature review 5

2.1 Situating teleology within science education 5

2.1.1 Science education in the current climate 5

2.1.2 Scientific explanations 6

2.1.3 Teleological explanations 7

2.1.4 Summary of section 8

2.2 History of teleological explanations about the natural world 8

2.2.1 Artificialism as teleology 9

2.2.2 Intuitive reasoning in developmental psychology 10

2.2.3 Misconceptions in education 11

2.2.4 Summary of section 13

2.3 Teleological explanations about different categories 14

2.3.1 Selective vs. promiscuous teleology 15

2.3.2 Using hierarchical classification 17

2.3.3 A proposed method of classification 18

2.3.4 Summary of section 22

2.4 Influences on teleological explanation 22

2.4.1 Influence of intentional artefacts 22

2.4.2 Influence of religion 23

2.4.3 Influence of evolution 25

2.4.4 Influence of language 29

2.4.5 Parental explanations 32
2.4.6 Summary of section 33
2.5 Constructing teleology 34
2.5.1 Naïve ideas 34
2.5.2 Types of teleology 39
2.5.3 Acceptability of teleology 42
2.5.4 Summary 48
2.6 Measuring teleology 50
2.6.1 Question wording 50
2.6.2 Questioned topic 52
2.6.3 Type of teleology 52
2.6.4 The construction of teleology 53
2.6.5 Teleology in natural discourse 53
2.6.6 Summary 54
2.7 Limiting teleological explanation 54
2.7.1 Teleology and the curriculum 55
2.7.2 Countering teleological explanation 57
2.7.3 Summary 58
2.8 Interventions in science education 59
2.8.1 Children's actual and preconceived understanding 59
2.8.2 Big ideas in science education 60
2.8.3 Evidence evaluation 60
2.8.4 Dialogic teaching 62
2.8.5 A discussion orientated intervention 63
2.8.6 Thinking skills programmes 65
2.8.7 Thinking skills programmes and the National Curriculum 68

2.8.8 Summary 69

3. Methodology 71

3.1 Introduction 71

3.2 Philosophical framework 72

3.3 Methods 73

3.3.1 Real world research 73

3.3.2 Experimental methods 74

3.3.3 Corpus linguistics 77

3.3.4 Statistical analyses 78

3.4 Working with gatekeepers (schools and parents) 79

3.5 Working with children 80

3.6 Ethics 81

3.7 Summary 82

4. Study 1: Influences on young children's teleological explanations for natural phenomena 83

4.1 Introduction 83

4.2 Method 84

4.2.1 Design 84

4.2.2 Materials 85

4.2.3 Participants 87

4.2.4 Procedure 88

4.2.5 Measures 88

4.3 Results 90

4.3.1 Influence of question wording 90

4.3.2 Influence of questioned topic 91

4.3.3 Type of teleology 94

4.3.4 Examples of children's explanations 97

4.4 Discussion 103

4.4.1 Influence of question wording 104

4.4.2 Influence of questioned topic 105

4.4.3 Type of teleology 107

4.4.4 Conclusion 108

5. Study 1b: Naturally occurring discourses surrounding natural phenomena 111

5.1 Introduction 111

5.1.1 Corpus linguistics 113

5.1.2 Conceptual metaphor 114

5.1.3 Study aim 115

5.2 Method 115

5.2.1 Sample: building the corpus 116

5.2.2 Procedure: Searching the corpus 117

5.2.3 Measures 117

5.3 Findings 120

5.3.1 Keyword frequencies 121

5.3.2 Keyword concordances: explicit teleological and scientific explanations 122

5.3.3 Keyword collocates: themes in discourse around natural phenomena 125

5.3.4 Conceptual metaphor 132

5.3.5 Conclusion 135

6. Study 2: Improving children's evaluation of explanations for the existence of natural phenomena 137

6.1 Introduction 137

6.2 Intervention design 138

6.2.1 Methods to structure group discussion 138

6.2.2 Terminology 140

6.2.3 Assessment 140
6.2.4 Selection of assessment and session discussion topics 141
6.2.5 Summary of intervention design 142
6.3 Method 142
6.3.1 Pilot study 142
6.3.2 Main study 145
6.4 Results 152
6.4.1 Data checking 152
6.4.2 Analysing children's judgements 152
6.4.3 Analysing children's rationales 153
6.5 Discussion 154
6.5.1 Assessment results 154
6.5.2 Evaluation of the intervention 157
6.5.3 Conclusion 158
7. Study 3: Improving children's evaluation of explanations for the existence of organisms' traits 161
7.1 Introduction 161
7.1.1 Evolution and teleological explanation 161
7.1.2 The evolutionary scope of Study 3 163
7.1.3 Study 3 aim 165
7.2 Method 167
7.2.1 Preliminary study 167
7.2.2 Main study 182
7.3 Results 189
7.3.1 Data checking 189
7.3.2 Children's judgements 190

7.3.3 Children's rationales 191

7.4 Discussion 193

7.4.1 Assessment results 193

7.4.2 Intervention evaluation 195

7.4.3 Conclusion 196

8. Discussion 197

8.1 Measuring teleology 197

8.1.1 RQ1: Does question wording influence children's explanations about the existence of natural phenomena? 197

8.1.2 RQ2: Does questioned topic influence children's tendency to provide a teleological response? 198

8.1.3 SQ1: What types of teleology do children display when they provide a teleological explanation? 200

8.1.4 SQ2: Do children's explanations about natural phenomena result from fragmented or coherent knowledge? 202

8.1.5 RQ3: Does the natural discourse children experience around natural phenomena influence children's explanations for their existence? 204

8.1.6 Summary 205

8.2 Limiting teleology 206

8.2.1 RQ4 & RQ5: Do discursive activities about the appropriateness of scientific explanations support children in developing an understanding of appropriate and inappropriate explanations for the existence of natural kinds? 206

8.3 Limitations 208

8.3.1 Study 1 208

8.3.2 Study 1b 209

8.3.3 Study 2 209

8.3.4 Study 3 210

8.4 Future directions 210

8.5 Final remarks 211

9. References 213

10. Appendices 235

List of Figures

Figure 2.1. A method of classification represented in three ontological trees..... 237

Figure 6.1. Example assessment Concept Cartoon for Study 2: Scientific explanation for why there is wind. 149

Figure 6.2. Example lesson Concept Cartoon for Study 2: Discussion prompt for why there are waves. 149

Figure 6.3. Lesson plan for Study 2, intervention session 4 – Why are there waves. 150

Figure 7.1. Example assessment (Female) Concept Cartoon for Study 3: Want-based explanation 1 for why reindeer have big antlers. 186

Figure 7.2. Intervention session 1 Concept Cartoon for Study 3: session prompt for why rabbits have long ears. 186

Figure 7.3. Generic lesson plan for Study 3 187

List of Tables

Table 2.1: *Types of teleology derived from literature* 40

Table 2.2: *Types of intervention in relation to the two levels of scientific thinking* 61

Table 4.1: *Pilot interview questioned topics split by question set* 86

Table 4.2: *Coding rubric for children's responses at two levels* 89

Table 4.3: *Mean number of scientific responses split by question wording and question set order* 91

Table 4.4: *Percentages* of teleological response split by questioned topic and question wording* 93

Table 4.5: *Questioned topics ranked by percentage of teleological responses, split by question wording* 93

Table 4.6: *Goodness-of-fit test results for response type by questioned topic* 94

Table 4.7: *Percentage score for type of teleology advocated split by question wording* 95

Table 4.8: *Percentages of relational-teleological responses* by questioned topic and question wording* 96

Table 4.9: *Questioned topics ranked by percentage of relational-teleological responses, split by question wording* 96

Table 4.10: *Goodness-of-fit test results for type of teleology by questioned topic* 97

Table 5.1: *Rubric for coding keyword concordances* 119

Table 5.2: *Frequencies of the keywords in the corpora* 121

Table 5.3: *Occurrences of teleological and scientific discourse in the corpus by keyword* 122

Table 5.4: *Number of design- vs. relational- teleological concordances within the corpora* 124

Table 5.5: *Number of raw and refined collocates by keyword* 126

Table 5.6: *Categorised collocates for day (Tokens = 3438, Types = 163)* 127

Table 5.7: *Categorised collocates for night (Tokens = 2435, Types = 141)* 128

Table 5.8: *Categorised collocates for rain (Tokens = 44, Types = 51)* 128

Table 5.9: *Categorised collocates for rainbows (Tokens = 34, Types = 6)* 128

Table 5.10: *Categorised collocates for snow (Tokens = 593, Types = 76)* 129

Table 5.11: *Categorised collocates for storms (Tokens = 14, Types = 4)* 129

Table 5.12: *Categorised collocates for waves (Tokens = 29, Types = 3)* 129

Table 6.1: *Positive and negative aspects of different discussion activities* 140

Table 6.2: *Coding rubric for judgement of statements* 144

Table 6.3: *Pre- and post-test assessment statements* 148

Table 6.4: *Coding rubric for children rationales of their judgements* 151

Table 6.5: *Mean judgement scores at pre- and post-test, by experimental group (Max. score = 12)* 153

Table 6.6: *Mean rationale scores at pre- and post-test, by experimental group (max. score = 12)* 154

Table 7.1: *Explanations for the existence of trait Y, with function P, in species X* 169

Table 7.2: *Candidate traits and their functions* 171

Table 7.3: *Coding rubric for judgement of statements* 175

Table 7.4: *Traits using during the assessments – did children focus on function?* 176

Table 7.5: *Traits using during the sessions – did children focus on function?* 177

Table 7.6: *Comparison of old and new want-based explanation wording* 179

Table 7.7: *Comparison of old and new evolutionary explanation wording* 180

Table 7.8: *Discussion topics used in the main study* 185

Table 7.9: *Coding rubric for children rationales of their judgements* 188

Table 7.10: *Correct judgements and rationales at pre-test by explanations type* 190

Table 7.11: *Mean judgement scores[1] at pre- and post-test, by experimental group (Max. score = 8)* 191

Table 7.12: *Mean rationale scores[1] at pre- and post-test, by experimental group (Max. score = 8)* 192

1. Introduction

1.1 Conception of thesis

My background when starting this PhD was that of Early Years education, the education and care of children from birth to 8-years-old. Following a BA in Early Childhood Studies, I enrolled on an Early Years Post-Graduate Certificate in Education. It was during the PGCE that I began to develop the ideas that led to the conception of this thesis. Reading about science education I was intrigued by references to children's teleological reasoning (the belief that things in the natural world have a specific purpose) and the problems this reasoning can lead to in science lessons. However, upon following up these references I was struck by the methods used to investigate teleology. Children were asked, for example, *what are clouds for?* This to me seemed biased, surely children would say that clouds were *for something* if they were asked this question. This idea, that children could be led into providing teleological explanations became the focus of my PGCE dissertation.

Unfortunately, I was not trained in research methods at this time and was essentially left to my own devices to design the study. The study designed to test my theory was poor, it was underpowered, it was an "experimental design" with one independent variable that had only one level. The "statistical" analysis consisted of comparing the study results to prior work, which had used completely different methods, by simply looking at percentage scores. Although it must have shown some promise, my assessor suggested I leave teaching and focus on educational research. Instead, I went to Nepal.

It was up various mountains in Nepal that with the support of Professor Ainsworth (at that point head of postgraduate research and who subsequently became my primary supervisor) the proposal for this thesis was developed, internet connection depending. The initial focus on question wording remained but other themes began to emerge which are present in this thesis. For example, the notion that young children are more capable than the literature indicates, that they can and do understand scientific theories from a young age; the desire to conduct fieldwork in real world settings rather than in laboratories; the idea that the discourses children experience out-of-school could influence their use of teleological explanations; and finally, the desire to not simply catalogue children's teleological explanations but to actively support children's scientific development.

1.2 Overarching themes

There are a number of themes that run through this thesis. Firstly, that teleology is a collection of a number of different styles of explanation, related by their collective focus on *purpose* or *goal-directed behaviour* as rationales for why something exists in the natural world. Secondly, there is the argument that there needs to be a reconsideration of how children's use of teleological explanations is assessed. This reconceptualising of how teleological explanations are constructed

provides a new lens for attempts to limit children's use of such explanations. Thirdly, the 'correctness' of teleological styles of explanation depends, in part, upon the context in which they are used. This thesis focuses on children's use of teleological explanations in scientific contexts (i.e., in the science classroom). However, as children come to formal education with a range of preconceptions their experience of teleological forms of explanations (i.e., from social contexts) may not be comparable with the view of teleology in scientific contexts. Therefore, methods seeking to limit children's use of teleological explanations may need to develop children's understanding of these explanatory styles and their awareness of the context in which other individuals may use them.

1.3 Overview of chapters

Following this introductory chapter, Chapter 2 explores relevant literature to develop the foundations for the following studies. It positions teleology within children's science education, detailing key conceptions associated with teleological explanations. The chapter considers how and why children may use teleological explanations in scientific contexts, outlining current suggestions in the literature and proposing a number of new ideas. The chapter then provides a discussion of the two major themes of this thesis: how children's teleological explanations are measured, and how educators may attempt to limit children's use of teleological explanations. Research themes arise throughout and are developed into research questions towards the end of the chapter.

Chapter 3 outlines the methodology used to explore the research questions developed in the literature review. The philosophical framework underpinning this thesis, pragmatic post-positivism, is outlined and the main research methods, experiments and corpus linguistics, are detailed. The chapter ends with discussion about ethical issues, working with gatekeepers (schools and parents) and, most importantly, working with children.

Study 1 is described in Chapter 4. This study explored the measurement of children's (aged 5- to 8-years-old) use of teleological explanations for natural phenomena (e.g., night, rain or waterfalls). It investigated the way in which questions are worded (question wording) and the topics of questions (questioned topic) can influence children's use of teleological explanations. Furthermore, theoretical aspects of children's use of teleology were explored (both different forms of teleological explanations and the construction of teleological ideas). The findings from this study suggest a need to reconceptualise the measurement of children's use of teleology.

Chapter 5 details Study 1b. This study emerged as a follow-up study to explore aspects of the results of Study 1, hence its designation. Study 1b involved using methods from corpus linguistics to investigate the natural discourse children experience about natural phenomena. The idea underpinning this investigation was that both a) the specific explanations children receive and b) the discussion themes surrounding natural phenomena will influence children's explanations about the existence of natural phenomena, with a particular focus on teleological explanations. The

results of this corpus investigation are used to argue that children's use of teleological explanations is influenced by their natural discourse, that is, the themes associated with natural phenomena manifest in the content of children's explanations.

Chapter 6 and Chapter 7 both explore a method of limiting children's use of certain teleological explanations by developing children's understanding of what are appropriate forms of explanation in scientific contexts. These two chapters focus on two different age ranges (5- to 7-years-old and 8- to 9-years-old) and two different areas in which children use teleological explanations (for the existence of natural phenomena and the origin of organisms' traits). Both studies findings support the argument that by developing children's awareness of why certain forms of teleological explanation are inappropriate in scientific contexts educators can limit children's use of these explanations.

Finally, Chapter 8 summaries the four studies and discusses their implications for both theory and practice. This chapter also details the limitations of the thesis and outlines areas for future study.

2. Literature review

2.1 Situating teleology within science education

This initial section briefly situates the work in the thesis within the field of science education to contextualise the forthcoming sections. Furthermore, it defines two major terms used in this thesis: *scientific explanations* and *teleological explanations*. The definitions provided in this section are purposely broad. Teleology, in particular, is further refined and defined in relation to a number of subtly different forms of teleological explanation. However, a broad definition is initially required to provide adequate background and to explore concepts associated with all forms of teleological explanation.

2.1.1 Science education in the current climate

A formal science education has long been valued in the British educational system. In 1867 the British Academy for the Advancement of Science called for curricula reform, it argued that science should be taught to develop a scientific way of thinking in the general population (Layton, 1981). While there have been, and still are, certain debates around the purpose of science in the curriculum (e.g., teaching scientific method vs. scientific knowledge), it is normally believed to fulfil two purposes a) the training of specialised scientists and researchers and b) the growth of a scientifically literate public (Smith, 2010). This thesis deals primarily with the second of these two aims: the development of citizens who are scientifically literate, who understand the nature of science and can apply this to real-world contexts. This understanding of the nature of science is defined with regard to the *achieving Qualified Teacher Status* reference series (a series designed to support the development of UK pre-service teachers), in the attempt to ensure that the content of this thesis is applicable to the teaching profession. Therefore, understanding the nature of science is defined as recognising that:

- ‘Science is about understanding the world around us.
- Scientific understanding is based on accumulating and interpreting evidence for proposed theories or models.
- Science ideas are often tentative but the “big ideas” are well established.
- Science is a result of human endeavour and thus scientific interpretations may be influenced by moral, ethical, political and commercial factors.’
  (Sharp, Peacock, Johnsey, Simon, Smith, Cross and Harris, 2014, p.10).

The last point of the definition, competency in recognising that others’ scientific explanations should be supported by evidence and can be shaped by certain beliefs or ideologies, is particularly relevant in the current cultural climate. It is argued that such an understanding of science could counter the rise of, or the rise of publicisation of, pseudo-scientific ideas (Goldacre, 2003); for example, denying climate change (Runciman, 2017), the flat Earth hypothesis (Burnett,

2016), Emotional Freedom Techniques (Singer & Nievod, 2003) or the use of Brain Gym in schools (Goldacre, 2014).

Early Years science education is vital to developing individuals' scientific literacy, as it forms the bedrock of children's scientific thinking skills (Greenfield et al., 2009; Zimmerman, 2007). However, young children are often believed to be underprepared for science lessons, in terms of reasoning skills and prior conceptual foundations (Chen, Hand & Norton-Meier 2017; Lehner & Schauble, 2006). These concerns are often attributed to the beliefs of teachers and children's parents (Keil, 1992), although the results of certain research projects (e.g., Coley, Arenson, Xu & Tanner, 2017; Piaget, 1927) would also suggest a similar developmental pattern for children. That is, certain intuitive methods of reasoning mean that they are underprepared or disadvantaged when learning about scientific concepts. One such form of intuitive reasoning, and the focus of this thesis, is teleological reasoning. This term is explored below but first a note on terminology. In the literature children's *reasoning* is usually based upon the *explanations* that children provide to various tasks. While the explanations children provide can indicate their underlying reasoning, they do not prove the existence of that pattern of reasoning. Hence, the terms scientific and teleological explanation/s are used, rather than scientific and teleological reasoning (unless referring to the results of prior studies). The following section provides a general definition of these terms, which are further developed throughout the thesis.

2.1.2 Scientific explanations

A good explanation (in line with Lipton, 2004) is defined as one which:

a) provides an understanding of a phenomenon, rather than simple knowledge of a phenomenon, abstracted as understanding Why B, rather than knowing That B (e.g., understanding why a tomato is a fruit, rather than simply knowing that it is a fruit);
b) is self-contained, therefore, does prompt a why regress, abstracted as A explains B but further understanding of A is not required (e.g., the statement *fruits contain seeds and a tomato has seeds* does not require a regress of further explanations to provide an understanding of why fruits have seeds);
c) could be in the form of a self-evidencing explanation, one where the phenomenon being explained forms part of the evidence for the correctness of the explanation, abstractly noted as A explains B but B justifies A (e.g., in the explanation *this sandwich has tomatoes on because I can taste tomato* the taste of tomatoes explains how the individual knows the sandwich has tomatoes on, and the presence of tomatoes on the sandwich justifies the act of tasting tomatoes).

The only style of explanation that fits all three criteria is a causal explanation. A causal explanation is argued to be effective at promoting understanding because a cause is the difference

between the occurrence and non-occurrence of a phenomenon (Lipton, 2004). Established causal explanations, in general and in the field of science, usually derive from *Inference to the Best Explanation* (or abduction): from the total of evidence a hypothesis is inferred which, if correct, best accounts for said evidence (Lipton, 1991; 2004). Furthermore, to qualify as a robust scientific explanation, the causal account must have reasonable empirical support (Cao, 2004).

2.1.3 Teleological explanations

During their Early Years education, children commonly use a style of explanation called teleological explanations to account for things in the world around them. Broadly, these are explanations which indicate that something exists for a specific purpose (e.g., rain is for watering the grass, puddles are for jumping in, scissors are for cutting paper) or that propose a goal-driven behaviour (e.g., individual reindeer *decided* to grow large antlers because they *wanted* to be better at fighting their rivals). For human-made artefacts (e.g., scissors) this form of explanation are appropriate; however, with regard to the natural world it is argued that these forms of teleological explanations can imply a purpose where there is none.

In a range of studies, children aged 3- to 10-years-old have been shown to primarily provide teleological explanations to account for things in the natural world (e.g., Carey, 1985; Kampourakis, Palaiokrassa, Papadopoulou, Pavlidi, & Argyropoulou, 2012a; Keil, 1994; Kelemen, 1999a, 1999b, 1999c; Opfer & Gelman, 2001; Smith, Cowie, & Blades, 2003). Therefore, unless signposted the forthcoming discussion is situated within this age band. However, it should be noted that the use of teleological explanations has been claimed to continue, albeit with less prevalence, well into adulthood (see for example Friedler, Zohar, & Tamir, 1993; Kelemen & Rosset, 2009). Furthermore, while the majority of work on teleology is conducted in either American- or British-English speaking countries, there is evidence that teleological explanation is not specific to the English language. Children and adults who do not speak English or have English as an additional language have been found to use teleological explanations, including individuals from China (Rottman, et al., 2017), Greece (Kampourakis et al., 2012a; Kampourakis, Pavlidi, Papadopoulou, & Palaiokrassa, 2012b), Israel (Zohar & Ginossar, 1998), Mexico (Kelemen, Callanan, Casler & Perez-Granados, 2005), Romania (Casler & Kelemen, 2008) and Peru (Sánchez Tapia et al., 2016).

Research into children's teleological explanations usually investigates a number of categories (what constitutes a category is returned to in section 2.3). To investigate a certain category (e.g. natural objects) prior work usually asks children to account for a number of what are termed here *exemplars:* items which are used as typical examples of the category. Three common reoccurring categories (and exemplars) are organisms and their appendages (giraffes or rhinos' horns), natural objects (clouds or pointy rocks) and artefacts (watches, hammers or watches' hands). Investigation of children's teleological explanations in a range of different categories (e.g., organisms, natural objects or natural phenomena) has resulted in two competing theories about how

children apply teleology. Children are viewed as either being selective in their application of teleology, applying it only to certain categories but not others (Kampourakis et al., 2012a; Keil, 1995) or promiscuous, applying it to anything and everything (Kelemen, 1999c, 2012). This is returned to in section 2.3.1.

Kelemen argues that the tendency for children to provide teleological explanations to questions about things in the natural world means that they prescribe some form of intentional design to natural kinds, in that they are viewed to exist or have been created for a purpose, to serve a function or to achieve a goal (Kelemen, 1999c). From this perspective it is considered that teleological explanation is indicative of a preconception that things in the natural world exist for a purpose. However, children's preconceptions can be particularly difficult for teachers to change (Peacock, Wright, Johnsey & Sharp, 2011). Consequently, evidence indicating that children have a preference for teleological explanations is a cause of concern for educators. For this reason, teleology is argued to be a major problem when teaching and learning about causality in the natural world, particularly with regard to evolution (Coley et al., 2017; Evans, 2013; Hanke, 2004; Kampourakis, 2014). The 'acceptability' of teleology (i.e., if and when teleological explanation is an appropriate explanatory style) is considered further in section 2.5.3.

2.1.4 Summary of section

This section briefly explored the thesis themes of teleology and science education. Two types of explanation pertinent to children's science education were defined. Scientific explanations are causal explanations, which are logical and supported by evidence. Teleological explanations are where something is explained with regard to a specific purpose for its existence or where goal-driven behaviour is attributed to organisms. The use of these explanations (suggesting a teleological preconception of the natural world) is generally considered to be problematic in science education, as they suggest a misunderstanding about the role of causality in the natural world. That is, children either a) are aware of a causal explanation but prefer a teleological one; or b) lack knowledge of a causal explanation therefore use a teleological one. Preconceptions of this type (e.g., that teleological explanations are appropriate in science education) can be a cause for concern in science education as they can be particularly difficult for educators to change (e.g., see Coley et al., 2017; Evans, 2013).

2.2 History of teleological explanations about the natural world

Traditionally, in the field of developmental psychology, teleological explanation is considered to be one of a number of intuitive reasoning styles. This section elaborates on this notion detailing a number of other intuitive styles of thinking (artificialism, anthropomorphism and essentialism) which are closely related to and at times overlap with teleological explanation. Furthermore, the 'misconception movement' in educational research is outlined. It is argued that the study of intuitive

reasoning and learners' misconceptions in science education represent two fields studying the same phenomena.

2.2.1 Artificialism as teleology

One of the first major studies into children's scientific thinking was conducted by Piaget (1929). In an iconic publication Piaget proposes that until around 11-years-old children think in an artificialist manner, which shares certain key elements with teleology. Essentially, artificialism is the belief that natural objects are artificial and are made, be it by human-kind or supernatural entities, to perform certain functions. However, the creator of the object may not necessarily be human but something which humans have created. An example of this is human-made fires that create smoke that then creates clouds.

Piaget proposed that children's scientific development follows a number of stages, with artificialist thinking progressing though three stages. This stage theory view of children's development, suggesting there are segmented, linear, non-reversible stages through which children move through, is now controversial (see Sutherland, 2000). The stance taken here is not that there are fixed age-stages but that a stage represents a style of thinking and that children can move, both forwards and backwards, through these stages. Essentially they represent an attempt to categorise a fluid continuum of development (Shayer 2003; Tryphon & Voneche, 1996).

In Piaget's spectrum children's scientific thinking moves from the first stage, which is fully artificialist and invokes humans or supernatural entities for the creation of objects and organisms that serve humans or animals, to the third stage, in which natural objects and organisms are naturally formed rather than made but may still serve some purpose (Lovell, 1961; Piaget, 1929). Piaget's research into artificialism presents three key points:

- Children of the same age usually gave similar responses and while children's answers shared commonalities explanations were idiosyncratic;
- Artificialist responses extend, in part, over more than one stage;
- Even in the last stage children still made artificial references when giving a correct answer.

There have been numerous criticisms of Piaget's work (Donaldson, 1987; Smith, Cowie, & Blades, 2003; Sutherland, 2000); however, as referred to in this chapter, there is considerable evidence that children use explanations which assign functions to biological entities and natural objects. While what Piaget's termed artificialist explanations are suggested to have a different origin to teleological explanations (see section 2.4 for discussion on origins), the actual explanatory style of artificialism (a focus on purposes or goals) is the same as teleological explanations, which are argued to represent a form of intuitive reasoning (Kelemen, 1999b; 2004).

2.2.2 Intuitive reasoning in developmental psychology

In the latter half of the 20th century there was considerable focus from developmental psychologists on what is termed intuitive reasoning or intuitive theories. These are predictable patterns in individuals' naturally occurring explanations, which are informed by their personal theories about the world. Individuals' intuitive theories are argued to support and at times act as boundaries to their cognitive development (Coley et al., 2017; Gelman & Legare, 2011). One major form of intuitive reasoning is teleology. Furthermore, while there are a number of these intuitive reasoning styles and they are often explored separately, they can be considered to overlap or be used in conjuncture when children provide explanations for natural kinds (Gelman & Legare, 2011; Polling & Evans, 2002, Tamir & Zohar, 1991; Zohar & Ginossar, 1998). Two intuitive reasoning styles which are different but related to teleology are *anthropomorphism* and *essentialism.*

Anthropomorphism is the use of explanations where there is the attribution of life and human-based psychological processes (e.g., feelings or goal-driven behaviour) to both non-human organisms and inanimate objects (both natural and artificial). For example, stating that *the wind was trying to blow the trees over,* a child worrying about a teddy left at home because it might feel lonely or the majority of depictions of animals in children's cartoons such as *Winne-the-Pooh*. This form of explanation and reasoning is also termed anthropocentric thinking (see Coley et al., 2017).

In relation to children's use of anthropomorphic explanations, there are indications that the majority of anthropomorphic language in Early Years settings does not come from children but from teachers (Thulin & Pramling, 2009) and from children's literature (Markowsky, 1975). However, there is evidence that children do use these explanations from around 4- to 10-years-old (Carey, 1985; Hantano & Inagaki, 1994; Inagaki & Hantano, 1987; Piaget, 1929). While there is limited research past the age of 10-years-old (Coley et al., 2017), there are tentative suggestions that high-school students (15- to 17-years-old) use anthropomorphic explanations to support their biological understanding (Zohar & Ginossar, 1998).

There are considered to be a number of societal benefits to anthropomorphic reasoning, for example, improved animal welfare and conservation (Urquiza-Haas & Kotrschal, 2015). However, in scientific contexts their use is problematic, as it either a) imbues objects with life and cognitive capabilities, or b) attributes human-based cognition to organisms' behaviour (Davies, 2010; Gallant, 1981; Wynne, 2007). These anthropomorphic forms of explanations overlap with teleological explanations. For example, *rain is for watering the grass* is a teleological statement but it also imbues *rain* with a life force, as it now has a *job* (watering grass) which relates to anthropomorphism. The anthropomorphic conflation of human cognitive processes with organism behaviour and its relation to teleology is returned to in section 2.5.3.3.

Essentialism is the use of explanations which indicate that individuals or items within a group (e.g., all starlings within a murmuration or all clouds within a soufflé) all share some essence (e.g., a starling essence or cloud essence). This essence defines the properties of individuals within

the group. In terms of advocating essentialist explanations, Polling & Evans (2002) assessed children's (aged 6- to 12-years-old) and adults' preferred explanations for why individuals within a species group stay together (i.e., are often found together). Older children (8- to 12-years-old) and adults primarily chose essentialist explanations when discussing types of plants and insects, but showed a preference for teleological explanations for ungulates and intentional (desire-based) accounts for humans. This suggests that as children age they show an increased tendency to predict behaviour or action based on group membership, that members within a group all have some form of shared essence.

This essentialist reasoning can help children with categorisation of entities and then prediction of qualities based upon group membership (Gelman & Legare, 2011). However, problems arise when entities are overgeneralised due to a perceived internal essence. For example, in terms of evolution, essentialism could lead to the belief that all organisms in a species are physically the same, that there is little or no variation in traits, resulting in difficultly in understanding the concept of continuous variation of traits within a species population (Gelman & Rhodes, 2012).

2.2.3 Misconceptions in education

While there was an investigation of intuitive reasoning in the field of developmental psychology, in educational research there was a trend to cataloguing children's misconceptions in relation to scientific concepts during the latter half the 20th century (1970s – 1990s, Taber, 2014). Although the amount of research that focuses solely on cataloguing children's misconceptions has diminished, this research area remains influential in terms of using this catalogue of misconceptions to inform research and practice about conceptual change (diSessa, 2014), particularly in initial teacher training literature (e.g., see Allen, 2014).

The misconception movement involves the core idea that any learner at some point could have some form of misconceived or misconstrued naïve ideas, as no individual is an expert in every field (Murphy & Alexander, 2013). Through investigation and collation of wide-ranging misconceptions (those common to the majority of children) it was believed that educators would be better prepared to identify children's misconceptions – misconceptions that can negatively influence their reception of new ideas (Gilbert, Watts & Osborne, 1982; Thomaz, Malaquis, Valente & Antunes, 1995). Children's misconceptions could consequently be explicitly targeted and challenged (Clough & Driver, 1985; Sharp et al., 2014), for example, through teacher guidance and argumentative interventions; by discussing their misconceptions alongside teaching of the correct conception; and developing children's understanding about the nature of science (Driver, Newton & Osborne, 2000; Treagust, Chittleborough & Mamiala, 2002; Thomaz et al., 1995). The idea is that children realise how (in relation to the correct conception) and why (in relation to the nature of science) their naïve model is in conflict with the scientific model, therefore their misconception can

be superseded. In summary, those cataloguing children's misconceptions did so due to a belief that science curricula needed to be shaped by children's current conceptions and misconceptions (Posner, 1982).

However, children's misconceptions, especially in Early Years settings, can be particularly varied for different scientific concepts (Sharp et al., 2014). That is, children come to formal education with a wide range of personal experiences regarding the world around them, experiences which can be considerably different from their peers. When attempting to explain a concept, children within a class will each draw upon their individual experiences resulting in a varied range of prior ideas about a single concept, within a single class. Consequently, children's misconceptions cannot only be assumed from a collective bank of misconceptions, such as those presented in initial teacher training textbooks. While these banks are useful, educators must also assess the children in their care to find their children's individual (mis)conceptions. While the identification of children's specific conceptions prior to formal education can be extremely useful in terms of developing informed teaching and learning situations (Leach, Driver, Scott and Wood-Robinson, 1992), Early Years teachers often do not attempt to survey the initial ideas of the children in their care before planning lessons, possibly to the detriment of children's education (Kambouri, 2016; Kambouri & Briggs, 2013). Therefore, while there is much literature on the benefit of using children's misconceptions to inform teaching, based on work by Kambouri, it does not appear to always translate into practice.

This lack of translation to practice may have roots in the *alternative framework* literature. This indicates that children hold alternative frameworks about the world, which are alternative in that they differ from a scientific position. These frameworks are informed by children's experiences and consequently children may struggle with scientific concepts which are counter intuitive to their alternative framework of the world (Driver, 2008). Thus, as children's alternative concepts exist within a personal framework of the world, a specific misconception cannot be challenged in isolation but rather the framework from which this misconception arises must be addressed. Driver, Asoko, Leach, Mortimer & Scott (1994) suggest that the way in which these frameworks are challenged requires mediation by educators who must tailor activities to suit the needs of individual learners. Consequently, the lack of assessment of children's initial ideas, as found by Kambouri, may not indicate that teachers do not attend to children's misconceptions, rather they address misconceptions at an individual level, when appropriate.

While misconceptions about certain scientific concepts can be numerous and idiosyncratic, in that they result from children's individual alternative frameworks, other scientific concepts may have a smaller number of misconceptions associated with them because they may relate to common experiences which children share. For example, the impetus misconception (the notion that when a body is moving it has an intrinsic force keeping it moving, Hestenes, Wells & Swackhamer, 1992) the octet framework (the idea that atoms react in order to have a full valence shell, Taber, 2014),

and, as documented in this chapter, teleological explanations about the natural world. While specific teleological explanations may vary for specific things within the natural world (e.g., children may claim that rain is for making muddy puddles, for filling up bodies of water, for providing drinking water, for washing cars or for watering the grass) the core principle that nature has a purpose or goal is a wide ranging misconception. Therefore, challenging children's use of teleological explanations may not need in-depth investigation by teachers prior to instruction, as teachers would be seeking to address the misconception that nature is purposeful. Idiosyncratic explanations for certain examples of things in the natural world could be addressed by teachers on an individual level and when necessary.

Research only cataloguing, rather than challenging, children's misconceptions is now limited (diSessa, 2014), as is the attempt to completely replace the initial misconception with the correct idea (Duit & Treagust, 2003). Rather current research into children's initial ideas focuses on designing suitable teaching and learning environments to foster gradual conceptual change across a range of misconceptions (Vosniadou & Skopeliti, 2017). Furthermore, there is a thread of work which considers misconceptions to not be ideas that will inhibit the learning of correct concepts but rather may be beneficial starting points for the development of appropriate scientific concepts (Evans, 2018; Smith, diSessa & Roschelle, 1993). The notion of challenging and reframing children's use of teleological explanations is returned to in section 2.7. Finally, in order to represent the literature this section has focused on what are usually termed 'misconceptions' which may result from children's alternative frameworks. From this point the term preconception is used in the thesis, rather than misconception. Preconception is preferred as these are children's ideas before they have received formal education, on such matters as causality in the natural world. Therefore, it is argued that the term misconception is not appropriate as children have not been taught the correct conception.

2.2.4 Summary of section

While different movements, the misconceptions (now referred to as preconceptions) movement and theories of intuitive reasoning appear to overlap in terms of what is actually being investigated (children's ideas before formal education), although not with regard to implications arising from these investigations. The examples of preconceptions which were found in the field of educational research could be categorised by different styles of intuitive reasoning. Similarly, the evidence for styles of intuitive reasoning, as documented in the field of developmental psychology are examples of preconceptions detailed in educational research. For example, the preconception that individual organisms wanted to evolve (e.g., some peacocks *decided* to grow large colourful tails because they thought it would be good for scaring away predators) represents teleological (goal-directed) and anthropomorphic intuitive styles of explanations. Another example would be explanations where the bonding of atoms is discussed in relation to the individual atoms wanting to form covalent or

ionic bonds (e.g., 'atoms combine to form molecules so that they can get the number of electrons that they need to become more stable', Talanquer, 2013, p. 1420), showing overlap of the octet preconception and teleological explanation. In this example, the student's explanation for how/why chemical reactions occur could be considered evidence for both a) that students often have an octet preconception (i.e., atoms combine with other atoms in such a way as to produce a 'full' valence shell) and b) that students use teleological styles of explanations to account for chemical reactions (i.e., atoms want to react for the specific purpose of achieving a 'full' outer shell). As can be seen both accounts for the student's idea have considerable overlap, for this reason it is argued that these two areas of study are exploring similar phenomena from different positions.

Considering that the two areas of study (intuitive theories and preconceptions) have considerable overlap, research into teleological explanations rarely spans both fields, rather work in developmental psychology and education is usually informed by prior research in the same area. Throughout this thesis research from both developmental psychology and education is explored to provide a wider background for children's use of teleological explanations. This synthesis of literature from both disciplines should provide new and original ways of researching and challenging children's use of teleological explanations.

2.3 Teleological explanations about different categories

Research investigating teleological explanations often focusses on a number of different categories (e.g., artefacts, natural objects or organisms), to do so children are asked about different examples of these categories (e.g., chairs, rocks or zebras). These 'examples' are treated as *exemplars*, that is they are considered to be typical examples which can be used to represent children's thinking about the category. Investigation of children's explanations for exemplars within different categories has been used as evidence to support theories describing how children reason about different categories, depending upon the research focus. Surveying prior research (including Carey, 2000; Kampourakis et al., 2012a; Kampourakis et al., 2012b; Kelemen, 1999b; 1999c; 2003) indicates that there are three common categories:

- organisms, for example, *an oak tree is for climbing, the wings of a bird are for flight* or individual rhinoceros decided to grow large horns for defence;
- non-living objects, for example, *a hollow in a hillside is for an animal to hide in, rain is for making muddy puddles*, or *day is a signal for people to wake up*;
- artefacts, for example, *a saw is for cutting wood* or *a bed is for sleeping in* (teleological explanations for artefacts are justified as they have a designer and serve a purpose).

This section explores the use and implications of investigation by these categories. Firstly, it outlines a current debate on when teleological explanations are used (selective vs. promiscuous

teleology). Secondly, what constitutes a category is explored. Finally, this section outlines a systematic method for classifying different kinds.

2.3.1 Selective vs. promiscuous teleology

Investigation by category has resulted in two competing theories about how children apply teleology. Children are viewed as either being selective in their application of teleology (Kampourakis et al., 2012a; Keil, 1995) or promiscuous, applying it to anything and everything (Kelemen, 1999b, 2012). In perhaps the first major study into children's teleological explanations, Keil presented participants with multiple choice questions about exemplars of natural objects and organisms to ascertain how children, aged 5- to 7-years-old, applied teleological explanations across categories. The questions asked children to choose between two options: if the exemplars had a) a teleological explanation or b) a causal explanation. For example, if the green colour of plants or the green colour in emeralds was a functional trait (teleological), the green helps them to multiply, or a physical trait (causal), the green is due to pigmentation. Children primarily chose causal options for natural objects and teleological answers for organisms (Keil, 1992; 1994; 1995). This, Keil claimed, indicated that children reason teleologically about organisms but not about natural objects. They are selective in applying teleology and do not unnecessarily attribute purpose to natural objects. Limitations of the study are acknowledged by Keil, mainly that a generic explanation may not fit all questioned topics: each participant could provide a different rationale for the same exemplar. This raises an issue with using multiple choice questions, a problem which is returned to later in section 2.6.1.

The proposal that teleological reasoning is only prevalent within the categories of organisms and artefacts was challenged by Kelemen (1999a). Sixteen 4- to 5-year-olds were asked *what is X for?* in relation to 22 different exemplars for three categories. Participants provided teleological rationales for artefacts (e.g., a clock), organisms (e.g., a lion) and non-living objects (e.g., a cloud), suggesting that young children assign purpose to all three categories, a notion dubbed by Kelemen as promiscuous teleological reasoning. While the results propose a broad application of teleology regardless of category, out of the 22 exemplars investigated 14 were organisms, 6 artefacts and only 2 exemplars were classed as non-living objects. This appears a disproportionate split of category examples, especially when considering the use of only two exemplars to propose a theory shift from selective to promiscuous teleology.

To validate these findings, Kelemen (1999a) presented hypothetical scenarios, using the same 22 exemplars, to 24 four- to five-year-olds, each scenario comprised of an individual stating that X was intended for a specific task, while another individual claimed that although X could be used for a certain task it was not specifically intended *for* that task. The results supported the original study (Kelemen, 1999a), in the follow-up study participants chose the *for* option ~80% of the time for artefacts, ~70% for natural organisms and ~85% for natural objects (means taken from graph

(Kelemen, 1999b, p. Fig 1 & 2)). Therefore, Kelemen argues this provides further evidence for the notion of promiscuous-teleology, in relation to children aged 4- to 5-years-old.

The prevalence of promiscuous-teleological reasoning across a wider age range is supported by additional studies by Kelemen (1999b; 2003), in which Grade 1, 2 and 4 students were given multiple choice questions about the appendages of four organisms (e.g., a reptile) and four natural objects (e.g., a pointy rock). Children were presented with two options for each of these exemplars, a teleological form of explanation and a scientific answer. The results indicated that participants subscribed function to both organisms and non-living objects, although, teleological explanations decreased with age. Furthermore, research with participants with English as an Additional Language has found a promiscuous application of teleology across a range of categories in China (Rottman et al., 2015), Romania (Casler & Kelemen, 2008) and Peru (Sánchez Tapia et al., 2016).

While the majority of recent research suggests a promiscuous use of teleological explanations, a selective application has been detailed in research using open-ended questions to investigate the teleological reasoning of 5- to 8-year-olds (Kampourakis et al., 2012a) and 7- to 8-year-olds (Kampourakis et al., 2012b). These studies focused on children's explanations for the colour and shape of two organisms, two natural objects and two artefacts. The age groups children were placed in (5- to 6-years-old; 6- to 7-years-old; 7- to 8-years-old) were negatively correlated with the number of teleological responses given by each group, that is, the older group provided the least teleological explanations and the younger the most. Children aged 5- to 6-years-old provided teleological explanations about the colour and shape of all three categories (e.g., ducks feet, stalagmites and scissors), whereas the majority of children aged 7- to 8-years-old employed teleology only for the shape of artefacts and organisms, not for natural objects. The results of these two studies support claims by Keil and Kelemen that children's tendency to provide teleological rationales is negatively correlated with their age. Furthermore, Kampourakis et al. (2012a) proposed that 5- and 8-years-old children moved from a promiscuous application of teleology to a selective application for natural objects. While the data does support this inference, only using two exemplars of natural phenomena could be considered an insufficient number of measures. As with Kelemen's first study, such a reduced set of examples surely cannot give credence to a theory about a whole category (i.e., natural objects). This is supported by Tao, Oliver, & Venville who found evidence of 8- to 9-year-old children providing teleological rationales for the Sun, the Moon (2011) and seasons (2013). Consequently, a wider investigation of examples within a category may be needed.

While there is an ongoing debate about the application of teleology (selective vs. promiscuous), the majority of studies would indicate a promiscuous application. However, both the studies by Kelemen and the opposing ones by Kampourakis et al. may not be a secure foundation for a promiscuous or selective view of teleology for children, due the limited number of exemplars used in these studies. Therefore, it is argued that there is need for a wider investigation to clarify

the categories, or exemplars of categories, for which children use teleological explanations. This argument is returned to in the development of research questions (section 2.6.2).

2.3.2 Using hierarchical classification

Although the use of categories is common in research on children's teleological explanations, exactly how these categories are formed in the first places appears to be strikingly absent in many studies. There is usually little information in the studies regarding definitions of the commonly used categories described above. The main defining features of these categories is the category label and the exemplars provided as examples or used as materials in the experimental studies. While these defining features are not ambiguous, the actual categories may not be robust, as certain exemplars may not be objectively placed within a category. For example, does *organisms* include humans? Are mountains classified as non-living objects? What about rain, fire or the aurora borealis? If exploration by some form of category is desirable, what is needed is a systematic method of classifying *natural kinds*. Keil proposed that 'there is no simple definition for natural kinds, but they are commonly thought to include such things as animals and plants, as well as elements and compounds' (Keil, 1992, p.25). The term natural kinds therefore incorporates a number of different categories of things which originate in the natural world, (e.g., the categories of organisms and natural objects), consequently, this overarching term is used throughout the thesis where appropriate.

There are a number of frequently used categories in the literature (organisms, natural objects and artefacts) and the construction of these categories appear to share a number of commonalities with work by Chi (Chi, 1992; 1997; 2008; Chi, Slotta, & Leeuw, 1994) into constructing ontological categories. In Chi's framework everything can be placed within a number of ontological categories. There are three main overarching categories: matter, processes and mental states. These main categories are called ontological trees. Each ontological tree has a number of subcategories, which in turn may have more subcategories. Therefore, any concept can be placed in a certain category, within a certain tree. Depending upon the granularity of categorisation, this category could be a broad 'higher level' category (e.g., dachshund placed within organisms) or a highly defined subcategory which has numerous parent levels (e.g., dachshund placed within dogs, within mammals, within animals, within organisms, within entities). Subcategories at the same level (e.g., subcategory level 2, see Figure 2.1) are considered to be parallel. While parallel categories across ontological trees do not share properties, parallel categories within trees can share properties but only in relation to their parent category. Placing a new concept within a tree results in it inheriting the qualities of the ontological category. Categorised items are considered static unless there is robust conceptual change (Chi, 1992; 1997; 2008; Chi, Slotta, & Leeuw, 1994). While it is not certain that previous research has used Chi's framework for classification by ontological

category, the way in which prior work has structured categories shares similar features to ontological classification and thus provides a rationale for future exploration.

Classification by ontological category is not without problems. For example, individuals' ontologies have been argued to be less stable than suggested by Chi, and therefore may be better understood as dynamic categories that alter depending upon context (Gupta, Hammer & Redish, 2010). While there may be debate around Chi's use of static-ontologies, this form of classification is used in this thesis to systematically classify categories as a) it provides a method of bounding research into natural kinds, so as to manage the scope of the investigation and b) it is a convention of research investigating children's teleological explanation to explore different categories, so abandoning exploration by category could make comparisons to earlier work difficult. Consequently, the term *categories* is used to denote categories used in prior work, while the term *ontological categories* refers to the categories derived from Chi's framework. The next section provides a detailed description of the different ontological categories used.

2.3.3 A proposed method of classification

Following suggestions from Chi (2008) it is maintained that non-living objects is an inappropriate category. It is argued that natural objects and natural phenomena are fundamentally different, the former being entities and the latter processes, and so they cannot be combined under the same parental heading. Consequently, a possible classification system for concepts about different kinds is proposed, see Figure 1. The classification system serves the needs of this thesis, hence the 'not discussed further' parts. While this taxonomy has been developed specifically for this thesis it is based upon representations by Chi (Chi, 1992, 1997, 2008; Chi, Slotta, & Leeuw, 1994).

As Figure 1 shows, concepts are split into three ontological trees: the highest form of classification. To avoid confusion the term ontological tree refers to the top level of classification in the tree (e.g., processes), not the sibling categories within a tree (e.g., constraint-based processes or natural entities) (Chi, 2008). Items within different trees do not share properties, nothing within the ontological category of entities can occur over time (process) or within the mind (abstractions). For example, a dog cannot occur over time, whereas a thunderstorm can. Likewise, a thunderstorm is not a tangible entity, it is an observable process. Furthermore, as items within trees share properties in relation to their parent category, a mushroom and a rock, for example, could be placed within the separate (level 2) parallel categories of organisms and objects, respectively, and, while they do not share the same properties in terms of life they do share the parent category property of being formed in nature.

This description of ontological classification is an idealised version where anything and everything can be easily categories into a distinct category; however, it is not always so simple. At times something may overlap two distinct categories and can, therefore, be thought of as a boundary

case. An area where boundary cases may occur is when differentiating between artificial vs natural entities or artificial vs natural phenomena. For example, square watermelons are natural entities that have been altered by human input to be grown into the shape of a cube, so could be considered to be a form of artificial entities due to the considerable human interaction. Another example could be the fog in Hong Kong, as fog is an example of natural phenomena and therefore will occur periodically under certain weather conditions. However, human activity (e.g., pollution or urban planning) has exacerbated the occurrence and intensity of fog in Hong Kong, as such Hong Kong fog could be considered to be both a natural and artificial phenomenon.

Basalla (1988) proposes the notion of naturfacts which are natural objects (or natural entities) that are used by humankind to serve a purpose. Examples would include rocks used as tools in prehistory, the thorny branches of Osage orange used as a precursor to barbed wire in colonial America, or, more recently, pinecones used as decorative ornaments. Usually these naturfacts are superseded by true artefacts, however, 'in many cases, separating an artefact from a naturfact is difficult' (Basalla, 1988, p. 50). Michl (2002) takes this further arguing that an artefact is not actually designed, rather that an artefact is the result of the continuous redesign of existing artefacts or of existing naturfacts. A redesigning process would, if starting with a naturfact (a natural entity), slowly redesign such an entity until a later iteration would no longer be classified as a natural entity but rather considered to be an artificial entity in its own right. This process outlined by Michl demonstrates how entities can transition through the boundary between natural and artificial entities.

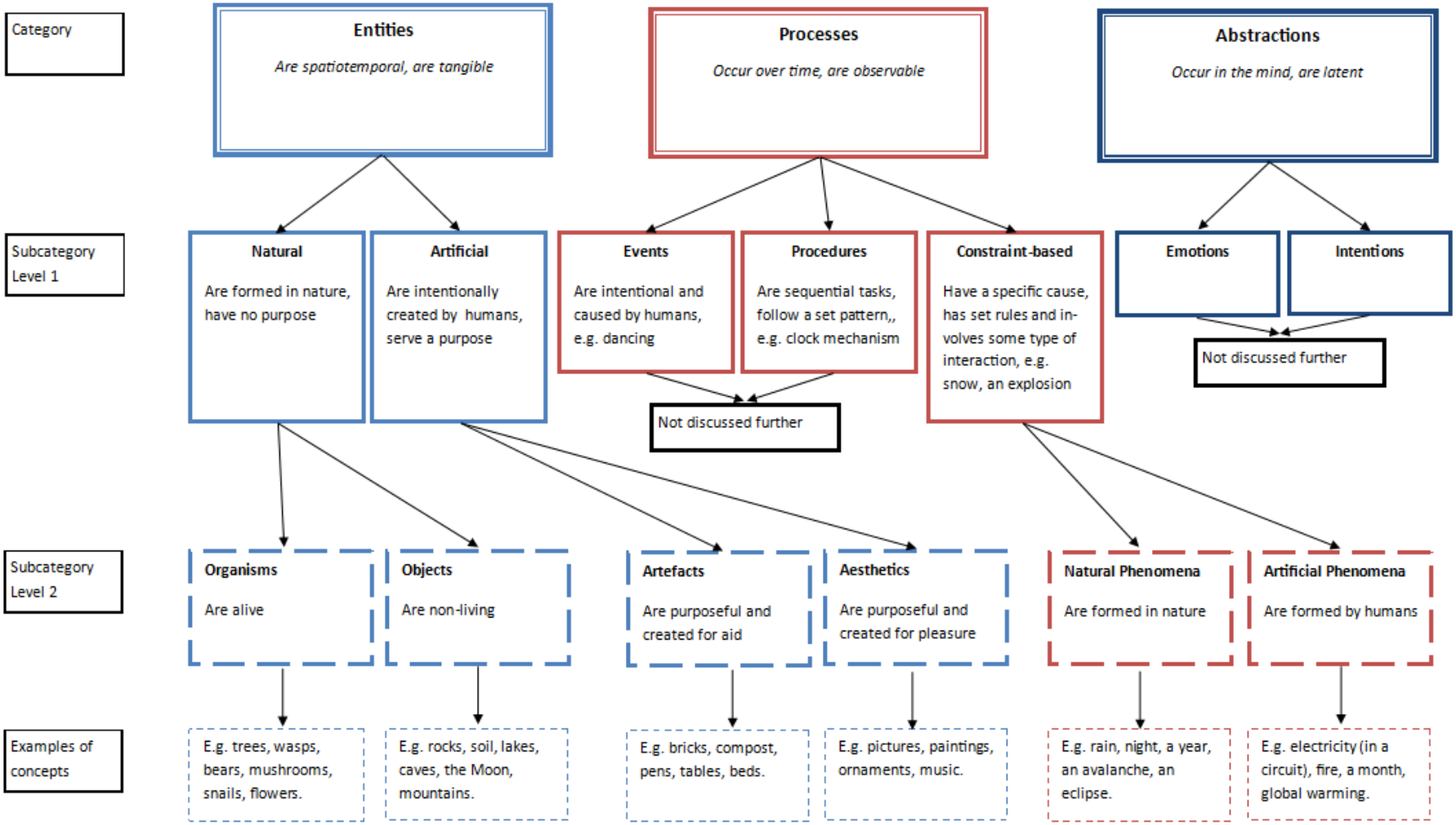
Ultimately a decision needs to be made as to where a boundary case will be categorised. It would be the responsibility of those discussing boundary cases to provide a rationale for their categorisation, where and if relevant boundary cases will be detailed. Through the remainder of the thesis when ontological categories are discussed it should be noted that certain things may not always be placed easily within a specified category.

Carey (1985) argues that a hierarchical structure of ontological categories such as the one developed here is problematic, as children do not always adhere to a hierarchical structure in their explanations about and classification of different kinds. At times Carey suggests children may collapse categories or may not have formed necessary parent or parallel categories which would lead them to make category mistakes. For example, realising that plants are susceptible to infections after previously believing that only animals can be sick (representing an absence of the parent category, *organisms*, for the sibling categories of plants and animals) or treating fire as an entity (representing a collation of organisms and natural phenomena). While it is not debated that children may make category mistakes and that the ontological classification system may not represent children's idiosyncratic systems, it is maintained that the system used in this thesis is fit for purpose. This system is not meant to be used by children to classify kinds, rather it is used as a research tool

to determine categories into which the researched examples should be placed without committing category mistakes.

Using Chi's ontological categories indicates that a distinction is needed between natural objects (e.g., rocks) and natural phenomena (e.g., wind). In previous work these categories have been collapsed into one category, often called non-living natural objects. Chi's work would suggest that natural objects are placed in the ontological tree of entities (tangible objects), whereas, natural phenomena are within the ontological tree of process (events which occur over time). This thesis maintains this distinction, treating the two as separate categories. Previous work has not made this distinction, therefore, there may need to be an investigation of children's explanations within each category. As mentioned exploration of categories should not rely on a small number of supposed exemplars, such as the one natural object and one natural phenomena used in Kelemen (1999a), the four types of natural objects used in Kelemen (1999b) or the two types of natural object used in Kampourakis et al. (2012a; 2012b). Consequently, a possible area of research is whether children display variation in explanatory style for a range of examples within a single ontological category. This could develop a broader picture of children's explanations for the ontological categories of natural phenomena and natural objects. Furthermore, it may provide evidence for the selective- versus promiscuous-teleology debate (discussed in section 2.3.1), as variation in levels of teleological explanation within an ontological category would suggest a selective application of teleology within a category. The development of this research aim is returned to in section 2.6.2.

Figure 2.1. A method of classification represented in three ontological trees



2.3.4 Summary of section

Research into teleological explanation usually investigates a number of different categories (often organisms, non-living natural objects and artefacts), which has resulted in two competing theories about children's use of teleological explanations. Either children are selective in their application, using it for the categories of organisms and artefacts, or they are promiscuous in their application, using teleological explanation for all categories. Two problems arose from this section. First, previous work has at times used a limited number of exemplars to represent an entire category. It was suggested that a small number of exemplars may not be representative, therefore, investigation of multiple examples of a single category may be required (this is returned to in section 2.6.2). Secondly, what constitutes a category in prior work was vague. Consequently, a systematic method of classification was outlined: classification by ontological category, as detailed in Figure 2.1. The three ontological categories primarily discussed in this thesis are organisms, natural objects and natural phenomena, all representing categories of things that exist or occur in the natural world. Section 2.3.2 defined Keil's (1992) term natural kinds, an overarching term which refers to anything which originates in the natural world. The three categories used in this thesis fall under the definition of natural kinds. Therefore, when appropriate and to avoid using these three category labels, the term natural kinds is used.

2.4 Influences on teleological explanation

There are three main factors detailed in the literature that are argued to influence children's use of teleological explanations: artefacts, religious education and evolution (teleology as an adaption for survival), all of which are explored in this section. This section also proposes two additional factors that may influence the use of teleological explanations. Firstly, from an evolutionary perspective, and following the argumentative theory of reasoning, teleology may be a consequence of reasoning. Secondly, from a sociocultural perspective, teleological explanations may be influenced by cultural patterns of language where teleological explanations are used in a non-literal manner. Finally, if teleology is a preconception that exists before formal education then parental explanation could be a considerable influence, therefore, evidence from studies into parents (or other adults) use of teleology are examined. The factors explained in this section are not considered to be competing influences on children's use of teleology but rather factors which may work independently or collaboratively to influence children's use of teleological explanations for natural kinds.

2.4.1 Influence of intentional artefacts

Kelemen (1999c) proposed that one explanation for children's use of teleological explanations about natural kinds is that they consider then to be a form of artefact, in that children treat natural kinds as belonging to the ontological category of artefacts. Kelemen argued this position from evidence showing that from a young age children are aware artefacts are used for a purpose to fulfil

a goal (e.g., see Hanna & Meltzoff, 1993) and will imitate such behaviour (Carpenter, Negell & Tomasello, 1998). Children's understanding about the word developed from these common and numerous experiences with artefacts could then be transferred onto natural kinds. Children's exposure to artefacts is undeniable, children grow up surrounded by a plethora of artefacts: toys, clothes, books, food utensils, etc. Even the interactions with artefacts children based in rural locations have can be considered to be more varied and numerous than their interaction with organisms and natural objects (Kampourakis, 2014). Consequently, this exposure could lead to children becoming intentional users of objects and transferring the idea that all artefacts have a purpose into the natural world, assuming everything is designed. (Carey, 1985; Evans, 2000; Kelemen, 1999c; Lovell, 1961; Tomasello, 1999). Kelemen (2012), the main proponent of this theory, argued that children employ a teleological framework about the natural world in the absence of specific knowledge and that this framework draws upon artefact design and intentional agency to explain the existence of natural kinds.

Empirical support for artefact usage influencing the use of teleological explanation is suggested by DiYanni & Kelemen (2005), who asked children should X be replaced if it could no longer complete a specific function. For example, should clouds be replaced if they no longer make rain? The results indicate young children, aged 5-years-old, suggest living and non-living things should be replaced and older children, 9-years-old, propose that organisms should be replaced if they no longer perform a function. Essentially, natural kinds are viewed as having the properties of artefacts, highlighting a conflation of natural kinds with artificial kinds. Further work by Kelemen & DiYanni (2005) found that, when asked to stipulate on the origin of the first X (e.g., the first mountain), children often suggest that the original example of a natural object or natural phenomenon (e.g., the first rock) was intentionally created for a specific purpose. The authors argued this is evidence of belief in intentional creation of natural kinds because children refer to a designing agent when postulating on the origin of natural kinds, in explanations such as 'he' or 'they' made X (e.g., mountains) for Y (e.g., people to go skiing). There is also a suggestion that because teleological explanations begin to be used around the same time as children begin to understand human-made artefacts are intentionally designed, that this correlation indicates the two are intertwined (Kelemen, 2012; Kelemen, Seston & Saint Georges, 2012).

2.4.2 Influence of religion

Teleological explanation may be affected by a religious upbringing, by teaching children to believe in a supernatural being. Western-religious tenets often imply that a supernatural entity created all natural kinds for a purpose, often for the benefit of humankind (Thomas, 1984). Adherence to such a belief structure shares parallels with teleological thought, with respect to explaining evolution and some natural phenomena. There is evidence suggesting that theists have a stronger tendency to support teleological statements than do atheists. For example, Banerjee & Bloom (2014) found

religious and non-religious adults both advocated teleological explanations and those with a supernatural belief did so at greater levels than atheists.

It is contended that individuals' views of the world are shaped by iconic and common forms of representations (Boudry Blancke & Pigliucci, 2014): that is, individuals' world views are structured by their intuitive reasoning, see section 2.2.2. Boudry et al. argue that teleological thinking can result in various forms of pseudoscientific belief, including creationism and intelligent design, leading to a reluctance to accept or even engage with scientific ideas (e.g., the theory of natural selection). This position is also argued by others, for example, Dawkins (2006), Hood (2009) and Shermer (2002).

It may be that religion and teleology exist in a cyclical relationship, each influencing each other. This is supported by research with over 1000 participants which investigated cognitive biases and supernatural beliefs and found that agreement with teleological statements correlated with participants' religious beliefs (Lindeman, Svedholm-Hakkinen & Lipsanen, 2015, see also Willard & Nornzayan, 2013). Linderman et al. stated that although there was a link between teleology and religion, insufficient evidence exists to confirm causality or directionality of the relationship. These projects have focused on adults' advocacy of teleology and religious affiliation but what about children's beliefs?

Evans (2000) found that American children attending state schools espoused significantly fewer teleological statements that their counterparts attending fundamental religious schools. However, a comparative study of teleological reasoning in America and Britain (Kelemen, 2003) (using the same methods as Kelemen, 1999b, discussed in section 2.3.1) indicated that the supposed more secular British children and more religious American children demonstrated similar levels of teleology. One possible explanation for discordance between these two studies is that Kelemen sampled an American and British state school for comparison, where religiosity may be similar, whereas Evans sampled state and fundamental schools, where overt religiosity differs.

Corriveau, Chen & Harris (2015) provide further support for religious influence on teleological explanation. In their study, children (aged 5- to 6-years-old) were asked to classify story characters as real or fictional, in religious (e.g., David vs. Goliath), realistic (e.g., *Goodnight Mister Tom*) and fantastical stories (e.g., *The Wizard of Oz*). Children from religious background judged religious, realistic and fantastical characters as real, whereas those from non-religious backgrounds only maintained realistic characters were real. The attribution of real-status to fantastical characters by religious children is suggested by Corriveau et al. to be caused by their religious education, as those children without religious upbringings systematically claimed the character in fantastical stories to be fictional. While this evidence may appear tangential to influences on children's use of teleology, it is postulated that if religious children transfer the properties of religious characters onto fictional characters then they may also transfer those religious tenets onto natural kinds. This is especially pertinent for those individuals from a

fundamentalist background (such as those in the study by Evans, 2000, detailed above) with a literal belief that a supernatural entity created the Earth and life for humankind, although this belief can be mediated by personal interpretation of religious texts (Pihlaja, 2013).

2.4.3 Influence of evolution

There are two evolutionary accounts for the origin of teleological explanation. Teleology may persist in the absence of knowledge about natural kinds due to:

a) It being an advantageous adaption – teleological thinking conferred a selective advantage to the ancestral population;
b) The argumentative theory of reasoning – teleological explanations provide convincing but poorly reasoned arguments that are used to convince others.

This section explores these two accounts in relation to influences on children's use of teleological explanations.

2.4.3.1 Teleology and evolution: an advantageous adaption

Dawkins (2006) posits that as teleology is widespread, humans are born with this form of intuitive reasoning (see section 2.2.2), therefore, this style of thinking must have conferred an evolutionary advantage. Dawkins claims this advantage, or one of the advantages, is the *intentional stance*. The intentional stance is when individuals use their prior beliefs and knowledge about an entity to infer the goal-driven behaviour of the entity and predict its behaviour or intention (Dennet, 1989). This intentional stance is argued to contribute to human survival. Dawkins (2006) uses the example of a tiger, upon seeing a tiger in the wild an individual would take an intentional stance towards the tiger and predict that it means to eat the individual. Thus, through prediction via the intentional stance the individual can take countermeasures that could lead to survival. This view is also proposed by Hanke (2004) who suggests that teleology may be an evolved method of reasoning as it allowed distant ancestors to anticipate, and avoid, the actions of purposeful predator behaviour. Consequently, from an evolutionary position teleology is caused by an intuitive reasoning pattern which assumes entities are imbued with intention, that they display goal-driven behaviour (teleological behaviour). Reaction to this predicated intention may be beneficial to survival in time-limited situations.

Tomasello (1999) expounds on this account proposing that a purposeful view of the world, in terms of designed objects (artefacts) and intentional behaviour of others, is a beneficial cognitive adaption. This adaption is further developed through a *ratchet effect* or *cultural ratchet*, which Tomasello proposes enables human culture to be shared between members of a group; the group then shares and develops this culture with new group members across generations. In essence, the ratchet effect describes how culture develops, that knowledge within a group is shared and accumulates over time. This ratchet effect would suggest that over time a pre-existing teleological

view of the world could further expand and develop through the ratchet effect. That as children learn from and develop through social interaction with their ancestors, across generations a teleological view of the natural world, and consequently, the use of teleological explanations would expand and increase.

The idea that teleological thinking is an advantageous adaptation, which is then mediated by the ratchet effect to result in teleological thinking and explanation that is increasingly used within a group, suggests an expanded explanation for the influence of evolution. Considering the nature-nurture debate, the contribution that nature and nurture make to children's cognitive development, the initial explanation in this section would indicate that teleology is primarily informed by nature, that is an evolved way of thinking that is beneficial for survival. However, adding the ratchet effect to this explanation for the influence of evolution on teleology, would indicate that teleology is informed both by nature (advantageous adaption) and nurture (the use of teleology is intensified through cultural evolution). Consequently, one possible cause of teleological explanation is that teleology is (or at least was) a beneficial intuitive reasoning style, which is then further mediated by cultural influences, such as the influence of intentional artefacts (section 2.4.1) religion (section 2.4.2), language (section 2.4.4) or parental explanations (section 2.4.5).

2.4.3.2 Teleology and evolution: a consequence of reasoning

This section considers a different explanation for the influence of evolution on teleological explanations. It posits that teleology may result as a consequence of reasoning not, as outlined in the previous section, a method of reasoning that was an advantageous adaption.

Reasoning has traditionally been assumed to have evolved in humankind to enable individuals to clarify and perfect their intuitive beliefs in order to become better decision makers. This is often thought to take place through the rational revision and correction of intuitions or intuitive methods of reasoning (Frankish, 2010; Mercier & Sperber, 2017), such as teleology. This view of cognition is often termed dual process theory, where there exists two types of processing. Type 1 is unconscious, evolutionarily old and fast acting, it can be thought of as intuitive thinking. Type 2 is conscious, evolutionarily new, is slow and deliberative, and limited to humankind. It can be thought of as rule-based, analytical reasoning. In dual process theory conceptions arising from Type 1 intuitions can be evaluated and shaped by Type 2 reasoning (Frankish, 2010). Consequently, this view of reasoning would imply that individuals should be competent lone reasoners, with the ability to recognise their incorrect intuitions and develop robust conceptions (Mercier & Sperber, 2017).

The dual process theory of reasoning has clear links with the misconception movement (section 2.2.3) where incorrect preconceptions need be appropriately challenged and are then revised or replaced by the learner. However, Mercier & Sperber argue that the dual process theory does not account for a number of flaws in individuals reasoning, including why individuals often

display confirmation bias and why their, at times inaccurate, beliefs are robust and routinely maintained in the face of contradictory evidence and arguments. For example, the resilience of those who promote alternative medicine, such as homoeopathy, contrary to the overwhelming evidence of its ineffectuality (Goldacre, 2008, Shuman, 1997) shows that an individual's belief can be robust even against overwhelming contradictory evidence. A recently developed theory of reasoning provides some explanation as to why people may believe and advocate such incorrect ideas.

Mercier & Sperber (2017; 2011) proposed that the main evolutionary function of reasoning (selecting and entertaining reasons to accept and maintain a concept) is argumentative: reasoning produces arguments used to convince another. From this view, reasoning has evolved in humankind as it is beneficial to human communication, it allows individuals to share arguments that develop their ability to communicate effectively, leading to them becoming more reliable individuals. Consequently, communication must be stable and beneficial to those involved, otherwise it would not take place. However, an interlocuer could benefit from communicating false information, and lies can be particularly difficult to detect, therefore, hearers must be *epistemically vigilant*. Epistemic vigilance is the process of filtering communicated information to detect falsehoods: two methods of which are the trust placed in the communicator and the coherence of the new information (i.e., does it make sense?). As a result, should an individual wish to convince others of the validity of their information they must either be highly trusted or produce a highly coherent argument to support their premise. To produce a particularly convincing argument an individual will select information that supports their position, while often ignoring evidence to the contrary. This asymmetrical construction can result in arguments which, while initially convincing, are poorly reasoned and may not stand up to critical evaluation by another.

This argumentative function of reasoning, in the context of human communication, has a number of testable predictions, supporting studies are not detailed here but are provided in Mercier & Sperber (2011; see also Claidiere, Trouche & Mercier, 2017; Mercier & Sperber, 2017; Mercier, 2013). The predictions are:

1. Reasoning should be effective at producing arguments, which are used to convince others to accept a premise. Reasoning should also enable individuals to critically evaluate the arguments of others. Numerous studies show people are proficient at evaluating others arguments by themselves or in group situations, and when in an argumentative situation individuals are skilled at producing arguments which support their position (see for example Kuhn, Ramsey & Toi, 2015).
2. An individual should display a broad confirmation bias, only attending to information supporting their position. This confirmation bias should only occur during the production of arguments, not the evaluation of arguments. The established prevalence of the confirmation bias is argued to result from the argumentative theory of reasoning as individuals should only seek out arguments which support their reasoning, not contradict

it; therefore, resulting in the blind avoidance of counter evidence and counter arguments. (see Nickerson, 1998, for discussion on the widespread phenomena of the confirmation bias).

3. As arguments are used to convince others, reasoning should be most effective and prevalent in group situations where there is a difference between members' preconceptions. Where the group share the same beliefs there will be little argument or critical consideration of beliefs. Participants in groups perform considerably better than individuals on reasoning tasks. In these group situations, individuals who have an incorrect conception are convinced to alter their position through convincing and coherent argument, not through trust in a supposed knowledgeable individual (see for example Moshman & Geil, 1998).

One complaint about the argumentative theory of reasoning is that it is a teleological stance to reasoning, where the ability to reason must have a purpose (Gouran, 2012). The implication being that a purpose should not be supposed for a trait which had developed via natural selection, rather that arguments should focus on the development of reasoning, not its function. However, considering explanations about evolution and function, an evolved trait, be it physical (e.g., eyes) or mental (e.g., reasoning), can be described as having a function, provided the mechanism which lead to the emergence of the trait is evolution (Ayala, 1970). This issue is returned to in more depth in section 2.5.3.

A question arising from the argumentative theory of reasoning is what constitutes good reasoning skills (Gouran, 2012), could it be a) ability to convince others or b) ability to detect and probe the flaws in others ideas? Mercier & Sperber's theory is that reasoning has evolved to do both a) and b), there need not be a definition of good reasoning skills merely that reasoning just does a) and b) depending on the context. However, as the argumentative theory predicts what can be considered poor reasoning (e.g., confirmation bias, gamblers fallacy or the sunk-cost fallacy, see Mercier & Sperber, 2011), good reasoning skills could be the ability to acknowledge and limit any resulting poor reasoning. That is, individuals may only attend to arguments that support intuitive or prior beliefs at the detriment of ignoring counter arguments.

With regard to influences on children's use of teleology, it could be considered to result from the consequences of the argumentative theory. From this view teleology would constitute a lazy form of argumentation for supporting individuals' views. Teleology would not be considered a method of reasoning, as suggested by the influence of intentional artefacts (section 2.4.1) or by the advantageous adaption evolutionary cause of teleological explanations (see section 2.4.3.1), but rather a poorly reasoned, biased argument to support prior preconceptions.

With the advantageous adaption cause of teleology outlined in the previous section, it would be expected that teleology would be coherent, as it is an evolutionary protective (against danger) and investigatory (in terms of finding beneficial purposes for objects) method of reasoning.

However, if resulting from the argumentative theory it would be expected that teleology would be more fragmented, as its use has resulted from a multitude of arguments to support poor reasoning (i.e., teleology will vary across contexts as it is used ad hoc to support various arguments). Therefore, children should incorrectly advocate teleology to support their preconceptions. The use of teleological explanations may then be bolstered via confirmation bias deriving from such sources as 1) experience of intentional artefacts (section 2.4.1), 2) religious influence (section 2.4.3) ad explanations resulting from 3) the influence of language (section 2.4.4) and 4) parental explanations (section 2.4.5). The argumentative theory would also predict that when children evaluate teleological explanations, presented by others in group situations, they should be capable of spotting flaws (i.e., a focus on purposes, rather than causes, to explain the existence of X).

The main implication from the argumentative theory of reasoning, for this thesis, is that individual reasoning can lead to poorly founded preconceptions, as individuals generally focus on arguments which support their intuitions and prior beliefs. However, Mercier & Sperber (2017) argue that in social contexts reasoning may flourish as individuals are particularly competent at evaluating the arguments of others. This should work to the benefit of those involved resulting in conceptual change. However, this is an optimal outcome, which is not influenced by the complexities of the real world where individuals may have certain goals and agendas when communicating. Consequently, group discussion and argument may be an effective method of countering children's use and acceptance of teleological explanations. This is not an endorsement for oppositional argument, as it is particularly difficult to argue individuals out of long-term, firmly held beliefs when arguing from opposed positions (Andriessen & Baker, 2014; Shermer, 2002). Rather it is a suggestion that discussion about children's preconceptions should be effective, as discussion could guide reflection about concepts that children may never have properly considered.

2.4.4 Influence of language

To date, little attention has been given to how language influences teleological thought. An exception to that claim can be seen in the view of Hanke, who argues that teleological language is a problem, he alleged that "biology is sick. Fundamentally unscientific modes of thought are increasingly accepted, and dominate the way the subject is explained to the next generation" (Hanke, 2004, p. 143). However, this view is not widely accepted as there is little evidence proffered by Hanke to support the assertion that teleological language leads to the acceptance of unscientific ideas, although he does detail numerous examples of teleological language in the communication of scientific ideas. While Hanke solely targets the domain of biology within scientific discourse there is evidence that teleological language is prevalent in all facets of life: the parent telling their child to go to bed because it is dark, children's books suggesting that rivers *help* to shape the land (Berenstain & Berenstain, 1997), or reference to goals and rules in thermodynamics (Wicken, 1981) and chemical reactions (Talanquer, 2007). Here is further evidence of what Hanke suggests, the use

of teleological language in discussion about natural kinds. These examples appear to indicate a trend for teleological explanations to relate the subject to human activity, to human purposes, even if the subject cannot have a purpose only a cause. However, on the other side of the spectrum to Hanke's views, Ruse (1989) and Zohar & Ginossar (1998) argue that purposeful language about function in biology is not problematic as teleological formations are often figurative explanations, which when used as a heuristic actually support children's developing understanding. This section further explores the idea that cultural language patterns may result in teleological explanation without a literal implication of purpose.

Language is a cornerstone of young children's interaction with the world, shaped by prevailing cultural attitudes and norms. Consequently, as language and thought are intertwined the type of language and the context both influence how children think and explain themselves (Alexander, 2006; Bruner, 1996; Sampson & Babarczy, 2013; Vygotsky, 1986). Essentially, if children are exposed to discourse that primarily discusses the natural world in a teleological manner it may, in turn, influence them in to provide teleological explanations when asked to account for the existence of things within the natural world (e.g. rocks or rain). Furthermore, certain categories of natural kinds may have higher proportions of teleological, or scientific, discussion in natural discourse than others. This may then result in these categories having a higher chance of receiving a teleological, or scientific, explanations to account for their existence.

It should be noted that a Whorfian view (that the structure of a culture's language shapes and restricts the way speakers conceptualise the world, Whorf, 1959) is not advocated here, culture is used in relation a certain child's experiences, as summarised by Bronfenbrenner's ecological model (1979). Individuals develop their ideas intersubjectively (Mercer, 2002), therefore, unless children grow up quarantined from society, sociocultural influences are unavoidable (Olson, 2003). This social development leads humans to imitate and assimilate the norms, values, practices and, importantly, language of the culture in which they are situated (Curtis & Pettigrew, 2009). Hoff (2006) argues that the subcultures of, for example, educational environments, family or peer groups, combine and influence what can be summarised as a child's idiosyncratic culture. The culture a child experiences will in turn influence the styles and patterns of language they use to explain themselves. This model shows how multiple influences from different subcultures can combine to develop what can be termed an individual child's culture, and how that culture then affects children's use of language. Furthermore, although each child's culture will be specific to them, children in similar situations will have commonalities in their cultures.

There is evidence that children employ metaphors within science, for example, in relation to the Ozone layer (e.g., the Ozone is a protective cover on the Earth, trapping harmful radiation, Christidou, Koulaidis & Christidis, 1997) or soil (e.g., a worm's job is to dig out tunnels, to find leaves to eat, Pramling, 2010), and that these metaphors are related to children's personal experiences (Jakobson & Wickman, 2007). While the majority of work on teleology is conducted

in either American- or British-English speaking countries, there is a body of evidence that shows that teleological explanations are used by those for whom English is an Additional Language, including individuals from China (Rottman, et al., 2015), Greece (Kampourakis et al. 2012a; 2012b), Israel (Zohar & Ginossar, 1998), Mexico (Kelemen, et al., 2005), Romania (Casler and Kelemen, 2008) and Peru (Sánchez Tapia, et al., 2016). Therefore, this use of teleological metaphors or language using teleological phrasing in a figurative manner may cross language barriers. Consequently, it is argued that if children are developing in a culture which employs teleological language to describe scientific concepts, then they are enacting the norms of society by propagating these styles of explanations. This notion is perhaps best articulated by Bruner & Haste (1987, p. 2):

> Through language, the child is quickly aided in her entry into *culture*: its metaphors, its kinds of explanation, its categories, and its ways of interpreting and evaluating events. These are not *invented* by the child; they are the common currency of the culture, the framework that determines the boundaries of the child's concepts.

Therefore, it is suggested that teleological explanations may at times be articulated because subcultures exert different levels of influence over children's thought and language use. It is discordance between subcultures use of language which may cause children to espouse figurative, or literal, teleological explanations.

It is proposed that in certain situations teleological language is used descriptively, in a non-literal manner, to support the explanation of both scientific and non-scientific ideas about natural kinds. This would be caused by cultural language patterns, where members of a subculture understand that it is a non-literal use of teleology. This style of explanation could increase the number of teleological explanations given to children by adults and, therefore, children who do not yet understand teleological language as figurative may employ teleological statements to describe natural kinds. One implication of this is that assessment of children's explanations would need to clarify if the teleological language children use is figurative or not. Similarly, if children do not understand this figurative use then this language may exacerbate teleological thought regardless of the appropriateness of the teleological explanation. Consequently, children would be espousing teleological statements with the fervent belief that something has a specific purpose.

The figurative use of teleology, therefore, presents two possible influences of cultural language use upon teleological explanations. Either children are influenced by the teleological explanations used in everyday and science contexts and come to actually believe that natural kinds have a specific purpose or they are mirroring the teleological language used by science to describe the natural world. Research, by Zohar & Ginossar (1998) indicates the latter as they found that high school students purposely used figurative teleological explanations as a heuristic, but research in relation to young children is lacking. An investigation of the teleological explanations children's experience in natural discourse would determine whether children perpetuate a cultural use of

teleological explanations, that is, do children reuse and recycle teleological explanations they encounter in everyday language. The related question of whether young children can understand that teleology may be acceptable in colloquial explanations, in social contexts but not in scientific contexts, is returned to in section 2.6.5.

2.4.5 Parental explanations

The problem with the proposal that purposeful language influences children's teleological explanations is that there is little evidence about what children's natural discourse entails for different examples of natural kinds. It would seem self-evident that young children's explanations are influenced by their conversations, particularly parent-child interactions, as children's parents are key in their conceptual development (Athey, 1990). Therefore, parents will influence children's scientific explanations, as they are children's first and most enduring educators (DCSF, 2008). Little attention so far has been given to how parents influence children's teleological explanations or how they may perpetuate the cultural teleological language discussed in section 2.4.4. It is adults who model to children how to use language to reflect, analyse and communicate ideas, and how language is modelled influences how children communicate (Mercer & Littleton, 2007).

Existing research suggests that parents provide causal, rather than teleological, explanations to children's questions about the natural world. Callanan & Oakes (1992) found 32% of adult explanations about the natural world were causal for 3-year-olds, 61% for 4-year-olds and 54% for 5-year-olds. A possible problem with this study is that it employed parental diaries to report children's questions, and parents' explanations, about *how things work* and *why things happen*. Focusing on these two questions forms could be considered a suitable method of investigating children's developing causal knowledge - the studies purpose. However, this project provides no information regarding teleological statements that can happen *in situ*, as opposed to child-initiated questions. For example, a child being told that the rain will water the grass. Consequently, teleological statements espoused by parents, without prompting from a child, will have been overlooked. Children may have been exposed to a lower percentage of causal explanations about scientific concepts than suggested by Callanan & Oakes.

Similar research by Kelemen, et al. (2005), specifically investigating teleological explanations given by parents, found that when parents respond to children's questions about *why things happen* or *how things work* parental responses are casual rather than teleological in nature. Furthermore, the number of overall teleological explanations provided by parents were low. For the category of natural objects and phenomena 28% of explanations were teleological, with 64% being causal, whereas for organisms the mean percentage responses were 15% teleological with 71% causal. The paper concluded that children's use of teleological explanations was not directly attributable to parental explanations, as there were twice as likely to be causal than teleological (Kelemen et al. 2005). However, this study has similar limitations to Callanan & Oakes (1992),

where the data collection method involved parents reporting children's *how things work* and *why things happen* questions via phone interviews. Again the focus on child-initiated questions negates any data regarding naturally occurring parent explanations or 'off-hand' teleological explanations (i.e., those which naturally occur in parent-child discussions).

The reported studies (Callanan & Oakes, 1992; Kelemen et al., 2005) into parent-child interaction only consider children's questions and their parents' responses. While this is a valuable method for uncovering specific examples of teleological and scientific conversation, teleological discourse around certain topics may not always result from child initiated questions. The impromptu discourse about a subject could provide different evidence of teleological conversation when compared to children's specific questions about the same subject. This suggestion is supported by Fender & Crowley (2007) who found that adults explanations that were not child-initiated influenced children's classification of the purpose of novel artefacts. Fender and Crowley suggest that these small fragments of explanation could accumulate to effect children's thinking and development of scientific classification models. Consequently, if teleological language is modelled to children, it could be propagated in their explanations about scientific concepts.

An example of parental-child dialogue influencing teleological explanation would be discussion between a parent and child about the topic of snow. This discussion may focus the snow being for a play activity (a common teleological explanation). However, a parent responding to the question *why is there snow?* is unlikely to respond with *it is for playing in*, posing a problem for using parental diaries as in Kelemen et al. (2005). Therefore, it is argued that investigating patterns in natural discourse around natural phenomena could provide different evidence than investigating children's questions and the resulting adult responses. Consequently, a method is needed which can explore children's natural discourse, is generalisable and is not limited to a specific element of children's talk. This research area is discussed further in section 2.6.5.

2.4.6 Summary of section

It may be that children's use of teleological explanations is influenced by artefacts, due to artefacts having an intended purpose; that it is a by-product of religious dogma, stating that a deity created the heavens and Earth for human-kind; or that this way of thinking was an advantageous adaption, in human evolution. However, this section proposed that teleology may be a consequence of reasoning, in line with the argumentative theory of reasoning. That is, teleological explanation is used in discourse as a poorly reasoned argument to support prior assertions, with the aim to convince another. This view of teleological explanation would suggest that teleology could flourish in individuals thinking but in social situations teleological arguments could be subject to critical analysis by group members. Thus, indicating that collaborative work may be appropriate to counter individuals' use of teleology.

An argument has also been made that teleological statements could be figuratively employed in certain subcultures, which may lead children to use or endorse a teleological style of explanation. The influence of language is one of a number of factors, including artefact-usage, religion and evolution, which may cause children to advocate teleological explanations. It is not suggested that teleological language causes children to *reason* teleologically, rather that teleological expressions are the linguistic norm in certain cultures and that children may simply be reproducing those. This view of teleological explanation, that it could at times be influenced by cultural patterns of language, represents a conceptual shift in the theory of teleology. Therefore, teleological explanation may not always be intended to suggest a specific purpose or goal-driven behaviour but at times used to suggest a relationship between an example of natural kind and its interaction with, or effect on, humankind. How teleology is constructed, in terms of different forms of teleological explanation is detailed further in the next section (2.5.2).

A likely source for this cultural, figurative use of teleology would be parental explanations. Although, the studies reviewed indicate little use of teleological explanation by parents. However, consideration of research methods suggests that the role of parental explanation in influencing children's use of teleological rationales may have been overlooked, as previous work may have omitted naturally-occurring teleological statements made by parents (or other adults). Consequently, different methods may be better suited to research teleology in natural discourse, this is explored further in section 2.6.5.

2.5 Constructing teleology

This section explores three facets of how the concept of teleology is theoretically constructed. Firstly, through the lens of naïve ideas (children's preconceptions) and conceptual change (the alteration of these preconceptions), teleology is investigated to see if it is the result of coherent or fragmented knowledge about the natural world. Secondly, section 2.4.4 indicated that there may be different forms of teleological explanation. Therefore, this section reviews the literature around teleology to define several different forms of teleological explanation. Thirdly, there is considerable debate surrounding when (contexts) and where (ontological category) teleology is appropriate. For this reason, the section will end with consideration about when teleological explanations are appropriate, with regard to the type of teleology explanation used and ontological category being explained. All three of these theoretical aspects of how the conception of teleology is constructed have different implications for understanding children's use of teleological explanation and for supporting conceptual change. Those implication are detailed throughout.

2.5.1 Naïve ideas

An initial note on terminology, naïve ideas (used non-pejoratively) are another name for preconceptions. Preconceptions has been used thus far as it is part of the lexicon of educators; however, the following reverts to naïve ideas as that is the dominant phrasing used in the current

conceptual change debate (although misconceptions has been dominant in the past, e.g., see Allen, 2014). The structure of naïve knowledge is highly debated and there exists a range of different stances. If this range of viewpoints is considered on a continuum, then at either end of this continuum are two positions. The furthest end of one side of this continuum suggest that naïve knowledge is coherent (it has a theory-like structure), whereas those researchers on the opposite end of the continuum would argue for a fragmented view of naïve knowledge (knowledge is formed of pieces of ideas that combine when needed). These two poles have different origins and implications for supporting conceptual change and are discussed in detail below. It should however be noted that these descriptive terms are used in reference to the extremes of a scale which helps summarise the key debates. However, in practice the arguments can be more nuanced. For example, many researchers do not consider naïve ideas to be one hundred percent coherent or fragmented but rather somewhere in between (diSessa, 2013).Moreover, both views share commonalities; for example, that individual's naïve ideas are dynamic and consistently evolve over time (Brown, 2010). There are some theories which can be argued to be almost directly in the centre of the continuum, such as Howe (1998). Howe proposes an action-based theory of cognitive growth in which children have a theory-like understanding of mechanisms (e.g., that clouds move and something must cause this movement) but a fragmented notion of variables that cause such mechanisms, which are derived from egocentric and linguistic experiences (e.g., that walking causes cloud movement). Therefore, such a position incorporates both coherent and fragmented perspectives by applying them to different aspects of naïve ideas. From this point forward these two terms, coherent and fragmented naïve ideas, refer to either end of a continuum; however, a more nuanced distinction will be recapped at certain key points in the thesis.

2.5.1.1 Knowledge-as-theory (coherent naïve ideas)

The dominant view in conceptual change literature is that individual's naïve ideas are coherent (Brown, 2010; Özdemir & Clark, 2007), or more aptly that if the structure of naïve ideas is considered to exist on a continuum then the dominant view is positioned towards one side of this continuum that is associated with a more coherent structure. Essentially, children's naïve ideas are part of structured concepts which cannot be isolated but rather form a web of ideas (Amin, Smith & Wiser, 2014). A key area of the debate surrounds the context in which children use certain naïve ideas (diSessa, 2013). Knowledge-as-theory would imply that children's naive ideas are limited in score and stable across numerous contexts: when asked to answer a question a group of similarly aged children would provide similar responses (Amin, Smith & Wiser, 2014).

Support for a core, coherent set of naïve ideas is provided by Ioannides & Vosniadou (2002) in a manuscript which typifies the stance of knowledge-as-theory. The term theory is used here in relation to an individual's explanatory framework. This framework is consistently applied by an individual across contexts (i.e., across questions dealing with the same concept) and should be

shared by the majority of members in a group, resulting in an explanatory system which displays considerable coherence (Ioannides & Vosniadou, 2002). In this study 105 children (aged 4- to 16-years-old) answered questions about forces. Children's responses suggested that children's conception of force was uniform and consistent across questions and individuals, although differences were apparent at various developmental phases. That is, children in similar age ranges provided a small number of similar explanations (naïve ideas) about force, and the older the children the closer their understanding of force became to the established scientific definition. Similar results, suggesting knowledge has a theory-like structure, have been obtained in other subject areas; for example, in evolutionary biology (Samarapungavan & Wiers, 1997), in motion (McCloskey 1983) and in astronomy (Vosniadou & Brewer, 1992).

In terms of implications for conceptual change, the coherence stance argues that conceptual changes takes place slowly over time. It works by presenting numerous counter arguments to an individual's current explanatory framework, until such a point when their current view is no longer viable in the face of counter argument. The individual's prior theory is then replaced with a more refined theory which may, at some point, coalesce with an established scientific theory (Ioannides & Vosniadou, 2002). In essence conceptual change works by challenging an individual's entire explanatory framework and replacing it with the correct, or more correct, framework.

2.5.1.2 Knowledge-in-pieces (fragmented naïve ideas)

Considering a continuum of viewpoints on the structure of naïve ideas, if one side focuses on a more coherent structure of naïve ideas than the other extreme of the continuum suggests a more fragmented structure. The description *knowledge-in-pieces*, coined by diSessa (2013), asserts that naïve ideas are a collection of fragments which combine when needed to form an idea. These fragments are often referred to as phenomenological primitives, or *p-prims*, which are often indescribable, microgeneralisations that arise from individual's experiences (diSessa, Gillespie & Esterly 2004; Davis, Horn & Sherin, 2013). These p-prims cannot, by themselves, be considered objectively correct, in relation to any established theory, until they are combined by an individual in a certain context. Furthermore, while they may be used in ways that are incompatible with scientific accounts, they may have utility in real world situations (diSessa, 2014). This aspect shares similarities with what is discussed as the cultural use of teleology (see section 2.4.4). The notion that p-prims can be combined differently across individuals and contexts, give rise to the principle that naïve ideas should be numerous and variable, they will not be coherent across a group of students.

Where the coherence position would imply that children's naïve ideas are stable across contexts, the fragmented stance suggests that children's naïve ideas will differ depending upon the context in which they are explaining, and when asked a question a certain group of children may provide a wide range of contradictory responses. Support for fragmentation of naïve ideas, with

regard to context, comes from diSessa, Gillespie & Esterly (2004) which is a quasi-replication of the aforementioned Ioannides & Vosniadou (2002), 30 children took part drawn from the same age-range. The procedure was broadly the same, the main difference being the inclusion of a number of additional questions which targeted the same concept (about physical forces) but within an altered context. This context change for the additional questions resulted in a considerable departure from Ioannides & Vosniadou derived set of core, coherent ideas for physical forces. diSessa et al. proposed that their dataset showed that children's naïve ideas were not compactly describable, as children's ideas varied considerably both within and across individuals. Furthermore, as part of a preliminary study which used the exact methods of Ioannides & Vosniadou (2002), diSessa et al. (2004, cited in diSessa, 2013, p.37) collected no responses from the children which demonstrated the core ideas of force advocated by Ioannides & Vosniadou. The same failure to replicate a supposed core set of naïve ideas was again achieved in relation physical forces (Clark, D'Angelo & Schlieght (2011), in biology (Southerland Abrams, Cummins & Anzelmo, 2001) and thermodynamics (Clark, 2006b). While currently not the dominant stance, this fragmented view is on the increase (Özdemir & Clark, 2007)

In terms of implications for conceptual change it is done through the reorganisation of knowledge fragments. This means that conceptual change takes time and requires multiple contexts, which address the multiple ways with which individuals see the world (diSessa, 2013); that is, through their multiple realities (see section 3.2 for discussion on ontological positioning). However, the benefit is that a naïve idea need not be completely demolished then rebuilt, rather through the gradual reframing of elements of children's ideas, educators can gradually 'chip away' at children's preconceptions which result in incorrect scientific ideas. From a simplified position, conceptual change for knowledge-as-theory can be thought of as arguing against and then replacing children's preconceptions with the *correct* concept, whereas conceptual change in knowledge-in-pieces can be thought of as the ongoing process of chipping away at and restructuring numerous knowledge fragments to represent, or more closely represent, the *correct* or target idea (Davis, Horn & Sherin, 2013).

2.5.1.3 Teleological naïve ideas: coherent or fragmented?

Teleological explanations can be thought of as preconceptions which arise when asked to explain why something occurs (e.g., why does rain happen? – for watering the grass) or the existence of a trait (e.g., why do reindeer have antlers – for digging up food). Consequently, teleological responses to these questions are an indicator that an individual has the naïve idea that things happen for a purpose. However, the question arises, *is teleology, when demonstrated, comprised of a coherent theory or a collection of fragments?* The answer to this question is dependent upon if there is a) within-child variation across contexts and b) across-child variation across contexts. In this situation, different contexts would be represented by different exemplars of

ontological categories. For example, rocks, soil or ice within natural objects; rainbows, fog or night within natural phenomena; spiders' silk, giraffes' necks or rhinos' horns within organisms. As discussed earlier, the terms coherent and fragmented are used to refer to opposing sides of a continuum. Therefore, what this question can be seen as asking is *what side of the continuum do teleological explanations fall, are they more coherent or more fragmented?*

Current thinking around teleology (e.g., studies by Kelemen, 1999a; Kampourakis et al., 2012a; and Keil, 1994, reviewed in sections 2.1.3 and 2.3.1) appears to indicate that children's use of teleological explanations is caused by theory-like naïve ideas. That is, these studies claim that children broadly apply a teleological naïve idea across ontological categories (e.g., natural objects or organisms) and teleology is consistent across individuals (e.g., children provide a small number of similar teleological explanations for a single exemplar). This is evidenced in the limited number of compactly describable teleological examples provided in papers and the general assumption that a limited number of exemplars can represent the teleological preconception for an entire ontological category.

It is maintained in this thesis that current research on teleological reasoning depicts it as a coherent naïve idea. That is, individuals provide similar levels of compactly describable teleological explanations across a range of discussion topics. This stance would predict that teleological ideas are highly resistant to change, due to them having a coherent, theory-like application. When a group of children is asked to provide an explanation for a range of contexts (e.g., different exemplars of natural phenomena such as lightning or wind), children's responses should be a) coherent within children (i.e., a child would provide similar levels of teleological explanations across the exemplars investigated) and b) coherent across children (i.e., all children should give a similar response for any given exemplar). Consequently, teleology would be a significant barrier to learning, as conceptual change would require the destruction of the teleological idea (purpose in nature) and the redevelopment of the correct explanatory framework (that nature is causal).

Whilst not currently advocated in the literature, following diSessa, it is proposed that a fragmented view of teleology might be advanced, in which teleological ideas would be more nuanced and contradictory (i.e., not stable for a category or an exemplar used to represent a category). Children's teleological ideas could be more susceptible to change as they constitute a loose collection of elements that can be reshaped one at a time. If this is the case, then when a group of children is asked to provide an explanation for a range of contexts (e.g., different examples of a single ontological category), children's responses should a) exhibit reasonable variation within children (i.e., a child would not provide similar levels of teleology across the range of exemplars used) and b) exhibit variation across children (i.e., a group of children would provide a range of purposeful responses for any given exemplar).

If teleology results from fragmented knowledge it may represent a diminished barrier to learning, as conceptual change would require the incremental reshaping of targeted knowledge

elements. It is viewed as a diminished barrier not because teleology is less of a problem but rather because a fragmented construction suggests a new way of tackling the barrier that is teleology. Teleological thought and explanation would not need to be entire challenged and wiped out, but through the gradual reframing of elements of children's ideas, educators could gradually 'chip away' at children's preconceptions which result in teleological explanation. This could be achieved through target interventions, for example where individuals learn when and where teleological explanation is appropriate (e.g., as learning heuristics, see Zohar & Ginossar, 1998).

This section has explored what constitutes a naïve idea in relation to conceptual change. The structure of naïve knowledge can be considered as a continuum with one side indicating a coherent structure and the other side suggesting a fragmented structure. Knowledge-as-theory (coherent naïve ideas) would suggest that children's ideas about a particular subject are compactly describable and demonstrate coherence both within and across a group of similarly aged children. Whereas, knowledge-in-pieces (fragmented naïve ideas) would indicate that children's ideas about a particular subject are numerous and varied, there would be reasonable variation both within and across a group of children. While teleology appears to currently be viewed on the continuum as a more coherently structured concept, an argument has been proffered that posits a more fragmented structural view of teleology. Counter to the knowledge-as-theory stance, this would predict children's teleological ideas are more nuanced and varied both across and within children for a selection of category exemplars. The knowledge-as-pieces stance would suggest that the teleology barrier to learning about the natural world could be tacked through a series of target interventions that sought to reshape separate elements of knowledge, in contrast to the traditional approach which would require teleological concepts to be demolished and the correct concept rebuilt in its place.

2.5.2 Types of teleology

A number of previous studies have attempted to classify different types of teleological explanation, although this is not a common strategy and there are numerous overlaps between different studies. Combining the sub-classifications of teleology used by other academics suggests four categories of teleological explanation, displayed in Table 2.1. In relation to the previous section, 2.5.1, these types of teleology are maintained regardless of whether teleology is the result of coherent or fragmented theories.

Table 2.1: *Types of teleology derived from literature*

Category	Explanation	Related literature
Design-teleology	A topic is explained in relation to a purpose or goal for which it was designed. The topic has the singular purpose of aiding/supporting something, the something can be itself or another entity; it acts upon an entity to produce a result. This explanation is not scientifically valid. For example, 'rain is for watering the plants' or 'day is for waking up the birds'	Woodfield, 1979; Zohar & Ginossar, 1998; Kelemen 1999c; Polling & Evans, 2002; Kelemen & DiYanni, 2003; Kelemen & Rosset, 2009; Kelemen, 2012
Religious-teleology (a subcategory of design-teleology)	This is a specific type of design-teleology. A topic is explained in relation to a purpose or goal for which it was designed, however, a god or other supernatural being is invoked as designer of the topic. For example, 'there is a storm because god is angry'	Woodfield, 1979; Kelemen & DiYanni, 2003; Lindeman el al., 2015
Functional-teleology	A topic is explained in relation to a function that it performs, within a system, or a goal which it pursues. What is paramount is that the function or goal is appropriate for the topic. The topic can be explained scientifically using teleological language. It is an appropriate scientific explanation. For example, 'an eagle's wings are for flying', 'a dairy cow is for milking'.	Woodfield , 1979; Zohar & Ginossar, 1998; Kelemen & Rosset, 2009; Kelemen, 2013
Relational-teleology	A topic is explained in relation to a purpose for which it can be used or an effect which the topic causes. The purpose or effect is human-centric. For example, 'snow is used by children for making snowmen' or 'night is a time period which people usually use for going to sleep'	ojalehto, Waxman, & Medin, 2013

Teleology has long been seen as problematic by philosophers of science (Ruse, 1989), although considering the sub-type of teleology, as shown in Table 2.1, this view refers specifically to design-teleology and its subcategory of religious-teleology. Design-teleology is argued by Hanke

(2004) to be a debilitating factor which restricts truly creative scientific thinking because it places the focus on purposes, resulting in children having the misconception that natural kinds have a teleological rationale. However, Hanke gives little indication of what 'creative' scientific thinking would encompass.

Design-teleological thinking is considered by Kampourakis (2014) to be the main conceptual barrier in understanding evolution, as individuals confuse artefact inspired design-teleology with organism functional-teleology. Essentially, children apply their conceptual framework about the creation of artefacts to the existence of natural kinds. This application of design-teleology echoes Piaget's (1929) notion of artificialism (section 2.2.1) and is often used as a rationale why young children struggle to learn causal explanations for the existence of natural kinds (Peacock et al., 2011).

The use of teleological explanation in relation to organisms and biological traits is debated. From a theoretical position, opponents argue that it should not be used or can be misleading (see Hanke, 2004; Kampourakis, 2014) whereas proponents suggest that functional-teleological explanations are necessary in discussions about evolution (see Amundson & Lauder, 1994; Ruse 1989). These arguments typically come from the philosophy of science or psychology. However, research in the field of science education indicates that high school students use teleological explanations for biological concepts without sacrificing scientific rigour and neither does the use of teleology in textbooks influence children to provide teleologically incorrect explanations (Zohar & Ginossar, 1998). Essentially, teleology can be used metaphorically, as a cultural method of description to aid understanding, rather than a stand-alone explanation.

Some teleological explanations may arise because children conflate cause with effect (Olson, 2003); for example, *she put her hood up because it was raining* (causal) with *it was raining because she put her hood up* (effect). This is supported by the relational-deictic framework (ojalehto et al., 2013), which suggests that when children provide teleological responses they are not stating that something has a designed purpose but merely relating it to a purpose (or multiple purposes) for which it can be used, or an effect that it causes. For example, how children use snow for making snowballs. This form of explanation is further referred to as relational-teleology. As the paper by ojalehto et al. is theoretical it provides no empirical evidence about the levels of teleological explanations which are relational-teleology, nor if children are even expressing this type of teleology.

This discussion of teleological explanation implies that they are easily coded into distinct categories, however, this may not always be possible. Certain examples of teleological explanations may be vague or provide little information to make a categorical decision as to which type of teleology is demonstrated. Consequently, the line between what is a design-teleological statement and a relation-teleological statement could at times be blurred. It may, therefore, be appropriate to think of teleological explanations as existing on a spectrum of teleology, where there will be a

graduated overlap between different categories of teleology. This is not to say that the categories that represent these two types of teleology overlap, rather that some examples of teleological explanation may be difficult to place in one category or the other. The main distinguishing feature of design- and relation-teleology is that the latter focus on how something is related to a human agent, a linguistic method for identifying this agent in vague teleological statements is outlined in section 4.2.5. From this point forward in the thesis types of teleological explanation (e.g., abstract categories such as relational-teleology) are treated as distinct categories of explanations, at the same time it is recognised that certain concrete teleological statements made by children (e.g., snow is for snowball fights) can be considered on a spectrum of teleology where they may lean more towards one category (e.g., relational-teleology) than towards another (e.g., design-teleology).

Investigation of types of teleology poses the question *what types of teleology do children use when they reason about different scientific concepts*. This is an important question because, as previously mentioned, certain types of teleology are appropriate whereas others are not (this is further elaborated in section 2.6.3). For example, if children use a high proportion of relational-teleological explanations this would not correspond with the traditional view that design-teleology in the dominant style of teleological explanation. Investigation of this type would show if children are confusing function with design or simply providing relational responses about scientific concepts. Should children show high levels of relational-teleology it would suggest that previous work may have overestimated children's tendency to provide design-teleological explanations. Furthermore, analysis of type of teleology also raises the question of whether children be taught to distinguish between different types of teleological explanations. Specifically, are children able to comprehend the contexts in which certain teleological explanations could be appropriate and when they are not. The notion that children could be taught to understand the appropriateness of teleological explanations is returned to in section 2.7.

2.5.3 Acceptability of teleology

Opinions on teleology are usually polarised. The use of teleological language in scientific contexts is viewed as either inappropriate and the cause of developmental difficulties or, alternatively, as a possible method of supporting individuals' understanding of scientific explanations about the natural world. This section reports the arguments surrounding the problems associated with teleological explanations before considering when teleology can be considered appropriate and, consequently, when children's teleological statements are correct. While the relevance of some types of teleological explanations for organisms is debated, there exists no such controversy for the existence of natural object or natural phenomena: teleology is absolutely incorrect for examples of these ontological categories. Furthermore, teleology is essential for artefacts, as they are human-made objects that are created specifically to serve a purpose. Consequently, the debate around teleology as problematic concerns teleological explanations about the ontological categories of

organisms and by extension organisms' traits. It is this debate around teleology in biology that the remainder of this section explores.

The following outlines two sides of this debate around the appropriateness of teleological explanation for organisms, an idea which is particularly contentions within scientific disciplines (Ayala, 1970). What is not discussed is creationist explanations for the existence of organisms, those which employ a religious form of teleology (Woodfield, 1979; Lindeman et al., 2015). There is limited evidence that children use this style of explanations in the literature. However, in a scientific context, a religious-teleological explanation would be considered inappropriate to account for the existence of organisms. Also not discussed is relational-teleology, explaining an organism or its trait by suggesting it exists because of something a human uses it for. For example, a cat (or the cat's fur) exists so people can stroke it. In these styles of teleological explanation the purpose may be justified, people may have a pet cat because they enjoy stroking it but it is not scientifically true, as a scientific account where a causal explanation would be required to account for the development of cats' fur (a complicated thing to explain but one that would appeal to arguments that hair and scales both develop from the same cellular structure, Di-Poï & Milinkovitch, 2016, that fur is beneficial for warmth, the sense of touch, Switek, 2014, and for cats' camouflage, Allen, Cuthill, Scot-Samuel & Baddeley, 2010).

2.5.3.1 Teleology as a barrier in biology

Teleology has long been seen as problematic by philosophers of science and by some within the scientific community (Ruse, 1989). Treating organisms and their traits as having a purpose is often treated as a hurdle or barrier to learning about evolution (Galli & Meinardi, 2011; Hanke, 2004; Kampourakis, 2014). Ribeiro et al. (2015) argued that teleology in biology is an *epistemic obstacle*, a form of knowledge or concept which is adaptive but results in incorrect answers which are resistant to contradicting evidence due to its apparent, but incorrect, explanatory power. This leads to teleological preconceptions becoming integrated with causal scientific theories in primary school and left unchallenged can then pose considerable problems to older children's science education (Kampourakis, 2009; 2014; Kelemen, Emmons, Schillaci & Ganea, 2014).

This results in another problem: the perpetuation of teleology. It is not only individual concepts within an ontological category that are viewed teleologically but the entire category. If one concept is considered to have a purpose then similar concepts are classified in the same way (Chi, 1992). Consequently, this may lead to misclassification being an effect of teleological thinking. The relationship between teleology and misclassification can be viewed as cyclical: misclassification causes teleological thinking and teleological thinking causes misclassification of similar concepts, *ad infimum* unless radical conceptual change occurs (Bloom & Weisberg, 2007; Chi, 2008).

Teleological explanations are particularly prevalent in biology and are proliferated in popular scientific writing because authors anthropomorphise animals and plants, to aid readers' scientific understanding, an example would be 'somatic cell divisions are used to make moral tissues, organs and instruments whose purpose is the promoting of germ line divisions (Dawkins, 1982, cited in Hanke, 2004, p.143). As mentioned, and discussed in further detail in section 2.5.3.2, many writers use this form of language as it may aid readers understanding. However, Hanke views this use of teleological language as unacceptable as it leads individuals to adopt unscientific models of thinking, although he does not provide evidence. What is discussed here is not Hanke's fundamental position but the notion that the use of teleological language leads to readers using teleological explanations, and without an understanding that this language is an explanatory aid the reader may develop a misconstrued understanding of a concept.

The use of teleological language can be problematic because communication between individuals relies on a co-constructed meaning of language and conceptualisation of concepts (Korman, Voiklis, & Malle, 2015), which permeates a culture and affects how members think about different subjects (Hoff, 2006). Words are not just used for the communication of ideas but for the process of idea formation. Clark (2006a) proposes that language creates *cognitive niches* which alter how people think and talk about different domains. Therefore, certain words or phrases can become signals signposting that certain objects share common abstract trails. This problem of subcultural use of teleological explanations was explained in section 2.4.4, which argued that there may be a problem when using teleological language without ensuring it is going to be construed as figurative usage. That is, the risk of miscommunication if an audience is not inducted within the subculture. Again, the problem with using teleological language if children do not view it as non-literal usage reoccurs.

There are a number of reasons why natural selection is one of the most difficult concepts in science education. One explanation is that children focus on individual organisms rather than populations, leading to the Lamarckian idea that evolution happens due to desire or action on behalf of an individual organism (Kelemen et al., 2014), a teleological style of explanation. The argument that teleology results in the incorrect assumption that evolution takes place due to the desire of an individual organism to change, or through an organism's goal driven behaviour, is also supported by Ruse (1989; 2000) and Ayala (1970). Kampourakis (2007) argued that this form of teleology is due to individuals' projecting their intentional thoughts onto organisms, as teleological explanations share commonalities with individuals' experience of their own purposeful behaviour. This projection of goal-oriented thinking on to organisms is form of anthropomorphism, which at times can be closely intertwined with teleological explanations (Wood-Robinson, 1995), as discussed in section 2.2.2.

Another argument for teleology being a core obstacle to learning about evolution, is that this style of explanation leads to the concept that organisms' change due to an evolutionary goal-

directed path (Galli & Meinardi, 2011). That is, the view of evolution as having a linear path towards a goal, the goal being providing beneficial, functional traits to organisms or producing perfect, terminal organisms (Ayala, 1970; Kampourakis 2007; Ribeiro et al., 2015; Ruse, 2005). Furthermore, goal-oriented teleology can detract from an important aspect of natural selection, which is the extinction of a species, as individuals only focus on the continued improvement of organisms (Kampourakis & Zogza, 2009; Ribeiro et al., 2015). Therefore, a goal-oriented view may result in a limited understanding of the scope of evolutionary theory.

Even though teleological styles of explanations may have a heuristic value (something which is returned to later). Kampourakis (2007; 2014) proposed that teleology can result in inaccurate conceptions of scientific ideas in biology, as the explanation focuses on the function of a trait rather than the underlying process that developed said trait. Focusing solely only the function of a trait can lead to problems if the explanation of the function a trait performs is conflated with the origin of the trait, leading to the notion that the trait originated *to perform that function* (Kampourakis 2007; 2014). For this reason functional-teleological explanations are argued to be inappropriate, if the explanation focuses only on the physiological description of a trait whilst omitting the explanation for the origin of the trait. Essentially, the argument against using teleology to explain the function of organisms' traits is that they are not completely scientifically accurate because the explanation for the origin of the trait is missing (Hanke, 2004; Ribeiro et al., 2015).

Looking at the teleology barrier and why it is inappropriate in scientific contexts results in three main explanations for needing to challenge this preconception. Teleology can result in:

1. the use of organisms' desires as the cause of evolutionary change;
2. evolution being thought of as having a predetermined path or natural selection as a force which has a goal;
3. accounting for the existence of organisms' traits solely in terms of their function, which is not appropriate in scientific contexts (although may be acceptable in non-scientific contexts).

2.5.3.2 Teleology as constructive

As above, many think teleological questions and explanations should not be used in scientific contexts (e.g., Ribeiro et al., 2015; Hanke, 2004). However, Ruse (1989) argues that this is extremely difficult and that even staunch critics of teleological thinking still employ teleological explanation, as it is a convenient and a natural style of explanation (also see Ayala, 1970; Allen 2009). For example, Darwin is claimed to be one of the most prolific users of teleological explanation in relation to evolution (Ruse 1989) as is Dawkins (Hanke, 2004), who was once the University of Oxford's Professor for Public Understanding of Science. Neither of these individuals are considered to be incompetent communicators of biological ideas, specifically evolution! This

raises the idea, initially discussed in section 2.4.4, that teleological explanations for organisms can be appropriate, that they may aid understanding provided that individuals receiving the explanations understand that they are used in a non-literal manner. This section further explores the notion that teleological explanations in biological contexts can be constructive.

Ruse (1989; 2005) considers teleological questions (e.g., *what is the purpose of ladybirds' spots?* or *what is the function of pigs' curly tails?*) to be an important heuristic to discover the origin of organisms' traits. The use of teleological questions, as the starting point for scientific investigation, is common among children, with research finding that young children's (3- to 4-year-olds) questions about animals and their traits primarily focused on the purposes and functions of the traits, rather than the purposes, behaviours and functions of the whole animals (Greif, Nelson, Keil & Gutierrez, 2006). This focus on the function and purpose of traits is understandable as biological traits often cannot be explained without reference to the function they provide for an organism (Amundson & Lauder, 1994). This focus on trait function is especially pertinent to those without a high-level of formal scientific training, such as children.

The use of teleology as a heuristic, as an efficient method of initial reasoning about the existence of traits, is argued to be beneficial to students (Zohar & Ginossar, 1989). However, for a teleological starting point to be effective, a causal answer must subsequently be found for the development of the trait and this is then combined with the original teleological starting point. This combination of teleological and causal explanations can result in an appropriate scientific rationale for the origin of a trait which uses a *teleological metaphor* to aid the explanatory power of the rationale (Ruse, 2000). With reference to the influence of language (section 2.4.4 and 2.4.5) what is being discussed here is a subcultural use of teleological explanation in a non-literal manner to support understanding. Although there is an assumption that the audience is aware of this cultural use of language, which may be problematic as mentioned in section 2.5.3.1. Therefore, the use of teleological language to support scientific explanations in biology is context dependent, the interlocutor needs to be aware of this cultural use of teleology if it is to support understanding.

A teleological metaphor is argued to be useful for developing understanding as it may simplify rather complex scientific process (Ayala, 1970), providing that the use of teleology is underpinned by natural laws and process, and does not rely on a creator, goal or desire of an organism (Ayala, 1970; Kampourakis, 2014; Ruse, 1989; 2005). Furthermore, the audience or discussants must understand the figurative, subcultural use of teleological explanation. This is shown to be possible in research by Zohar & Ginossar (1998) indicating that students can use teleological explanations without sacrificing scientific rigour; essentially children used teleology as an additionally method of description rather than a stand-alone explanation. Therefore, it is proposed that certain styles of teleological explanation may be useful to help account for organisms' traits if the teleological explanation refers to the benefit the adapted trait has provided to the species

(functional-teleology), not if the teleological explanation implies that the trait is a result of goal-direct behaviour (design-teleology).

2.5.3.3 Differing origins of teleology

After investigating the different views about teleology in relation to evolution, being a barrier to learning (section 2.5.3.1) or a support for learning (section 2.5.3.2), it seems initially that the opposing views discussed in these two sections treat teleology as one type of explanation. While the range of teleological explanations are usually only referred to as *teleology,* the two sides of the debate around the use of teleology in biology are primarily talking about different styles of teleological explanation and different contexts in which they are used, or more accurately different underlying themes for a teleological explanation. For example, the statement *giraffes have long necks for reaching the tops of trees* is teleological: it explains the existence of the neck by referring to a purpose for which they are used. However, to see if this is appropriate in scientific contexts there needs to be clarification of the underlying cause for this teleological rationale. Is the origin of giraffes' long necks due to a) a giraffe's desire to reach the leaves so it grew a long neck or b) continuous variation of neck length within the giraffe species, and the result that longer necked giraffes were more suited to their environment so lived longer and reproduced more than those with shorter necks.

The former (a) explains the existence of a trait by reference to a desire the organism had, equally the teleological explanation could have invoked natural selection as having a predetermined path or end goal to create that trait. This is the style of teleological explanation that is argued to be a barrier to understanding evolution, one where evolution is conceived as having a route to a specific goal, or that an individual organism can drive evolutionary change because of an innate desire. As discussed in section 2.2.2, this form of teleological explanation intersects with anthropomorphic explanations where the organism is imbued with human emotions and desires. This inappropriate use of teleology is referred to as design-teleology, when a trait is explained by reference to intentionality on behalf of an organism (or the *force* of natural selection).

The latter (b) explains the existence of a trait by partial reference to the process of natural selection, referring to continuous variation of traits and how a species can adapt over time to share a common trait. This style of teleological explanation is one that incorporates scientific elements. This explanation may use teleology as a starting point to aid comprehension but there is also evidence of knowledge about the underlying scientific process. This appropriate use of teleology is referred to as functional-teleology, when a trait is explained with reference to part of the evolutionary process which resulted in a trait capable of performing a certain function.

The above distinction between two forms of teleology, design-teleology and functional-teleology, suggest that functional explanations of organisms' traits are appropriate if the underlying mechanism is not a goal-directed, intentional process but rather alludes to the process of evolution

via natural selection (Ayala, 1970; Ruse, 2000). The definition of the two forms of teleology has been provided, however, perhaps the simplest method of distinguishing between them is to look at adaption. Ruse (2000) proposed that teleology explanation is appropriate for the traits of organisms when it focuses on the metaphor of adaptedness not adaptability. These metaphors are summarised as:

- Adaptedness – organisms are well adapted to a certain situation or environment because of natural selection and generational change.
- Adaptability – organisms are viewed as goal-orientated, they adapt because they desire to do so, they want to change to suit a certain situation or environment.

In relation to teleology, design-teleological explanations discus the adaptability of organisms, their goal or desire to adapt to future situations. For example, a number of individual anteaters deciding to grow long tongues because they needed they collect more food. Whereas functional-teleological explanations of anteaters' long tongues should focus on prior adaptations, how the species has changed over time to suit its environment. The desire or aim of an organism has no bearing on the process of evolutionary change.

One thing to note about the use of teleological explanations for organisms, humankind is able to meddle and direct evolutionary change, resulting in seedless bananas, sheep with luxurious fleeces and cows capable of producing vast amounts of milk, essentially evolution via human selection. These organisms and their traits are due to a goal-directed, purposeful process (e.g., dairy cows are for producing milk) which appear to be design-teleology because it is: humankind have actually designed these organisms to have a specific trait. Therefore, the function of the trait is explained via human controlled selection and inheritance of a certain trait. Organisms' traits resulting from human controlled section can be considered a form of artificial kinds, rather than natural kinds, as they have been artificially manufactured by humankind, rather than having their origin in natural processes. For artificial kinds, both artificial organism traits and artefacts, a design-teleological explanation would be entirely appropriate.

2.5.4 Summary

This section considered three aspects of how the concept of teleology is constructed. Firstly, the structure of children's knowledge about the natural world was investigated. It was proposed that if teleology is the result of what was termed coherent knowledge about natural kinds, it would be expected that when a group of children is asked to provide an explanation for a range of contexts (e.g., different category exemplars), children's responses should be a) coherent within children (i.e., a child would provide similar levels of teleology across a range of exemplars) and b) coherent across children (i.e., all children should give a similar response for any given exemplar). Conversely, if teleology is the result of what was considered fragmented knowledge, when a group of children is

asked to provide an explanation for a range of contexts (e.g., different exemplars of an ontological category), children's responses should a) exhibit reasonable variation within children (i.e., a child would not provide similar levels of teleology across a range of exemplars) and b) exhibit variation across children (i.e., a group of children would provide a range of different responses for any given exemplar). These positions, coherent and fragmented knowledge, are not binary categories, rather they indicate the furthest reaches of a continuum representing the structure of naïve ideas. Therefore, such terms are used to suggest that teleology is the result of naïve ideas that are more towards one end of the continuum than the other. Current research into children's use of teleological explanations appears to treat teleology as resulting from coherent knowledge, however, there is little work which specifically explores the structure of children's naïve ideas with regard to teleology. It is proposed that an investigation of this area is required, as the structure of children's knowledge would have differing implications for conceptual change. Briefly, knowledge-as-theory would indicate that conceptual change required the gradual removal or reshaping of prior explanatory frameworks with increasingly correct frameworks, whereas a knowledge-in-pieces approach would suggest that conceptual change would involve the recontextualising of various knowledge elements through targeted interventions. The development of this research aim is returned to in section 2.6.4.

Secondly, a number of different types of teleological explanation were defined, these definitions are used throughout the thesis. While different forms of teleological explanation are used in prior work, there appear to be little empirical research on children's use of different forms of teleological explanations (one exception would be Barnes, Evans, Hazel, Brownell & Nesse, 2017). It is proposed that further investigation of the type of teleological explanation is needed, as the form of explanation used has differing implications. For example, a design-teleological explanation about an example of natural phenomena would suggest a preconception that it existed for a specific purpose. A relational-teleological explanation about an example of natural phenomena could indicate the use of a cultural style of explanation (as discussed in section 2.4.4), for example, night is not actually claimed to exist for a specific purpose, but rather it is related to a reaction it causes or something for which is could be used for by humankind. The type of teleology children primarily use would determine the type of intervention needed to limit the use of teleological explanation. While these types of teleology are considered to be distinct categories in an abstract sense, actual concrete examples of teleological explanations could fall on a spectrum of teleology where there may be overlap between the categories. Consequently, it must be considered whether such examples lean more towards one category of teleology than the other.

Thirdly, this section determined the acceptability of teleological explanations in evolutionary contexts. For natural objects and natural phenomena it was outlined that any form of teleological explanation is unacceptable to account for the existence of examples of these

categories. With regard to the development of organisms and organisms' traits the use of teleology is more complex.

In opposition to those fundamentally against the use of any form of teleological explanation in biology (e.g., Hanke), it was argued that a simple functional-teleological explanation about the purpose of an organism's trait was acceptable, as a trait does serve a function for the organism. For example, it is appropriate to say that birds' wings are for flying or flowers' petals are for attracting pollinators. The distinction between appropriate and inappropriate teleological explanations is made at a deeper level than this simple functional statement. The distinction concerns the origin of organisms' traits. A functional-teleological explanation for the appearance of a trait in the population (through continuous variation) and its consequent prevalence (it conferred a selective advantage) would be acceptable. However, a design-teleological account where a functional trait served a purpose but originated due to a desire (goal-oriented behaviour) on behalf of a specific organism would be inappropriate. Furthermore, it was acknowledged that relational- and religious-teleological explanations for organisms or their traits may be culturally acceptable, although they would not be appropriate as scientific accounts. The distinction between acceptability of different styles of explanation is of particular importance to any work seeking to develop children's understanding of teleological explanation, and consequently is returned to in section 2.8.

2.6 Measuring teleology

Building upon a number of issues raised thus far, this section explores how prior work has sought to measure children's use of teleological explanations. The aspects of measurement discussed are question wording (the manner in which a question is worded), questioned topic (the topic a question focuses on), type of teleology (the style of teleological explanation children use), the construct of teleology (knowledge-as-theory vs. knowledge-in-pieces) and, finally, investigating teleology in natural discourse. Following the consideration of measurement techniques, a number of research questions are detailed.

The term *topic* is used to refer to the subject of a question which children are asked (e.g., the subject *night* in the question *why is there night?*). Therefore, in interview situations children could be asked to explain the existence of different topics, hence the construction *questioned topic*. When considering ontological classification of kinds, these topics would correspond to *examples* of natural kinds; for example, storms are an example of natural phenomena, rubies of natural objects and millipedes of organisms. Contrary to prior work, these are not considered to be *exemplars* of a category, as the example used may not accurately represent the entire ontological category.

2.6.1 Question wording

It is suggested that when exposed to teleologically leading language, questions and explanations, children may be prompted to provide purposeful explanations for the existence of natural kinds.

This idea raises possible problems with previous studies that could be considered to have employed leading questions, which place demands upon participants, signposting children to provide responses which contained purposeful or intentional phrases.

As discussed in section 2.3.1, research conducted by Keil (1992) employed multiple-choice options in the format of 'it is better for X to have Y' and 'Y helps there to be more X'. For example, plants are green because it helps there to be more of them. The phrases *better for X* and *Y helps* are suggestive signposts that could direct the participant into considering the subject in a teleological manner. *Better for* and *helps* imply that the subject, X, benefits from the support or aid or Y. Therefore, it is not controversial to suggest children may prefer a teleological explanation over a physical rationale.

The oft cited paper by Kelemen (1999a), investigating the scope of children's teleological thinking, asked children 'what is X for?' The word *for* here clearly implies a purpose and questions phrased in this manner may lead children into explaining the subject teleologically. Effectively, the question is placing a demand upon the participant, it is asking the child to provide a purposeful answer.

This demand characteristic, the participant acting as a good or faithful subject (Weber & Cook, 1972), can result in biased data as participants attempt to provide answers that the interviewer may want to hear (Orne, 1962). It could be argued with the question, *what is X for,* that what the interviewer wants to hear is a purposeful rationale for X. Young children are particularly susceptible to questions which suggest a 'correct' answer or which may lead them into providing a certain type of response. Research into the effect of leading questions in a courtroom context, by Bjorklund et al. (2000), showed that when presented with a leading-question almost 60% of 5-year-olds and 25% of 7-year-olds provided incorrect information, which corresponded to the bias in the leading-question. Therefore, suggesting that children can be easily influenced by the way questions are worded. Consequently, data gathered using, *what is X for,* may provide an unrepresentative picture of children teleological explanations. It should be acknowledged that Kelemen (1999a) does suggest that it could be leading but maintains that further research using fixed choice options, between teleological and scientific explanations, negates the error and supports her findings.

The use of fixed choice options raises two problems. Firstly, they do not measure what a child actually thinks but which of two options they prefer, or even which of two options they dislike the least. Therefore, fixed choice options are not an assessment of children's knowledge about scientific concepts but instead what type of option they will choose when lacking subject specific knowledge. It is an artificial test which is not representative of real world assessment. Kampourakis et al. (2012a) expand upon this arguing that if the purpose of the research is to discover children's actual idiosyncratic ideas then open-questions are a necessity. Secondly, if teleological language influences children propensity to provide purposeful explanations, children may choose a teleological explanation rather than a scientific account simply because, as discussed is section

2.4.4, teleology is a common method of explanation in the scientific subculture (Hanke, 2004; Ruse, 1989; Zohar & Ginossar, 1998).

The notion that demand characteristics may influence children's responses suggests that previous studies may be biased. Due to wording questions in a teleological manner children's propensity to provide teleological rationales may have been overestimated. The first research question attempts to investigate this concept by asking *does question wording influence children's explanations about the existence of natural kinds?* It is hypothesised that question wording will influence children's responses, with a teleologically leading treatment producing more teleological responses than a neutrally worded question. This research question will be investigated in Study 1. Due to the complex nature of acceptable teleological explanations for organisms in biological contexts (section 2.5.3.3), this study will focus on explanations for the existence of natural phenomena as teleological explanations for this ontological category are always incorrect.

2.6.2 Questioned topic

As discussed in section 2.3, previous work has investigated a small number of *questioned topics* within the ontological categories of natural objects and natural phenomena. Questioned topic refers to the subject of a question, that is, the thing which is under discussion (e.g., rain in the question: why is there rain?). The use of limited questioned topics suggests a view that certain examples of an ontological category can be used as exemplars of that category. Instances of limited questioned topic use include studies using four topics (Kelemen, 1999b; 2003), two topics (Kelemen, 1999a; Kampourakis et al., 2012a; 2012b) and one topic (Keil, 1992; 1995). Furthermore, the criteria for exemplar selection is often not discussed, there is no indication as to why, or even if, the exemplars used are representative of the entire ontological category.

It was argued that using a small number of exemplars to represent an entire ontological category would be problematic if there is within-category variation, that is, if levels of teleological explanation differ depending upon the question topic child are asked to explain. For example, children may be more prone to advocating a teleological rationale for the existence of rocks, than for the existence of soil. Consequently, the second research question explored in Study 1 is *does questioned topic influence children's tendency to provide a teleological response?* If there is large variation across questioned topic (even within a ontological category) then this has implications for a) theories that children's teleological reasoning is "promiscuous" and b) implies that caution needs to be taken interpreting results of studies that used only a limited number of questioned topics.

2.6.3 Type of teleology

Section 2.5.2 outlined a number of different forms of teleological explanation and argued that often research into children's use of teleological explanations does not make a distinction between different forms of teleology, treating all forms of teleological explanation as design-teleology (although an exception would be ojalehto et al., 2013). However, if children's teleological

explanations for natural objects or natural phenomena are not always indicative of design-teleology but rather relational-teleology. This may indicate that teleological explanations are not only due to a fervent belief that natural kinds have a specific purpose, rather that teleology may result from natural patterns of language use, as discussed in section 2.4.4. As these forms of teleology have different origins, design-teleology resulting from a purposeful view of nature and relational-teleology being influenced by natural language patterns, there use has different implications for conceptual change. Design-teleology represents an inaccurate prior belief which needs to be changed. Relational-teleology represents a form of discourse that may be acceptable in colloquial discourse but needs to be understood as inappropriate for use in scientific discussion. Therefore, the third research question asks *what types of teleology do children display when they provide a teleological explanation?*

2.6.4 The construction of teleology

The second and third research questions also provide an opportunity to consider the structure of children's naïve ideas about the origin of natural kinds: *do children's explanations about the natural world result from fragmented or coherent knowledge?* Evidence for the nature of teleology, it being the result of coherent or fragmented ideas about nature, will be informed by the previous two research questions. That is, by the levels and examples of teleological explanations provided across questioned topics. A coherent structure would imply that for a range of contexts (a number questioned topics representing examples of an ontological category) children's explanations should a) demonstrate considerable coherence within children (a child should be coherent in their teleological or scientific responses across a range of questioned topics) and b) demonstrate considerable coherence across children (for a single questioned topic all children should provide a small number of compactly describable responses). Whereas a fragmented structure would predict that a) children's responses would vary across contexts (they would not be stable in their propensity to provide teleological or scientific responses across a range of questioned topics) and b) there would be variation across children (for a single questioned topic there would be a range of response types given by a group of children). If children's explanations indicate a coherent structure of knowledge this would coalesce with and support current thinking. However, a fragmented view would imply that children's naïve ideas may be more context dependent and that conceptual change may be achieved through the use of targeted interventions.

2.6.5 Teleology in natural discourse

Section 2.4.4 and 2.5.2 proposed that teleological explanation, specifically the use of relational-teleology, is influenced by cultural patterns of language. It argued that teleological explanation may be used figuratively in non-scientific discourse to discuss natural kinds. For example, a parent stating that *nighttime is a signal for children to go to sleep* or claiming that *it is good that it is raining as it will water the grass*. Discourse of this type, where natural kinds are discussed in

relation to a consequence they have, could result in children conflating cause and effect when proving rationales for the existence of natural kinds. This would echo what in section 2.5.2 is termed relational-teleology, where something is explained in relation to a purpose for which it can be used or an effect which it causes.

There have been a small number of studies investigating parental explanations in relation to teleology (see section 2.4.5). Furthermore, in these studies data was collected via parental interviews, questionnaires and diaries, and the focus was on children's questions related to natural kinds not on the natural discourse surrounding these topics. It was proposed that this gap in the literature could be investigated by analysing the discussion which children experience surrounding natural kinds. Therefore, the fifth research question, concerning the measurement of teleology considers *does the natural discourse children experience about natural kinds influence children's explanations for their existence?* This question is further developed and explored in Study 2, Chapter 6.

2.6.6 Summary

Review of the existing literature regarding the measurement of teleological explanations has resulted in the development of five research questions. These are:

- Research Question 1 (Study 1) *does question wording influence children's explanations about the existence of natural kinds?*
- Research Question 2 (Study 1) *does questioned topic influence children's tendency to provide a teleological response?*
- Sub-Question 1 (Study 1) *what types of teleology do children display when they provide a teleological explanation?*
- Sub-Question 2 (Study 1) *do children's explanations about the natural world result from fragmented or coherent knowledge?*
- Research Question 3 (Study 1b), *does the natural discourse children experience about natural kinds influence children's explanations for their existence?*

These questions are further refined and explored in the relevant chapters (Chapters 4 and 5). Thus far the literature review and developed questions have primarily focused on the measurement of children use of teleological explanations, be it specific data collection methods, the structure of naïve ideas or possible causes of teleology. The remainder the literature review focuses on methods for actually limiting children's use of teleological explanations.

2.7 Limiting teleological explanation

This section considers current methods of limiting children's use of teleological explanations for the existence of natural kinds. Firstly, it surveys the National Curriculum for England and Wales to

determine if and when teleological explanation are countered. Secondly, it reviews studies with a specific aim to limit teleological explanation. Thirdly, it details two further research questions pertaining to methods of limiting teleological explanations.

2.7.1 Teleology and the curriculum

As teleological explanations are primarily used by children and are argued to be a barrier to scientific understanding (see section 2.1.3), it could be expected that attempts to limit children's use of teleology would begin in childhood. Specifically, that the curriculum would support children's scientific development in areas prone to teleological explanation. There is no indication in the current curricula for England and Wales, neither the Early Years Foundation Stage (DfE, 2016) nor the National Curriculum 2013 (DfE, 2013), that teleological forms of explanation should be explicitly discussed or countered in formal education. This, however, is to be expected. These two curricula are not prescriptive, in that they do not dictate specific learning objectives, rather they represent programmes of study and attainment targets for children. Therefore, while teleology does not currently feature in the curriculum, the need for teachers to limit these forms of explanation is inferred from the statutory requirements that children should be taught specific scientific concepts. Developing children's understanding of these concepts should, theoretically, lead to children using scientific explanations for natural kinds. That is not to suggest that conceptual change is simply fostered by outlining the correct concept, rather over time and through a range of teaching and learning activities, teachers can develop children's understanding of the correct scientific concept. The following details curriculum targets which relate to limiting teleological explanation in three ontological categories: natural objects, natural phenomena and organisms.

In terms of understanding scientific explanations for the existence of natural objects, there are limited occurrences of statutory attainment targets that would counter teleological explanations for the existence of natural objects. The majority of the science curriculum focuses on the properties of objects, not the creation of objects. The requirements at Year 3 (7- to 8-years-old) present one example regarding the formations of rocks and soils, where children should:

- 'Describe in simple terms how fossils are formed when things that have lived are trapped within rock
- Recognise that soils are made from rocks and organic matter' (DfE, 2013, p.158).

Another example at Year 2 (6- to 7-years-old) is the target that children 'should be taught to explore and compare the differences between things that are living, dead, and things that have never been alive'. This requirement could relate to the classification of natural kinds, with natural object having the shared property of purposeless entities. A deeper understanding of the existence for other natural objects, such as mountains or stalagmites may also come from the Key Stage 2 (7- to 11-years-old) geography curriculum where children should 'describe and understand key aspects of physical

geography, including: climate zones, biomes and vegetation belts, rivers, mountains, volcanoes and earthquakes, and the water cycle' (DfE, 2013, p.186).

There are a number of National Curriculum attainment targets which would present opportunity for children to learn about scientific accounts for the existence of natural phenomena. These include:

- At Year 1, 5- to 6-years-old (DfE, 2013, p.9), "observe changes across the four seasons" "observe and describe weather associated with the seasons and how day length varies";
- At Year 2, 6- to 7-years-old (DfE, 2013, p.10), "explore and compare the differences between things that are living, dead, and things that have never been alive";
- At Year 3, 7- to 8-years-old (DfE, 2013, p.18), "recognise that they need light in order to see things and that dark is the absence of light";
- At Year 5, 9- to 10-years-old (DfE, 2013, p.29), "use the idea of the Earth's rotation to explain day and night and the apparent movement of the sun across the sky";
- Again at KS2, 7- to 11-years-old, in geography (DfE, 2013, p.186), 'describe and understand key aspects of physical geography, including: climate zones, biomes and vegetation belts, rivers, mountains, volcanoes and earthquakes, and the water cycle'.

With regard to teleological explanations for the existence of organisms, curricula content seeking to limit these explanations would focus on evolution. Evolution it is not specifically mentioned in the national curriculum until Year 6 (10- to 11-years-old) when children should (DfE, 2013):

- "recognise that living things have changed over time and that fossils provide information about living things that inhabited the Earth millions of years ago"
- "recognise that living things produce offspring of the same kind, but normally offspring vary and are not identical to their parents"
- "identify how animals and plants are adapted to suit their environment in different ways and that adaptation may lead to evolution"

However, there are also links to adaptation and inheritance in Year 2 and Year 4, where children should:

- At Year 2 (6- to 7-years-old): "identify that most living things live in habitats to which they are suited and describe how different habitats provide for the basic

needs of different kinds of animals and plants, and how they depend on each other" – which links to how organisms have adapted to their habitats;

- "notice that animals, including humans, have offspring which grow into adults" – links to inheritance
- At Year 4 (8- to 9-years-old): "recognise that environments can change and that this can sometimes pose dangers to living things" – this links to how organisms may become adapted to their habitats, through natural selection.

While not specifically related to one of the three ontological categories, but rather spanning all three, the Early Years Foundation Stage (DfE, 2016) has an area entitled *understanding the world*. This curriculum area 'involves guiding children to make sense of their physical world and their community through opportunities to explore, observe and find out about people, places, technology and the environment' (p.8). Furthermore, 'children [should] know about similarities and differences in relation to places, objects, materials and living things. They [should] talk about the features of their own immediate environment and how environments might vary from one another. They [should] make observations of animals and plants and explain why some things occur, and talk about changes' (p.12). This curriculum area presents opportunities for children to learn about the existence of natural objects, natural phenomena and organisms.

These Early Years Foundation Stage and National Curriculum attainment targets all have to possibility to counter children's use of teleological explanations for the existence of natural kinds through teaching of the correct scientific explanation. Again it is stressed that conceptual change would not be expected from the simple presentation of the scientific explanations, but rather through appropriate teaching and learning activities across time.

2.7.2 Countering teleological explanation

Although the reviewed curricula suggest that children should not formally learn about evolution and natural selection until Year 6 (10- to 11-years-old) they should be exposed to notions related to adaptation at earlier stages: Year 4 (8- to 9-years-old) and Year 2 (10- to 11-years-old). There are learning aims that tangentially relate to the existence of natural objects and natural phenomena in the Early Years Foundation Stage (0- to 5-years-old) and Key Stage 1 (5- to 7-years-old) curricula, although specific undertaking the objects and phenomena do not occur until Key Stage 2 (7- to 11-years-old). For example, the formation of rocks in Year 3 (7- to 8-years-old) or the day/night cycle in Year 5 (9- to 10-years-old).

Nevertheless, there is work showing that young children can understand evolutionary principles. Kelemen et al. (2014) conducted a series of interventions where children (5- to 6-years-old and 7- to 8-years-old) took part in a joint child-teacher reading of a custom storybook detailing the adaptation of a species (elephants). The book focused on a change from elephants with variable

trunk sizes to elephants with primarily thin trunks. Following the 10-minute intervention, children showed significant gains from pre-test in terms of explaining trait change across generations for the species depicted in the book. Furthermore, they could apply the evolutionary principles to the development of traits in other organisms. This gain was maintained at delayed post-test, three months later. This study is evidence that young children can understand aspects of the process of natural selection, specifically in terms of population change across time.

Early exposure to ideas is arguably beneficial to learning about the same concepts comprehensively later in education (Peacock et al., 2011), particularly if subject matter is appropriately scaled, for example in a spiral curriculum (see Bruner, 1966). However, learning about scientific concepts is complicated once children's have formed preconceived ideas, as these can act as considerable barriers to conceptual change (diSessa, 2014). Especially, if children are only presented with counterevidence to their ideas, in such cases they would be unlikely to reject their preconception but rather demonstrate a strong confirmation bias (Sharp et al., 2014). This line of reasoning would indicate that simply teaching the correct conception may not always be sufficient to counter children's use of teleological explanations.

It is proposed that children's use of teleological explanations needs to be specifically tackled alongside the teaching of explicit scientific concepts to overcome the barrier of teleology. This idea has also recently been advocated by Barnes et al. (2017) in relation to University students' understanding of teleology and evolution. Henderson et al. (2015) argued that criticism of ideas is vital to children's science education, as rigorous appraisal of ideas enables children to develop evaluation skills, understand the nature of science and develop epistemic vigilance (see section 2.4.3.2). However, Henderson et al. (2015) claim that science education has become too focused on children constructing knowledge, that there is currently little opportunity for children to actually critique knowledge. While there are a number of intervention studies which have sought to teach children about the correct scientific conception to counter their use of teleological explanations (e.g., Kelemen et al., 2014, outlined above), there appear to be very little on children actually understanding the nature of teleological explanations. The vast majority of research into children and teleology have been snapshots of what children in certain age-ranges think about different ontological categories. Consequently, it is suggested that interventions specifically targeting children's understanding of why teleological explanations are inappropriate may be effective at limiting their use of such explanations to account for the existence of natural kinds.

2.7.3 Summary

Throughout this chapter it has been established that children use high levels of teleological explanation for natural kinds. However, this section indicates that there is little research and no actual curriculum guidance on countering the use of teleological explanations, other that teaching the correct conception. This imbalance led the development of RQ 4 and RQ 5. These two research

questions explore if discursive activities about the nature of explanations in science support children in recognising and understanding why teleological explanations are inappropriate for natural kinds. Study 2 (RQ 4) focuses on limiting the use of teleology for natural phenomena and Study 3 (RQ5) on limiting teleology for organisms. These interventions are similar in structure but different in terms of the ontological category focused on, which allows the intervention to target teleology across different categories of natural kinds[1]. The remainder of the literature review details successful science education interventions and pedagogical techniques to inform the design of Studies 2 and 3. The specific research questions are detailed in section 2.8.8.

2.8 Interventions in science education

While the majority of research into children's use of teleological explanations has provided a snapshot of what children know at certain *stages* (see section 2.7), there has been very little work on limiting children's use of teleological explanations. Studies 2 and 3 sought to explore how children learn to differentiate teleological and causal explanations, by investigating how to support children in learning about the elements of appropriate (scientific) and inappropriate (teleological) explanations. The following provides a rationale for conducting the interventions and situates them within the larger field of research into science education. It does this by considering three themes in the field of Primary Science Education: 1) the difference between children's actual and preconceived understanding; 2) the move from *small ideas* to *big ideas*; and 3) children's acquisition of scientific skills, specifically evaluation of evidence. Following this is an argument for the nature of the intervention (discussion-based) with an investigation of thinking skills programmes (specifically P4C) that have employed discourse-based interventions with positive results.

2.8.1 Children's actual and preconceived understanding

Klahr, Zimmerman & Jirout (2011) argued that the majority of curricula underestimate K-8 (5- to 15-year olds) children's ability to learn scientific concepts, as there is no consideration of the knowledge children already possess about the natural world. Duschl, Schweingruber & Shouse (2007) in their review of how children (aged 5- to 14-years-old) learn and are taught science found that children bring an extensive amount of scientific knowledge to the classroom and are much more competent than usually assumed. They can reason in a concrete manner about abstract scientific concepts, can make reasoned predictions and can evaluate evidence. Therefore, these reviews of curricula and classroom practice suggest children's potential to understand scientific concepts is often underestimated, adults may have a misconceived notion that a scientific concept or skill is beyond the grasp of a young child.

[1] While the omission of natural objects may be questionable, in the light of the results of Study 1 (Chapter 4, section 4.2.2.2) the category of natural objects was not viable.

An explanation for the disunity between adults' preconceptions about when children can learn scientific concepts and when children actually can learn such concepts is offered by Klahr et al. (2011). They suggest that viewing children as developing through stages (e.g., Piaget's three stages of Artificialism or Kelemen's shift from teleological to causal thinking) constrains and limits the teaching of science in the Early Years. This, perhaps, is one of the factors, including teachers' training and conflicting priorities, which contribute to Tu's (2006) finding that in preschool settings only 4.5% of the content was related to formal science teaching.

2.8.2 Big ideas in science education

One of the purposes of the Study 2 and 3 interventions is to contribute to developing children's scientific literacy, specifically in relation to enabling children to consider "something 'in a scientific way', seeing, for example, whether or not evidence has been taken into account in the explanation of an event or phenomenon, whether it makes sense in terms of related events or phenomena" (Harlen & Qualter, 2014, p.15). This is key because the aim of science education is not to simply impart scientific subject knowledge but to improve individuals' scientific literacy (Sharp et al., 2014; Millar & Osborne, 1998). Harlen & Qualter, (2014) propose that a major effect of attaining scientific literacy is that children move from thinking with *small ideas* (concepts related to specific thing, e.g., rainbows occur when the sun shines through rain) to being able to apply *big ideas* (theories that have a wide-ranging applications and can be used when encountering new things, e.g., natural phenomena have a cause).

This shift is highly beneficial to children's development as big ideas have a greater explanatory value than small ideas, as they are applicable to a greater range of concepts (Loxley, Dawes, Nicholls & Dore, 2010). Consequently, the interventions sought to enable a shift from small to big ideas, in order to facilitate abstraction about purpose in nature. Teaching about purpose in the natural world (i.e., for the ontological categories of organisms and natural phenomena) could have a larger impact than simply teaching children about the causes of several exemplars. This is because the big idea of *purposeless natural phenomena* or *the lack of goal-driven behaviour in species development* is more applicable to newly encountered knowledge, acting as a schema which could assist children in evaluating any further explanations they encounter or develop.

2.8.3 Evidence evaluation

Science is not simply knowledge focused. It is argued that there is both a need a) for children to learn scientific facts, and b) that science education must involve looking at styles of scientific thinking and how they differ to the normal everyday style of thinking (Sharp et al., 2014; Roth, 2014). 'Scientists to not believe ideas by authority; they believe ideas because of the evidence and arguments that support them' (Russ, Coffey, Hammer & Hutchison, 2009, p.877). Therefore, there is a need for teachers to develop children's understanding of the nature of science through

development of the three main scientific competencies: hypothesising, observing and evaluating (Klahr et al., 2011).

Table 2.2: *Types of intervention in relation to the two levels of scientific thinking*

	Type of scientific process		
Type of knowledge	Forming hypotheses	Designing and running experiments and observations	Evaluating evidence
Domain-specific	A	B	C
Domain-general	D	E	F

(Klahr et al., 2011, p.971)

Using Klahr and colleagues' categorisation of science education interventions (see Table 2.2) the proposed interventions are Type C interventions. The two separate proposed interventions explore knowledge that is domain-specific (causal and teleological explanations for the existence of natural phenomena or the development of organisms' traits) and the process of evaluating evidence. However, should children take part in both of the interventions they would be exploring multiple ontological categories (e.g., organisms and natural phenomena), thus the interventions would then be classified as a type F intervention. Previous type C and F interventions (looking at evidence evaluation) show that young children can be taught to distinguish between appropriate and inappropriate forms of evidence and recognise complex patterns in evidence (for summary see Klahr & Chen, 2003). Consequently, the aim of the interventions *to improve children's evaluation of causal and teleological explanations (possible evidence) for the existence of natural kinds* is feasible in relation to current literature.

Understanding how to evaluate evidence allows children to critically test their ideas, leading to refinement of concepts and the understating that explanations given about, for example natural phenomena are not automatically correct (Osborne, Collins, Ratcliffe, Millar & Duschl, 2003). This is further supported by Chin & Osborne (2008) who proposed that when children ask questions they collate, investigate and assess their knowledge in relation to others knowledge. Therefore, when learning to interpret evidence and form conclusions, children need to look at patterns in data, how their explanations relate to other individuals, or societal, explanations. This shared investigation of competing explanations is perhaps best facilitated through group discussion where children can share and asses others' ideas (Harlen, 2006; Mercier & Sperber, 2017), allowing children to develop their knowledge, discussion skills, self-assessment, argumentation and evidence evaluation (Treagust & Tsui, 2014). For example, Howe, Rodgers & Tolmie (1990) found that when children (aged 8- to 12-years-old) engaged in group discussion about how to interpret evidence this led to improved conceptual understanding, providing children's preconceptions differed pre-

discussion. The following section expounds on this argument through consideration of dialogic teaching.

2.8.4 Dialogic teaching

In classroom environments Loxley et al. (2010) recommend dialogic teaching to enable effective discussion based activities in science, as do Harlen & Qualter (2014), Smith (2010), Gurgeon, Hubbard, Smith & Dawes (2012). The key principle of dialogical teaching is that talk in the classroom is not only the method of assessment or the method delivering information but the foundation of learning. Talk is the key way in which children actively construct meaning with others (Alexander, 2006). There are clear links here to the argumentative theory of reasoning (section 2.4.3.2), that group discussion enables individuals to clarify and criticise the evidential potential of others explanations. Dialogic teaching is related to *exploratory talk*, collaborative talk in which individuals share and critically discuss ideas with an aim to develop a consensus (Mercer & Howe, 2012; Mercer & Littleton, 2007), but with more focus on teacher guidance. While the benefits of exploratory talk have been shown to be numerous (e.g., see Howe & Abedin, 2012), teacher-led dialogic teaching was considered more appropriate to the proposed interventions to ensure children would be encouraged to consider the implications of teleological explanations. In summary, dialogic teaching involves collaborative, critical exploration of ideas, in a teacher guided group setting where all contributions are allowed and related to personal experience (if relevant), ideas are questioned and challenged, statements are justified, and reflection is encouraged (Loxley et al., 2010; Alexander 2006).

Dialogic teaching is a supporting pillar of *Philosophy for Children* (P4C, Smith, 2010) and may be beneficial in the science classroom because of its capacity to explore an issue in-depth, enabling different views to be considered and a shared understanding to be established (Harlen & Qualter, 2014; Gurgeon et al., 2012). This is especially pertinent when various social or cultural factors could influence children's preconceptions and their thinking about ideas under discussion (as outlined in section 2.4.4). Mercer (2008), for example, highlights the notion that personal religious beliefs may well influence individuals thinking about scientific concepts and, consequently, inhibit opportunity for conceptual change. Mercer further notes that while research into conceptual change refers to the importance of talk as a method for explaining concepts, it ignores talk as a tool for reflection. Specifically, that appropriate dialogue is actually a driving force behind individuals' critical consideration of conceptions, and ultimately, conceptual change. Research into the link between developing understanding of scientific concepts and dialogic activities indicates that after a series of 12 lessons, specifically focusing on talking skills (e.g., critical questioning and negotiation), Year 5 children in the treatment group showed significantly

greater improvement on science assessments (with a large effect, $\eta^2_p = .29$[2]) than children who were assigned to a no treatment control (Mercer, Dawes, Wegrif & Sams, 2004).

This capacity for dialogic teaching to foster conceptual change, particularly with regard to ideas that can be influenced by socio-cultural factors, makes it beneficial as a theoretical underpinning for the proposed interventions. This is because children's teleological explanations can be idiosyncratic and rooted in personal contexts, consequently, the type of talk involved in dialogic teaching could enable investigation of children's specific ideas and the development of shared views about purpose in nature.

In dialogic teaching it is the teachers' role to facilitate the discussion, encouraging children to elaborate upon and explain their answers while valuing and facilitating contribution from all (Harlen & Qualter, 2014). However, it is important to note that while dialogic teaching provides a classroom culture in which high quality talk is valued and expected (Smith, 2010b) it is not simply a check list of methods. Instead it is an attitude of teachers, a way of behaving, which enables teachers to have multiple methods which satisfy the main principles of dialogic teaching. Consequently, to promote effective talk the proposed interventions needed to incorporate the underlying principles of dialogic teaching. These suggest that classroom talk should be:

- "*collective*: teachers and children address learning tasks together, whether as a group or a class;
- *reciprocal*: teachers and children listen to each other, share ideas and consider alternative viewpoints;
- *supportive:* children articulate their ideas freely, without fear of embarrassment over 'wrong' answers; and they help each other to reach common understandings;
- *cumulative*: teachers and children build on their own and each other's ideas and chain them into coherent lines of thinking and enquiry;
- *purposeful*: teachers plan and steer classroom talk with specific educational goals in view." *(Alexander, 2006, p.36).*

2.8.5 A discussion orientated intervention

The previous section mentioned the importance of dialogic teaching and group discussion in relation to developing children's skills in evidence evaluation. This section elaborates upon this outlining why scientific discourse is vital in science education.

Discourse is vital to developing conceptual understanding because it is though talk that young children are exposed to, and interact with, scientific concepts (Treagust & Tsui, 2014; Loxley et al., 2010), as language is the tool with which children explain, share and analyse ideas (Mercer

[2] This effect size measurement may be Eta squared or Cohen's *f* rather than partial-Eta squared, as the effect size measure is not specifically detailed in the manuscript. Partial-Eta is assumed as it is the most frequently reported effect size for ANOVA. If the measure is Eta squared it still represents a large effect size, however, if it is Cohen's *f* it represents a medium effect size.

& Howe, 2012; Fisher, 2007; Smith, 2001; Vygotsky, 1986; Bruner, 1996). It is for this reason that Loxley et al. (2010, p.37) maintain that learning to talk in a scientific manner can be a powerful way to develop children's ability to 'question, assert, explain, hypothesise, [and] reflect'. Section 2.4.3.2 outlined the argumentative theory of reasoning, which proposes that group discussion and argument allows human reasoning to flourish. In these social situations individuals can become particularly competent at evaluating the arguments of others (see section 2.4.3.2). Consequently, for the Study 2 and 3 interventions to develop children's scientific knowledge and skills then it may be beneficial for children to engage in discussion with others to develop a comprehensive understanding of a concept; the concept being the role of purpose in the natural world.

It is often acknowledged that in-depth talk is effective in supporting children's development (Alexander, 2006; Bruner, 1996; Mercer, 2002; Sampson, 2007; Sampson & Babarczy, 2013; Vygotsky, 1978, 1986). However, the majority of teacher-child talk is limited to Initiation-Response-Feedback (IRF) questioning where children's answers usually last less that 5-seconds, with less that 15-seconds of thinking time. IRF does not allow children to develop their thinking skills, merely an opportunity remember a correct answer (Gurgeon et al., 2012; Wellington & Osborne, 2001). A systematic review of research on classroom dialogue (Howe & Abedin, 2013) found that when research focused on observing current classroom practice then the IRF model of dialogue was found within educational settings from pre-school to secondary school. Specifically with regard to primary school science classrooms, Treagust & Tsui (2014) found that the IRF interaction model was abundant within these settings. Therefore, they advocate a move to questioning for higher level thinking with teacher-student dialogue, children questioning teachers, and co-constructed discourse and debate about scientific issues, rather than simple knowledge testing questions. Examples of this include interventions and schemes of work developed by Kuhn and colleagues (e.g., Crowell & Kuhn, 2014; Jewett & Kuhn, 2015; Kuhn, 1989; Kuhn, Ramsey & Arvidson, 2015), where students engaged in collaborative discourse and argumentation to develop scientific thinking skills. The results of which showed that this format of scientific activity can greatly improve children’s competencies in explaining their reasoning and evaluating evidence, particularly in relation to considering multiple variables when problem solving. The proposed interventions incorporated the style of classroom discourse indicated by Kuhn, and Treagust & Tsui's (2014) recommendations, with the aim to inject the much needed discussion into primary science classrooms.

Loxley et al., (2010) proposed that discourse-based science activities are particularly effective when word meaning can cause confusion. They cite an example of children discussing pupillary response and their confusion due to the word *open* being decontextualised, as it normally relates to doors opening on hinges. A relevant example would be the use of the words *for, wants, needs* or *helps* in relation to children's teleological explanations. Consequently, group discussion can provide a platform for children to explore additional meanings of words and develop more

effective ways of expressing an idea. This quality of discourse to clarify and develop a shared understanding of expressions is particularly pertinent to the proposed interventions. For example, with the Study 2 intervention there is a need for children to differentiate between design-teleology and relational-teleology for the existence of natural phenomena. That is, while explanations about natural phenomena could seem to suggest *X is for Y*, what is actually meant is that *X can be used to do Y*. Therefore, it is proposed that discourse-based activities may be beneficial in developing the idea that the purposeful words are not used literally and are not appropriate for scientific accounts (see section 2.4.4). Due to the benefits of a discussion-oriented intervention outlined in the section and section 2.8.4 on dialogic teaching, the following section explores discourse-based classroom strategies, Cognitive Acceleration in Science Education (CASE) and Philosophy for Children (P4C), which were drawn upon to design the overall structure of the Study 2 and 3 interventions.

2.8.6 Thinking skills programmes

Initially two existing thinking skills programmes were considered as a basis for the interventions: CASE and P4C. These are collectively referred to as *learning to think* or *thinking skills* programmes which draw heavily on similar underlying principles. CASE activities are usually conducted with data and may include experimental activities (Shayer, 2003), which may not be feasible for looking at teleological explanations. Although, P4C was not specifically designed for science education, during P4C sessions children usual explore philosophical and abstract concepts (Smith, 2010b), to which teleology belongs.

The *CASE @ KS1* study (Adey et al., 2002; Venville et al., 2003) involved 14 experimental and 8 control Year 1 (5- to 6-years-old) classes. The experimental group received 29 CASE sessions based on schemas of concrete operations, (e.g., causality) over one school year. It was not the role of the CASE @ KS1 project to teach specific scientific concepts but to show that developing early exposure to scientific methods and forms of inquiry "has a direct bearing on later development of scientific thinking" (Venville et al., 2003, p.1318). Venville et al. (2003) found that during the intervention lessons children demonstrated seven classifications of specific *good thinking* behaviour, which was argued to be beneficial to cognitive development: *explains, highlights discrepancy, adopts a new idea, demonstrates, thinks/works collaboratively, asks questions* and *other useful strategies*. The average occurrence for these seven strategies was considerably higher for the treatment group than for the control group. For example, the mean occurrence of children providing an explanation in a single session were 28 explanations in the experimental group and 6.9 explanations in the control group. Venville et al. (2003) suggested that the increased use of these thinking skills was due to the CASE intervention, which follows as the treatment lessons were specifically designed to develop these seven thinking skills and the control lessons where not. Therefore, there is evidence that CASE can promote these thinking skills, although as the measurements of occurrences of specific thinking skills are averaged across participants, it is

unclear if a small number of children contributed a large number of occurrences of thinking skills or if occurrences were equally distributed between participants.

In a different analysis of the same CASE @ KS1 project, Adey, Robertson & Venville (2002) analysed quantitative data pertaining to the above study. A pre-/post-test design was used with a matched-control group, the assessment consisted of two tasks based on Piagetian protocols, specifically conservation and drawing tasks. The results showed that both experimental groups improved from pre- to post-test, however, the CASE group showed significantly greater improvement in terms of intervention specific understanding (d = 0.47) and knowledge transfer (d = 0.43), both indicating considerable score differences between the treatment and control group.

Oliver & Venville (2015) proposed three explanations which appear to contribute to the effectiveness of CASE. Firstly, from a socio-cultural perspective CASE is effective because the group discussion-oriented sessions allow children to listen to, and collectively construct and criticise ideas (there are strong parallels here with dialogic teaching, see section 2.8.4). Secondly, due to the metacognitive element of CASE, as metacognitive activities support children in considering the nature of scientific ideas and their own thoughts. Therefore, leading to an improvement in children's inhibition of their initial ideas, that is, the ability to ignore an initial idea whilst considering alternatives, which, consequently, allows for the consideration of more complex and advanced concepts. The third proposal for the effectiveness of CASE focuses on the long term and transfer effects. Oliver & Venville argued that neurological research provides additional insights into teaching and learning. Using Dubinsky, Roehrig & Varma's (2013, cited in Oliver & Venville, 2015, p.384) core concepts, including the notion of plasticity, they suggested that because children undertake frequent CASE lessons they are re-using, and consequently, strengthening the neural pathways associated with the lessons schemas. This leads to retention of more sophisticated concepts and ways of thinking.

While CASE appears to be advantageous for children's scientific development, CASE was not appropriate for the interventions in this thesis due to differences in aims of CASE and the thesis interventions. Primarily, CASE programmes adhere to a lesson structure that is developed in relation to a specific selection of activities, or *pillars*. One of these pillars, social construction, involves children working together to construct explanations about the scientific concept under investigation (Adey et al., 2002; Venville et al., 2003). This development of possible explanations for the existence of natural kinds would not have been desirable in the proposed interventions, as the aim of the interventions was to evaluate the explanations of others, not to develop new rationales for the existence of natural kinds. While the interventions were not modelled on the content of CASE sessions, certain practical considerations were taken from *CASE @ KS1*. This including timing of activities and grouping of children (Seleznyov & Adhami, 2015).

Philosophy for Children (P4C) forms part of the framework which underpins this intervention. The P4C method draws on the Socratic method and Vygotsky's notion of socially

constructed meaning (there are also links with dialogic teaching, section 2.8.4, and the Argumentative Theory of Reasoning 2.4.3.2) to provide children a space to think critically, develop reasoning and questioning skills, and explore complicated philosophical and abstract concepts (Lipman 1980; Smith 2010b). It should be noted that the P4C programme does not focus on philosophical schools of thought (e.g., realism or idealism) but rather philosophical competencies, such as interrogation and critique of ideas, collaboration and argument construction (Gorard, Siddiqui & See, 2015; Tagget et al., 2005).

While in practice P4C is usually used to debate social and religious ideas, Smith (2010) advocated that P4C could allow children to develop an understanding of the abstract nature of scientific concepts. He proposed that children's scientific education primarily focuses on the effects of scientific ideas (i.e., the effects of friction or gravity) which often leave children with little understanding about the actual concept itself. Smith asserts that by periodically allowing children time to reflect on these concepts and discuss theirs and others ideas, children may gain a deeper understanding of concepts. This proposal for developing a wider scientific literacy is shared by a number in the scientific community (see section 2.8.1).

The P4C session structure has five main principles (Smith, 2010b) that were incorporated in to the design of the Study 2 and 3 interventions:

1. Circle time style structure with an antecedent (e.g., a concept cartoon);
2. Sessions take place once a week for up to an hour;
3. Children set the question, this a key principle, although at times the teacher may set the question (as in the proposed intervention);
4. Discussion involves the 4 C's of thinking, critical, caring, creative and collaborative;
5. There is no need for a consensus on the issues.

Trickly & Topping's (2004) meta-review found a consistently moderate standardised effect (0.43) for P4C in improving factors such as children's (aged 5- to 15-years-old) reasoning skills; although the reviewed studies were not conducted with young children (i.e., 4- to 7-year-olds). However, a recent longitudinal study conducted with 8- to 11-year-olds (Gorard et al., 2015) suggested that P4C may have little benefit in terms of transferable skills. The study involved 48 schools in which children were tracked through Key Stage 2 (7- to 11-years-old), of these 48 schools 22 took part in the P4C intervention and the other 26 acted as a waitlist control group. The statistical evaluation of the programme focused only on the Year 4 and 5 children (8- to 10-years-old) who received the P4C lessons over the course of an academic year. Children's development was measured in relation to their Key Stage 2 (KS2) achievement in line the National Curriculum for England and Wales and the standardised Cognitive Ability Test 4th edition (CAT4). The results of the evaluation showed that following the intervention the treatment group showed significantly greater gains than the control group, in relation to KS2 achievement and CAT4 scores. However, the associated effect sizes for KS2 progress in reading (g = 0.12), writing (g = 0.03) and maths (g

= 0.10) do not reach the common 0.2 threshold considered to be a small effect (Cohen, 1988). The results were similar for children's CAT4 gain scores, which again only show minor effect sizes; for example, overall CAT4 gain (g = 0.07), verbal CAT4 subscale (g = 0.08). Gorard et al. (2015, p.21) contended that they cannot 'state that these gains are definitely the result of P4C, but overall there is some promise.' Although thought to be little benefit in terms of pupil attainment in English and mathematics, P4C is a suitable influence for developing the intervention because the proposed interventions aim is not to improve transferable skills (e.g., English and mathematics progress, the measured outcomes in Gorard et al.) but rather to improve children's reasoning about, and interrogation of, ideas, as advocated by the Tagget et al. (2005) meta-review.

Investigation of P4C has outlined a number of benefits to the underpinning theory, pedagogical techniques and related educational outcomes. For this reason the proposed interventions are inspired by and borrow certain elements from CASE and P4C (further discussion of their influence is presented in the relevant chapters, sections 6.2 and 7.2). While these thinking skills programmes have been shown to be effective, a broader question relates to their compatibility with the National Curriculum for England and Wales, as fundamentally these are interventions which are designed to ultimately be used by teachers, in schools. The following section explores why thinking skills interventions are appropriate in context of the English National Curriculum.

2.8.7 Thinking skills programmes and the National Curriculum

Nisbet (1993) maintains that any curriculum is insufficient if it does not include the teaching of thinking skills. The 1999 Primary National Curriculum for England and Wales (DfEE, 1999) stated that thinking skills should be encouraged in a number of areas (Adey & Shayer, 2002). However, that curriculum is now redundant. In the 2011 white paper (DfE, 2011) for the 2013 National Curriculum for England and Wales it was outlined that schools should develop their own specialisms as regards to thinking skills programmes in order to encourage lifelong learning. This could have resulted in an improvement from the 1999 National Curriculum, as schools would have autonomy to incorporate and develop programmes such as CASE and P4C. Unfortunately, this white paper guidance is missing from the published 2013 National Curriculum (DfE, 2013). Furthermore, the original advice in the 1999 National Curriculum, that thinking skills should be developed, has also been removed from the current 2013 curriculum. Even the word *thinking* has almost halved from 23 to 12 occurrences (standardised frequencies = 4.37×10^{-4} to 2.29×10^{-4}).

Considering the decreased focus on thinking skills in the curriculum from 1999 to 2013, is there a place for thinking skills programmes in primary schools. One of the three core aims of the National Curriculum science programme of study (DfE 2013, p.3) is that children should "develop [an] understanding of the nature, processes and methods of science through different types of science enquiries that help them to answer scientific questions about the world around them". While not often considered to be a scientific method of investigation, discursive-exploration of scientific

ideas with peers is a key element of the scientific method (Numbers and Kampourakis, 2015). Interventions such as CASE, P4C and the proposed intervention studies would provide a suitable environment for children to engage in this type of scientific inquiry.

Discussing classroom discourse, Mercer & Howe (2012) highlight that goal-directed collaborative dialogue supports children's development, however, this type of talk is lacking in settings. While seemingly underemployed in schools, the National Curriculum does highlight the importance of speaking and listening in helping children "in making their thinking clear, both to themselves and others" (DfE, 2013, p.4). Additionally, the Key Stage 1 (5- to 7-years-old) guidance (2013, p.146) states "the principle focus of science teaching in Key Stage 1 is to enable pupils to experience and observe phenomena, looking more closely at the natural and human-constructed world around them", including grouping and classifying entities, and noticing patterns. These curriculum aims correlate with the underlying aim of the proposed interventions. That is, to encourage children to think about claims and evaluate evidence for teleological statements, as this may support the shift from small to big ideas. Therefore, it is argued that while developing children's thinking skills may not be specifically prescribed in the 2013 National Curriculum for England and Wales, it is an implication of the curriculum's aims. Furthermore, support for these types of lessons comes from academic calls for thinking skills programmes, for example, see Oliver & Venville (2016).

2.8.8 Summary

The educational rationale behind the proposed interventions is to improve children's scientific literacy, to assist them in considering things in a scientific way, and to move from small ideas (explanations related to specific concepts) to big ideas (theories that have a wide ranging applications and can be used when encountering new things). This shift is important as big ideas (e.g., all natural phenomena must have a causal explanation) have a greater explanatory value than small ideas (e.g., rain must have causal explanation), as they are applicable to a greater range of concepts. Therefore, the Study 2 and 3 interventions sought to enable this shift from small to big ideas; as teaching about the lack of purpose and goal-driven change in the natural world could have a wider-ranging impact than simply teaching children about the causes of several exemplars.

Harlen (2006, p.165) advocates four points for developing children's scientific ideas:

1. "Extending children's experience", that is, teachers need to gradually introduce new concepts;
2. "Helping children to test their ideas" through either discussion or experimentation, which allows a move from small ideas which explain specific events to big ideas which explain general events;
3. "Providing access to alternative ideas" to encourage cognitive conflict;

4. “Promoting communication, dialogue and reflection”, enabling children to socially construct their ideas and to develop a shared understanding of concepts.

The proposed interventions used these four guiding principles while incorporating ideals, theory and methods from dialogical teaching and P4C. The design of the sessions is discussed further in sections 6.2 and 7.2. The Study 2 intervention investigated the research question *do discursive activities about the appropriateness of scientific explanations support children to develop an understanding of appropriate (causal) and inappropriate (teleological) explanations for the existence of natural phenomena?* The Study 3 intervention had the aim of *do discursive activities about the appropriateness of scientific explanations support children to develop an understanding of appropriate (evolutionary) and inappropriate (design-teleological) explanations for the existence of organisms' traits?* These questions are elaborated on and further developed in the relevant study chapters.

3. Methodology

3.1 Introduction

This chapter describes the theoretical and certain practical aspects of the methodology used in this thesis. Due to the incremental approach used for the empirical work the contents of this chapter focus on the overarching themes regarding the methodology. The actual methods pertaining to individual studies are discussed in the specific study chapters.

A number of research questions arose from the literature review, which were addressed by four studies. Study 1 explored two main research questions:

- RQ1 – *does question wording influence children's explanations about the existence of natural phenomena?* It also addressed two sub-research questions.
- RQ2 – *does questioned topic influence children's tendency to provide a teleological response?*

There were also two sub-questions in Study 1:

- SQ1 – *what types of teleology do children display when they provide a teleological explanation?*
- SQ2 – *do children's explanations about natural phenomena result from fragmented or coherent knowledge?*

Study 1b was initially designed to explore the question *does the natural discourse children experience about natural kinds influence children's explanations for their existence?* However, following Study 1 this research question was refined in relation to focus on a specific ontological category and consequently asked

- RQ3 – *does the natural discourse children experience about natural phenomena influence children's explanations for their existence?*

Two intervention studies (Studies 2 and 3) addressed similar research questions using similar methods.

- RQ4 (Study 2) – *do discursive activities about the appropriateness of scientific explanations enable children to develop an understanding of why causal explanations are appropriate and teleological explanations are inappropriate to account for the existence of natural phenomena?*
- RQ5 (Study 3) – *do discursive activities about the appropriateness of scientific explanations enable children to develop an understanding of why evolutionary explanations are appropriate and design-teleological explanations are inappropriate to account for the existence of organisms' traits?*

This chapter firstly details the philosophical framework which underpinned the exploration of these research questions. Secondly, it broadly outlines the methods used to a) collect and b) analyse the data needed to answer these questions. Thirdly, certain aspects of working with children and schools are reflected upon, these are issues which are relevant the all of the experimental studies conducted. Finally, this chapter outlines the decisions and practices with regard to the ethical treatment of participants and data.

3.2 Philosophical framework

The philosophical framework underpinning this thesis is that of pragmatic positivism or perhaps more aptly pragmatic post-positivism, due to the accepted concept that research does not discover a standard model of truth but rather certain approximations of truth (Crook & Garrett, 2005; Robson, 2011). The following deconstructs the term pragmatic post-positivism.

The overall guiding principle behind the research methodology is that decisions should be made pragmatically: the most appropriate research method is what works best to answer the question (Biddle & Schafft, 2014; Johnson & Onwuegbuzie, 2004; Johnson, Onwuegbuzie, & Turner, 2007). This pragmatic approach to research has resulted in a philosophical framework which does not dictate research methods nor the type of data collected. The main consideration was that the methods used and data collected are those which are needed to effectively answer the research questions (Benton & Craib, 2011; Feilzer, 2010).

While this thesis follows a pragmatic style, certain post-positivist stances influence the research. Consequently, the thesis is underpinned by certain epistemological and ontological views. Principally that knowledge about the natural and social world can be systematically discovered and, while many different ideas can explain the same concept, a theory about a concept is the idea which most effectively explains said concept. Furthermore, the research is conducted with the post-positivist epistemological belief that to construct a theory, data should be investigated deductively, logically and in an objective manner (Mack, 2010; Pring, 2000), although a theory would only be applicable to certain individuals in similar contexts (Schwandt, 2000). In relation to ontological positioning reality is viewed as independent and external to individuals. However, an individual may construct their own personal reality in relation to idiosyncratic experiences, therefore, research participants can have multiple understandings of the same event: multiple realities (Potter, 2000; Stake, 2010). This is not to state that every multiple reality is true merely that individuals believe their reality to be true.

As a consequence of combining a pragmatic approach with a post-positivist epistemology and ontology, the research is conducted with three main philosophical underpinnings. Firstly, decisions about research methods should be made by considering what methods and form of data are the most appropriate to answer the research questions. Secondly, theories about reality are those which are created in a logical and systematic manner, and most effectively explain a certain concept.

Thirdly, a theory is a generalised view of knowledge due to its construction from individuals' multiple versions of reality (Benton & Craib, 2011; Muijs, 2011; Robson, 2011), therefore, results are only applicable to certain groups of people in similar contexts.

3.3 Methods

3.3.1 Real world research

Real world research is often associated with lower levels of internal validity but higher levels of external validity. This is not to suggest that these two forms of validly have a perfect inverse relationship, rather that design choices that seek to improve external validity may pose a threat to internal validity (Shadish, Cook & Campbell, 2002). While rigorous standards of all forms of validity and reliability should be strived for this is not always practical in real world settings, such settings can be messy and multifaceted (Robson, 2011). This can lead to problems such as when additional confounding variables are not accounted for in the research design (Shadish, Cook & Campbell, 2002). However, it is argued that relaxing control over variables particularly those related to the selection and history of participants (which may lead to lower levels of internal validity) is worthwhile to ensure that research is grounded in context and actually represents the multiple facets of human interaction and learning, as it is better to potentially sacrifice aspects of internal validity to better portray participants (Robson, 2011). Therefore, the experimental studies in the thesis are conducted within school settings, thus have high levels of external validity: they are generalisable to similar populations (Shadish, Cook & Campbell, 2002). This may have led to issues associated with internal validity.

Two common threats to internal validly are the *selection* and *history* of participants. The former is related to individual participant differences and how participants are assigned to treatment groups (e.g., a number of high performing participants could be assigned to the same treatment group), this selection threat can result in group differences being incorrectly attributed to a certain treatment. This threat is minimised by random-assignment of participants (see relevant chapter sections). The later threat concerns the history of participants during experiments, suggesting that certain factors may influence participant performance that are outside of the scope of measurement for the experiment. Often such history threats are usually minimised by laboratory-based studies that can control for outside influences (Shadish, Cook & Campbell, 2002), such a stringent control of variables is not always possible in real world research. For example, data collection for Study 2 had to take place in two schools at two different times of the school year, which may have introduced a confounding variable of parent, school or teacher. Furthermore, schools that participate in educational research are often slightly above average, for instance a school may have slightly lower than average number of children receiving Free School Meals – a common benchmark for inequality. It is expected that participating schools will fall on the upper side of typical English schools, as schools that are on the lower side of typical, such as those in special measures, are

unlikely to have the time or space to accommodate such studies. While these limitations are acknowledged (and explored within relevant chapters and section 8.3) they are considered acceptable limitations due to the broader goal of producing applied research.

3.3.2 Experimental methods

As the thesis is conducted within a post-positivist framework, a methodology that employed testable, repeatable methods was desirable. From this stance, and from a pragmatic view, an experimental approach was the obvious choice for the main studies (1, 2 and 3), as it enabled hypothesis testing and causal inference within a (reasonably) controlled context (Wilson & Maclean, 2011). A controlled design was paramount in Study 1 due to the focus on specific linguistic cues (i.e., forms of question wording used to investigate RQ1) and was important for conducting efficacy trials for the interventions used in Studies 2 and 3.

Often criticised for not being able to be used to explore contextual factors, the use of experimental methods is viewed with suspicion by some within the field of educational research (Mishler, 1979), as experiments are often used in the attempt to provide a generalised view of a phenomena rather than an understanding of complex individual characteristics (Robson, 2011). However, as discussed in relation to real world research and applied findings, this was the focus of the thesis: a generalised picture of how factors influence children's explanations (Study 1) and a controlled test of teaching methods (Study 2 and 3). The thesis focus was not an in-depth study of certain individuals' thinking. Furthermore, experiments are usually criticised by those following subjectivist methodologies who, for example, view experiments as dehumanising to participants (Oakley 2000, cited in Robson, 2011, p. 84). This is a problem that experiments may have to overcome in education: they are conducted within a philosophical framework which differs from those who would criticise the approach. As the thesis was conducted with pragmatic underpinnings and sought to provide evidence of how various factors influence children teleological explanations (Study 1) and if children can be taught that teleological explanations are inappropriate (Study 2 and 3) an experimental approach was the most suitable research design. Therefore, while the approach could be criticised by some within the field of educational research, an experimental methodology was the best method to answer the research questions. The use of an experimental approach for the three studies allowed for generalizable results and limitation of confounding factors.

One of the main methods in reducing confounding variables (in Studies 1, 2 and 3) was the use of a control group, which can greatly assist in causal inference (Field & Hole, 2003; Shadish, Cook & Campbell, 2002). Study 1 used a repeated-measures design, which meant that the participants received both treatments and acted as their own control group. This cancelled out the confounding factors associated with variation between groups of participants in independent group designs (King, Keogh & Verba, 2004). That is, comparing two treatments which were experienced by all participants controls for between-participant variation. The main problem associated with

repeated-measures designs is *order* or *carry-over effects*: the initial treatment influencing participants' performance during the latter treatments. This testing threat to internal validity was limited using counterbalancing (Shadish, Cook & Campbell, 2002; Wilson & McLean, 2011). Specific details pertaining to counterbalancing are provided in Chapter 4 (section 4.2.4).

Studies 2 and 3 used pre-/ post-test, control group designs. In both studies a true experimental design was used (randomisation within a class), as opposed to a quasi-experimental design (one treatment class and one control class), as it had the advantage of removing or reducing the selection threat to internal validity (Robson, 2011; Shadish, Cook & Campbell, 2002). Therefore, limiting the influence of the class teacher (as multiple classes of children were required) and theoretically limiting participant bias (Wilson & McLean, 2011). However, it is acknowledged that the decision to use randomisation could have led to rivalry and demoralisation (Shadish, Cook & Campbell, 2002), as children in the control groups were aware they are not involved in the interventions, and bleed effects, as there was no way to stop children in the different groups discussing the intervention activities.

There are also a number of negatives associated with a pre-/post-test assessment. Pre-testing can lead to practice effects and fatigue effects, as participants receive the same questions for pre-test and post-test, which decreases internal validity (Wilson & MacLean, 2011). However, this was decided to be an acceptable risk. The main method used to improve the internal validity in Studies 2 and 3 was through use of a control group. By adding a control group internal validity was increased, as any relative gain to the experimental group which is not mirrored in the control group is arguably attributable to the treatment (Shadish, Cook & Campbell, 2002; Wilson & McLean, 2011).

There are different forms of a control group, in medical research types of control groups are well defined (e.g., head-to-head comparisons or untreated control) but these control group labels are rarely used (Karlsson & Bergmark, 2015). In social science research control group terminology appears poorly defined and also rarely used. Bryman (2016), Field & Hole (2013) and Cohen, Manion & Morrison (2011) simply refer to *control groups* where a group does not receive the treatment; however, Robson (2011) labels these situations as *no-treatment control groups*. There are also arguments that educational interventions should be tested against current best practice (see Goldacre, 2013) in order to test not just what works but what works better than the competition.

Considering types of control groups, introduces the notion that control group could receive no treatment, a 'business-as-usual' treatment, or comparable treatment control. For example, in a test of an intervention to improve spelling skills, children in the treatment group could be compared against a group that received a) no formal input of any kind on spelling skills (no treatment), b) normal input in terms of school curriculum related lessons on spelling skills (business-as-usual) or c) another comparable intervention for improving spelling skills (comparable treatment).

Using this terminology Studies 2 and 3 used no treatment control groups, as the content covered in the interventions was not statutory curriculum content for the age-ranges investigated, nor was there any comparable treatments which the interventions could be contrasted against (i.e., no treatments which specifically sought to develop an understanding of teleological explanations). Consequently, in the thesis, a no treatment control was used to control for the effects of repeated testing (i.e., carry over effects from pre- to post-test) and maturity (i.e., children changing over time due to factors not associated with the treatment). While this situation can provide information regarding the efficacy of the interventions, it does not allow claims to be made about the *relative* efficacy of the intervention. That is, the trials can show if the interventions are successful in their aims but not *how* successful they are relative to other approaches. The methods used in the interventions may be the best possible method to challenge children's preconceptions or the methods may be one of the worst possible ways (although while still producing an effect).

As discussed, one aim of the thesis was to conduct research which was applicable to real world settings, therefore threats to generalisability, to external validity, must be considered. Robson, 2010, after LeCompte & Goetz, 1982, p.91) outlines four main threats:

- Selection – the findings only relate to the participants investigated (e.g., results may be biased by only sampling participants with an interest in scientific theories);
- Setting – the findings are context dependent (e.g., results may be biased by a school with a specific focus on science education);
- History – historical or participant specific events may influence participant performance (e.g., results may be biased by participants being home tutored in scientific theories);
- Construct effects – the constructs investigated only relate to the group investigated (e.g., results may not be generalizable to the general population if only children in the 1st or 99th quantile of science achievement are sampled).

These threats are usually limited through representative sampling (Newby, 2010). Therefore, for all experimental studies conducted the samples are described in detail (in the relevant chapters) to allow the reader to compare the sample with the general population. Participating schools did not appear to significantly differ from the national average in terms of pupil statistics (English as an Additional Language, Free School Meals and Special Educational Needs) so are considered representative, consequently, the results are argued to be generalizable.

A final note on experimental methods, all experiments are subject to certain limitations, some of the general limitations of the experiments in this thesis have been discussed above, however, there were limitations which were specific to certain studies. These limitations (and the limitations of Study 1b) are not discussed in the study-specific chapters, except in so far that they may affect the interpretation of the specific results, but rather together in the final main discussion

(Chapter 8) of the thesis. This collation of limitations was done to provide a concise set for the overall thesis.

3.3.3 Corpus linguistics

A question arising from the literature review was (RQ3) *does the natural discourse children experience about natural phenomena influence children's explanations for their existence?* Corpus linguistics was used as a method to answer this question or more precisely two component questions: (CQ1) *what specific examples of teleological and scientific explanations do children experience about natural phenomena?* and (CQ2) *what are the themes associated with natural phenomena in the discussions children experience?* This section explores the rationale and theory behind the use of corpus linguistics, whereas Chapter 5 details the creation of the corpus and the specific steps taken to analyse the corpus data.

Corpus linguistics is the analysis of a corpus for patterns of language use. A corpus is a large collection of transcripts, which when combined usually have word counts over 1,000,000. These transcripts are often not collected by a single researcher or research team but sourced from various online databases (e.g., UK Data Service, 2017). The term *transcript* can refer to any form of linguistic data, for example, transcribed spoken language in the form of interviews or observations, or written language in the form of newspaper articles or student essays.

The primary reason for using corpus linguistics was that it facilitated the systematic analysis of large amounts of naturally occurring language (Gries, 2006). The benefits of using real world data were explored in section 3.3.1. This form of data was particularly important for Study 1b as data originating from, for example, laboratory studies or structured interviews will be shown, in Chapter 4, to be influenced by factors such as question wording. Corpus linguistics also had the advantage of generalisability, as corpora are representative of the discourse in the general or a target population the results are applicable to similar populations (McEnery, Xiao & Tono, 2006). A further benefit was the use of secondary qualitative datasets that were originally costly to collect and transcribe (Fielding & Fielding, 2000). Therefore, by building a specialised corpus a large amount of data was available that would not have been possible to collect within the context of this thesis.

A corpus can be explored in various ways; one method is to search a corpus for a number of *keywords* (Stubbs, 1995), for example, different exemplars of natural phenomena. This would result in the frequency of occurrence for each keyword along with their *concordances* and *collocates*. Concordances are the keywords in context, essentially the keyword in the sentence in which it appears in the corpus. Keyword concordances are used in Study 1b to explore (CQ1) *what specific examples of teleological and scientific explanations do children experience about natural phenomena?*

Collocates are words which appear beside or near a keyword with a higher frequency than they occur elsewhere in the corpus (Stubbs, 1995). For example, light could have the collocate of *feather*, that is, the two words appear together more frequently than suggested by chance, perhaps due to the idiom *as light as a feather*. Exploring and categorising collocates can show general patterns of language associated with a keyword, which can subsequently be used to suggest themes in discourse (Stubbs, 1995). In relation to teleological discourse of, for example, night a proportion of collocates may be associated with the theme of sleep (a common teleological explanation for night) without night having any overt teleological concordances. Therefore, suggesting that the way night is talked about, it being in the context of sleep related words, could influence children's explanations for the existence of night. Consequently, collocate investigation could support the answering of CQ2: *what are the themes associated with natural phenomena in the discussions children experience?*

3.3.4 Statistical analyses

The main statistical analyses used for the experimental studies in this thesis was ANOVA and ANCOVA, along with a number of post-hoc tests including *t*-tests, Cochran's *Q* and Chi-square goodness-of-fit (these are explored further in the relevant sections). The analysis of collocates in the corpus linguistics study relied on the use of the Mutual Information statistic (see Gries, 2009) which is detailed in section 5.2.3.3. Before analyses were conducted data were checked for various assumptions including independence of observations, homogeneity of variances and normality.

Normality of variables, that the data have a normal distribution, is one of the assumptions of ANOVA and ANCOVA. However, social research often produces messy, non-normal datasets (Abelson, 1999). Common practice when data deviates from normality would be to use non-parametric tests (Field, 2013). The use of such tests often comes at the cost of low of power to detect an effect, especially when using small datasets (Coolican, 2009) such as the datasets used in this thesis. Glass, Peckham & Sanders (1972) argued that the use of non-parametric test is often not warranted in these situations, as researchers treat the normality assumption as a tick box rather than considering the implications of what happens when assumptions are violated, and to what degree must they be violated before the test is unsuitable.

Following various statistical modelling Glass, Peckham & Sanders (1972) concluded that skewed distributions analysed using ANOVA have very little impact on significance levels or on statistical power, providing the test is two-tailed and the distribution is not wildly non-normal (i.e., the kurtosis and skew are within certain limits suggested by the paper). In a meta-review on using ANOVA with non-normal distributions, Harwell, Rubinstein, Hayes & Olds (1992) found considerable support that this form of analysis is robust to non-normal data providing that variances are equal.

In this thesis the Shapiro-Wilk test (Shapiro & Wilk, 1965) is used as a test of normality when using ANOVA and ANCOVA. Where the variables were found to significantly differ from normality, the group variance, skew and kurtosis of the distributions was considered. Providing that these measurements were not pathologically skewed the analysis proceeded. Where appropriate, variables which significantly differed from normality are indicated by footnote.

While outliers have been shown to influence parametric tests (Zimmerman, 1994), in the experimental studies where exploration of variables suggested an outlier it was not initially removed if it was not considered extreme nor considered to be illegitimate. Furthermore, testing was conducted with and without the outlier (Kruskal, 1960); if the results were not significantly different then the outlier was retained.

3.4 Working with gatekeepers (schools and parents)

Participant recruitment was consistently a problem thoughout the studies, a problem often faced in educational experimental studies (Rochelle, *et al.*, 2014). The primary problem with recruitment at the school level was a lack of response from schools. In the studies schools were initially approached via email to the head teacher, which contained information about the study and a call for participants. If there was no response to this initial email the school was contacted via telephone to inquire if the email had been received. A small number of schools were initially contacted by a third party whom had a large number of contacts in local institutions, this individual recommend the school to take part in the research studies. Following recommendations by the third party the contact process was the same as the above.

While the main barrier to recruitment was at the school level, the parental level caused recruitment problems in the initial stages of the thesis. During Study 1 opt-in consent forms were sent to the parents of 270 children; one form was returned. Following this opt-in consent forms were sent to the parents of 64 children in a different school; eight forms were returned. The head teachers in both of these schools had continually had problems with parents returning permission forms. Therefore, the decision was made (with the full oversight of the School of Education Ethics Committee) to use opt-out consent forms during further recruitment rounds in Study 1 and in the latter studies.

Opt-out consent was only used with the permission of the school. Providing the head teacher and class teachers agreed to use of the opt-out method, combined information letters and consent forms (see Appendix 1 for a Study 1 example) were sent to the parents / carers of all children in the classes where the teachers and head teacher had agreed the research could be conducted. The consent forms contained an opt-out clause, which indicated that parents who did not want their child to participate were asked to return the form to the school. Following Dockett et al. (2012), in order to ensure that parental concerns could be addressed, data collection took place a minimum of two

weeks later to provide adequate time for parents to contemplate their child's involvement in the study and to discuss with the researcher or teacher about the project.

3.5 Working with children

This section discusses the decisions made and practicalities of working with children, while section 3.6 specifically covers ethics, some ethical issues are discussed in this section where relevant; for example, securing children's consent. The elements covered here pertain to all three experimental studies conducted in the thesis. Study 1b did not involve working children, therefore in this section when *the studies* are mention it refers to only the three experimental studies (Studies 1, 2, and 3).

Recruitment at the child level is deceptively easy within school settings, researchers could simply use their power as an adult (who was also viewed as a teacher) to tell children to participate, without asking for consent or explaining the research (Lundy & McEvoy, 2012; Woodhead & Faulkner, 2000). From an ethical and a child-rights perspective this is unacceptable. While this is an exaggerated example, it is possible to slip into patterns of behaviour that mirror the outdated practices described. The following discusses the methods and problems surrounding obtaining children's consent during the studies, in relation to ethics and child-rights.

Children's consent is the main form of agreement required for research with children to take place (Conroy & Harcourt, 2009). The United National Convention on the Rights of the Child outlines that children's views should be taken into account in relation to their age and maturity (Article 12, United Nations, 1989). Contrary to some suggestions, children are perfectly capable of consenting from a young age if the study purpose and requirements are appropriately explained (Dockett et al., 2012). In order to provide children with sufficient time to think about their involvement in the research, during all studies it was outlined to the whole cohort (e.g., a class of children) what involvement in the research would entail and an overview of the research purpose. Furthermore, at the beginning of the data collection interviews, the purpose and aims of the studies were reiterated and children were asked to provide consent. They were also informed that at any point they could stop and leave the activity. To ensure clarity and simplicity of explanation, presentation of the study outline was refined with children's help during the piloting stages before each main study.

Children have the right to make decisions on events which affect them (Article 12, United Nations, 1989). However, understanding that it is their decision can be problematic as there are power structures in schools that often result in children obeying adult commands (Woodhead & Faulkner, 2000). This could be exacerbated if the researcher is viewed as a teacher or if the research activity resembles a traditional school activity. Both of these points were true for all studies. To counter inadvertently coercing children to take part, during the studies children were periodically reminded that the researcher was not a teacher, that it was a research study not a lesson, and it was their decision to be involved.

While children may provide consent at the onset of the interview their willingness to be involved may falter (Flewitt, 2005). Consequently, children's consent must be monitored throughout the data collection (BPS, 2014). To ensure children's on-going consent throughout the studies, they were monitored for signs of dissent (e.g., loss of interest in the interview or providing large numbers of don't know answers) and the sessions were stopped if necessary (Dockett, Einarsdóttir, & Perry, 2009; Dockett et al., 2012). Observing dissent can be problematic, for example, a signal of unease may not indicate withdrawal of consent but rather restlessness. Consequently, an adapted version of Winnicott's squiggle game was used when children appeared uncomfortable, children were then asked if they wished to continue. When used, this "pause" in the session allowed children a break and all with whom it was used agreed to continue the session.

The squiggle game was an adapted version of Winnicott's (1971) psychological assessment aid. Originally designed to explore children's thoughts and feelings, but also to engage children, the game involved an adult and a child taking turns to draw squiggly lines on paper, the child then used these squiggles to create a picture, which Winnicott argued could be used to discuss children's emotions. Berger (1980) proposed that the game is effective because it is a fun, simple activity which quickly engages and initiates discussion with children. During the studies conducted in this thesis the squiggle game began with the interviewer drawing a squiggle on a slice of paper, the participant then had 30 seconds to turn the squiggle into a drawing. The interviewer and the participant then discussed the drawing. As mentioned this game was used to allow participants a break during sessions if they appeared uncomfortable. However, the activity was also employed as an icebreaker.

Icebreakers are often needed in research with young children to develop the participant-interviewer relationship; thereby, creating a conversational environment and reducing possible shyness (Freeman & Mathison, 2009; Hatch, 2009; Hogan, 2005). This icebreaker was used in all pilot and main studies. Following the icebreaker children took part in the study specific activities which are detailed in the relevant chapters.

3.6 Ethics

All research was conducted within the guidelines advocated be the University of Nottingham (UoN, 2015) and the British Educational Research Association (BERA, 2011). As a consequence parents had the right to withdraw their children at any point.

Data were anonymised and kept confidential with only the researcher and supervisory team having access. Data is stored on secure servers, in line the Data Protection Act (1988). D(ata will be kept for a minimum of seven years.

Due to working with children there was a potential risk to the participants involved. However, all preliminary, pilot and main studies conducted were reviewed by the University's ethics committee which deemed there to be no risk of harm to children involved.

Consent in all the studies was first sought from the head and class teachers in each school, then opt-out consent forms and information sheets were sent to the children's parents. Opt-out consent was only used with the permission of the school. Providing the head teacher and class teachers agreed to use of the opt-out method, information letters and consent forms (see Appendix 1) were sent to the parents / carers of all children in the classes where the research was to be conducted. The consent forms contained an opt-out clause, where parents who did not want their child to participate were asked to return the form to the school. In order to ensure that parental concerns could be addressed, data collection took place a minimum of two weeks later to provide adequate time for parents to contemplate their child's involvement in the study and to discuss with the researcher or teacher about the project (Dockett et al., 2012).

This opt-out practice has been used in a number of other educational studies (see a review of opt-out methods by Dockett et al., 2012) and is only considered to be provisional consent because children may not wish to participate when invited to be interviewed in the school (ESRC, 2010; Harcourt & Conroy, 2011). As stated multiple times in literature, children are perfectly capable of agreeing to take part in an interview from a young age if it is appropriately explained to them (Dockett et al., 2012; Flewitt, 2005; Freeman & Mathison, 2009). The process regarding obtaining and monitoring children's consent was outlined in section 3.5.

3.7 Summary

This chapter has explained the methodology used in this thesis to explore the research questions raised in the literature review. Firstly, the philosophical framework used was outlined, that of pragmatic post-positivism with the derived belief that research needs to be informed by, be applicable to and solve problems that exist in real world settings. Secondly, it was detailed that Studies 1, 2 and 3 used an experimental design, and corpus linguistics was used in Study 1b. Furthermore, the use of statistical analyses as the main method of data analyses throughout this these was detailed and justified. Thirdly, this chapter explained issues surrounding working with gatekeepers (schools and parents), working with children and ethical considerations. The main decisions being the use of opt-out consent, the continued verification of children's consent and the use of icebreakers with children.

4. Study 1: Influences on young children's teleological explanations for natural phenomena

4.1 Introduction

The literature review (Chapter 2) explored current thinking around children's explanations for the existence of natural kinds, with a particular focus on children's teleological explanations. The use of these explanations is considered to be problematic when children come to learn about causal, scientific accounts (see section 2.1.3). Current studies would suggest that children primarily use teleological explanations to account for the existence of natural kinds, including organisms, natural objects and natural phenomena (see section 2.3.1). The study detailed in the chapter aimed to address a number of research questions that arose from the literature review (see section 2.6). Specifically, those related to the measurement of teleological explanations. The first main research question was *does question wording influence children's explanations about the existence of natural kinds?* It was hypothesised that question wording will influence children's responses, with a teleologically leading treatment producing more teleological responses than a neutrally worded question. This was due to previous manuscripts using potentially leading questions to investigate children's ideas. For example, the often cited manuscript by Kelemen (1999a) which used the question wording w*hat is X for?* A question that may place a demand on a participant to provide a purposeful rationale for X. If the hypothesis above is correct it would indicate that certain studies will have overestimated children's tendency to provide teleological rationales. This would imply that educators and researchers need to avoid leading questions or statements, and that teleology may not be the barrier it is claimed to be.

The second main research question relates to questioned topic, RQ2 asks *does questioned topic influence children's tendency to provide a teleological response?* This question arose after consideration of the number of exemplars used in previous work. A number of studies (e.g., Kampourakis 2012a, Kelemen 2003) only used a couple of exemplars (e.g., clouds) to account for children's reasoning for an entire ontological category (e.g., natural phenomena). However, a small number of exemplars may be misleading if they are not representative. If there is large variation across the examples used (even within a ontological category) then this has implications for a) theories that children's teleological reasoning is "promiscuous" and b) implies that caution needs to be taken interpreting results of studies that used only a limited number of questioned topics.

Sub-question one explores the type of teleology children are espousing. Specifically, it asks *what types of teleology do children display when they provide a teleological explanation?* This question originated from consideration of the type of teleology discussed in manuscripts: teleology is usually only discussed as design-teleology (although an exception would be ojalehto et al., 2013). If children's teleological explanations are not always indicative of design-teleology, but suggest

relational-teleology, it may indicate that teleological explanations are not only due to a fervent belief that natural kinds have a specific purpose, rather that teleology may result from natural patterns of language use.

The second main question and the first sub-question also provides an opportunity to consider the structure of children's naïve ideas about the origin of natural kinds. Thus the second sub-question is *do children's explanations about the natural world result from fragmented or coherent knowledge?* A coherent structure would imply that for a range of contexts (a number of questioned topics) children's explanations should a) demonstrate considerable coherence within children (a child should be coherent in their teleological or scientific responses across a range of questioned topics) and b) demonstrate considerable coherence across children (for a given questioned topic all children should provide a small number of compactly describable responses). Whereas a fragmented structure would predict that a) children's responses would vary across contexts (they would not be stable in their propensity to provide teleological or scientific response across a range of questioned topics) and b) there would be variation across children (for a single questioned topic there would be a range of response types given by a group of children). If children's explanations indicate a coherent structure of knowledge, this would coalesce with and support current thinking. However, a fragmented view would imply that children's naïve ideas may be more context-dependent and that conceptual change may be achieved through the use of targeted interventions, in which educators could gradually chip away at children's preconceptions that result in teleological explanation.

4.2 Method

4.2.1 Design

Chapter 3 outlined the reasons for using an experimental approach and the pragmatic-post-positive framework underpinning this thesis. Study 1 used a repeated-measures design to address the research questions. A 2x2 mixed design was used to test the influence of question wording on children's explanations. The first factor, question wording, was within participants and had two levels: the open-treatment (why is there X?) and the leading-treatment (what is X for?). As each child was asked both types of question, the second factor, treatment order, determined if their first question set comprising of either the leading- or open-questions. That is, half the children were randomly assigned to receive a block of open then leading questions whereas the other half had a block of leading then open questions. As each treatment consisted of five questions, in total children were asked about ten examples of natural phenomena, and so each participant received questions about the ten examples in a randomised order. As question set order and order of questioned topic were randomised separately, across the sample questioned topics used were approximately split between the two treatments (e.g., roughly half of the total questions about rain were asked using the leading-treatment).

4.2.2 Materials

4.2.2.1 Question wordings

Review of the literature reveals there are two main methods previous studies have used to inquire about children's use of teleological explanations: open-ended questions and multiple-choice questions. For this consideration of potential question wordings multiple-choice options have been rephrased as open-ended questions. For example, Kelemen's (1999c) scenario, with a puppet stating that *X is for Y* and another saying *X is not for Y it just does it,* can be reworded as the question *is X for Y?* Consideration of the question wordings used in previous studies (Kampourakis et al., 2012a; Keil, 1992, 1994; Kelemen, 1999a, 1999b; Polling & Evans, 2002; Tao et al., 2011) shows that there are two main forms of question wording: is the question-subject (X) for something and if so what, and, why does the question-subject (X) have this appendage (Y). These two question types, *is the question-subject for something and if so what* and *why does the question-subject have this appendage* (i.e., *what is the purpose of a certain appendage?*), can be combined into one main question format: *what is X for?* A question format which seeks to inquire about the purpose of X, be X an organism, natural object, natural phenomena, artefact or a specific part of those. Use of the question wording *what is X for?* allows for comparison with Kelemen's (1999a) highly cited and influential paper.

The question wording *what is X for?* is further referred to as the leading-treatment, as it is possibly teleologically-biased and may lead children into providing a teleological response. To allow for comparison with the leading-treatment there needs to be an open-treatment: a question wording which is not biased, it does not lead children into providing a teleological or scientific response. After consideration of various open question wordings the one decided upon was *why is there X?*, as it is a general question: it can be used with any example of natural phenomenon or natural object. When asking about different examples the question may have to be adjusted slightly. For example, through changing *is* to *are* for pluralised examples (e.g., why *are* there rainbows), or adding *the* to the question subject (e.g., *the* Sun), this is not considered to alter the meaning of the question wording.

4.2.2.2 Questioned topics

A preliminary study was conducted to identify suitable examples of natural phenomena and natural objects for discussion. The goal was to identify examples for which children could articulate a response irrespective of accuracy and so was not concerned with whether children's responses were teleological or scientific. To start with a wide pool of potential candidates were identified from previous studies (DiYanni & Kelemen, 2005; Kampourakis et al., 2012a; Keil, 1994; Kelemen, 1999a, 1999b; Piaget, 1929; Polling & Evans, 2002) along with additional relevant examples generated from discussion with colleagues. This candidate set was reduced by excluding examples that could be appropriately answered by a teleological response. For example, organisms' traits

were removed, as a trait (e.g., a bird's wing) can be described scientifically using functional-teleology. Artefacts were also disregarded, as a design-teleological answer is required. Finally, organisms were also omitted because domesticated animals, those in captivity and plants in gardens could be appropriately explained in functional-teleological terms, leaving a small selection of viable examples. This resulted in natural objects (non-living, natural entities, e.g., rocks, soil, mountains) and natural phenomena (time-constrained, natural events or processes) as the most suitable ontological categories.

Twelve 6- and 7-year-olds (female = 6) were interviewed in groups of four. Group interviews were used to investigate 20 possible questioned topics, as shown in Table 4.1. Children were settled with an adapted version of the squiggle game (Winnicott, 1979) to quickly build a relationship between the children and the interviewer. Therefore, creating a conversational environment and reducing possible shyness in children (see Methodology, Chapter 4, for further discussion on working with young children). All participants enjoyed the squiggle game and became progressively more talkative, indicating it to be an appropriate icebreaker for the main study. Following the icebreaker children were informed about the purpose of the interview: to develop an understanding about what children think about different scientific concepts. Furthermore, children were presented with the following statement:

> *Now some of these questions may be tricky but that's OK, because lots of people have different ideas about the answers. However, I think you all might have some really good ideas and I want to know what you all think. If you believe something different to your friend, that's OK; your answer is just as good. Remember this is not a test, and if you can't think of any answer to a question that's fine just say you don't know.*

This statement was included to discourage children from repeating other children's answers or trying to guess a 'correct' answer.

Table 4.1: *Pilot interview questioned topics split by question set*

Question Set 1		Question Set 2	
clouds	the Sun	snow	rain
day	rocks	waterfalls	mountains
storms	waves	darkness	night
the Moon	light	wind	stars
rainbows	icebergs	rivers	sand

The pilot interview consisted of two question sets, shown in Table 4.1, for each example of natural phenomenon and object children were asked *why is there X?*. Children were asked about Question Set 1, then participants had a short break from the questions and played another round of the squiggle game before resuming the interview with Question Set 2. After each question (e.g., *why is there snow?*) each child was encouraged to briefly give their initial response to provide an initial picture of thinking, the child who responded first was varied. Following the initial response children were asked to elaborate upon their answer or reflect on the comments made by another child. This was done to create a discussion around the responses and encourage reflection. After a few questions most participants naturally built upon the responses of other children, or changed their answer to reflect the same idea as another child. However, this did not happen for all participants for all questions: sometimes children would stick firmly with their original response. Immediately following each question set, children were asked about how difficult the questioned topics had been. This was done to compare participants' ability to provide an articulate response with their perceived difficulty in answering the question. Finally, children were asked about how they would improve the overall interview process for children in the future; for example, number of questions or discussion time. The interviews ended by thanking children for their involvement. All pilot interviews were audio-recorded.

After consideration of participants' responses, children's opinions about question difficulty and interviewer observations, a number of the 20 possible examples of natural phenomena and natural objects were eliminated. In particular, natural objects, such as rocks, proved particularly difficult for children to provide any response to, which is surprising considering the focus on natural objects in previous work (see Kelemen, 1999b; 2003; Kampourakis et al., 2012a; 2012b). Consequently, the remaining examples for the main study were all within the category of natural phenomena: darkness, day, light, night, rain, rainbows, snow, storms, waterfalls and waves.

4.2.3 Participants

The participants were 66 children, aged 68- to 104-months-old, recruited from three classes within the same primary school from a large city in the Midlands of the UK. This school is single-form entry, is judged to be performing well on national scales, and is above the national average for both pupils from minority ethnic groups and those entitled to free school meals. The sample consisted of 22 children from each of Year 1, Year 2 and Year 3 (5- to 8-years-old, M = 86 months, SD = 10.84, Female = 34). No child was classified as having Special Educational Needs (SEN) or English as an Additional Language (EAL).

4.2.4 Procedure

Children were settled using an adapted version of Winnicott's squiggle game (Winnicott, 1971). Following the icebreaker children were informed that the interview would be recorded and presented with the following statement:

> *Now, some of these questions may be tricky but that's OK, because lots of people have different ideas about the answers. However, I think you might have some really good ideas and I just want to know what you think. Remember this is not a test, there are no right answers, it is just what you believe. And if you're not sure about an answer, just say 'I don't know', OK?*

This statement was included to discourage children from making up responses or trying to guess a 'correct' answer.

After the introduction participants then received five questions using the leading-treatment or open-treatment, dependent upon treatment group. Following these questions, children played another round of the squiggle game before resuming the interview with the five remaining questioned topics using the counterpart question wording. If children took a substantial amount of time to respond to a question the question was repeated, then were offered the choice to either 'have some more thinking time or move onto the next question'. Probes were only used to add information to obscure responses. For example, *can you tell me a bit more about Y?,* Y being a summary of a participant's initial ambiguous response. The entire interview process lasted around 15 minutes for each child. All interviews were audio-recorded and took place in a quiet communal area outside of the classrooms. The study adhered to the BERA ethical guidelines and was approved by the University's ethics board.

4.2.5 Measures

Children's responses were coded on two levels (Table 4.2). At Level 1, answers were coded as *teleological*, *scientific* and *other*. At Level 2, teleological responses were sub-classified to allow for investigation of the types of teleological explanations children provide. As functional-teleology is not possible for natural phenomena, the data was coded into *design-teleology* and *relational-teleology,* using the Level 2 rubric in Table 4.2.

Table 4.2: *Coding rubric for children's responses at two levels*

Level 1			
Grade	Teleological (0)	Scientific (1)	Other (NA)
Explanation	The topic is given a purpose, a function or a goal, which aids or supports either itself or another entity.	A scaled down, causal explanation for a topic, answer does not contain teleological notions. It may not be scientifically correct.	Descriptive responses (participant provides a description of the topic), Non sequiturs (participants answer is an unrelated comment, it is not related to the topic), Don't know (participants state they 'don't know' the answer or are 'unsure') or No answer (participant provides no response).
Example	'There are waterfalls so wild animals have a place to drink' or 'there is night to tell us to go to sleep'	'It snows because the clouds get cold and all smushed up together'	
Level 2: sub-classification of teleological responses			
Grade	Design-teleology		Relational-teleology
Explanation	A topic is explained in relation to a purpose or goal for which it was designed. The topic has the singular purpose of aiding/supporting something, the something can be itself or another entity; it acts upon an entity to produce a result which is not a scientifically valid.		A topic is explained in relation to a purpose for which it can be used or an effect which the topic causes. The purpose or effect is human-centric.
Example	'Rain is for watering the plants' or 'day is for waking up the birds'		'snow is used by children for making snowmen' or 'night is a time period which people usually use for going to sleep'

Whilst theoretically different, in terms of being distinct abstract categories, actual examples of design-teleology and relational-teleology could prove difficult to distinguish. Firstly, as there is

no published coding scheme to differentiate the two. Secondly, as discussed in section 2.5.2, actual examples of teleological explanations (e.g., children's statements) can fall on what was termed a spectrum of teleology. As such it needs to be determined whether a vague teleological statement leans more towards one category of teleology than another. Consequently, the approach developed for this thesis classifies statements based on design-teleology as those that discuss natural phenomena performing actions which aid or support something: the topic being explained will act on an agent. (e.g., rain is for watering the plants) Whereas, explanations based on relational-teleology will discuss something as being used by an agent to perform an action. In such an account an agent will act on the topic being explained (e.g., snow is for children to make snowmen with). However, this raises an issue, children explanations can be brief and not provide sufficient information to distinguish between these two types of teleology; for example, the response that *snow is for making snowmen*, could be considered to be either type; therefore, provoking a dilemma when coding.

Linguistic analysis provided a method of solving this dilemma. Explanations employing design-teleology, accounting for natural phenomena as aiding or supporting something, will have the questioned topic as the subject of the answer, and the subject will be acting upon an object. For example, in the response *rain is for watering the grass*, the subject (*rain*) is acting on (*watering*) the object (*grass*). Conversely, relational-teleological explanations, those which account for natural phenomena as something which an agent uses, will introduce an agent into the response via a subject complement (e.g., 'you', 'we' or 'I') or implicitly adding an agent via a verb, which implies an agent (e.g., 'making', 'throwing' or 'learning'). For example, in the statement, *snow is for making snowmen*, the subject complement (*for making snowmen*) implicitly implies an agent via the verb (*making*); therefore, *snow* is used by an agent for a subjective purpose (*making snowmen*).

To establish the extent of inter-rater reliability, 10% of the data, (66 responses) were coded by a second researcher. Cohen Kappa indicated a very good level of agreement ($K = .885$, $p < .001$ for Level 1 coding and $K = .841$, $p < .001$ for Level 2 coding).

4.3 Results

4.3.1 Influence of question wording

Overall, 69.2% of responses were teleological, with the leading-treatment questions eliciting 80.0% teleological responses and the open-treatment 58.4%. However, 72 responses out of the possible 660 were coded at Level 1 as *other*. Those were then removed from the data, resulting in 588 valid responses to the 10 questions, 457 (77.7%) of which were teleological and 131 (22.3%) were scientific. Splitting the data by treatment shows the leading-treatment received in 264 (92.1%) teleological answers, whereas 193 (68.2%) teleological responses were given to the open-treatment.

A 2x2 ANCOVA was conducted with the independent variables of question wording (leading- and open-treatment) and question set order (first question set-leading or first question set-

open). The dependent variable was the summed responses to the leading- and open-questions, resulting in a participant score of 0-5 for each treatment. The covariate was participants' age in months, which was mean centred to avoid compromising the main effect (Enders & Tofighi, 2007). Percentage means are displayed in Table 4.3.

Table 4.3: *Mean number of scientific responses split by question wording and question set order*

Question wording	Percentage of scientific responses (*SD*)		
	First question set-leading (n = 33)	First question set-open (n = 33)	Overall (n = 66)
Leading	7.88% (15.76)	7.88% (12.18)	7.88% (13.97)
Open	26.06% (26.68)	37.56% (33.07)	31.82% (30.37)

Results indicate that question wording had a significant effect on children's responses, F (1,63) = 44.57, $p.$ < .001, η_p^2 = .41, as a significantly greater number of teleological responses were given in the leading-treatment condition (92.1%) than in the open-treatment (68.2%). Participant age also predicted responses: older children provided fewer teleological answers, F (1,63) = 4.40, $p.$ = .041, η_p^2 = .06, although there was no interaction between question wording and age, F (1,63) = 0.01, $p.$ = .962, η_p^2 < .01. Question set order did not significantly influence responses, F (1,63) = 1.72, $p.$ = .179, η_p^2 = .02, nor was there a significant interaction between question wording and question set order, F (1,63) = 2.59, $p.$ = .093, η_p^2 = .04. The test results suggest that children were not primed to provide teleological or scientific responses if the first question set they received was a leading- or open-treatment, respectively.

4.3.2 Influence of questioned topic

The proportion of scientific responses varied between the examples of natural phenomena used as questioned topics, ranging from 10.5% for waves to 49.2% for rainbows. Cochran's Q test was used to investigate if there was a difference in the amount of teleological responses for the ten questioned topics, across question wordings. The maximum corrected measurement η^2_Q (Serlin, Carr & Marasuilo, 1982) is provided as a measure of effect size. Due to missing data (those coded as 'other"), this test was conducted with a sample size of n = 34. The results suggest that there is a significant difference between the responses provided to the ten questioned topics, Q (9) = 33.29, $p.$ < .001, η^2_Q = .11.

Exploring the influence of topic with the full Study 1 dataset raises a problem. The data in Study 1 was collected using two question wordings, a leading- and open-treatment, and the leading-treatment was shown to increase children's tendency to provide teleological explanations.

Therefore, the full dataset (leading- and open-treatments combined) will contain biased data. The following investigates this bias by exploring the Study 1 data by questioned topic and question wording. *Combined dataset* refers to the data gathered using the leading- and open-treatments, and *leading dataset* and *open dataset* refer to the data gathered using the leading- and open-treatments, respectively.

Considering children's teleological and scientific responses to each questioned topic by question wording (Table 4.4) and the ranked order of questioned topics by question wording (Table 4.5) suggests a considerable change in the rank order of questioned topics when comparing the combined and open datasets. Although the tables show that when moving from the combined dataset to the open dataset the percentages of teleological responses to every questioned topic decreased, the actual rank position of four questioned topics also changed. The ordering of storms, waves, rain and snow (positions 6-9) were reversed. Comparing the combined dataset to the leading dataset also shows considerable rank position shifts. Consequently, using the combined dataset may not provide an accurate estimate of the number of teleological explanation children provide for the ten questioned topics. For this reason the open dataset will be used in the following analysis.

Table 4.4: *Percentages* of teleological response split by questioned topic and question wording*

	Question wording					
	Open and Leading		Open		Leading	
Topic	*n*	% teleological	*n*	% teleological	*n*	% teleological
Darkness	58	82.8%	29	69.0%	29	96.6%
Day	60	85.0%	31	74.2%	29	96.6%
Light	62	83.9%	33	72.7%	29	96.6%
Night	62	88.7%	33	81.8%	29	96.6%
Rain	59	74.1%	30	63.3%	29	89.7%
Rainbows	59	50.8%	28	25.0%	31	74.2%
Snow	61	77.0%	38	68.4%	23	91.3%
Storms	56	64.3%	25	32.0%	31	90.3%
Waterfalls	57	89.5%	32	87.5%	25	92.0%
Waves	54	75.9%	18	50.0%	36	88.9%

*Remaining percentages are the scientific explanations.

Table 4.5: *Questioned topics ranked by percentage of teleological responses, split by question wording*

	Question wording		
Rank*	Leading and Open	Open	Leading
1	Waterfalls	Waterfalls	Darkness†
2	Night	Night	Day†
3	Day	Day	Light†
4	Light	Light	Night†
5	Darkness	Darkness	Waterfalls
6	Storms	Snow	Snow
7	Waves	Rain	Storms
8	Rain	Waves	Rain
9	Snow	Storms	Waves
10	Rainbows	Rainbows	Rainbows

*topics ranked in descending percentage order, †topics share the same rank

Consequently, to further investigate if certain questioned topics received more teleological (or scientific) responses than other questioned topics, ten (one per questioned topic) chi-square goodness-of-fit tests were used. As previous work has only investigated a limited number of questioned topics, there is no published work that can be used to predict the effect of questioned

topic on response type (the split of teleological to scientific responses). Consequently, the expected value was set to the weighted-mean level of teleological and scientific responses across the ten questioned topics from this study (teleological : scientific = . 624 : .376), which is a conservative method as a questioned topic is being compared against an average which includes itself. Thus, this analysis can identify which (if any) questioned topic deviates from this average. Crammer's V (φ_c) is provided as a measure of effect size; results are shown in Table 4.6.

Table 4.6: *Goodness-of-fit test results for response type by questioned topic*

		Percentage of responses				
Topic	*n*	Teleological response	Scientific response	χ^2 (df = 1)	*p.*	φ_c
Darkness	29	69.0%	31.0%	0.53	.465	0.14
Day	31	74.2%	25.8%	1.84	.175	0.24
Light	33	72.7%	27.3%	1.50	.220	0.21
Night	33	81.8%	18.2%	2.44	.118	0.27
Rain	30	63.3%	36.7%	0.01	.915	0.02
Rainbows	28	25.0%	75.0%	16.68	< .001	0.77
Snow	38	68.4%	31.6%	0.59	.443	0.12
Storms	25	32.0%	68.0%	9.84	.002	0.63
Waterfalls	32	87.5%	12.5%	8.60	.003	0.52
Waves	18	50.0%	50.0%	1.18	.278	0.26

The results show that the responses for darkness, day, light, night, rain, snow and waves did not significantly differ from the mean number of responses to the ten questioned topics. However, the observed values for rainbows, storms and waterfalls significantly different from the mean responses for the ten questioned topics. Rainbows and storms resulted in less teleological and more scientific responses than the average. While, for waterfalls children showed a tendency to provide more teleological explanations than the average value for the ten questioned topics.

4.3.3 Type of teleology

To investigate the type of teleology espoused by participants the 457 teleological responses were coded at Level 2 (see Table 1). The Level 1 categories of *other and scientific* were removed to provide an indication of preference for type of teleology. Coding at Level 2 resulted in 133 responses displaying design-teleology and 324 indicating relational-teleology. Table 4.7 shows the percentage scores for type of teleology advocated, overall and by treatment type, and indicates an approximate 3:7 split for design-teleology and relational-teleology. This appears consistent across

question wordings, however, the previous section (4.3.2) indicates that children's explanations cannot be averaged across an ontological category: there is an influence of questioned topic. Consequently, there needs to be an analysis of type of teleology by questioned topic.

Table 4.7: *Percentage score for type of teleology advocated split by question wording*

	Design-teleology	Relational-teleology
Overall (n = 457)	29.1%	70.9%
Leading-treatment (n = 264)	30.3%	69.7%
Open-treatment (n = 193)	27.5%	72.5%

Again, there is a problem due to the influence of question wording. Exploring children's design- vs. relational-teleological responses to each questioned topic by question wording (Table 4.8), and the ranked order of questioned topics by question wording (Table 4.9), suggests a limited change in the rank order of questioned topics when comparing the combined and open datasets. Table 4.9 shows that when moving from the combined dataset to the open dataset the actual rank position of only two questioned topics changes. The positions of snow and waterfalls (positions 5 and 6) were reversed. Comparing the combined dataset to the leading dataset also shows a larger number of rank position shifts. While the change from combined to open datasets is minimal, for consistency the open dataset will be used in the following analysis.

Table 4.8: *Percentages of relational-teleological responses* by questioned topic and question wording*

	Question wording					
	Leading and Open		Open		Leading	
Topic	*n*	% relational	*n*	% relational	*n*	% relational
Darkness	48	81.3%	20	80.0%	28	82.7%
Day	51	92.2%	23	95.7%	28	89.3%
Light	52	82.7%	24	87.5%	28	78.6%
Night	55	81.8%	27	85.2%	28	78.6%
Rain	45	44.4%	19	47.4%	26	42.3%
Rainbows	31	64.5%	7	57.1%	23	65.2%
Snow	47	72.3%	26	73.1%	21	71.4%
Storms	36	38.9%	8	0.00%	28	50.0%
Waterfalls	51	72.5%	28	67.9%	23	78.3%
Waves	41	61.0%	9	55.6%	32	62.5%

*Remaining percentages are the design-teleological explanations.

Table 4.9: *Questioned topics ranked by percentage of relational-teleological responses, split by question wording*

	Question wording		
Rank*	Leading and Open	Open	Leading
1	Day	Day	Day
2	Light	Light	Darkness
3	Night	Night	Light†
4	Darkness	Darkness	Night†
5	Waterfalls	Snow	Waterfalls
6	Snow	Waterfalls	Snow
7	Rainbows	Rainbows	Rainbows
8	Waves	Waves	Waves
9	Rain	Rain	Storms
10	Storms	Storms	Rain

* Topics ranked in descending percentage order, †topics share the same rank

To further investigate this preference for relational-teleology, as above[3], ten (one per questioned topic) chi-square goodness-of-fit tests were used to determine any preference for response type (design-teleology or relational-teleology) for each questioned topic. As there is no published work that can be used to predict the type of teleology children are espousing, the expected values were set using the weighted-mean levels of design- and relational-teleology (design-teleology : relational-teleology = .350 : .650). Results are shown in Table 4.10.

Table 4.10: *Goodness-of-fit test results for type of teleology by questioned topic*

Topic	*n*	Percentage of responses		χ^2 (*df* = 1)	*p.*	φ_c
		Design-teleology	Relational-teleology			
Darkness	20	20.0%	80.0%	1.99	.158	0.32
Day	23	4.3%	95.7%	9.52	.002	0.64
Light	24	12.5%	87.5%	5.36	.021	0.47
Night	27	14.8%	85.2%	4.86	.028	0.42
Rain	19	52.6%	47.4%	2.58	.108	0.37
Rainbows	7	42.9%	57.1%	0.19	.665	0.16
Snow	26	26.9%	73.1%	7.93	.005	0.55
Storms	8	100.0%	0.0%	14.83	< .001	1.36
Waterfalls	28	32.1%	67.9%	0.10	.747	0.06
Waves	9	44.4%	55.6%	0.35	.555	0.20

The results show that the responses for darkness, rain, rainbows, waterfalls and waves did not significantly differ from the mean of the ten questioned topics. However, the observed values for day, light, night, snow and storms are significantly different from the mean. Teleological explanations for storms showed a preference for design- over relational-teleology. Whereas, day, light, night and snow mirrored the overall mean preference for relational-teleology but with a higher proportion of responses than the mean of the ten questioned topics.

4.3.4 Examples of children's explanations

The following details examples of the responses provided to different questioned topics. This is provided for two reasons. Firstly, to ground the results in real-world examples rather than a focus on observed and estimated statistics. Secondly, to provide examples of children explanation for

[3] Again Cochran's Q test was performed using the full dataset to test for any initial difference in response type across questioned topics and question wordings. Due to missing data (data coded as 'other' and the removal of scientific responses), Cochran's Q test was conducted with a sample size of $n = 8$. The result suggests a significant difference, $Q(9) = 20.87$, $p. = .013$, $\eta^2_Q = .29$. However, this finding is tentative due to the loss of sample size caused by listwise deletion.

each questioned topic to help address the sub-research question regarding the coherence or fragmentation of children's naïve ideas about the existence of natural phenomena. The types of responses are outlined for the ten questioned topics in turn.

4.3.4.1 Examples for darkness

Children's scientific explanations for darkness indicate a common, coherent framework. The majority of the scientific accounts centred around the absence of the sun, because 'the sun goes lower and lower until it is night-time', or 'cos it is night-time and the sun is not showing.' There was also evidence of the Earth's rotation in some answers: 'the Earth goes in difference positions […] so there will be darkness in different places and sunlight in different places', or 'cos the moon goes around the Earth, […] it front of the Sun at night and that makes it dark.' Furthermore, a small number suggested that darkness levels alter with the seasons, 'when it is spring it is kind of darkness but when it is autumn there is a lot, a lot of dark […] it gets really dark.'

The teleological accounts demonstrated less coherence. With five forms of teleological rationales contradictorily suggesting that darkness was for:

- Primarily as a time period or indicator for people to sleep or rest; for example, 'we can go to sleep and we, so we don't see any light', 'darkness is for, like, people to sleep' and 'so we can go to sleep' (relational-teleology).
- Time as a time for nocturnal animals, for example, 'so the night animals get a turn to come out' and 'for more nocturnal cos bats are nocturnal animals' (design-teleology)
- Some children also mentioned darkness being for producing a scary environment, 'for being scared' and 'for Hallowe'en […] it gets all dark and spooky and when you go to peoples' houses it is scary' (relational-teleology).
- One participant also mentioned it being for celestial bodies 'so the stars can come out' (design-teleology).
- One child also associated darkness with evil 'there is darkness because some people think, like, bad thoughts […] and you know, like, when it is night time I thought that was god fighting with the devil'. (religious-teleology)

4.3.4.2 Examples for day

Children's scientific explanations for day indicate a common, coherent framework. The majority of the scientific accounts discussed the rising of the Sun as the cause of day. For example, it is day 'cos when, after night, the Sun comes back up and the moon goes down', because 'in the morning the day starts again […] when it is day time it is a little bit light, then another nit of light, then more light,' and 'cos the sun comes out with larva falling out of it.' One participant also made reference

to the rotation of the Earth: day is 'because the Sun moves around the Earth […] which makes the Sun reflect into different places.'

Children's teleological accounts for day demonstrated a lack of coherence across-participants, with four forms of teleological explanations.

- The first suggesting that day was for telling people to wake up (relational-teleology), effectively some sort of natural alarm clock. For example, 'if you don't have morning you will sleep all day', 'day is for being awake and you go to sleep in the night' and 'if you didn't have daytime you would have to sleep all though'.
- The second proclaiming that day is a period for human activity (relational-teleology), although the type of activity did vary being either Play-based activities (e.g., 'so we can play' or 'for us to be active and run about') or Work-based activities (e.g., 'daytime is for people to go out, like, if you need to go out, like to work' or 'cos you got to go to school, you got to learn words').
- The third associating light with aiding human vision (relational-teleology), for example, 'so you can see and people don't have to walk in the night-time cos you can't see very well in the night', if we didn't have day it would just be dark and we couldn't see […] what we were doing' and 'so you can see'.
- The fourth indicates that light was there to produce sunshine for the earth (design-teleology).

4.3.4.3 Examples for light

Children's scientific explanations for light indicate a common, coherent framework. The majority of the scientific accounts have the Sun as the cause of light. For example, there is light because 'in the sun it makes light', 'the Sun brings us lightness from the ring around it', and 'cos the Sun [beams] come down.' One participant also suggested fire as the cause of light 'light is made out of fire.'

The teleological accounts demonstrated some coherence. With one major form of teleological rationale, which suggested that light was for aiding human vision (relational-teleology). For example, 'it helps us see where we are going so we don't bump into stuff', 'to see properly', 'so they can see better cos if it is, like, dark you'd be, like, bashing into lamp posts'. Although, a handful did indicate that light was designed for animal vision (e.g., so the birds can see where they are going), and one participant did state that light was for supporting plant life, 'so the plants can grow'.

4.3.4.4 Examples for night

Children's scientific explanations for night indicate a common, coherent framework. The majority of the scientific accounts centred on the absence of sunlight as the cause of night. For example, the

Sun setting: ‘cos when the sun goes down, when, it creates darkness, it creates, erm, that is called night’; or the Earth rotating: ‘the Earth is turning and when our part of the country is dark it is night, then when it is near the Sun it is day’; or the moon acting as a filter ‘the sun goes behind the moon and makes the night and the moon shine.’

The teleological accounts also demonstrated a coherent approach to night, with almost all teleological rationales suggesting that night was a signal for humans to rest and sleep ‘cos you need to sleep’, because you have to go bed’, ‘when you go bed and if there wasn’t no night-time you couldn’t, erm, go sleep and then in the morning if you wanted to go somewhere you wouldn’t be bothered cos you hadn’t been sleep’. However, a couple of children claimed night is for nocturnal animals (e.g., ‘so mice can run out and get food’, so, like, the insects can come out’) and for celestial bodies (e.g., ‘for the stars and moon to come out’).

4.3.4.5 Examples for rain

Children’s scientific explanations for rain indicate a common, coherent framework. The majority of the scientific accounts suggested rain to be caused by ‘the sky, the sky makes storms and rain’; by clouds combining, ‘the clouds make rain […] sometimes they squash together and the rain comes out’; due to seasonal change, ‘when it is past summer, the summer goes and it turns rainy’; by steam ‘cos the steam goes up to the clouds […] and it makes the clouds bigger and it turns into rain’; and, finally, rain is ‘just the weather’.

The teleological accounts demonstrated a more fragmented system of naïve ideas. With five different teleological rationales suggesting that rain was:

- For watering the plants (design-teleology); for example, ‘it makes the grass grow’, ‘to help plants grow’, ‘it is for keeping all the plants, grass and flowers healthy.’
- For animals (design-teleology); for example, ‘animals can drink it if there is a little hole’ or ‘for the fish cos it goes in the pond.’
- For playing with (relational-teleology); for example, ‘people can play in it’, in case you want to jump up and down in puddles’, ‘because if you like jumping in the rain with your feet.’
- For washing things (design- and relational-teleology) ; for example, ‘it’s to get more water’, to get water […] the rain can go on in the sea and you can carry it with a bucket’, ‘rain is for, like, washing your car and washing stuff.’
- Due to the supernatural (religious-teleology); for example, ‘it is cos god is crying’, ‘because, like, when god is sad all the water comes down.’

4.3.4.6 Examples for rainbows

Children’s scientific explanations for rainbows indicate a common, coherent framework. The majority of the scientific accounts discussed the interaction of sunlight and rain as the cause of

rainbows. For example, ‘the sun and the rain make the rainbow, I think because the water’s got, like, a reflection inside it, the Sun shines on it’, ‘sun and rain makes rainbows […] when the both come out at the same time a rainbow appears’ and ‘because it rains and suns at the same time, which makes, which makes the rainbows’. However, a small number only referred to rain as a cause, ‘cos when it rains there are rainbows’, and others suggested clouds as the cause of rainbows, ‘the sun shines through the clouds and the clouds make the rainbow’.

The limited number of teleological accounts for rainbows makes it difficult to explore the structure of naïve ideas but those available indicate two types of rationale. Rainbows are:

- equated to a natural art form for human enjoyment (relational-teleology) ‘because it is colourful and you can look at them and sit down’, ‘for joy’, and ‘for making people happy.’
- associated with the cultural adage of being supernatural treasure markers (design-teleology) ‘some people might think that that there is treasure at the end of the rainbow’ and ‘if you follow the rainbow it could lead you to something really special, like treasure’.

4.3.4.7 Examples for snow

Children’s scientific explanations for snow indicate a common, coherent framework. The majority of the scientific accounts mentioned temperature, for example, ‘because it is so cold’ and ‘when it is cold it turns really cold, then it snows’. Although a large number of answers discussed both temperature and rain, ‘the rain gets really cold and falls out of the sky’ and it is very cold, the rain comes down and it freezes and becomes snow’. Other accounts referred to snow clouds, ‘the clouds are filled with, like, snow, like the rain has been there for ages so, I think it turns into snow and falls down as snow’, and some replies added a seasonal cause, ‘cos if it is winter you have snow, you know rain makes snow.’

The teleological accounts demonstrated a fragmented view of snow with six teleological rationales. Snow was:

- Seen as a substance which enables Christmas (relational-teleology), ‘it is for celebrating Christmas’, ‘there’s snow cos you know […] it is, like, gods birthday’, it needs to be snowing or no one would know it was Christmas’.
- To provide a play environment ‘because you need to play snowball fights’, so people can go sledging’, ‘it is because, basically, children could build snowmans and play snowball fights.’
- For supporting animal life (design-teleology) ‘so the polar bears can get some rest’, ‘so some animals can live in snow’.

- For filling up human made bodies of water (relational-teleology), ‘so when it melts you can use the water in water fountains, ponds, lake and sometimes people chuck it in the sea and then add a bit of salt. Put a bit of salt in it and chuck it in the sea’.
- For cooling humans down (relational-teleology), ‘it is because […] otherwise everyday it would be hot and we’d would get burnt’.

4.3.4.8 Examples for storms

Children’s scientific explanations for storms indicate a common, coherent framework. The majority of the scientific accounts discussed centred on heavy rain as the cause of storms. For example, ‘if it is raining and the clouds are together then sometimes storms come’, ‘storms are, like, you know how the rain comes down […] if it’s pouring down loads and loads and loads you get a thunderstorms, like if it is raining too much’ and ‘because the rain goes dead hard’. However, a large number of other answers made reference to storms being caused by thunder or lightening in addition to rain. For example, ‘you’ve got lightening and clouds shooted lighting down’ and ‘because sometimes, the clouds get rid of wet, like really stiff wet, and it makes thunder’. Furthermore, there was one mention to seasoning influencing storms, ‘there are storms, like, in November […] the weather, it changes in the year’, and one child proposed that storms are ‘just cos there is weather really’.

The teleological accounts demonstrated a less coherent structure. With four teleological rationales suggesting that storms were:

- For producing water (design-teleology), ‘o make more rain, so you can get more water’, ‘if there is a really big storm there will be a flood’.
- For aiding the Earth (design-teleology), ‘it is quite good for the Earth to have storms’.
- Used by plant life (design-teleology), ‘we have rain especially for the plants’, ‘to water the flowers’, ‘the flowers need rain, cos sometimes the plants need a lot’.
- Caused by the mood of a supernatural being (design-teleology), ‘it is because when god woke up in a bad mood’, ‘when god cries, yeah, that is rain. Then when god, like, shouts that’s thunder’.

4.3.4.9 Examples for waterfalls

Children’s scientific explanations for waterfalls indicate a common, coherent framework. The majority of the scientific accounts centred on water falling from rocks as the cause of waterfalls. For example, ‘there are just, like, rocks, big rocks, you know, like in a river, […] that leads to a waterfall’, ‘water just piles up in there and when there is nowhere else to move around it just does down’ and ‘because there is a huge part of water somewhere, and, erm, there is a rock and […] the water just comes from above down the rocks’.

The teleological accounts demonstrated little coherence. With five teleological rationales suggesting that waterfalls are:

- For providing humans with water (relational-teleology), ‘for clean safe water’, ‘so people can have showers’, ‘so you can get water’.
- For providing animals with water (design-teleology) ‘for the fish to drink in and for the fish to swim’, ‘to make water for the animals that are in the wild’.
- Seen as a natural art form (relational-teleology), ‘to make places look nice’, ‘to make the place look even nicer with a little waterfall or a big waterfall’, so you can have a look at them’.
- For play (relational-teleology), ‘when it is, erm, hot there is a waterfall and people sit in it to get cold’, ‘so you can play in it’, ‘so people can swim in it.’
- To move water around the Earth (design-teleology), ‘to put water in rivers’, ‘so it goes into the river’.

4.3.4.10 Examples for waves

Children’s scientific explanations for waves indicate a common, coherent framework. The majority of the scientific accounts discussed centred around wind as the cause of waves. For example, ‘because the wind blows the water and makes the waves’, ‘the ocean, when it is windy, it’ll go upwards and downwards so the waves go up and downwards’ and if there is wind it blows them [waves] back and forwards’. While one child mentioned the gravitational pull of the moon in their response, ‘because of the moon […] it makes the waves’ and another refed to water displacement as a cause ‘when there is a big splash in the sea it creates a big, like, thing which comes towards us and that is called waves’.

The limited number of teleological accounts for waves makes it difficult to explore the structure of naïve ideas but those available indicate two types of rationale. Rainbows are:

- To provide support to animals (design-teleology), ‘the fish can swim better when there are waves’, ‘it there was no waves […] the animals that live in the sea won’t be able to move.’
- For playing in (design-teleology), ‘so people can swim better’, ‘waves are for making the sea more fun’, ‘so people can go on the skateboard, like surfing on it.’

4.4 Discussion

This study had three aims, to investigate 1) the influence of question wording on children's tendencies to provide teleological explanations about the existence of natural phenomena; 2) the influence of questioned topic within the category of natural phenomena, on children's propensity to provide teleological rationales; and 3) the type of teleology, design or relational, that children use to account for natural phenomena. Before addressing these questions, a clear finding is that

regardless of question wording, questioned topic and age, children demonstrated a considerable inclination to provide teleological explanations for natural phenomena. The following explores the findings pertaining to the three research questions and the implications arising.

4.4.1 Influence of question wording

The data support the hypothesis that question wording influences children's responses to questions about scientific concepts. Children provided a significantly greater number of teleological responses (91.4%) to the leading-treatment (what is X for?), whereas the open-treatment (why is there X?) received 64.3% teleological answers. Therefore, questions phrased using teleological-language can exacerbate children strong tendency to provide purposeful explanations for questions about natural phenomena. There was no significant interaction between question wording and question set order, although the partial-eta shows a small effect size. There does appear to be a trend that children who received the leading-treatment first to provide lower numbers of scientific responses when they received the open-treatment. This possible trend would have made the test of the hypothesis more conservative. Considering children's age, an increase in age did produce differences in the responses participants gave, as children's age increased they provided less teleological and more scientific responses, which is to be expected and supports findings by Kelemen (1999b, 2003) and Kampourakis et al. (2012a; 2012b). However, regardless of age, participants predominantly provided teleological responses to the leading-treatment, while the open-treatment question received less teleological and more scientific answers. Consequently, the leading question wording was found to increase children's tendency to provide teleological explanations across the entire age range investigated.

The findings regarding question wording raise problems for the interpretation of previous work into children's teleological explanations Research using teleologically-biased questions or explanations to elicit children's explanations about the natural world is likely to have been influenced by the impact of purposeful-statements. For example, 'does Y help X?', 'is it better for X to have Y?' (Keil, 1995), 'what is X for?', 'is X for Y, or not?' (Kelemen, 1999a), 'is X useful for something?' (Kampourakis et al., 2012a). Even if children have been given the option to decline to answer, suggest it was a strange question or eschew an explanation that contained a teleological bias, they are still exposed to teleologically-biased language. Children can be prompted by the leading phrases used in leading and multiple-choice questions or may refrain from stating that do not know an answer if they could relate natural phenomena to a purpose for which it could be used, thereby echoing the relational-teleology derived from the relational-deictic framework (ojalehto et al., 2013).

From an educational point of view this finding indicates that the language used to phrase questions or explanations about scientific concepts can place demand characteristics upon respondents. Employing purposeful language (willingly or unwillingly) leads children into

providing teleological explanations. This is problematic as children's ability to talk, and think, about scientific concepts rests, at least in part, on their understanding of scientific terminology (Dawes, 2004, Mercer & Littleton, 2007). Consequently, educators must model for children how to use language to question, analyse and communicate ideas. This is particularly pertinent to young children as the focus of Early Years science education 'is to enable pupils to experience and observe phenomena, looking more closely at the natural […] world around them' (DfE, 2013, p. 146). The implications for assessment are clear; the findings support calls for the use of open questions in research into children's explanations (e.g., see Kampourakis et al., 2012a). In the biology classroom functional-teleology can be appropriate with young children; for example, learning about the role of petals on a flower. However, it is not appropriate in relation to Earth Sciences, where children should not be inadvertently prompted into explaining natural phenomena as purposeful or agentic, and therefore, the use of teleological-leading language is a cause for concern.

It is suggested that the use of teleological explanations are influenced by exposure to artefacts (Carey, 1985; Evans, 2000; Kampourakis, 2014; Kelemen, 1999c) and by religious education (Banerjee & Bloom, 2014; Evans, 2000; Lindeman et al., 2015), which leads to robust conceptual misclassification (Carey, 1985; Chi, 1992, 2008). Demand characteristics placed upon children by teleologically-leading language can now be added to this list. While not an independent cause it is claimed to be an exacerbating factor. Teleologically-worded questions may intensify children's tendencies to provide teleologically explanations, to propose purpose where there is none.

4.4.2 Influence of questioned topic

The results show that questioned topic can have a considerable influence on response type. Even using the conservative method of comparing explanations for a questioned topic against a mean which includes said questioned topic, three out of the ten examples of natural phenomena significantly differed to the mean numbers of responses given for natural phenomena. For night and waterfalls children were more likely to provide teleological explanations than was suggested by the category mean, whereas rainbows and storms received significantly more scientific explanations than was suggested by the mean value for natural phenomena. Therefore, suggesting that teleological reasoning is not consistent within an ontological category, although, further research would be needed to confirm this is replicable for other ontological categories, for example, natural objects.

This finding has two main implications. First, it could be inferred that children's explanations for natural phenomena support the 'promiscuous' teleology stance. 58.4% of children's total explanations, when averaged across questioned topics, were teleological, close to Kelemen's (1999a) findings of ~70% teleological explanations for natural objects. However, that conclusion would be misleading. When averaged across questioned topics children may employ

teleology for the category of natural phenomena but the results of this study suggest that their application of teleology is more nuanced. Children provided significantly different numbers of teleological explanations depending upon the example of natural phenomena they were explaining. Within the category of natural phenomena, children displayed selective-teleology, not promiscuous-teleology. Consequently, caution needs to be taken interpreting the results of studies if they used only a limited number of supposed exemplars to represent a category, as depending upon the examples investigated studies may have over, or under, estimated children's teleological tendencies.

The second implication relates to the structure of knowledge about the natural world. As mentioned earlier there is no work on the structure of children's teleological naïve ideas, the preconception that nature has a purpose. It was postulated that the answer to this question is dependent upon if there is a) within-child variation across contexts and b) across-child variation across contexts. With knowledge-as-theory predicting considerable coherence in responses across examples of natural phenomena and across children, and knowledge-in-pieces suggesting considerable variation in response type.

Looking at the data for response type (teleological vs. scientific, and design- vs. relation-teleology discussed in the next section) it would appear that, firstly, for the range natural phenomena investigated children's responses exhibited reasonable variation. Individual children varied in response type across examples of natural phenomena used, children did not demonstrate a coherent view of explanations for the existence of natural phenomena. As indicated in the results a given child did not only use only scientific explanations, or relational-teleology or design-teleology to account for all ten examples of natural phenomena, rather children a range of explanatory styles depending upon the natural phenomena they were discussing. Thus demonstrating considerable within-child variation across contexts.

Secondly, for any given example of natural phenomena there was little coherence in explanatory style, that is, the participants provided a range of contradictory rationales to account for a single questioned topic. As indicated in the descriptive results regarding children's explanations (section 4.3.2), children's scientific accounts were broadly coherent (with usually one or two causal explanations per example of natural phenomena) across the ten examples of natural phenomena. However, for darkness, day, rain, snow, storms and waterfalls the teleological explanations across-participants indicated a range of contradictory purposeful rationales for their existence: there was little coherence. (See sections 4.3.4.1 - 4.3.4.10 for examples). Therefore, demonstrating considerable across-child variation across contexts. Even for the examples of natural phenomena where there was some coherence in teleological response (light and night), there were still a small number of outlying explanations. Consequently, these two complementary results suggest that children's naïve ideas about the existence of natural phenomena may be the result of fragmented knowledge.

Further exploration into the structure of naïve knowledge in other ontological categories is needed to confirm such findings. However, children's naïve ideas about the causes of natural phenomena are argued to be a form of fragmented knowledge. If children's knowledge about natural phenomena is fragmented, this would indicate a diminished barrier to learning. This statement follows from diSessa's argument that conceptual change would require the incremental reshaping of targeted knowledge elements. It is viewed as a diminished barrier not because teleology is less of a problem but because a fragmented construction suggests a new way of tacking the barrier that is teleology. Teleological thought and explanation does not need to be entirely eradicated in young children's discourse. Rather through the gradual reframing of elements of children's ideas, educators could gradually chip away at children's preconceptions which result in teleological explanation. This could be achieved through targeted interventions, for example where individuals learn when and where teleological explanation is appropriate (e.g., learning heuristics, Zohar & Ginossar, 1998).

4.4.3 Type of teleology

With regards to type of teleology, the results show that, overall and by treatment type, there is an approximate 3:7 split for design-teleology and relational-teleology. Consequently, for the examples of natural phenomena investigated children's teleological explanations are more likely to be relational than to exhibit design-teleology. That is, on a spectrum of actual examples of teleological explanation the majority of children's statements were either clearly codable, or lent sufficiently towards, the area of the spectrum that is classified as relation-teleology. While the term *lent sufficiently towards* may appear subjective the number of such cases was minimal, and the coding is supported by strong interrater reliability scores. Again, using the conservative method of comparing a questioned topic against a mean which includes said questioned topic, there is an influence of questioned topic. When compared to the mean number of responses for natural phenomena, certain examples were significantly more likely to be explained via relational-teleology (day, light and night) and others (storms) by design-teleology. This variation in type of teleological explanation across individuals and questioned topics supports the claims made above about the fragmented structure of naïve knowledge about the causes of natural phenomena.

These results indicate that the majority of children's teleological explanations advocate relational-teleology rather than design-teleology, and the finding is applicable to the majority of natural phenomena investigated. As this study only explored natural phenomena, further research is needed to establish if there is a preference for relational-teleology as opposed to design-teleology in other ontological categories. However, the results of this study do provide empirical support for the suggestions made by ojehlato et al. (2013) and the notion that children use relational-teleology. The implication of this finding is that teleology may be less of a developmental problem than previously proposed. If children are using relational-teleology, instead of the usually referred to

design-teleology, it suggests that when children provide teleological explanations that may not always be expressing a fervent belief that something exists *for a specific purpose* but rather discussing it in relation to something *for which it could be used*. Perhaps, therefore, relational-teleology is a cultural method of describing natural phenomena, a natural pattern of language, similar to the explanations given by high-school students in the study by Zohar & Ginossar (1998). This again would suggest that teleological explanation is a diminished barrier to learning about causality in the natural world. It is not necessarily a specific belief that X exists for a purpose, but rather a style of discourse about X. That is, children participate in discussion around natural phenomena which considers them in relation to an effect which they have, or a purpose for which they can be co-opted. For example, a child may not believe rain is for watering the grass but may use this explanation due to a prior discussion where rain is discussed in relation to it watering the grass. Consequently, it is possible that relational-teleology indicates that some teleological explanations may be due to certain patterns of language whereby children relate natural phenomena to a purpose for which it could be used. Further empirical work exploring children's naturally-occurring discourse is needed to support this argument.

4.4.4 Conclusion

Has there been an overestimation of young children teleological tendencies? Children do display a strong bias towards providing teleological explanations for natural phenomena. However, the results of this study suggest a more complicated picture.

Firstly, children can be influenced by the way in which questions are worded, resulting in an overestimation of children's teleological responses. Secondly, children's teleological explanations are not consistent within the ontological category of natural phenomena, with certain questioned topics being more likely to be explained teleological, or scientifically, than others. Thirdly, sub-categorisation of teleological explanation suggests a more nuanced understanding of teleology. The majority of examples of natural phenomena provoked responses classified as relational-teleology. This finding does not match the commonly used definition of teleology, which implies design-teleology. Consequently, when children espouse these purposeful explanations they may simply be using teleological language as a descriptive tool to explain how a topic could be used (similar to many popular science books, (Hanke, 2004), and educational textbooks, (Zohar & Ginossar, 1998; Talanquer, 2007)). The variation in response type (both scientific vs. teleological and design- vs. relational-teleology) suggests that naïve ideas, children's preconceptions and knowledge about natural phenomena may have a fragmented structure. A fragmented view of knowledge might imply that teleological explanation can be limited using target interventions, focusing on the big idea that nature is causal not purposeful, a concept which can be applied to all natural phenomena. A question raised from these findings is how can we develop methods to limit children's acceptance and use of teleological explanations for the existence of natural kinds?

Confining the study to the category of natural phenomena is a limitation of this study, and further research is needed to confirm if the results are applicable to different ontological categories. However, taken together these findings suggest that teleology may not be the major barrier it has been claimed to be, but the high number of teleological responses for both question wordings and all questioned topics certainly suggests it is still a considerable hurdle.

5. Study 1b: Naturally occurring discourses surrounding natural phenomena

5.1 Introduction

The findings of Study 1 suggest that there are 1) fluctuations in the levels of teleological explanations children provide within the category of natural phenomena; and 2) a tendency for children to provide relational-teleological over design-teleological explanations for the existence of natural phenomena. Regarding the first point, children provided significantly more teleological responses for certain natural phenomena (waterfalls and night) and significantly more scientific responses to others (rainbows and storms). It is theorised that the fluctuation in these levels could, in part, be due to how children discuss these topics in social contexts, as in England and Wales there is no specific formal education about scientific explanations for the existence of natural phenomena in the age range investigated (DfE, 2015; DfE, 2014). Therefore, children preconceptions are likely to be formed by their informal discussions, that is, the natural discourse about natural phenomena which children experience. Natural discourse is defined here as spontaneous dialogue, (e.g., impromptu unstructured discussion between a child and a peer or an adult), not conversations which are predetermined, structured and take place for a specific reason, such as talk between child and teacher in a science lesson.

In relation to the second point, the preference for explanations which display relational- over design-teleology, it was suggested that relational-teleology may be partially caused by the natural discourse, the cultural styles of explanation for natural phenomena. It could be a style of explanation which children have experienced, which subsequently leads them to advocate similar explanations. For example, *there is night so people can go to sleep*, in which night is being used by people for going to sleep (demonstrating relational-teleology). The underpinning argument of these two points, that teleological explanation is partially influenced by language, counters traditional accounts for the existence of teleological explanations (see section 2.4.4). As this inquiry aims to provide further explanation for the results of Study 1 it is not considered a separate study but rather an extension and, therefore, is designated as Study 1b.

This explanation for children's use of teleological explanations having a partially linguistic origin is based upon sociocultural theories of development. This notion was explored in the literature review (section 2.4.4). In summary it was argued that as language and thought are intertwined (Bruner, 1996), that is, the form of language used in certain contexts creates a child's idiosyncratic culture (Hoff, 2006), which can influence how children explain themselves. For example, if children are exposed to discourse that primarily discusses an example of natural phenomena in a relational-teleological manner it may, therefore, result in children providing relational explanations when asked to account for its existence. Furthermore, certain examples of

natural phenomena (e.g., rainbows or waterfalls) may have higher proportions of teleological, or scientific, discussion in natural discourse than other examples, which may then result in children providing higher numbers of teleological, or scientific, explanations to account for their existence.

The problem with the above proposal is that there is little evidence about the nature of children's natural discourse about natural phenomena. It would seem self-evident that children's explanations are influenced by their conversations, particularly parent-child interactions, as children's parents are key in their conceptual development (Athey, 1990), such preconceptions would then manifest in formal schooling, (e.g., in science education, Howe, Devine & Tavares, 2013).

There are few studies looking at naturally occurring teleological discussion, although a specific focus on natural phenomena appears to be limited, one example investigated naturally occurring questions and answers (Kelemen et al., 2005). The study explored the spontaneous questions of US-based Mexican descent-children (3- to 5-years-old) and their parents' responses, in relation to different ontological categories (organisms, natural objects and phenomena, artefacts, and social behaviour). It is the latter part which is of interest: what types of explanation (teleological or scientific) did parents' provide to children's questions about natural kinds. Data was collected by interviewing children's parents, in person and via phone calls, during these interviews parents were asked to report any questions their children asked about *why things happen* and *how things work*, and their responses. In preparation for these interview parents were required to keep diaries of their children's questions and their answers. Training was provided to ensure they understood the task. The results suggested that when parents respond to children's questions about *why things happen* or *how things work* parental responses are causal rather than teleological in nature and the number of overall teleological explanations provided by parents was low. For the category of natural objects (e.g., rocks) and phenomena (e.g., clouds) 28% of explanations were teleological, with 64% being causal whereas for organisms (e.g., horses) the mean percentage of responses were 15% teleological and 71% causal. Kelemen et al. concluded that children's use of teleological explanations is not directly attributable to parental explanations.

The reported study (Kelemen et al., 2005) into parent-child interaction only considers children's questions and their parents' responses. While this is a valuable method for uncovering specific examples of teleological and scientific conversation, teleological discourse around certain natural phenomena may not always result from child initiated questions. The impromptu discourse about natural phenomena could provide different evidence of teleological conversation when compared to children's specific questions. For example, discussion between a parent and child about snow may focus on the snow being used for play (a common teleological explanation in Study 1); however, a parent responding to the question *why is there snow?* is unlikely to respond with *it is for playing in*, posing a problem for using parental diaries as in Kelemen et al. (2005). Therefore, it is argued that investigating patterns of natural discourse around natural phenomena will provide

different evidence than investigating children's questions and the resulting adult responses. Consequently, a method is needed which can explore children's natural discourse around natural phenomena, is generalisable and is not limited to a specific element of children's talk. Corpus linguistics provided this method and was used to investigate the research question arising from the literature review: (RQ3) *does the natural discourse children experience about natural phenomena influence children's explanations for their existence?* This question was supported by two component questions, (CQ1) *what specific examples of teleological and scientific explanations do children experience about natural phenomena?* and (CQ2) *what are the themes associated with natural phenomena in the discussions children experience?*

5.1.1 Corpus linguistics

Section 3.3.3 of the methodology explored the rationale and theory behind the use of corpus linguistics, this section revisits that discussion recapping key terminology and theory relevant to this chapter. *Corpus linguistics* is a method of systematically investigating patterns of language within a collection of natural discourses (Gries, 2006). *Natural discourses* are examples of speech or writing which are not created for a purpose; for example, an impromptu conversation between two individuals would be natural discourse, whereas a prepared speech would not be natural discourse. Corpora are usually explored using a number of *keywords*, in this study the examples of natural phenomena used in Study 1 were used as keywords. Keyword investigation results in the *frequency* of occurrence for the keywords, their concordances and collocates. *Concordances* are the keywords in context, essentially the keyword in the sentence in which it appears in the corpus. Keyword concordances should provide evidence to answer the component question (CQ1) *what specific examples of teleological and scientific explanations do children experience about natural phenomena?*

Collocates are words which appear beside or near a keyword with a higher frequency than they occur elsewhere in the corpus (Stubbs, 1995). For example, *light* could have the collocate of *feather*, that is, the two words appear together more frequently than suggested by chance, perhaps due to the idiom *as light as a feather*. Exploring and categorising collocates can show general patterns of language associated with a keyword; therefore, allowing for the emergence of themes in discourse (Stubbs, 1995). This form of analyses allows the investigation of CQ2, *what are the themes associated with natural phenomena in the discussions children experience?* In relation to teleological discourse, for night a proportion of collocates may be associated with the theme of sleep (a common teleological explanation in Study 1) without night having any overt teleological concordances. Therefore, if the themes emerging from collocate analysis show similarity with the Study 1 responses, it could suggest that the way individuals talk about natural phenomena could influence children's explanations.

5.1.2 Conceptual metaphor

Collocates of significance are usually either a literal reflection of the natural world (e.g., grass having the collocate green, as in green grass) or represent a metaphoric aspect of a concept (e.g., time having the collocate wasting, as in you are wasting my time). These metaphoric collocates can be used to explore possible conceptual metaphors associated with a keyword, or target domain (Deignan, 2005). These metaphors are not linguistic metaphors, in a poetic sense, but rather conceptual metaphors. Conceptual metaphor theory (Lakoff & Johnson, 1980) is the practice of discussing a concept (referred to as a target domain) in reference to another concept (source domain); for example, TIME (target domain) IS MONEY (source domain), as in *you are wasting my time*, *I've invested a lot of time* or *that mistake cost us two days*. Target domains are usually abstract and lack definition, whereas source domains are tangible and consist of concrete examples, therefore, essentially conceptual metaphors can add understandable structure to abstract concepts (Evans & Green, 2006; Kövecses, 2002).

Lakoff & Johnson (1980) argue that conceptual metaphors shape thought and styles of explanation, essentially that naturally occurring discourse can be shaped by conceptual metaphors. Mainly because the vast majority of natural discourse is actually non-literal and metaphoric, as opposed to the common preconception that our spoken communication is literal and non-metaphorical (Gibbs, 1994; Evans & Green, 2006). The use of the term metaphor can at times be misleading, metaphor in everyday usage often refers to the implied comparison of two categories (e.g., *I am a rock*, *fields of gold*, or *you are my sunshine*), these examples are often termed resemblance metaphors and feature in conceptual metaphor theory (Evans & Green, 2006). However, the majority of linguistic research focuses on figurative natural discourse (e.g., *our marriage is on the rocks*, Lakoff & Johnson, 1980, p.45), that is neither poetic nor implying a relationship between two categories, a marriage cannot literally be on the rocks. Exploring these types of non-literal associations, as opposed to the more traditional resemblance metaphors, can reveal conceptual metaphor maps associated with a concept, for example, LOVE IS A JOURNEY (Evans & Green, 2006). Conceptual metaphors often structure explanations when explaining complicated and abstract ideas (e.g., explaining the existence of natural phenomena to young children) and can become conventionalised for certain concepts due to the need for clarity and recognition of familiar patterns of language to aid communication (Deignan, 2005). Therefore, collocates associated with natural phenomena can be investigated for themes which may suggest related conceptual metaphors. These conceptual metaphors may help to explain the use of teleological explanations for natural phenomena, in that the conceptual metaphors may show considerable similarity to children's responses in Study 1.

5.1.3 Study aim

The results of Study 1 show that there is fluctuation in the levels of teleological explanations for natural phenomena and that children show a preference to provide relational- over design-teleological explanations for the existence of natural phenomena. It is theorised that the cultural patterns of language which children experience could influence their explanations for the existence of natural phenomena. Corpus linguistics was used as a method to provide evidence for or against this theory, due to the methods ability to systematically investigate large amounts of natural discourse by exploring not only specific occurrences of keywords (concordances) but also the themes surrounding keywords (collocates). Consequently, this investigation used a corpus-based approach to answer the question (RQ3) *does the natural discourse children experience about natural phenomena influence children's explanations for their existence?*

This question was supported by two component-questions explored in this study that help to answer the main research questions. One such question (CQ1) was *what specific examples of teleological and scientific explanations do children experience about natural phenomena?* Coding concordances for scientific and teleological content allowed for exploration of the forms of explanation children encounter about the existence of natural phenomena. Further coding of teleological concordances into design- and relational-teleology allowed for corpus and Study 1 comparison in relation to the type of teleology children experience.

The other component-question was (CQ2) *what are the themes associated with natural phenomena in the discussions children experience?* Coding collocates into categories (e.g., the category of play in relation to an example of natural phenomena) allowed for comparison of collocate categories with the responses in Study 1. For example, using snow for play related activities was a common relational-teleological explanation in Study 1, therefore, if children are influenced by natural discourse, play related words should be significant collocates of snow. Furthermore, categorising collocates allowed for inference of conceptual metaphors for natural phenomena (see 5.1.2). If the themes emerging from collocates and the derived conceptual metaphors share commonalities with the responses in Study 1, it would support the suggestion that natural discourse influences children's use of teleological explanations.

5.2 Method

The following discusses how the corpus was created, the procedure used to search the corpus and the measures used to assess keyword frequency, concordances and collocates. To recap keywords are the words or phrases under investigation in a corpus analysis. Concordances are the keywords in context, essentially the sentence in which the keyword appears. Collocates are words which appear near a keyword with significantly higher frequency then they appear elsewhere in the corpus. Furthermore, in the remainder of this chapter the phrase *keyword/s* is used to refer to the occurrences of example/s of natural phenomena within the corpus. *Tokens* indicates the total number of words

in the corpus or in a specified section of the corpus (e.g., collocates for day). *Types* indicates the number of word types in the corpus or specified section. Tokens and Types are at times standardised to allow for comparison of frequencies between different keywords of natural phenomena.

5.2.1 Sample: building the corpus

Only children's spoken discourse was investigated in this study, children's experiences of spoken discourse, as opposed to written discourse, were investigated for two related reasons. Firstly, spoken and written language are dramatically different in terms of their purpose and structure (Wilkinson, 1971), with the transient nature of spoken language resulting in it being tailored to a child's knowledge and in situ reactions (Schallert, Kleiman & Rubin, 1977). Therefore, if written discourse were explored in may need to be a separate or comparative investigation. (*The Oxford Children's Corpus* would provide a suitable corpus of texts written for and by children, see Wild, Kilgarriff & Tugwell, 2013). Secondly, written language can not be examples of natural discourse. They may have been created for a specific purpose and, consequently, may bias the corpus; for example, non-fiction books which explain the weather.

It is acknowledged that an increasing proportion of the spoken language children experience will come from technology and the media, for example, television and video games, and these experiences will influence children's development to some extent (Smith, Cowie & Blade, 2015). However, this type of spoken discourse is not included in the corpus because a) transcribed examples of television programmes or video games are limited and copyrighted, and b) these programmes or games will have a purpose behind the dialogue so are not natural discourse.

As there were no suitable current corpora of children's spoken discourse one had to be created for this study. Three online databases were searched for suitable transcripts: *CHILDES*, *Talkbank* and the *UK Data service*. These databases were selected as they provided transcribed examples of children's talk with other children and adults. To be included in the children's corpus the data had to be a) spoken discourse from children aged between 3- and 10-years-old (the general age range for children advocating teleological explanations based on the literature review), and b) naturally occurring spoken language (not from interviews or laboratory studies). The decision was made to expand the age range from the one investigated in Study 1 (5- to 8-years-old), as children's early experiences significantly influence their knowledge and explanations (Curtis & Pettigrew, 2009; DfE, 2016; MacNaughton, 2008). This expanded age-range also had the advantage of including a larger number of transcripts, which resulted in a larger corpus of children's spoken English and, therefore, a more representative study (McEnery & Wilson, 2001).

There were 58 datasets which met the criteria, 56 from *CHILDES* (MacWhinney, 2000) and 2 from the *UK Data service* (ESRC, 2016). The datasets span the period of 1962 to 2001, although the majority are from the 1990's, and are either British- (59.9% of tokens) or American-English (40.1% of tokens). Combining both British- and American-English sources had the benefit

of creating a larger corpus. While there are acknowledged differences between British- and American-English, for example, spelling of certain words or grammatical differences related to specific word usage (see BYU, 2017), there is little to suggest that the discourse around natural phenomena should differ in English speaking countries (e.g., America, New Zealand, United Kingdom, etc.). Although if this study had investigated discussion about the existence of organisms and their traits then a difference may have been expected between counties due to religious factors (see Evans, 2000; but also Kelemen 2003, section 2.4.2). Combined, these datasets form a 7,941,053 word (types = 38017) corpus of children's spoken English. However, the majority of transcripts do contain information regarding the speaker (e.g., *CHI:* or *MOT:*). This contextual information will have increased the word count of the corpus, therefore, the total word count is not completely accurate. It is argued that this is not a problem, as a) it adds contextual information to allow for inferences about the corpus data; b) this is an exploratory study and so slight inaccuracies regarding word counts is tolerable; and c) not a corpus comparison where Std. Freq. would need to be accurate, in this corpus all Std. Freq. are affected by the same additional information.

5.2.2 Procedure: Searching the corpus

The corpus was searched for occurrences of the ten natural phenomena used in Study 1, darkness, day, light, night, rain, rainbows, snow, storms, waterfalls and waves. For each of these keywords their frequencies in the corpus were determined, following this their concordances were investigated and coded, lastly the keyword's collocates were categorised. Following advice from Stubbs (1995) collocates were searched for in a span of five words to the left and right of the keyword. The measures for concordance and collocate coding are outlined below.

Keyword occurrences were searched for using the keywords in their singular and plural form, for example, *night* was searched for using 'night' and 'nights'. Furthermore, keyword lemmas were included in the search terms, for example, the lemmas ‘nighttime’ and ‘nightime’ were included for the keyword *night*. Search terms were case insensitive. As some of the data in the corpus included irrelevant information for this study, for example information regarding the speaker (e.g., *CHI:* to indicate a child speaking), these instances were omitted from the collocate results. The software used to search and analyse the corpus was *AntConc* (Anthony, 2014).

5.2.3 Measures

5.2.3.1 Keyword frequencies

The frequencies of each keyword are provided along with their standardised frequencies, the percentage occurrence in the corpus. This information is presented to summarise the corpus and to provide context to the following analyses, as it situates the concordance investigation within the corpus showing the occurrences of keywords which were not analysed, (i.e., those which were not teleological or scientific).

5.2.3.2 Keywords concordances

As this study was interested in specific examples of teleological and scientific discourse (see CQ1), concordances were only coded into these two categories; the remainder of the concordances were ungraded. The rubric was the same as the one used for Study 1 (Table 5.1, Level 1). Frequencies of teleological and scientific talk in relation to the keywords overall number of concordances are provided, along with the ratio of teleological to scientific discourse and relevant examples. Firstly, the overall split of scientific to teleological concordances is provided. Secondly, this data is split by keyword as Study 1 showed considerable influence of questioned topic upon explanatory type. Thirdly, teleological concordances were further coded to determine if they suggested design- or relational-teleology (Table 5.1, Level 2), which again was explored amalgamated and split by keywords. 10% of the keyword concordances (958) were coded by a second coder. Cohen Kappa indicated a high level of agreement ($K = .81$, $p < .001$ for Level 1 coding and $K = .89$, $p < .001$ for Level 2 coding).

Table 5.1: *Rubric for coding keyword concordances*

Level 1		
Grade	Teleological	Scientific
Explanation	The topic is given a purpose, a function or a goal, which aids or supports either itself or another entity.	A scaled down, causal explanation for a topic, answer does not contain teleological notions. It may not be scientifically correct.
Examples	‘There are waterfalls so wild animals have a place to drink’ or 'there is night to tell us to go to sleep’	'It snows because the clouds get cold and all smushed up together' or 'there is night because the earth is spinning around and when the bit we are on is dark then we have night time
Level 2: sub-classification of teleological responses		
Grade	Design-teleology	Relational-teleology
Explanation	A topic is explained in relation to a purpose or goal for which it was designed. The topic has the singular purpose of aiding/supporting something, the something can be itself or another entity; it acts upon an entity to produce a result which is not a scientifically valid.	A topic is explained in relation to a purpose for which it can be used or an effect which the topic causes. The purpose or effect is human-centric.
Examples	‘Rain is for watering the plants' or 'day is for waking up the birds'	‘Snow is used by children for making snowmen' or 'night is a time period which people usually use for going to sleep'

5.2.3.3 Keyword collocates

Collocates were explored to investigate the general patterns of talk around the keywords (see CQ2), to determine if certain keywords are associated with teleological themes; for example, night may associated with the theme of sleep through the collocates of *sleeping, bed, tired* and *pillow*. The Mutual Information statistic (MI) was used to assess the strength of the keywords’ collocates. As MI is not a significance test but a measure of association, to be considered linguistically interesting collocates required a freq. ≥ 3 and MI ≥ 3 (Stubbs, 1995; McEnery, 2006). In line with common practice (e.g., McEnery 2006; Sardinha, 2006) all function words (words with minimal lexical meaning which are used to structure sentences) returned as collocates were removed and are not

discussed. This is due to function words providing little information for the purpose of this study. Examples of function words include determiners (*the, that*), conjunctions (*and, but*) pronouns, (*I, she*), quantifiers (*some, neither*), prepositions (*behind, on*) and adverbs (*then, why*). Furthermore, some collocates were mispronunciations (e.g., *alwight* for *alright*), in these instances if the correct spelling of the word is a collocate the mispronunciation was treated as a lemma, if the correct spelling was not a collocate the mispronunciation are not reported. This is because incorrectly spelt words which only occur near a keyword will always achieve a significant collocation statistic (i.e., MI ≥ 3) because they only appear near that keyword (Stubbs, 1995), the only way to ensure that word is a true collocate is if the correct spelling is also returned as a significant collocate. There was a considerable decrease in the remaining collocates after function words are removed, but that is not suppressing when 'function words account for less than one-tenth of 1 percent of your vocabulary but make up almost 60 percent of the words you use' (Pennebaker, 2011, p.1).

To investigate the themes emerging from the collocates, for each keyword the resulting collocates were grouped into categories. Collocate categories were inductively created by grouping approximately 25% of the data then applying these categories to the remaining collocates. To ensure collocates were correctly categorised their concordances were taken into account, where relevant information pertaining to collocates concordances is provided. 20% of the keyword collocates (189) were coded by a second coder in to the created collocate categories for each keyword, Cohen's kappa indicated a good level of agreement (K = .786, $p < .001$). Reported for each collocate category are tokens (the number of occurrences of all collocates within that category), types (the number of unique collocates within the category) and percentage of tokens and types for each category within a keyword. The percentage statistics allowed for consideration of the size of a collocate category, to check for dominance of a certain collocate category for a given keyword.

5.3 Findings

This chapter deviates from the structure used in Studies 1, 2 and 3, to report the results. Rather than the traditional separate results and discussion sections, the chapter combines both these sections under the heading 'findings'. A combined approach is used as a presentation of solely statistical results from corpus analyses can be difficult to interpret. Therefore, the combination of statistical results with a discussion about the interpretation of results should provide an easier to interpret chapter. Consequently, this section is split into four main areas. Firstly, a report on the keyword frequencies to provide contextual information about the composition of the corpus. Secondly, details of the concordance investigation to answer the question (CQ1) *what specific examples of teleological and scientific explanations do children experience about natural phenomena?* Thirdly, presentation of collocate categories to show emerging themes. Fourthly, use of collocate themes to propose conceptual metaphors. The third and fourth area together provide evidence to answer the

question (CQ2) *what are the themes associated with natural phenomena in the discussions children experience?*. This chapter concludes with a final summary of how these findings contribute to the question (RQ3) *does the natural discourse children experience about natural phenomena influence children's explanations for their existence?*

5.3.1 Keyword frequencies

After beginning the analysis it was apparent that the keywords of darkness and light were particularly problematic, the concordances for these keywords did not relate to natural phenomena. Rather they referred to the descriptive words *dark* and *light.* The word *darkness* did not even feature in the corpus. Additionally, when not used in a description, *light* referred to artificial phenomena, specifically, light bulbs. Consequently, these keywords are removed from any further investigation, as analysis of their concordances and collocates would yield no information for the present study. Table 5.2 shows the frequency of the keywords in the children's corpus, along with a standardised frequency: the percentage of occurrence in the corpus. This information is provided to detail the composition of the corpus and to allow for comparison with future work.

Table 5.2: *Frequencies of the keywords in the corpora*

Keyword	Freq.	Standardised Freq.
Day	3858	48.58% $x10^{-5}$
Night	2836	35.17% $x10^{-5}$
Rain	1205	15.17% $x10^{-5}$
Rainbows	210	2.64% $x10^{-5}$
Snow	1156	14.56% $x10^{-5}$
Storms	113	1.42% $x10^{-5}$
Waterfalls	31	0.39% $x10^{-5}$
Waves	171	2.15% $x10^{-5}$

5.3.2 Keyword concordances: explicit teleological and scientific explanations

Coding of the concordances, the sentences or phrases in which keywords appeared, resulted in many concordances which were not actually discussions about the keyword, but rather dialogues which refer to the keyword as a marker of time. For example, *remember when it snowed, remember the time there was a big storm* or *in two days' time*. A large number of concordances also were simple comments on the state of the weather. For example, *is it still raining?, it's meant to snow soon* or *look there's a rainbow.* There were, however, some concordance which included teleological or scientific discussion. The raw (number of concordances) and standardised frequencies (raw frequency as percentage of total keyword concordances) of scientific and teleological discourse are displayed in Table 5.3.

Table 5.3: *Occurrences of teleological and scientific discourse in the corpus by keyword*

	Teleological concordances		Scientific concordances	
Keyword	Freq.	Std. Freq	Freq.	Std. Freq
Day	5	0.69%	4	0.55%
Night	32	2.40%	13	0.98%
Rain	6	0.21%	7	0.25%
Rainbows	-	-	4	0.33%
Waterfalls	-	-	1	0.88%

The frequencies of the concordances show that in the corpus there are low levels of teleological and scientific discussion around the keywords. Although the mean occurrences of teleological and scientific concordances shows a dominance for teleological over scientific talk, at 59.7% (M = 8.6 concordances) to 40.3% (M = 5.8 concordances), respectively. This mean concordance split is similar to the split found in Study 1 for the open question wording (62.3% to 37.6%).

Unfortunately, the overall finding that there were limited concordances which were specific examples of teleological or scientific explanations makes any comparison with children's responses in Study 1 difficult. Therefore, this section merely details the concordances found in order to still answer the question (CQ1) *what specific examples of teleological and scientific explanations do children experience about natural phenomena?* Although there will be implications for the final discussion and the answering of the main question (RQ3) *does the natural discourse children experience about natural phenomena influence children's explanations for their existence?*

To provide context to this concordance investigation, a selection of scientific and teleological concordances are provided. Examples of scientific concordances include:

- For Day – '[Child] when the sun comes up it will be tomorrow'; '[Adult] the sun comes out during the day'; '[Child] yesterday it was day when the sun wasn't down'.
- For Night – '[Father] night is when the sun goes down, the sun came up this morning'; '[Child] hey it is night. [Mother] how can you tell? [Child] because look outside it's starting to get dark'; '[Mother] they found out why it's dark on one side of Earth and it's light on the other … because when it's night time Earth's in a big shadow'.
- For Rain – '[Child] it was so rainy because clouds were raining ... the clouds make it rain'; [Father] how does rain get made Ross? [Child] there's water down here and well it goes up into the air and … once it goes all the way up to the clouds they get foggy .. and then they gotta wash out and … drops of water must come out'; '[Mother] rain comes out of clouds'.
- For Rainbow – '[Child] if there's sun and rain then a rainbow comes out', '[Child] There might be a rainbow now because if it rains and the sun comes out sometimes there's a rainbow'; '[Mother] oh the sun's come out now, there might be a rainbow'.
- For Waterfalls – '[Mother] that's a waterfall ... the water goes from the top of the rocks down to the bottom'.

Examples of teleological talk include:

- For Day – '[Mother] you just go to preschool during the day and mommy goes to work and daddy goes to work'; '[Mother] you have to go to school every day'; '[Child] do we have to wake up now? [Mother] yeah cause it's daytime isn't it'.
- For Night – '[Child] everybody sleeps at night time? [Mother] uhhuh. [Child] why? [Mother] so they can rest their bodies'; '[Adult] everybody has to go to sleep for the night don't they [child] Yup'; '[Mother] it's bedtime … because it is nighttime'; '[Child] night is when you go to bed'.
- For Rain – '[Child] god let it rain'; '[Child] god made the rain', '[Mother] the plants they all need rain so it's nice when it rains because the plants and the trees like rain, they can't grow without rain'.

Table 5.4: *Number of design- vs. relational- teleological concordances within the corpora*

Keyword	Freq. design-teleology	Freq. relational-teleology
Day	0	5
Night	6	26
Rain	6	0

Teleological concordances were also coded for design- vs. relational-teleology, Table 5.4 shows the frequency of relational- and design-teleological concordances. The keywords of rainbows and waterfalls were removed as they had no teleological concordances. The majority of teleological discourse around the keywords of day and night was relational-teleology. Comparing the split of design- to relational-teleological concordances found in the corpus (25.6% : 74.4%) with the overall split found in Study 1 (35.0% : 65.0%), shows considerable similarity. Although, the limited keywords (3 out of 8) and the overall limited teleological concordances for day and rain make direct comparisons with Study 1 problematic, especially when considering the variation in explanatory type influenced by questioned topic (see section 4.4.3).

Examples of relational-teleological discussion include: for day, '[Child] it's sunny time, it's day time, it's time to get up' or '[Father] we go to work everyday'; and for Night, '[Mother] children have to go to sleep … because it is night time' or '[Adult] it is night time now … you have to go to sleep'. Conversely for Rain, all of the discourse was suggestive of design-teleology. For example, '[Child] god made the rain' or '[Mother] the plants they all need rain'. Night also had a small number of teleological concordances which demonstrate design-teleology. For example, '[Adult] you've got something else that needs to come out at night time … stars' or '[Father] moles come out at night … and mouses … they all have to come out at night'.

Overall there were low occurrences of specific teleological and scientific concordances for the keywords in the corpus, and consequently, limited number of design- and relational-teleological concordances. Only three keywords (day, night and rain) had concordances which outright demonstrated teleological discourse and five keywords (day, night, rain, rainbow and waterfalls) had concordances which discussed scientific explanations for their existence. Therefore, although limited, this suggests children do experience some explicit teleological and scientific explanations around at least some examples of natural phenomena, via their peers, parents or other adults. This finding influences the interpretation of the results in Study 1: it helps to account for the levels of scientific explanation provided by the participants. It provides further evidence that children may hear scientific explanations outside of formal education, as specific knowledge about the causes of particular natural phenomena is not a requirement of the National Curriculum (DfE, 2015) or Early Years Foundation Stage (DfE, 2014).

The levels of scientific and teleological discourse in the corpus differ to the levels of scientific and teleological explanations found in Kelemen et al. (2005). The findings of that study suggested that 28% of explanations to children's why questions about natural objects and natural phenomena are teleological and 64% are scientific, which would suggest that this corpus investigation would have a higher number of scientific concordances than teleological ones. However, this was not echoed in the keyword concordances, few of which were scientific and the overall pattern was actually reversed (with approximately 60% teleological to 40% scientific). An explanation for the discrepancy is the forms of the data. The corpus data is naturally occurring talk, whereas that manuscript explored parents' responses to children's spontaneous why questions. Although the aims of these two studies were different, it is proposed that corpus linguistics provides a more holistic view of the discourse children experience around natural phenomena, due to the methods ability to explore vast quantities of naturally occurring language (Gries, 2006) which is representative of the discourse in the target population (McEnery, Xiao & Tono, 2006).

The concordance analysis sought to answer the question (CQ1) *what specific examples of teleological and scientific explanations do children experience about natural phenomena?* Overall both the split of teleological vs. scientific concordances (59.7% : 40.3%) and the design- vs. relational-teleological concordances (25.6% : 74.4%) were extremely similar to the split of children's responses in Study 1 (62.3% : 37.6%; 35.0% : 65.0%). Possibility indicating that children's preconceptions are formed, in part, by their informal discussions around natural phenomena. However, such a claim would be tentative. The concordance analysis found a low number of specific examples of teleological and scientific discussion about natural phenomena. Furthermore, the exclusion of the majority of examples of natural phenomena originally used for keywords is problematic. Study 1 found a considerable influence of questioned topic, and consequently, an argument based on evidence from a limited number of keywords is untenable. If more teleological concordances could be analysed from other corpuses perhaps the variation for other examples of natural phenomena could be explored. Therefore, it is proposed that the concordance investigation provides suggestive evidence that children experience teleological and scientific explanation about natural phenomena but the current evidence is insufficient to argue for a strong connection between the two.

5.3.3 Keyword collocates: themes in discourse around natural phenomena

The following summarises the keywords' collocates, that is, the words which appear significantly more frequently near the keywords than elsewhere in the corpus. Table 5.5 details the *raw collocates* (those with frequency ≥ 3 and MI ≥ 3) and the *refined collocates* (those remaining after removal of function words). The refined collocates were grouped into categories to investigate emerging themes surrounding the keywords. For each keyword the total number of collocates

(Tokens), number of original collocates (Types) and relevant examples are provided. As discussed darkness and light are not investigated as the returned concordances and collocates do not refer to the actual natural phenomena under investigation. Furthermore, as all collocates for the keyword *waterfalls* came from one transcript, it is unrepresentative of natural discourse, subsequently, this keyword is removed from the collocate analysis.

Table 5.5: *Number of raw and refined collocates by keyword*

	Raw collocates		Refined collocates	
Keyword	Types	Tokens	Types	Tokens
Day	2606	38052	234	4871
Night	517	19934	265	5766
Rain	365	100811	121	1456
Rainbows	111	1591	28	215
Snow	349	9864	143	1505
Storms	75	754	10	57
Waterfalls	25	175	2	12
Waves	114	1320	14	93

The categorisation of collocates resulted in nine themes that were used across the eight keywords: activity, description, entertainment, greeting, location, names, natural phenomena, posturing and time period. However, a number of these categories provide little information to analyse themes in discussion around natural phenomena or to propose conceptual metaphors related to teleological explanations. Consequently, the category of *description* was removed from the following analyses, as descriptive terms for natural phenomena provide little information for the purposes of this investigation. *Entertainment* was removed as entertainment collocates referred to books, music or televisual media in which natural phenomena were discussed. As these are not examples of natural discourse, rather scripted narratives, they are not appropriate for the current investigation. The category *names* was removed as collocates detailing children's names offer little information regarding the content of the discussion, as it would be anticipated that a corpus of discourse involving children would include a wide selection of names. Finally, the category *natural phenomena*, where the keyword collocated with itself or another keyword, provides little information other than the notion that examples of natural phenomena are rarely discussed separately. It is to be expected that keywords will collocate with other examples of natural phenomena as they often may not occur in isolation (e.g., rain happens at night, in the day, before and during rainbows, and forms part of storms). This removal of uninformative collocate categories

has resulted in five collocate themes, which are activity, greeting, location, posturing and time period. Full details of all keywords' collocates categories are provided in Appendix 2.

Where appropriate some collocate categories were sub-categorised, for example *activity* had the subcategory of *play* (for snow) and *sleep* (for night). Subcategories were used where a group of collocates' types and tokens constituted a large proportion of a parent category but did not qualify as a separate category. These sub-categories were used to indicate pervasive types of collocates, for example the subcategory *sleep* contained over half the *activity* tokens for night. Table 5.6 – Table 5.12 show the collocate categories for each keyword, with the number of tokens and types for each collocate category, along with the percentages of tokens and types each category represents. Also included in the tables are a number of collocate examples. Where sub-categories are outlined, the percentages relate to the total number of collocate tokens and types, not to the number of collocate tokens and types of the parent category.

Table 5.6: *Categorised collocates for day (Tokens = 3438, Types = 163)*

Category	Sub cat	Tokens	Tokens as % of total	Types	Tokens as % of total	Examples
Activity		1018	29.61%	67	41.10%	*Work*, *School* and *Tea*
	Play	104	3.03%	12	7.36%	*Toys*, *Kite* and *Cricket.*
Location		116	3.37%	12	7.36%	*Home*, *Nursery* and *Beach.*
Time period		2304	67.02%	84	51.15%	*Tomorrow*, *valentines* [day] and *Sunday*.

Table 5.7: *Categorised collocates for night (Tokens = 2435, Types = 141)*

Category	Sub cat	Tokens	Tokens as % of total	Types	Tokens as % of total	Examples
Activity		232	9.53%	86	60.99%	*Bath, Working* and *Dinner*
	Play	49	2.01%	8	5.67%	*Playing, Cowboys* and *Game*
	Sleep	681	27.97%	39	27.66%	*Bed*, *Dream* and *Sleep*
Greeting		590	24.23%	10	7.09%	*Bye, Good* [night] and *Shh*
Location		81	3.33%	18	12.77%	*Dublin*, *Home* and *Stay*
Time period		1532	62.92%	27	19.15%	*Last* [night], *Friday* and *Christmas*

Table 5.8: *Categorised collocates for rain (Tokens = 44, Types = 51)*

Category	Sub cat	Tokens	Tokens as % of total	Types	Tokens as % of total	Examples
Activity		135	30.41%	21	41.18%	*Umbrella, Indoors* and *Plants*
	Play	29	6.53%	4	7.84%	*Play, Pretend* and *Happy*
Location		81	18.24%	7	13.73%	*Outside, Inside* and *Park*
Posturing		58	13.06%	5	9.80%	*Think, Might* and *Hop*
Time period		170	38.29%	18	35.29%	*Afternoon, Summer* and *Night*

Table 5.9: *Categorised collocates for rainbows (Tokens = 34, Types = 6)*

Category	Tokens	Tokens as % of total	Types	Tokens as % of total	Examples
Activity	34	100%	6	100%	*Making, Painting* and *Finished*

Table 5.10: *Categorised collocates for snow (Tokens = 593, Types = 76)*

Category	Sub cat	Tokens	Tokens as % of total	Types	Tokens as % of total	Examples
Activity		334	56.32%	42	55.26%	*Suit, Boots* and *Walking*
	Play	208	35.08%	26	34.21%	*Snowman, Snowballs* and *Sledge*
Location		58	9.78%	7	9.21%	*Outside, Newcastle* and *Everywhere*
Posturing		93	15.68%	11	14.47%	*Think, Wonder* and *Guess*
Time period		108	18.21%	16	21.05%	*Winter, Christmas* and *Remember*

Table 5.11: *Categorised collocates for storms (Tokens = 14, Types = 4)*

Category	Tokens	Tokens as % of total	Types	Tokens as % of total	Examples
Time period	14	100%	4	100%	*Coming, Night, Nine* [O'Clock]

Table 5.12: *Categorised collocates for waves (Tokens = 29, Types = 3)*

Category	Tokens	Tokens as % of total	Types	Tokens as % of total	Examples
Activity	29	100%	3	100%	*Jumping, Over* and *Wave*

The keywords' collocates compared to the concordances offers a different picture of the interplay between natural language and teleology. To explore this proposal collocate categories are examined. Collocates are words which are significantly more likely to appear next to a keyword than other words in the corpus, and consequently, can be used to categorise and infer patterns of talk around a certain keyword or phrase. The following details collocate categories to assess common themes shared by different natural phenomena. This analysis is conducted to answer the question (CQ2) *what are the themes associated with natural phenomena in the discussions children experience?*

A common collocate category was *location* shared by four keywords (day, night, rain, snow). While it could be suggested that these *location* collocates are simply describing the location of natural phenomena, the category is subtly different in that *location* collocates refer to memories

of natural phenomena. For example, *remember when it snowed when we went to Newcastle* or *we're going to spend the night in Dublin*. When the relevant concordances are investigated, these *location* collocates often appear in situations where the keyword is being used to distinguish or remember a situation, that is, natural phenomena are used as a marker to indicate a specific event in the past or an event yet to occur. The collocate category of *posturing*, shared by two keywords (rain and snow), supports this notion of using natural phenomena as a marker. For this category collocates were in concordances where individuals were posturing on the possible occurrence of a keyword (e.g., *do you think it will rain while we're on holiday* or *do you think it will snow this Christmas*). These *posturing* collocates suggest discourse around some examples of natural phenomena is often related to a future event intertwined with an example of natural phenomena. These two collocate categories share similarities with certain Study 1 explanations which discussed natural phenomena with regard to specific events, such as day being a time to go to school, snow being related to Christmas and rain being a time for wet play time.

The common collocate category *time period* indicates a trend for natural phenomena to be involved in conversations marking events or action in relation to the keywords. There were six keywords which had *time period* as a common category (day, night, rain, snow, storms and waves). For the keywords day, night, storms and waves this was a major collocate category representing a large proportion of the keywords' collocate tokens and types. This dominant shared category is a clear indicator of discourse surrounding natural phenomena discussing the keywords in relation to a specific event, whether past, present or future. It could be inferred from this that certain examples of natural phenomena are used as markers of time, to assist in the recall of memories (e.g., *remember the time is snowed a few Christmases ago*) or to structure events which have yet to happen (e.g., *there is going to be a storm on Friday afternoon you will need to take an umbrella when you go out*). Using day and night to mark periods of time is to be expected as they are actual time periods, in these situations the discourse is describing a time of (e.g., *it's on Monday night, a day in April*) or type of (e.g., *Bonfire night*, *Mother's day*) day or night. However, this is unusual for the other natural phenomena, as there are not time periods which are named based on other natural phenomena (an exception may be *April showers* referring to the occurrence of spring rain in England). Rather what is common in the corpus is for the natural phenomena to measure a time period or to be used as an indicator of time to measure the duration of the natural phenomena. For example, *I had to stay inside for the length of the storm* or *the storm lasted for three hours*. Again, this discourse associating natural phenomena with time-describing collocates is argued to indicate that natural phenomena are used as markers of past and future events or action.

The final shared collocate category is *activity*, shared by six of the keywords (day, night, rain, rainbows, snow, waves) and being a major category for day, night, rain, snow, and waves. The high number of collocate tokens and types and the prevalence of this category across the majority of the keywords suggests it to be a dominant category in terms of representing the discourse children

experience around natural phenomena. That is, natural phenomena are often discussed in situations where there is also reference to an activity or task. For example, linking day with going to school, night with going to sleep or snow with making snowmen.

This significant co-occurrence of natural phenomena and related activities in the same dialogue may help to account for children's teleological explanations in Study 1 and more generally. A proportion of children's teleological explanations could be because when lacking an actual explanation they refer to previous conversations around said phenomena which discussed it with regard to a certain activity. The previous examples, depict activities and phenomena which were linked by children in their teleological explanations in Study 1 (e.g., linking day and school, night and sleep, and snow and snowmen, see Study 1, section 4.4.2 for further discussion). Regarding the difference between levels of design-teleology and relational-teleology, the majority of activity collocates are related to activities in which humans are using a natural phenomenon to perform an action (e.g., using snow to make snowmen) or the natural phenomena is causing an individual to perform an action (e.g., putting a coat hood up due to rain). These activities are those which a relational-teleological explanation would refer to, they would not be classified as design-teleology. Therefore, it is argued that these activity related conversations indicate a relationship between the natural discourse associating natural phenomena with human related activities and the dominance of relational-teleology in Study 1. This is further supported by two subcategories which are worthy of note, as they pertain to the relational-teleological explanations in Study 1.

The first being *play*, this was a subcategory of *activity* shared by five keywords (day, night, rain, snow, and waves). This collocate subcategory indicates a large proportion of talk around these keywords is related to play-based activates, and therefore, is associating the natural phenomena with a specific act of play. This could provide evidence for natural discourse influencing children's relational-teleological explanations in Study 1. A clear example would be the large number of play related terms that collocated with Snow, which mirrors the large number of responses in Study 1 suggesting that Snow is used for play. This is particularly apparent in regard to snowball fights and building snowmen which were both a common explanation in Study 1 and a common play activity in the corpus collocates.

The second subcategory was *sleep*, this was only present in one keywords' collocates: night. While this is not a shared category between keywords it is commented upon because of the strong link it provides between natural discourse and children's teleological explanations. The vast majority of teleological explanations for night in Study 1 referred to night being a time when people go to sleep, with a small number claiming night was for nocturnal animals. A large number of sleep related collocates (27.97%) were also found in the corpus for the keyword of night. For example, *sleeping*, *bed* and *dreaming*. The large proportion of sleep related collocates for night and the high number of sleep-based teleological explanations in Study 1, it is argued, indicate a link between

how night is discussed in children's everyday conversation and their teleological explanations for night.

Considering the collocate categories together there are a number of conclusions which could be drawn as to the themes in natural discourse surrounding natural phenomena. A teleologically related discourse is suggested by collocate categories including *location* and *posturing* with their linking of natural phenomena and past or future events. *Time period* supports this assertion, suggesting natural phenomena are often used to mark or deliminate times of events. Perhaps due to the very nature of natural phenomena, being time-constrained events, they are ideal folk-markers of time and duration. Finally, the category *activity* (including the subcategories *play* and *sleep*) has clear teleological links, specifically to the relational-teleological explanations in Study 1. Consequently, in relation to the question (CQ2) *what are the themes associated with natural phenomena in the discussions children experience?*, it is proposed that the discourse children experience around natural phenomena often has the theme related to a certain activity, or a specific time period. These themes in discourse around natural phenomena are further explored in the following section.

5.3.4 Conceptual metaphor

As discussed collocate exploration is a method which may indicate conceptual metaphors. Conceptual metaphors are patterns of speech which sculpt talk and explanation about certain concepts. For example, when discussing time it is often metaphorically treated as money, hence the conceptual metaphor TIME IS MONEY. There currently exists no conceptual metaphors related to natural phenomena, the existing master list (see Lakoff, Espenson & Schwartz, 1991) has a backlog lasting years in terms of uploading agreed upon conceptual metaphors; therefore, some may exist but have not yet been confirmed. Using the collocate categories for the keywords three conceptual metaphors are proposed: TIME AS NATURAL PHENOMENA, NATURAL PHENOMENA ARE ACTIVITIES and NATURAL PHENOMENA ARE A RESOURCE IN PLAY. These are not considered the only possible options but are simply three which are apparent from this corpus investigation. These conceptual metaphors are used as evidence to answer the question (CQ2) *what are the themes associated with natural phenomena in the discussions children experience?*

TIME AS NATURAL PHENOMENA

The first proposed conceptual metaphor is TIME AS NATURAL PHENOMENA. This metaphor would entail discussion surrounding natural phenomena use certain examples of natural phenomena as a marker for events. These events or actions could take place in the past, present or future. In these situations a specific natural phenomenon would be mentioned in order to assist in the recollection of an event (e.g., remember that Christmas when it snowed a lot), to indicate a length of time (e.g.,

we have to stay in the classroom until it stops raining), to signal a specific time or event (e.g., Father's day, Burns night), or delimitate an action (e.g., you need to sleep at night). As discussed above, the support for TIME AS NATURAL PHENOMENA is evident in the collocate categories of *time period, locations,* and the *activity* subcategory of *sleep*. In situations where this metaphor is used, individuals are employing their concrete experience of natural phenomena to understand or think about the more abstract notion of time.

This conceptual metaphor would mean that a proportion of teleological explanations for natural phenomena would associate them with specific events or time periods. Examples of these types of teleological explanations in Study 1 would include linking, for example, day and playtime, snow and Christmas, night and sleep. This claim is supported by the results in of the corpus investigation and children's responses in Study 1. For certain natural phenomena (day, night, snow, storms) there were a high proportion of teleological explanations that also referred to time periods in Study 1, this mirrors a high proportion of the corresponding keywords' collocates in the corpus investigation.

NATURAL PHENOMENA ARE ACTIVITIES

The second proposed conceptual metaphor is NATURAL PHENOMENA ARE ACTIVITIES. This metaphor would suggest that discourse around natural phenomena would refer to an example of natural phenomena and a specific activity. This would involve the phenomena being used to perform an activity (e.g., children using waves to push them along), saving individuals' time or effort to perform a task (e.g., the rain will water the garden), or having an influence on an individual resulting in them performing an action (e.g., needing to put on waterproofs to go out in the storm, or it being night time so children have to go to sleep). As outlined above, support for NATURAL PHENOMENA ARE ACTIVITIES is indicated in the collocate categories of *activity*, including the subcategories of *sleep.* In situations where this metaphor is used, individuals are relating their experiences of specific activities to the occurrence of natural phenomena, they are using their concrete experiences to explain the existence of the more abstract natural phenomena.

This conceptual metaphor would mean that a proportion of teleological explanations for natural phenomena would associate them with a range of activities, primarily those which are human-centric and, therefore, would be classified as relational-teleology. Examples of these types of teleological explanations in Study 1 would include night as a time for children to go to sleep, rain means children need to wear waterproofs or use an umbrella, or storms are for washing the cars. This claim is supported by the results of the corpus investigation and children's responses in Study 1. For certain natural phenomena (day, night, snow and waves) there were a high proportion of teleological explanations in Study 1 which also referred to activities that children perform with, or are due to, the example of natural phenomena, this is reflected by a high proportion of the corresponding keywords' collocates types in the corpus investigation.

NATURAL PHENOMENA ARE A RESOURCE IN PLAY

The third proposed conceptual metaphor is NATURAL PHENOMENA ARE A RESOURCE IN PLAY, this is very closely related the second metaphor NATURAL PHENOMENA ARE ACTIVITIES, although it contains and extra source category. This metaphor would suggest that discourse around natural phenomena would refer to a specific play activity and how the natural phenomenon is used in play. This would involve it being used by children to facilitate a game, (e.g., children using snow to have a snowball fight), or undertake a play related activity (e.g., rain being used by children to make mud pies). Support for NATURAL PHENOMENA ARE A RESOURCE IN PLAY is indicated in the sub-collocate categories of *play* within *activity*. In situations where this metaphor is used to explain natural phenomena individuals would relate their experiences of play to the occurrence of natural phenomena. That is, they would be using their concrete experiences of incorporating a natural phenomenon into their play to explain the existence of the more abstract natural phenomena.

This conceptual metaphor would mean that a proportion of teleological explanation for natural phenomena would specifically associate with the human-centric activity of play, therefore, would be classified as relational-teleology. Examples of these types of teleological explanations in Study 1 would include snow being used for playing games with, muddy puddles caused by rain being a thing for children to play and jump in or waves are for people play jumping over them. This claim is supported by the results of the corpus investigation and children's responses in Study 1. For certain keywords (rain, snow and waves) there was both a) a high proportion of collocates related to play; and b) a high proportion of teleological explanations in Study 1 referring to play activities and natural phenomena.

In summary these three conceptual metaphors further contribute to the collocate analysis. The derived metaphors develop the notion that there are a number of themes in children's natural discourse which relate natural phenomena and teleological ideas. The themes children experience suggest that they may experience or use metaphors which discuss TIME AS NATURAL PHENOMENA, indicate NATURAL PHENOMENA ARE ACTIVITIES or that NATURAL PHENOMENA ARE A RESOURCE IN PLAY. It is proposed that the conceptual metaphors indicate that the way children conceptualise natural phenomena will influence how they talk about natural phenomena. The three metaphors all correspond to the majority of Study 1 responses, particularly those which were coded as relational-teleology. Therefore, supporting claims made about the influence of language upon children's use of teleological explanations, in essence that teleology may be partially caused by natural patterns of language (see section 2.4.4).

5.3.5 Conclusion

This study aimed to address the question *does the natural discourse children experience about natural kinds influence children's explanations for their existence?* It was proposed that this research question could support or refute the sociocultural notion that teleology has a partially linguistic origin: children's use of teleological explanation is influenced by the cultural norms of language use which they experience (outlined in section 2.4.4). Support for such an argument would provide evidence as to why children showed a preference to use relational- rather than design-teleology in Study 1. Corpus linguistics was used due to the method's ability to focus on naturally occurring language around natural phenomena, as was theorised that the natural discourse children experience may influence their explanations for the existence of natural phenomena.

Following investigation of the component question (CQ1) *what specific examples of teleological and scientific explanations do children experience about natural phenomena?*, a clear finding was the limited number of specifically teleological and scientific concordances, but they were present. There were few scientific explanations in the corpus, which runs counter to Kelemen et al. (2005) finding that 64% of answers to children questions were causal. However, this could be due to compromised reliability for parental diaries, causes by sanitised accounts, where information is altered to fit the project (Bytheway & Johnson, 2002) and because their study explored answers to specific questions. Singling out specific questions and answers is not the scope of natural dialogue and the nature of the two studies are not necessarily comparable.

The limited number of teleological and scientific concordances found in the corpus, including the inability to explore such explanations for half of the ten natural phenomena originally investigated, makes claims based on the concordance analysis untenable. The results of this study would indicate that children experience a low number of specific teleological and scientific explanations about natural phenomena. Therefore, there is little evidence to suggest that the specific explanations children experience in natural discourse influence their use of teleological or scientific explanations for the existence of natural phenomena.

The collocate analysis allowed for exploration of CQ2: *what are the themes associated with natural phenomena in the discussions children experience?* Categorising the collocates suggests there were a number of key themes which emerged, chiefly *activities (*including sub-categories of *sleep* and *play)*, *greetings*, *location* and *time period*. These collocate categories were used to propose three conceptual metaphors related to natural phenomena, which children use or experience. The metaphors were TIME AS NATURAL PHENOMENA, NATURAL PHENOMENA ARE ACTIVITIES and NATURAL PHENOMENA ARE A RESOURCE IN PLAY. Conceptual metaphor theory would indicate that these metaphors would structure how individuals discuss related domains (Lakoff & Johnson, 1980). Therefore, if children's are being exposed to these conceptual metaphors then their explanations about natural phenomena would discuss natural phenomena in relation to time, activities or their relevance to play. Children's explanations in Study 1 support this notion,

the majority of children's teleological explanations in Study 1 were proposing that examples of natural phenomena exist for the purpose of signifying a time period, indicating an activity or for use in play. Consequently, it is argued that in the natural discourse children experience there are a number of themes associated with natural phenomena. These themes would suggest that natural phenomena are associated with time, activities and play. It is proposed that these themes influence children to provide teleological explanations based upon them when explaining the existence of natural phenomena, which is supported by the results of Study 1.

In summary, children experience limited specific examples of teleological or scientific explanations for natural phenomena in their natural discourse. However, discussions around natural phenomena involve a number of themes, themes which are closely related to the teleological explanations children provided in Study 1. This relationship between themes in discourse and children's teleological explanations is especially strong for relational-teleological explanations, those where something is explained in relation to a purpose for which it could be used by human-kind. Therefore, it is proposed that children's tendency to provide teleological explanations, particularly those which demonstrate relational-teleology, is in part influenced by the natural discourse which they experience. That is, the natural discourse children experience about natural phenomena influences children's explanations for their existence.

This finding indicates that teleology may be partially socially constructed, as it may be influenced by cultural patterns of discourse, such as those detailed in this chapter. The claim that children's teleological explanations are shaped in part by the natural discourse they experience is supported by sociocultural theories of learning and development. That is, children's experiences, their differing subcultures that have shaped their conceptions of the world, will influence their styles of explanation and patterns of language use (Hoff, 2006). While the association between the discourse children experience and the explanations they provide may seem axiomatic, this is the first know evidence supporting this link in relation to children's teleological explanations about the natural world. This result is of importance as it suggests that children's use of teleological explanation may be influenced by natural patterns of language, in addition to children's innate tendency as suggested by Kelemen (1999c) and Kampourakis (2014).

6. Study 2: Improving children's evaluation of explanations for the existence of natural phenomena

6.1 Introduction

The literature review (Chapter 2) indicated children have a broad bias to advocate teleological explanations to account for the existence of natural phenomena (e.g., night, rain or snow). For example, *there is wind for blowing the leaves of trees* or *there is night so people know when to go to sleep*. The findings of Study 1 suggested that although teleology may not be as much as a barrier as it is claimed to be, the high number of teleological responses certainly suggests it is still a considerable hurdle – a hurdle that the corpus investigation suggests is partially constructed by the natural discourse children encounter around natural phenomena. A question raised from these findings (and in the literature review, section 2.7) is whether methods can be developed which limit children's acceptance and use of teleological explanations for the existence of natural phenomena. The study described in this chapter aims to answer this question.

Most of the research into children's use of teleological explanations has provided a snapshot of what children know at certain age-related stages (see section 2.7), there has been much less attention on limiting children's use of teleological explanations. This study sought to explore how children learn to differentiate teleological and causal explanations, by investigating how to support children in learning about the elements of appropriate scientific, and inappropriate teleological, explanations. Section 2.8 outlined the educational rationale behind this proposed intervention: to improve children's scientific literacy, to assist them in considering things in a scientific way, and to move from small ideas (explanations related to specific concepts) to big ideas (explanations that have a wide ranging applications and can be used when encountering new things). This shift is important as big ideas (e.g., all natural phenomena must have a causal explanation) have a greater explanatory value than small ideas (e.g., rain must have causal explanation), as they are applicable to a greater range of situations. Therefore, this intervention sought to enable this shift from small to big ideas; as teaching about the lack of purpose in relation to natural phenomena could have a wider-ranging impact than simply teaching children about the causes of several examples of natural phenomena.

It was decided that a discussion-based approach would be used to promote evaluation of evidence, to inform, and foster the development of children's big ideas about science. This was due to the success of discussion-based programmes, for example, P4C, in teaching children about abstract and philosophical concepts. The proposed intervention incorporated the ideals, theory and methods from P4C and specifically the form of discussion used in P4C – dialogical teaching, to develop a discussion-based intervention. The design of the sessions is discussed further in section 6.2. The intervention had two main intertwined aims:

1) To improve children's judgements about explanations for the existence of natural phenomena;
2) To improve children's understanding of, and skill in articulating, the elements of appropriate and inappropriate explanations for the existence of natural phenomena.

These educational aims sought to answer the research question (RQ4) *do discursive activities about the appropriateness of scientific explanations enable children to develop an understanding of why causal explanations are appropriate and teleological explanations are inappropriate to account for the existence of natural phenomena.* The focus age range was Key Stage 1 (5- to 7-year-olds).

6.2 Intervention design

Sections 2.8 and 6.1 provided a rationale why the intervention was discussion based, this section describes two lesson elements that promote group discussion, with reference to the focus age range (5- to 7-year-olds). These two methods are Concept Cartoons and puppets; each method is advocated by Harlen & Qualter (2014), Loxley et al. (2010), Lunn (2006), Sharp et al. (2014), Smith (2010), Sutton (1992) and Wellington & Osborne (2001). Also included is the guidance which set the 'rules' for the group discussion. Following this the use of terminology in the intervention is considered and, finally, the lesson elements which were piloted are presented.

6.2.1 Methods to structure group discussion

By taking elements from dialogic teaching and P4C, group discussion should (Alexander, 2006; Sharp et al. 2014; Smith, 2010b):

- be less structured, (e.g., it is not the standard IMF structure but instead one where conversation flows between all participants);
- allow children to respond to others ideas;
- allow children to ask others questions;
- be less formal, children should be allowed to contribute easily (e.g., there is less need for ‘hands-up’);

The following two lesson elements (Concept Cartoons and puppets) represent activities which embody these four principles of effective group discussion. Concept mapping was considered, it involves brainstorming children’s ideas then using that map as a discussion aid which can be added to and altered in light of discussion. It is argued that concept mapping provides an excellent medium to structure children’s ideas and develop a shared understanding of multifaceted ideas, such as purpose in nature (Anderson-Inman & Ditson, 1999; Guastello, Beasley &Sinatra, 2000; Novak & Cañas, 2006; Novak & Gowin, 1984; Pearson & Somekh, 2003). However, due to concept maps free-flow nature it may have been difficult to interject teleological explanations if none arose through the activity. The use of puppets and Concept Cartoons ensured that teleological ideas would be presented and discussed.

Concept Cartoons are typified by a cartoon depicting a certain scientific concept (e.g., causality, temperature change or materials and their properties) and, usually, four figures offering explanations about the science. There is also a blank speech bubble to encourage children to contribute their thoughts. For example, one shows a snowman with children offering statements such as '*don't put the coat on the snowman. It will melt him*', '*it will keep him cold and stop him melting'* and *'I don't think the coat will make any difference*'. These contradictory statements then form the starting point for discussion (Naylor & Keogh, 2000; Keogh & Naylor, 2012). Concept Cartoons have proven a popular method with children (Keogh & Naylor, 1999) and are claimed to provide an excellent starting point to develop argumentation skills (Chin & Teou, 2009) and generate discussion about complicated scientific concepts (Keogh & Naylor, 2012). Furthermore, they are often used as a starting point in thinking skills programmes such as P4C, especially in relation to philosophy of science (Smith, 2010b). Usually Concept Cartoons are aimed at, and developed for, upper primary level. However, can be used with young children to generate the same benefits (Morris, Merrit, Fairclough, Birrell & Howitt, 2007).

The PUPPETS project (Naylor, Keogh, Downing, Maloney & Simon, 2007) investigated the effect of puppets taking part in child conversations, in relation to engagement and child talk about science. The findings showed that puppets can enhance engagement, act as a stimulus for talking and reasoning, and engage reluctant speakers. Furthermore, the findings indicated that there was no influence of age on the effectiveness of puppets with children aged 4- to 11-years-old. All children explained more and provided more justification for their answers when working with puppets compared to typical classroom talk (Naylor et al., 2007; Simon, et al., 2008). Perhaps the benefits of puppets are due to the puppets acting as a 'less knowledgeable' group member, compared to the teacher, reminiscent of the iconic 'naughty teddy' study by Donaldson and McGarrigle (1974). Indeed, Simon et al. (2008) suggest that puppets are effective because they are controlled by the teacher but act as though the lesson is child-led. This arises because the puppet is the teacher, with their pedagogical and subject knowledge, but behaves as a member of the child's peer group, a member who needs the child to be the expert.

Table 6.1: *Positive and negative aspects of different discussion activities*

Method	Positives	Negatives
Concept Cartoons	Statements presented on the cartoon are predetermined, so must be discussed.	Poorly designed cartoons may prompt children to agree with certain statements more than others
Puppets	Explanations to the puppet can consolidate children's ideas and previous discussion.	Possible confusion if puppeteer acts as teacher rather than a conduit for the puppet

Table 6.1 shows the positive and negative aspects of the two outlined methods. Puppets and Concept Cartoons can start with the antecedent of a teleological or causal statement, in these two methods the teacher has full control over the direction of the talk. Furthermore, Concept Cartoons and puppets immediately encourage evidence evaluation, as there is no time needed for children to describe and elaborate their ideas before discussion. For these reasons, and others presented in Table 6.1, Concept Cartoons and puppets were considered suitable discussion-based activities for the intervention. Puppets were not used for promoting group discussion, rather for the statement evaluation and summary plenary at the end of each intervention session.

6.2.2 Terminology

The use of terminology in science education is often a contentious issue. While there is an argument for introducing terminology from the beginning (see Sharp et al., 2014), for example, words that can have multiple meanings for descriptive scientific language (e.g., vibrating for trembling), this can lead to confusion. Nitz, Ainsworth, Nerdel & Prechtl (2014) provide supporting empirical evidence, finding that a large number of scientific terms or complicated vocabulary can be overwhelming to students: in their research participants who reported their science lessons had a high number of linguistic terms developed lower representational competence (the ability to use, make and evaluate scientific representations) and lower content knowledge. This supports the findings of Merzyn (2005) and Wellington & Osborne (2001). Therefore, Harlen & Qualter (2014) propose that until a child has grasped a concept there is no need to learn specific terminology (e.g., teleological or photosynthesis), instead they suggest using the language that children employ in their descriptions. The intervention followed this stance and did not use the terminology *teleology* or *teleological reasoning*, but adopted terms such as *purpose* and *purposeful explanations*.

6.2.3 Assessment

When working with young children there is an even greater need to tailor the data collection method to the participants' abilities, for example, accounting for reading or writing ability (see methodology section 3.5 for further discussion). The method should not be one that is difficult or above children's

skill level, as this could lead to misleading data (Docket Einarsdóttir & Perry 2009; Scott, 2008). Using written assessment for this age range (5- to 7-year-olds) may be suitable for some children but others could struggle due to their writing ability. Therefore, the actual writing could take precedent over the answers content. Furthermore, children of this age can provide a more detailed verbal answer, as their talk is more advanced than their writing ability. (Scott, 2009). Freeman & Mathison (2009) do not advocate the use of written methods with children until children are 8-years-old, and only after pre-assessing that their ability matches the demands of the task. However, there is a problem with using verbal methods of assessing children: task effects associated with working memory. Verbal interviews with young children, if using overly abstract or multifaceted questions, can lead to high working memory demands, as young children can only hold and operate a limited number of units, or chunks, or information at one time (Gathercole, Pickering, Ambridge & Wearing, 2004; Siegler, 1998). Consequently, Christian & James (2000), O'Kane (2000) and Scott, (2008) suggest the use of pictures in interviews to make verbal questions more concrete.

Concept Cartoons, outlined above (section 6.2.1), are cartoons in which characters offer statements about a scientific concept. While providing an excellent activity to foster group discussion, it is also proposed that Concept Cartoons can provide an effective method to assess children's judgement of statements about the existence of natural phenomena. This was done by presenting children with multiple Concept Cartoons about a concept with each cartoon only having one statement about the topic under discussion, children then evaluated the explanation as good or silly (appropriate/ inappropriate) and were asked to provide a rationale for their judgement. Presentation of one explanation at a time is less demanding in terms of working memory demands and may also encourage a more rigorous appraisal of the explanation, compared to presenting multiple statements on a single cartoon. The assessment Concept Cartoon included an image of the natural phenomena under discussion that should have assisted in making the verbal questions more concrete.

6.2.4 Selection of assessment and session discussion topics

Assessment and session discussion topics were taken from Study 1 and its pilot study, possible topics included clouds, darkness, day, light, night, rain, rainbows, snow, storms, waves, waterfalls and wind. As these examples of natural phenomena are the basis of the intervention discussion, both in the assessments and main sessions, they are referred to as discussion topic/s or discussed topic/s. After consideration darkness and light were removed due to their similarity to night and day, storms was swapped for the Study 1 preliminary study questioned topic of lightning due to the majority of children's responses for storms in Study 1 simply referring to lightning, night was also removed as the explanation for night is effectively the counter explanation for day. This resulted in 8 possible discussion topics for the intervention study (clouds, day, lightning, rain, rainbows, snow, waves, and wind). Taking into account the ranking of levels of teleological explanation, clouds,

rain, rainbows and waves were selected as the intervention discussion topics and day, lightning, snow and wind were selected for the assessment discussion topics. The split was created by taking into account the levels of teleological explanations given to each discussed topic, with the assessment discussed topics receiving more teleological responses than the lesson discussed topics. Essentially, this split was an attempt to have assessment discussed topics, as those which children would have the least prior knowledge about.

6.2.5 Summary of intervention design

In summary, it was decided that Concept Cartoons with individual statements would be used to assess the efficacy of the intervention and specific scientific terminology would be limited. The intervention design was underpinned by the structure of P4C and dialogic teaching (section 2.8.4 and 2.8.5). It begins with an introduction to the session discussion topic before proceeding to a group discussion about the appropriateness of various teleological and scienctific explanations, supported by Concept Cartoons. Finally, children evaluated statements offered by a puppet. With regard to frequency and groupings for the intervention sessions, it was decided that children would undertake the four weekly intervention sessions, in groups of four, this was modelled on the frequency and grouping used in CASE and P4C programmes (Adey, et al., 2002; Smith, 2010b; Seleznyov & Adhami, 2015; Venville et al., 2003,).

6.3 Method

6.3.1 Pilot study

6.3.1.1 Pilot aim

A pilot was conducted to assess the practicality of and refine the overall intervention and the pre-/post-test assessment methods. Lesson elements were assessed to ensure they generated adequate group discussion and the pre-/post-test methods were investigated to ensure they sufficiently measured the target knowledge. Children's reflections about the activities and information about timings were collected for both.

6.3.1.2 Pilot participants

The participants were eight Year 2 children, aged 6- to 7-years-old, recruited from the same class. No child had SEN or EAL. The school was a two-form entry from a large city in the Midlands of the UK. It was judged to be performing excellently on national scales, and is below average for both pupils from minority ethnic groups and those entitled to free school meals.

6.3.1.3 Pilot design

The pilot study had a one group, pre-/ post-test experimental design, a control group was not used as an efficacy test of the intervention was not the pilot aim. The pre- and post-test assessment was conducted through individual interviews, following the post-test children were asked to feedback about the assessment method. The intervention lessons were conducted with children in groups of four. During the final intervention session children were asked to feedback about the intervention

sessions. All interviews and lessons were audio-recorded and conducted in a communal area to produce a more relaxed and sociable environment.

6.3.1.4 Pilot procedure

The pre-test assessment began with participants being told about who the interviewer was, where they had come from and the purpose of the interviewer's visit: 'to talk and think about some different science topics'. Participants were then asked to consent to the interview (outlined below). Following this the participants took part in an icebreaker activity. This was an adapted version of Winnicott's squiggle game (Winnicott, 1971) used to quickly develop the participant-interviewer relationship; thereby, creating a conversational environment and reducing possible shyness (Freeman & Mathison, 2009; Harcourt & Conroy, 2011; Hogan, 2005). Participants were shown two example Concept Cartoon assessments about night, one with a teleological statement and one with a scientific statement. These were read by the interviewer who explained why they were appropriate/inappropriate explanations for the existence of night. Participants were then informed that 'I have some more of these cartoons here and all I want you to do is to tell me if they are a good or a silly explanation for why there is something'. For each of the 12 Concept Cartoon assessments, the discussion topic was stated, the explanation read and the participant asked 'do you think that is a good explanation for why there is X?' Following their response, children were asked 'why do you think it is a good/silly explanation?' The introduction, icebreaker and example were not used in the post-test assessment, children were merely reminded about the pre-test were they had to think about good and silly explanations. Following the post-test participants were asked to comment on how they found the task, if it was too long or short, and what they would change if it was done with other children.

Each of the four intervention sessions had the same three part format, with the exception of the final session where children were asked to comment on how they found the sessions, if they were too long or short, and what they would change if they were done with other children. The three parts were:

1. Introduction (2-3 minutes): children were told that they would be talking about the discussion topic of clouds, rain, rainbows or waves (depending on the session), specifically focusing on why it happens and how it is made. The children were then encouraged to offer their own explanations and comments upon the discussion topic.
2. Discussion activity (10 minutes): They were then presented with a Concept Cartoon for the discussion topic, which showed three statements about its existence. These statements were taken from the most popular answers collected in Study 1, children were also encouraged to offer their own explanations. Children were encouraged to discuss the statements in relation to if it accounted for the existence of the discussion topic, with the teacher prompting children to consider if the discussed topic had a purpose.

3. Puppet plenary (4-5 minutes): Children were introduced to the puppet who provided additional teleological and scientific explanations about why there is X. Participants were encouraged to decide if it was a good or a silly statement, then asked to tell the puppet why it was a good or silly idea.

6.3.1.5 Pilot measures

For the assessment, children's judgements were coded using the rubric in Table 6.2 and analysed for pre- to post-test change. Children rationales were not analysed but used to develop a coding scheme for the main study (section 6.3.2.6). For the interventions sessions, children's comments and the researcher's reflections are discussed and alterations to the study are provided.

Table 6.2: *Coding rubric for judgement of statements*

	Child's judgement	
Statement	Good	Silly
Design- or relational-teleology	Score = 0	Score = 1
Scientific	Score = 1	Score = 0

6.3.1.6 Pilot results

During the pilot study children's average score improved from 57.2% (*SD* = 20.0%) correct judgements at pre-test to 81.3% (*SD* = 17.0%) correct judgements at post-test. A score increase of 24.1 percentage points (95% CL 8.83% to 39.0%). A dependent *t*-test was conducted and the results showed a significant score difference, $t\ (8) = -3.75$, $p. = .007$, with a strong effect size, $d_{rm} = 1.67$. Therefore, the results suggest that following the pilot study intervention, children's understanding of the elements of appropriate and inappropriate explanations improved considerably.

6.3.1.7 Pilot changes

Considering the researcher's and children's reflections on the sessions, two main changes were made the structure of the intervention sessions. Firstly, asking the children openly about the discussion topic in the introduction was problematic: once an explanation was offered, other children usually echoed the same explanation. Therefore, the introduction went straight into the Concept Cartoon discussion. Secondly, during the main activity, children often over explained their ideas, thereby limiting the opportunity for others to speak. This can, and was, remedied by moving the discussion along when children were repeating their idea or going 'off topic'. However, there was occasional difficulty in balancing timings when a) the discussion needed to move on and b) when a child needed to develop their idea. This remained a problem throughout the main study and was resolved depending upon the specific case.

In summary, conducting the pilot to assess to practicality of and refine the intervention and assessment process resulted in the following:

- The pilot suggested that the intervention may be effective at developing children's understanding of inappropriate teleological explanations;
- Children's ideas should not be investigated during the introduction, as it limits discussion of teleological statements;
- The discussion topics used for the four sessions are appropriate but were reordered as rainbows, rain, clouds and waves;
- Children should be encouraged to elaborate but not excessively;
- Use of the puppet resulted in an effective plenary activity;
- Children need to be prompted to foster discussion between themselves not simply talk to the teacher.

6.3.2 Main study

6.3.2.1 Aim

The aim of Study 2 was an efficacy trial of the intervention which sought:

1) To improve children's judgements of explanations about the existence of natural phenomena;
2) To improve children's understanding of, and skill in articulating, the elements of appropriate and inappropriate explanations about the existence of natural phenomena.

Following piloting the exploration of children's ideas about the discussion topic during the introduction of the intervention sessions was removed. Apart from this one alteration the intervention sessions and assessment structure remained the same as discussed in section 6.3.1.3.

6.3.2.2 Design

A 2x2 mixed experimental design was used, the within-groups variable was assessment, with two levels (pre-test and post-test), the between groups variable was experiment group, with two levels (intervention group and control group). A no treatment control group was used, as there was no comparable treatment to compare the intervention against.

6.3.2.3 Participants

53 children, aged 5- to 7-year-old, took part in the study. Participants were drawn from two schools, 25 Year 1 children took part from school one (Female = 9, English as an Additional Language = 8, Special Educational Needs = 0) and 28 Year 2 children took part from school two (mean age = 84.5 months old, *SD* = 9.44, Female = 11, EAL = 3, SEN = 3). School one (Year 1 children) was a single-form entry Catholic primary school in a large city in the East Midlands of the UK. The school was judged to be performing outstandingly on national scales, and was above the national average for both pupils from minority ethnic groups and those entitled to free school meals. Due to school one policy, children's date of births were not provided, therefore, specific analysis by children's age in

months was not conducted. School two (Year 2 children) was a single-form entry infant school in a large city in the East Midlands of the UK. The school was judged to be performing well on national scales, was above the national average for both pupils from minority ethnic groups and below the national average for pupils entitled to free school meals. In each school, children were randomly assigned to either the intervention group (27 children, Year 1 = 13, Year 2 = 14, Female =14, EAL = 6, SEN = 2) or the control group (26 children, Year 1 = 12, Year 2 = 14, Female = 6, EAL = 5, SEN = 1).

6.3.2.4 Procedure

The procedure was similar to the pilot study, pre-test assessment began with a child being told about who the interviewer was, where they had come from and the purpose of the interviewer's visit: 'to talk and think about some different science topics'. Participants were then asked if they were happy to answer some questions about different scientific topics and for the conversation to be recorded. Winnicott's squiggle game (Winnicott, 1971) was used to develop the participant-interviewer relationship; thereby, creating a conversational environment and reducing possible shyness (Freeman & Mathison, 2009; Harcourt & Conroy, 2011; Hogan, 2005). Participants were shown two example Concept Cartoon assessments about night (Appendix 3), one with a teleological statement and one with a scientific statement. These were read by the researcher who explained why they were appropriate/inappropriate explanations for the existence of night. Participants were then informed that 'I have some more of these cartoons about different topics and all I want you to do is to tell me if they are a good or a silly explanation for why there is something'.

All the children (both intervention and control group) were assessed before and after the intervention using twelve questions across four discussed topics (day, snow, lightening and wind). The introduction, icebreaker and example were not used in the post-test assessment, children were merely reminded about the pre-test were they had to think about good and silly explanations. In each school the order in which participants were selected to take part in assessment was arbitrarily decided. For each discussed topic children received a randomised mix of teleological and scientific statements (see materials, section 6.3.2.5). These assessment sessions lasted, on average, seven minutes. All assessment was audio recorded and transcribed for coding. For each statement children were asked:

- To judge the explanation as good (appropriate) or silly (inappropriate) – this allowed for the assessment of children's initial judgements of explanations;
- To provide a rationale for their judgement – this allowed for the assessment of children's understanding of the elements of appropriate or inappropriate explanations.

To recap, the four intervention sessions (on rainbows rain, clouds, and waves) comprised of:

- Introduction: children (in groups of three or four) were presented with a Concept Cartoon that displayed several possible incorrect explanations about a certain discussion topic;
- Main discussion: children discussed the statements on the Concept Cartoon with teacher guidance, at no point were children told their ideas were correct;
- Plenary: children summarised the session to a puppet, the puppet then offered children new statements to judge as good (appropriate) or silly (inappropriate).

The children in the treatment group were randomised, within each school, into groups of three or four for each of the four sessions. These group sessions lasted, on average, 15 minutes for each group of children. Year 2 data collection took part in the latter half of the Spring term (March - April), Year 1 data collection took place at the end of the Summer term (June - July). The data collection in both schools took place across seven weeks, with assessment sessions taking place the first and seventh weeks and interventions sessions taking place during an afternoon on the second, third, fifth and sixth weeks. Due to different classroom layouts, Year 1 assessment and intervention sessions took place in a quiet area of the classroom, whereas Year 2 activities took place in a shared communal area.

6.3.2.5 Materials

Table 6.3 shows the four discussion topics with the corresponding explanations used in the assessment Concept Cartoons. An example assessment Concept Cartoon in provided in Figure 6.1, an example lesson Concept Cartoon in Figure 6.2, along with an accompanying lesson plan in Figure 6.3.

Table 6.3: *Pre- and post-test assessment statements*

	Statement type ("There is X …")		
Topic	Design-teleology	Relational-teleology	Scientific
day	for telling the birds to wake up	so children can go to school	because the Earth is spinning around and when the bit we are on faces the sun, then it is daytime
lightning	for lighting up a dark, stormy sky	so people can use it to power electric appliances, like the TV or lights	because parts of clouds bump together and make electrical charge, and when the cloud gets full of electric charge, a lightning bolt comes out the cloud
snow	for telling some animals to hibernate over winter	so children can make snowmen	because it is really cold and the rain drops freeze and turn into snowflakes
wind	for blowing the leaves of trees	so children can fly kites	because the air is moving, in the sky hot air moves up and cold air rushes into the space left

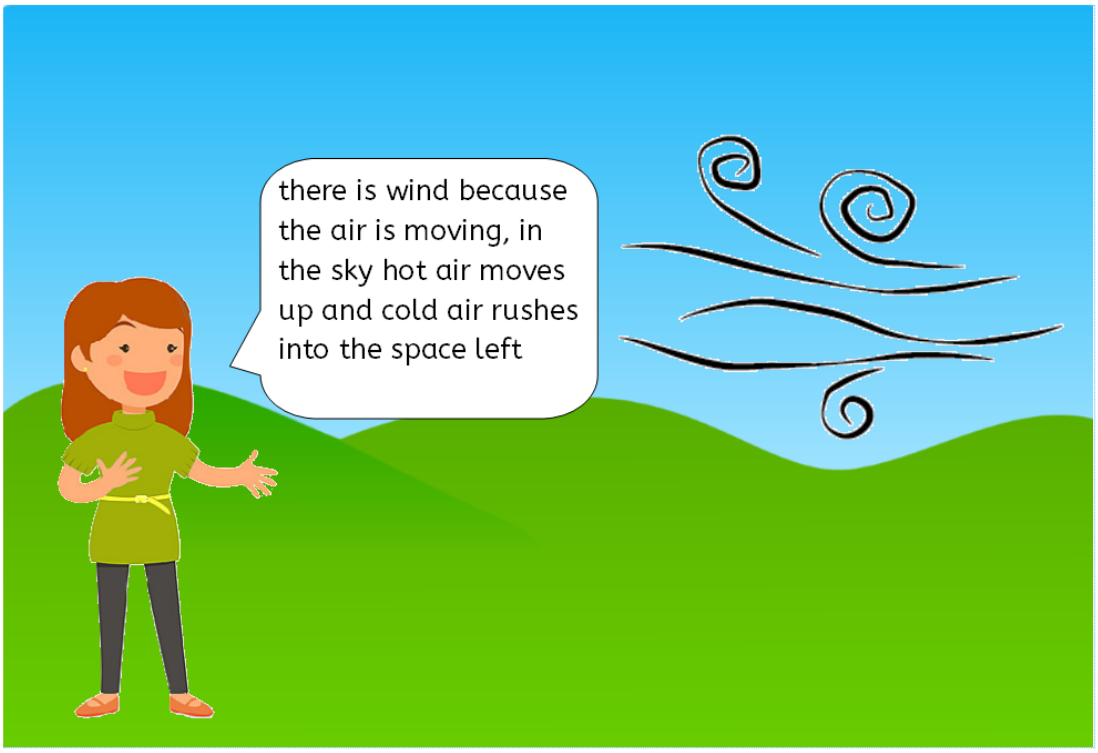


Figure 6.1. Example assessment Concept Cartoon for Study 2: Scientific explanation for why there is wind.



Figure 6.2. Example lesson Concept Cartoon for Study 2: Discussion prompt for why there are waves.

Groupings	Group 1:	Date:
Session	Four: Waves	Time:

What to do:

Introduction (5 minutes)

Remind chn about why I am here - talking about why things happen
Introduce children to the topic: **Waves**.
Show and explain the concepts cartoon
It shows some people thinking about why there is X – (*What do you think about what the people are saying?)*
Write down chn's ideas

Main discussion (10 minutes)

Continued to discuss the topic - use questions to guide the discussion.
Encouraged chn to consider:

- What is the topic was associated with;
- What causes it to happen;
- How to people/animals use the topic.

Introduce topics from box if children struggle

Evaluation of statements (5 minutes)

Introduce Prof. Bananas – he has missed the previous discussion!
Encourage chn to tell the puppet what they have been taking about.
About the cartoon's ideas for *why there is* X. **what did the children decide?**

Prof. Bananas has some ideas about why there is X.
Chn to offer advice upon ideas, i.e. why it was a good or silly idea
Use statements in the box and notes of chn's ideas
Chn to use thumbs up /down for good/silly

Resources: Recorder, puppet, puppet bag, cloud concept cartoon, A3 paper for children's ideas, pens

Questions to ask:

- What do you think? (to new child);
- Did everyone hear that, is there anything anyone wants to add?;
- So you're saying ……? (to add clarification to idea);
- How does this relate/differ to X's idea before? (to link ideas back);
- That reminds me of an idea I heard a while ago, that … (to add more examples);
- Who thinks this a good idea?;
- Are there any problems with this idea?;
- Is that the reason why is it happens or just something we/animals could use it for?;
- Take your time/ have a little think/ is there anything you want to add/ can you tell us a little bit more about X?;
- Can anyone think of anything else X is used for? (to search for counter examples);

Examples for discussion / Statements for the puppet

Design	For helping fish swim
	For helping to move boats
	For pushing sand/stones onto the beach
	For hiding sea animals, like sharks
Relational	So people can surf/jump over them
	So people can swim quicker
	So people can relax listening to the relaxing sound
Scientific	Because the wind blows across the sea, making the water move as waves
	Because of the pull from the moon
	Because of splashes by people, animals, boats

Figure 6.3. Lesson plan for Study 2, intervention session 4 – Why are there waves.

6.3.2.6 Measures

Children's judgments were coded using the rubric in Table 6.2, their rationales for their judgements were coded using the rubric in Table 6.4. Approximately 10% of the data, (126 responses) were coded by a second researcher. Cohen's Kappa indicated a good level of agreement ($K = .787$, $p < .001$).

Table 6.4: *Coding rubric for children rationales of their judgements*

Score	Explanation type	Example
1	The rationale for the judgment of a statement discusses the style of explanation used in the statement. That is the statement is claimed to be good because it offers a causal explanation to why/how X happens. The statement is silly because it does not offer a causal explanation to why/how X happens. Instead it states what X is used for Y (RT) or states X is for Y when it is not for anything (DT)	[There is snow because of rain freezing] is good cos it tells you how the snow happens [Wind is for flying kites is silly] because it is just telling you what children can do, not why there is wind
0	Any other form of rationale, including but not limited to, a non-response; Rewording: The rationale for the judgment of an statement is a repeated or reworded version of the statement; Teleological: The rationale for the judgment of a statement discusses the topic of the statement in a teleological manner. The event taking place in the statement (or an outcome) is discussed in relation to a purpose it serves for an agent; Axiomatic: The rationale for the judgment simply states the statement is true/false, no additional information is provided.	[Snow is for animals to hibernate is good] because there is snow to tell animals when they need to go to sleep [There is snow because of rain freezing] is good cos then then people can have snowball fights [There is snow because of rain freezing is good] cos that is a fact

6.3.2.7 Ethics

The study was conducted within the ethical guidelines from the University of Nottingham (UoN, 2015) and the British Educational Research Association (BERA, 2011). After consent had been provided by the schools' head teachers opt-out consent forms were sent to parents. The rationale for opt-out consent is discussed in section 3.4. In summary the opt-out consent was only considered to be provisional consent; therefore, after being informed about the purpose of the interview

children were asked if they agree to take part. If they dissented, the interview would not have taken place. The interview would have been stopped if a child indicated dissent through discomfort or withdrawal, although this did not happen.

6.4 Results

6.4.1 Data checking

To conduct the following analyses two variables were created: judgment score and rationale score, these were participants' mean score for their 12 judgements or 12 rationales; each with a maximum score of 12. To ensure there was no difference between the Year groups' judgements and rationales at pre-test, two one-way ANOVAs were conducted. The independent variable was group, with 4 levels (Year 1 control, Year 1 treatment, Year 2 control, Year 2 treatment), the dependent variable was judgement score or rationale score, for the respective tests. The four groups judgment scores were not significantly different, F (3, 52) = 1.83, $p.$ = .154, η^2 = .101, nor were there rationale scores, F (3, 52) = 2.67, p = .057, η^2 = .141. Consequently, the following analyses combined the treatment (Year 1 and 2 treatment groups) and control (Year 1 and 2 control groups) groups.

To explore the influence of discussed topic, which was shown to be a factor in levels of teleological reasoning based on Study 1 (section 4.3.2), Cochran's Q tests were used to explore differences in the levels of correct judgments at pre- and post-test for both the treatment and control groups. These tests are used to assess if children are treating the twelve statements about four discussed topics of natural phenomena as a homogeneous group or if there is variation in their responses levels between the 12 statements. Cochran's Q showed that at pre-test both groups responses were consistent across the statements (Control: Q (11) 14.02, $p.$ = .231, η^2_Q = .024; Treatment: Q (11) = 4.49, $p.$ = .953, η^2_Q = .007), these results were mirrored at post-test (Control: Q (11) 11.03, $p.$ = .441, η^2_Q = .018; Treatment: Q (11) = 18.63, $p.$ = .068, η^2_Q = .031). The same procedure was conducted with children's rationale scores, again showing consistent responses at pre-test (Control: Q (11) 5.13, $p.$ = .925, η^2_Q = .008; Treatment: Q (11) = 3.23, $p.$ = .987, η^2_Q = .005) and post-test (Control: Q (11) 11.72, $p.$ = .385, η^2_Q = .020; Treatment: Q (11) = 7.55, $p.$ = .753, η^2_Q = .012). Taken together this descriptive and statistical analyses suggest that children were not influenced by discussed topic, there was no influence of statement type (teleological or scientific) or discussed topic upon children's judgements or rationales. Therefore, the planned analysis for the efficacy test of the intervention can be conducted without concern for questioned/discussed topic bias that was found in Study 1.

6.4.2 Analysing children's judgements

A 2x2 mixed ANOVA was conducted to investigate the influence of the intervention on children's ability to correctly judge statements about the existence of natural phenomena. The within-groups independent variable was assessment, with two levels (pre-test and post-test), the between-groups

independent variable was experimental group (control or treatment), the dependent variable was participants' judgement score.

The analysis revealed a main effect of assessment (pre-test vs. post-test) on children's judgement scores, F (1, 51) = 43.92, p. < .001, η_p^2 = .463. There was also a main effect of group (control vs. treatment), F (1, 51) = 12.32, p. < .001, η_p^2 = .195, and a significant interaction between assessment and group upon children's correct judgments, F (1, 51) = 103.45, p. < .001, η_p^2 = .670

Table 6.5: *Mean judgement scores at pre- and post-test, by experimental group (Max. score = 12)*

	Control (n = 26)	Treatment (n = 27)
Pre-test score (*SD*)	6.04 (2.34)	5.26 (2.36)
Post-test score (*SD*)	5.04 (2.71)	10.00 (2.16)

To further investigate these results, multiple pairwise comparisons were conducted, Bonferroni corrections were applied to control for multiple tests. The independent *t*-tests showed no significant difference between the groups pre-test scores, t (51) = 1.20, p. = .932, d = .33, however, there was a significant difference between the groups' post-test scores, t (51) = -7.35, p. < .001, d = 2.02. The paired *t*-tests showed a significant difference in the control groups judgment score from pre- to post-test, t (25) = 3.09, p. = .019, d_{rm} = 0.38, and there was a significant difference in the treatment group's judgement score from pre- to post-test, t (26) = -10.35, p. < .001, d_{rm} = 2.08. These results and the group means in Table 6.5 show that following the intervention the treatment group had a substantial significant increase in their correctly judgements (from 43.8% to 83.3% correct), whereas the control group had a significant decrease in their correct judgements (from 50.3% to 42.0% correct).

6.4.3 Analysing children's rationales

A 2x2 mixed ANOVA was conducted to investigate the influence of the intervention on children's understanding of the elements of appropriate, or inappropriate, explanations about the existence of natural phenomena. The within-groups independent variable was assessment, with two levels (pre-test and post-test), the between-groups independent variable was experimental group (control or treatment) , the dependent variable was participants' rationale score. The analysis showed a significant main effect of both assessment, F (1, 51) = 41.70, p. < .001, η_p^2 = .450, and group, F (1, 51) = 8.83, p. = .004, η_p^2 = .148. There was also a significant interaction between assessment and group, F (1, 51) = 53.34, p. < .001, η_p^2 = .511.

Table 6.6: *Mean rationale scores at pre- and post-test, by experimental group (max. score = 12)*

	Control *(n* = 26)	Treatment (*n* = 27)
Pre-test score (*SD*)	2.38 (3.44)	2.07 (3.30)
Post-test score (*SD*)	2.03 (3.54)	7.70 (4.05)

To further investigate these results, multiple pairwise comparisons were conducted, again Bonferroni corrections were applied to control for multiple tests. The independent *t*-tests showed no significant difference between the groups pre-test scores, t (51) = 0.33, p. > .999, d = .09, however, there was a significant difference between the groups post-test scores, t (51) = -5.41, p. < .001, d = 1.48. The paired *t*-tests showed no significant difference in the control group's rationale score from pre- to post-test, t (25) = 1.08, p. > .999, d_{rm} = 0.09, and there was a significant difference in the treatment group's rationale score from pre- to post-test, t (26) = -7.58, p. < .001, d_{rm} = 1.50. These results and the group means in Table 6.6 show the control group did not significantly differ in their rationale scores from pre- to post-test (from 19.8% to 16.9%). However, following the intervention the treatment group showed a substantial significant increase (from 17.2% to 64.1%) in rationales scores. Examples of correct rationales include:

- [There is snow for telling some animals to hibernate over winter] is silly because "hibernate is like a science word but it doesn't tell us how [snow] is generated";
- [There is wind for blowing the leaves of trees] is silly because "leafs get blown off when the wind has to come, but it is not telling us about the wind"
- [There is day so children can go to school] is silly because "it is not telling us how it is made, it is just telling up about children going to school";
- [There is snow so children can make snowmen] is silly "because you might want to make snowmen but it is not for that and it's not telling you how it is made".
- [There is day because the Earth is spinning around and when the bit we are on faces the sun, then we have daytime] is good "because it tells you why we have daytime"
- [There is lightning because parts of clouds bump together and make electrical charge, and when the cloud gets full of electric charge, a lightning bolt comes out] is good "cos it is telling us how the lightning is happening, cos all the clouds bump together"

6.5 Discussion

6.5.1 Assessment results

The main aim of the intervention was to improve:

1) Children's initial judgements about teleological and scientific explanations about the existence of natural phenomena;

2) Children's understanding of the elements of appropriate and inappropriate explanations about the existence of natural phenomena.

The results show that the intervention was successful in developing these two intertwined areas. Young children can be taught to recognise that teleological explanations are inappropriate to account for the existence of natural phenomena. In relation to the first of the two aims, for the children assigned to the treatment group their average score for their correct judgements about the existence of natural phenomena improved from 43.8% to 83.3%, whereas the control groups average score slightly decreased from 50.3% to 42.0%. The results show that the condition children were assigned to had a significant effect upon their judgement score from pre- to post-test and that both of the groups' judgment score changes were significant. Essentially, after the intervention there was a significant improvement in the intervention group's judgements of explanations even when the discussed topics themselves had not been covered during the four teaching sessions. Following group-discussion about the four intervention discussed topics, the treatment group showed improvement in their judgements, claiming that teleological statements were inappropriate, and causal statements were appropriate, explanations for the existence of natural phenomena.

The significant difference in the control groups' scores from pre- to post-test scores (50% correct judgements to 42% correct) could be due to classroom grouping. The control group received no input or feedback between the pre-test and post-test. Therefore, if they viewed the treatment group as receiving additional support they may believe that there are not involved in the intervention because their original responses were correct. Essentially the control group believed the intervention group needed further tutoring, whereas they did not, and consequently, stuck with, and even increased, their preference to endorse teleological statements. Support for this explanation comes from studies finding children are aware of how and why they are grouped (Hallam, Ireson & Davies, 2010) and that Year 2 children (6- to 7-years-old) assigned to higher level groups, in this case the one that did not need additional input, increased in their academic self-confidence (Weinstein, 1976).

The intervention was also successful with regard to the second of the two aims: to improve children's understanding of the elements of appropriate, or inappropriate, explanations about the existence of natural phenomena. For the children assigned to the treatment group their average score for their understanding of the elements of appropriate explanations for the existence of natural phenomena improved from 17.2% to 64.1%. Whereas the control groups average score slightly decreased from 19.8% to 16.9%. The analyses showed that the condition children were assigned to had a significant effect upon their rationale score from pre- to post-test, although only the treatment group's rationale score change was significant. After the intervention, there was improvement in the treatment groups' understanding of the elements of appropriate or inappropriate explanations for the existence of natural phenomena. For example, following the intervention when children in the treatment group suggested that, the statement day *is due the rotation of Earth* is good, they

provided a greater range of ideas which included awareness of appropriate explanation elements, such as ‘because it is actually telling us how there is day’ or similarly when suggesting that wind is not for flying kites ‘because it's not why it happens, people just use it for that’. These types of statements suggest a developing understanding that appropriate scientific explanations are causal (e.g., the explanation must suggest how something happens), and that teleological explanations are inappropriate (e.g., people just use natural phenomena for something, they are actually not *for* that).

One of the underlying themes of the intervention was the shift from small to big ideas, from knowledge of ideas and concepts related to specific events to knowledge of ideas and concepts that have wide ranging applications and can be used when encountering new things (Harlen & Qualter, 2014; Loxley et al., 2010). It was proposed that using the intervention to teach children about purpose and natural phenomena would have a larger impact than simply teaching children about the causes of several exemplars. This is because the big idea of purposeless natural phenomena is more applicable to newly encountered knowledge, acting as a schema to assist children in evaluating future explanations. As the assessment used investigated different discussion topics to those discussed in the intervention it means that children were not remembering what had been discussed. Instead the children were beginning to assimilate the big idea that explanations for all natural phenomena must be causal, not teleological, and were applying this concept to new examples of natural phenomena. Furthermore, they were beginning to provide rationales for their judgements based upon awareness about the elements of appropriate, causal, scientific explanations.

At pre-test, the total sample correctly judged 46.0% of the teleological statements as inappropriate and 46.2% of the scientific explanations as appropriate. These proportions of correct judgements are close to chance, so it could be claimed to not provide an adequate view of children’s teleological reasoning. However, children’s rationale scores are not chance based, and show low levels of understanding, with 18.4% of rationales about teleological statements and 17.5% of rationales for scientific statements demonstrating awareness of the elements of appropriate, and inappropriate, explanations. Consequently, children’s initial scores suggest they have limited knowledge of elements of effective scientific explanations and, possibly, a bias to endorse teleological explanations for the existence of natural kinds, which is supported by Kelemen (1999b, 2003), Kampourakis (2012a; 2012b) and Study 1. These teleological preconceptions should be difficult to change, if the teleological conception is robust and systematically applied by students (diSessa, 2014). Those that advocate stage- or theory-theories would claim this to be true (e.g., Piaget's three stages of artificialism or Kelemen's age-related shift from teleological to causal thinking) (diSessa, 2014; Klahr et al., 2011). However, the results of Study 1 indicate a more nuanced version of teleology. It is argued that one of the reasons children are viewed as having teleological preconceptions which are resistant to change is because no research has yet attempted

to specifically develop children's understanding of why teleological explanations are not appropriate for the natural world.

The results of Study 2 reveal that children's teleological preconceptions were not inherently difficult to change, a finding which runs counter to current thinking on children and teleology. This study shows children's ability to evaluate the statements dramatically improved after rather limited instruction, approximately 40 to 50 minutes. From this study, it seems that children are much more competent than usually assumed, with support they are able to reason in a concrete and abstract manner about scientific concepts, and evaluate evidence from a young age.

6.5.2 Evaluation of the intervention

This section argues that the success and the simplicity of the intervention is due to its theoretical underpinnings and the activities used. It discusses the programmes influencing its development (P4C), the methods used to foster critical discourse in the intervention (Concept Cartoons and puppets), the assessment of children, and finally the place for this intervention in the curriculum.

The intervention drew on the theoretical underpinnings of P4C, these influences may have contributed to the effectiveness of the intervention. For instance P4C draws upon Vygotskian ideas of socially constructed meaning, which is argued to be one of the explanations for the programme's effectiveness (Smith, 2010). This explanation is particularly relevant for the intervention in this study. As demonstrated in Study 1, the majority of children's teleological explanations take the form of relational-teleology: explaining the existence of something with regards to how it can be used, or a behavioural effect it causes in people. Furthermore, the corpus linguistics investigation (Study 1b) suggested that children's teleological explanations have a partially cultural origin, children are influenced by the natural discourse they experience. Taken together these findings indicated that some of children's teleological explanations are socially constructed. Allowing children time to explore and develop these socially constructed ideas has strong parallels with the highly effective dialogic teaching (e.g., see Alexander, 2006) and the argumentative theory of reasoning (Mercier & Sperber, 2016), where children have opportunities to consider different views and collaboratively construct a shared understanding (Grugeon et al., 2012; Mercer et al., 2009). Therefore, it is proposed that part of the intervention's success is due to it proving children with space and time to critically reflect on, and discuss, preconceived ideas in a supportive group environment.

The use of Concept Cartoons and puppets to foster discussion was already evaluated following the pilot study (see section 6.3.1.6). The Concept Cartoons again provided an excellent starting point to encourage the children to discuss and evaluate possible explanations. As in the pilot, children initially struggled with the discussion format but became more accepting of a) having to talk to their peers, not to the teacher, and b) there being no correct answer to the Concept Cartoon activity. Although, a discourse analysis of the sessions transcripts would be needed to confirm this

observation. Children's initial hesitation with this style of open-ended critical discussion has also been noted in similar Cognitive Acceleration programmes (Seleznyov & Adhami, 2015).

Similar to the pilot findings, children were at their most critical when evaluating the puppet's statements, and at their most coherent when explaining what was inappropriate about the teleological explanations. The children treated the puppet as a less knowledgeable peer, with whom they needed to be the expert and offer support, echoing the findings of Simon et al. (2008) and evoking children's behaviour in the study by McGarrigle & Donalson (1974). Contrary to the findings of Naylor et al. (2007) there were two Year 1 children, one likely encouraged by the other, who were strongly opposed to the puppet activity, both refusing to engage with the puppet and claiming the activity to be childish. However, this was far outweighed by the positive reaction from the majority of the sample, some of whom would argue with and chastise the two who would not take part in the puppet role play activity. Consequently, while there may have been a small minority who disliked the puppet activity, for the overwhelming proportion of children the puppet activity provided an excellent foil for the researcher to engage children in evaluating and criticising possible explanations.

The intervention activities demand no lengthy intervention-specific training for a teacher to use. These activities are within the skill set of all primary teachers, the intervention sessions are similar to circle time activities, a key lesson format in the repertoire of every Early Years teacher. Therefore, once a teacher understands the aims and purpose of the intervention it should theoretically be easy to incorporate into their teaching practice. However, would this intervention be practical in schools: does it fit with the prescribed curriculum? Section 2.8.7 addressed this question finding that the English National Curriculum (DfE, 2013) encourages children to develop their thinking skills and to be assisted in understanding scientific methods which allow for investigation of the world. Critical collaborative-discussion about ideas is a core scientific activity (Henderson et al., 2015; Thompson, 2001). Furthermore, this intervention represents one of the thinking skills programmes which are needed in classrooms (Henderson et al., 2015; Oliver & Venville, 2017).

6.5.3 Conclusion

In summary, the intervention was one of the types of studies reviewed by Klar & Chen (2003) that show that young children can distinguish between appropriate and inappropriate forms of evidence and recognise complex patterns in evidence. Consequently, the aims of the intervention, to improve children's evaluation of scientific and teleological explanations (possible evidence) for the existence of natural phenomena, was within children's abilities. Furthermore, as the assessment investigated different discussion topics to those discussed in the intervention this means that children were not remembering what had been discussed. Instead the children were beginning to

assimilate the 'big idea' that explanations for all natural phenomena must be causal, not teleological, and were applying this concept to new examples of natural phenomena. These findings suggest teleology may not be such a major barrier to teaching causality in Early Years education as has been previously assumed. When provided with limited instruction, children were more capable than the research currently indicates. The results of this study reveal that children's teleological preconceptions were not inherently difficult to change, a finding which runs counter to current thinking on children and teleology. The following chapter aims to determine if the intervention can be used with examples of different ontological categories, specifically in relation to organisms and evolution.

7. Study 3: Improving children's evaluation of explanations for the existence of organisms' traits

7.1 Introduction

Study 1 indicated that teleology may not be the major barrier it is claimed to be, in relation to understanding scientific concepts, but the high number of teleological responses for both question wordings and all questioned topics certainly suggests it is still a considerable hurdle. While the results of Study 1 pertained to natural phenomena, the influence of the possibility context independent factors of leading-question wording and lack of exemplars within an ontological category suggests that teleological explanation may also have been overestimated for organisms' traits. Furthermore, the Study 2 intervention, designed to limit children's endorsement of teleological explanations for the existence of natural phenomena, proved to be highly effective. A question raised from Study 2 was *can the intervention be used to teach children about styles of explanation for other ontological categories?* This chapter explores this question in relation to teleology and organisms' traits.

Section 2.5.3 of the literature review considered the scientific appropriateness of the different teleological explanations for organisms and their traits. The following briefly recaps this discussion before outlining Study 3, which aimed to limit children's endorsement of teleological explanations for the development of organisms' traits.

7.1.1 Evolution and teleological explanation

There are a large number of studies which suggest that children (aged 5- to 10-years-old) primarily use, or advocate, teleological explanations to account for the existence of organisms and their traits (see for example, DiYanni & Kelemen, 2005; Kampourakis et al., 2012a; 2012b; Kelemen, 1999a; 1999b; Polling & Evans, 2002; Sánchez Tapia et al., 2016). While the topic of evolution is not taught in the national curriculum until Year 6 (10- to 11-years-old) children's should be exposed to notions related to adaptation at earlier stages: Year 2 and Year 4 (6- to 7-years-old and 8- to 9-years-old, DfE, 2016). This early exposure to these ideas is argued to be beneficial to learning more comprehensively about evolution later in education (Peacock et al., 2011). However, learning about evolution is particularly difficult as children's preconceived ideas can act as considerable barriers to conceptual change (diSessa, 2016), as when only presented with counterevidence or formal instruction, children usually demonstrate a strong confirmation bias (Sharp, 2014). Nevertheless, there is work showing that young children (5- to 6- and 7- to 8-years-old) can understand evolutionary principles. Following a series of interventions, involving children reading a custom storybook detailing the adaptation of a species, from pre- to post-test assessment children were significantly better following the intervention at explaining trait change across generations

(Kelemen et al., 2014). This study is proof that young children can be taught to understand the process of natural selection, in terms of population change across time.

There are two main sides to the debate about the appropriateness of teleological explanations for organisms' traits (e.g., tigers' stripes are for camouflage). One side views these functional-teleological explanations as inappropriate (see for example, Galli & Meinardi, 2011; Hanke, 2004; Kampourakis 2014; Ribeiro et al., 2015), as their use could result in:

1. the use of organisms' desires as the cause of evolutionary change;
2. evolution being thought of has having a predetermined path or natural selection as a force that has a goal;
3. accounting for the existence of organisms' traits solely in terms of their function, which is not wholly scientifically correct.

Whereas the other position is that teleological exposition is extremely difficult to remove from biological explanations, as it is a convenient and a natural style of explanation (Allen 2009; Ayala, 1970; Ruse, 1989). The researchers that take this position argue that explanations for organisms' traits which focus on the function and purpose of traits are understandable, as biological traits often cannot be explained without reference to the function they provide for an organism (Amundson & Lauder, 1994). This argument seems especially pertinent for those without a high-level of formal scientific training, such as children. Following this position, there is a number of studies showing that children can use teleological explanations as a learning heuristic without sacrificing scientific rigour (Zohar & Ginossar, 1998); essentially using teleology as an additional method of description rather than a stand-alone explanation.

The literature review (section 2.5.3.3) argued that these two positions appear to have different origins, and that confusion may result from the overarching term *teleological explanation*. Therefore, what is required is sub-categorisation of teleology to define the appropriateness of an explanation. Those arguing teleology is a barrier discuss these forms of explanation in relation to organism desires or goal-directed behaviour. For example, a tiger decided to grow stripes so that it could use them to camouflage better to catch prey. This version of teleology for organisms' traits, *design-teleology*, suggests that the trait was designed, and created, by the organism (or other entity) to serve a specific purpose. Conversely, those proposing that teleological explanation can be supportive to understanding evolution, discuss these forms of explanation in relation to trait function and natural selection. For example, tigers with more pronounced stripes were able to camouflage better in their environment so could catch more prey. This version of teleology for organisms' traits, *functional-teleology*, suggests that the trait serves a function for the organisms and that the presence of the trait conferred a selective advantage. Therefore, it was argued that teleology may be useful to help account for organism traits if the explanation refers to the benefit the adapted trait has provided to the species (functional-teleology), not if they imply that the trait is a result of goal-direct behaviour (design-teleology).

7.1.2 The evolutionary scope of Study 3

There were a number of considerations relating to the evolutionary content of Study 3, namely should there be a focus on the origin of parts (traits) vs. wholes of organisms, should organisms be discussed singular vs. plural, what taxonomic level should be explored (e.g., fish or insects, plants or animals). Some of these considerations were simple to resolve. Intervention sessions and assessment had to refer to organisms in the plural form, as evolution considers species change not individual change; therefore, referring to individual organisms is misleading (Kampourakis, 2014). Furthermore, design-teleological explanation can also use plural forms as they may describe a small group of animals deciding or wanting to change.

It was decided that Study 3 would focus on the evolution of organisms' traits as opposed to the whole organisms, as a number of studies (e.g., Kampourakis et al., 2012a; Kelemen, 1999a; 1999b) have found different levels of teleological explanation for traits vs. wholes, with the former usually resulting in higher levels of teleological explanation. Consequently, the evolution of traits can be considered an area that requires more immediate intervention. Furthermore, if a child had no prior evolutionary knowledge then the focus on assessing explanations about organisms' traits is a more achievable initial learning goal, as the evolution of a single trait is a more focused discussion, compared to the evolution of an entire species.

With regard to the type of organism, plants vs. animals (vs fungi vs. protozoa), consideration of the literature (see literature review, section 2.3) suggests that the sub-category of animal life was most appropriate, as teleology as a problem is usually discussed in relation to animal evolution, not plant evolution. That is not an argument that teleological explanations are not used in relation to plant life, rather that as animals are more often discussed in the literature then they provide a suitable starting point for work attempting to limit the use of teleological explanations for the existence of organisms' traits.

Considering the evolutionary scope resulted in the Study 3 intervention exploring explanations for the existence of animals' traits. The last main consideration was the issue of trait selection. Previous studies often move between specific or general traits, between a specific animal part (e.g., a giraffe's neck or a rhino's horn) and a generic animal part (e.g., eyes or scales), without concern for typicality of traits. Currently, for specific or general traits there is no evidence that levels of teleology differ in the reviewed literature; however, there is also no evidence that they do not, it has simply not been investigated. The following section considers the existing research in relation to the impact of focusing on either specific or general traits.

Pop-science books tend to focus on distinctive organism traits, for example, Grasset's (2015) *How the Zebra got its Stripes* explores evolutionary explanations for the existence of the female hyenas' penis, giraffes' long neck and the gait of gazelles. Biological textbooks also use

atypical cases of evolution, often favouring the evolution of the modern horse or giraffes' necks (Gould, 1992). Children's books follow a similar pattern, often commenting upon rare traits in the natural world; for example, Kipling's (1908) *Just So Stories* unscientifically account for the existence of elephants' trunk and leopards' spots amongst other specific traits. This raises a number of questions: what is typically the focus of evolutionary questions and explanations? Is it better to focus on children's explanations for the existence of what could be considered specific organisms' traits – those possessed by a single species (e.g., giraffes' long necks) or focus on what could be considered generic traits – those possessed by multiple species (e.g., legs, wings, tails, etc). Are traits such as zebras' stripes is some way more iconic, more recallable, more salient. Phylogenetics (the study of evolutionary history and relationships between species) provides the necessary terminology. What has been discussed as specific traits, could be considered examples of *apomorphic* traits: derived traits which are particular to, and can often be used to define, a certain species. Apomorphioc traits will not be shared by two taxa in relation to their most recent common ancestor. Generic traits are examples of *synapomorphic* traits: ancestral characteristics which are shared by a number of different species and taxon and so can be used to define a clade of organisms. Synapomorphic traits are found in the most recent common ancestor for two different taxa. There is little research into if children respond differently when asked to account for what are referred to as apomorphic (specific) or synapomoprhic (general) traits.

Research by Tunnicliffe (1995, cited in Tunnicliffe & Reiss, 1999, p. 142) suggests that young children focus on *salient* or unfamiliar appendages when discussing animals. In a study children (5- to 14-year-olds) named organisms by their anatomy 87% of the time, that is, when asked to name a depicted organism and probed for why they named it that, most children used striking or salient features to classify the organism and consequently name it (Tunnicliffe & Reiss, 1999). Tunnicliffe & Reiss argue that this focus on salient traits results in these features becoming the ones by which children recognise and define an organism or group of organisms. Therefore, suggesting that salient traits may be better at eliciting comment and discussion than generic traits, as a child will be discussing the salient trait in relation to the specific organism under discussion, not a generic trait which they could associate with multiple organisms.

The salient traits explored by Tunnicliffe are usually examples of apomorphic traits. Consequently, Study 3 focuses on the origin of apomorphic, or salient, traits of organisms (classified as animals). However, there may be examples of apomoprhic traits which are not considered by children to be salient, furthermore, numerous examples of what children consider salient traits is lacking. One possible source may be children's story books and books designed to teach children about science. Therefore, various children's books were explored, along with previous studies into teleological explanation, for potential organism traits which are both potentially salient and apomorphic. These were used to construct a candidate set for piloting (see section 7.2.1.3).

7.1.3 Study 3 aim

Taking the preceding argument into account, the intervention aim was to develop children's understanding of what can be considered appropriate (functional-teleological/evolutionary) and inappropriate (design-teleological) explanations for the existence of organism traits. Children need to account for organisms' traits by invoking adaptedness (causal, evolutionary change) not adaptability (goal-orientated behaviour or desire on the part of the organism). This focus of the intervention is in line with calls to ensure early on that children do not use design-teleology for organisms, but rather understand the natural processes behind their existence to avoid teleological preconceptions becoming entrenched (Kampourakis 2014; Ruse, 1989; Galli and Meinardi, 2011) and, therefore, more robust when learning science at a higher level (Evans et al., 2016; Kampourakis 2007; Wood-Robinson, 1995).

The Study 2 discussion-based intervention format was considered to provide a suitable medium, as teaching and learning about evolution is more effective when children can discuss the evolutionary concepts in relation to their preconceptions (Kampourakis & Zogza, 2009). The vast majority of the intervention structure remained unchanged, but incorporated the above distinction between the two types of teleology. However, there were three further issues arising that were considered.

The first was assessing children's judgements of statements. It was decided that children would still assess Concept Cartoon statements but that these would include a statement about the function of a trait, which was supported by either a design-teleological explanation (the trait exists because of a desire on part on the organisms) or a functional-teleological explanation (the trait exists due to evolution). Relational-teleology and religious-teleology were not investigated. In reference to terminology, in this thesis, functional-teleological explanations explicitly underpinned by evolutionary concepts can be considered *evolutionary explanations*. Evans et al. (2016) refers to these forms of design-teleological explanations as *want-based explanations* (as the animal in some way *wants* to change). To avoid confusing functional explanations for a trait (functional-teleology with no reference to evolutionary concepts) with the actual development of a trait (an evolutionary explanations what makes reference to the function of a trait for a species), *trait function* is used to refer to the function of a trait, whereas the term *evolutionary explanation* is used to refer to the scientific mechanism which led to the development of a trait in a species. Furthermore, to avoid confusion with design-teleological explanations in Study 2 (where natural phenomena existed for a purpose) with the design-teleological explanations in this study (where organisms' traits exist because individual animals wanted to develop certain traits), *want-based explanations* is used to refer to these goal-oriented design-teleological explanations.

The second was the evolutionary scope of the intervention. It was decided to focus on a scaled down version of natural selection (e.g., focusing on the adaptation of one trait to an environment across a limited time span) and not to include multiple aspects of evolution (e.g.,

mutation or isolation). Support for a scaled down, but directed and precise version of natural selection is advocated by Kampourakis & Zogza (2009) and Kelemen et al. (2015). As want-based explanations deal with the emergence of a trait in an ancestral population (that a group of individuals decided to develop a trait) it was decided that the evolutionary explanations should have a similar focus. Therefore, the focus of the evolutionary explanations was how a species developed a trait, which is due to continuous variation of a trait in the ancestral population (i.e., some organisms had less pronounced versions of the trait, others had more pronounced versions but most were in the middle). This continuous variation of trait in the ancestral population would then be shaped by natural selection: the pronounced version of the trait conferred a selective-advantage. Furthermore, the existence and extent of a trait being shaped by continuous variation within a population is one of the major misunderstandings of natural selection (Gregory, 2009), and therefore, an area for educational interventions.

Using these forms of explanation there were two levels of explanation involved in Study 3, 1) the origin of a trait and 2) the prevalence of a trait. In want-based explanation the origin of a trait was number of individuals deciding to grow the trait. In this situation there would be other individuals who did not decide to grow the trait, which would mean that an explanation is needed to account for prevalence of the trait in modern animals. This explanation for the prevalence can be the same as in the evolutionary explanations (natural selection). Consequently, in these two forms of explanation the origin of a trait in the ancestral population was explained by either:

a) A want-based explanation – a group of organisms developed the trait as they wanted it to serve a specific purpose;
b) An evolutionary explanation – in a certain species there was continuous variation of a trait (some organisms had less pronounced versions of the trait, others had more pronounced versions but most were in the middle).

The explanation for the prevalence of the trait in the modern species was the same for both origins: the presence (of the pronounced version) of the trait conferred a selective advantage to those organisms, in that environment.

The third consideration was the age-range of the children. At Year 2 (6- to 7-years-old) children should have some instruction about how animals are suited to their environment (adaption) and about life cycles (inheritance), therefore, discussing why organisms are adapted to their environment (through survival and inheritance) should not introduce too many new principles (e.g., mutation). This would therefore discount Year 1 (5- to 6-years-old) as they would have had no mandatory instruction on these principles. However, Year 2 were discounted due to national examinations scheduled during the proposed data collection. Consequently, Year 3 (7- to 8-years-old) and 4 (8- to 9-years-old) were the target population for the Study 3 intervention, as they should have had some related science education and from a practical positon were available.

In summary, this introduction recapped the reclassification of types of teleology for the existence of organisms and their traits:

- Want-based explanations (or design-teleology) refers to explaining the origin of trait in relation to an evolutionary goal or desire of animals to change. Can be seen as the animals wanting to change due to self- or social-serving benefits.
- Evolutionary explanations (or functional-teleology) refers to origin of trait in relation to continuous variation of a trait in an ancestral population (some organisms had less pronounced versions of the trait, others had more pronounced versions but most were in the middle).

The proposed intervention has the aim of answering: *does a discussion-based intervention improve children's understanding that evolutionary (functional-teleological) explanations are appropriate, and that want-based (design-teleological) explanations are inappropriate, for the existence of organisms' traits?*

7.2 Method

This section outlines the design of the intervention, its preliminary study and the methods used in the main study to assess the efficacy of the intervention.

7.2.1 Preliminary study

7.2.1.1 Intervention design

Due to the effectiveness of the intervention format in Study 2 the main intervention design was maintained. The main change was an alteration to the plenary aspect, in Study 2 the plenary included children responding to a puppet who offered them statements to judge. While puppets have been argued to be an effective tool for promoting group discussion (see Keogh & Naylor, 2007) with the target age-range (8- to 9-year-olds), they were not used in Study 3. This was due to consultation with the class-teacher for the children involved in the main study, who suggested that the use of puppets would be problematic with the children (echoing the issue with two children from Study 2). Therefore, the plenary involved the same activity (children quickly evaluating statements), however, the researcher read out the statements rather than the puppet. Furthermore, it was decided that four intervention sessions were not required given the large effects shown in Study 2 and the school timetables. The decision was made to reduce the number of intervention sessions from four to two.

In terms of explanations, this study aimed to contrast want-based explanations with evolutionary explanations of the origin of traits within a species. For these types of explanation the origin of a trait is explained by either:

a) A want-based explanation – organisms developed the trait to serve a specific purpose;
b) An evolutionary explanation – In a population there was continuous variation of a trait.

In both cases the explanation for the prevalence of the trait in the current species is the same for both origins: the presence (of the pronounced version) of the trait conferred a selective-advantage to those organisms, in that environment. The wording of these two types of explanation is shown in Table 7.1, in an abstract and example form.

Table 7.1: *Explanations for the existence of trait Y, with function P, in species X*

	Want-based explanation	Evolutionary explanation
Abstract form	X's ancestors did not used to have Y. All of them were born without Y. One day some X wanted Y. They thought that Y would be better for P. So they all grew Y to do P. The Xs with Y did P more and so survived to have more babies. Their babies also had P. Over a long time all Xs were born with Y, which are better for P.	X's ancestors had a diminished version of Y. When they were born, there was variation in Y. Some Xs had less pronounced Y, some had more pronounced Y, but most were in the middle. Those with pronounced Y were better at P. The Xs with Y did P more and so survived to have more babies. Some of their babies also had P. Over a long time all Xs were born with Y, which are better for P.
Example for Giraffes' necks	Giraffes' ancestors used to have short necks. Every giraffe that was born had a short neck. One day some giraffe wanted a longer neck. They thought that a long neck would be better for eating the leaves at the tops of trees. So they grew long necks so they could eat more food. The long necked giraffes ate more food and so survived to have more babies. Their babies also had long necks. Over a long time all babies were born with long necks, which are for eating more leaves.	Giraffes' ancestors used to have short necks. When they were born, there was variation in neck length. Some giraffe shorter necks, some had longer necks, but most were in the middle. Those with longer necks were better at eating leaves at the tops of trees. The long necked giraffes ate more food and so survived to have more babies. Some of their babies had even longer necks. Over a long time all giraffes were born with long necks, which are for eating more leaves.

7.2.1.2 Preliminary aim

Due to the influence of questioned topic upon children's explanations (see Study 1), a preliminary study was conducted to find suitable discussion topics for the main study. As in Study 2, the terms

discussion topic/s and discussed topic/s are used to refer to the example/s of organisms' traits under discussion in the intervention assessment and main sessions. Suitable discussion topics were those where children did not focus their response solely on the function of a trait (i.e., what the trait is claimed to be used for) and did not disregard an explanation solely because they view the trait function as unbelievable. Furthermore, the study also trialled and allowed for refinement of the content of the intervention sessions.

7.2.1.3 Discussion topic candidate set

A candidate set of traits was selected from literature (DiYanni & Kelemen, 2005; Kampourakis et al., 2012b; Keil, 1994; Kelemen, 1999a, 1999b; Piaget, 1929; Polling & Evans, 2002), books aimed at children and through discussion with colleagues. These included an organisms' trait with two different functions. For example, that *giraffes' necks* had the trait function of 1) for reaching leaves at the tops of trees and 2) for watching out for danger. Table 7.2 shows the candidate set of traits and their functions. There are at times multiple proposed evolutionary explanations for the existence of a trait (in terms of beneficial trait function) with little consensus about the correct one, or a suggestion that the trait may serve multiple functions (Grasset, 2016). Therefore, the beneficial trait functions used in Study 3 are those which have been suggested in scientific work. Although they may not be the preferred evolutionary explanation, they are possible examples of trait functions that may have conferred a selective-advantage to those organisms. The selection criteria for trait function was simply that the function is *feasible*, no claim is made that they are the true, or only, trait function that led to the prevalence of the trait in the modern species.

Table 7.2: *Candidate traits and their functions*

Organism (X)	Specific species	Trait (Y)	Trait Function 1 (P)	Trait Function 2 (Q)
			Trait is better [insert explanation] and so survived longer to have more children	
Anteaters	Myr-mec-ophaga tri-dactyla	long tongues	for eating ants from their nests. The long tongue anteaters were better at getting more food to eat	for keep all the germs away. The long tongue anteaters were better at cleaning themselves and not getting ill
Butterflies	Danaus-plex-ippus	colourful wings	for warning predators they were poisonous. The colourful butterflies were better at warning predators and not being eaten	for camouflage on bright flowers. The colourful butterflies were better at hiding from predators and not being eaten
Camels	Camel-usdrom-edarius	hump	for storing fat. The small hump camels were able to use the stored up fat when there was no food	for storing water. The small hump camels were able to use the stored up water when there was no water
Elephants	Loxo-donta-africana	trunks	for quickly finding food. The long trunk elephants were better at getting more food to eat	for spraying cold water on their bodies. The long trunk elephants were better at keeping cool in the hot weather
Flamingos	Phoen-icopter-us roseus	long legs	for finding food as they could look at a wider area. The long legged flamingos were better at getting more food to eat	for keeping their bodies out of the mud. The long legged flamingos were better at keeping clean and germ free
Giraffes	Giraffa-camel-opar-dalis	long necks	for eating leaves at the tops of trees. The long necked giraffes ate more food	for watching out for predators. The long necked giraffes were better at spotting danger

Horses	Equus-ferus-caballus	hoof	for quickly running away from predators. The horses with hooves were better at escaping from dangerous predators	for kicking during fights with enemies. The horses with hooves were better at winning fights
Leopards	Leo-pardus-pardalis	spots (rosette)	for camouflage to help sneak up on prey. The spotty leopards were better at sneaking up and catching more prey	for camouflage to hide from predators. The spotty leopards were better at hiding from predators
Lions (males)	Panthera Leo	mane	for protection during a fight. The male lions with manes were better at not getting hurt	for looking big and scary to scare away rivals. The male lions with manes were better at scaring away rivals
Reindeers	Rangifer tarandus	antlers	for fighting enemies. The reindeers with antlers were better at wining fights and not getting hurt	for digging in the snow for food. The reindeers with antlers were better at finding more food
Rhinos	Rhinoceros uni-cornis	horn	for fighting enemies. The rhinos with horns were better at wining fights	for digging in the dirt for food. The rhinos with horns were better at finding more food
Zebras	Equus grevyi	stripes	for confusing predators. The stripy zebras were better at escaping from predators	for keeping mosquitos away. The stripy zebras were bitten less and got less illness
Jellyfishes	Chry-saora fus-cescens	stinging tentacles	for capturing prey. The jellyfish with stinging tentacles were better at catching more food to eat	for fighting predators. The jellyfish with stinging tentacles were better at wining fights with predators

Octopuses	Octopus vulgaris	suckers	for capturing fish. The octopuses with suckers were better at catching more food to eat	for breaking shellfish apart. The octopuses with suckers were better at breaking shells apart to get more food
Peacocks (male)	Pavo cristatus	colourful feathers	for scaring predators. The male peacocks with colourful feathers were better at scaring away dangerous predators	for attracting female peacocks. The male peacocks with colourful feathers were better at attracting females and so had more children.
Rabbits	Oryct-olagus cuni-culus	long ears	for hearing for danger. The rabbits with long ears were better at listening out for predators	for keeping cool. The rabbits with long ears were better at keeping cool in the hot weather
Snakes	Vipear aspis	poisonous fangs	for capturing prey. The snakes with poisonous fangs were better at catching more food to eat	for fighting enemies. The snakes with poisonous fangs were better at winning fights with enemies
Spiders	Argiope aurantia	silk	for building webs to catch prey. The spiders with silk were better at catching more food to eat	for anchoring themselves to surfaces. The spiders with silk were better at attaching themselves to high up places so they didn't fall off
Squirrels	Sciurus caro-linensis	fluffy tail	for balancing on thin branches. The squirrels with fluffy tails were better at balancing in high up places so they didn't fall off	for keeping warm. The squirrels with fluffy tails were better at keeping warm when it was very cold weather

7.2.1.4 Preliminary design

The preliminary structure was a shorted version of the intervention structure used in Study 2, comprising of an introduction, a main discussion of a Concept Cartoon and a plenary where children quickly evaluated additional statements. The pre-/post-test procedure remained the same. Combined the study structure was:

- Pre-test – children evaluate 12 statements across three discussion topics (four statements per topic)
- Session 1 – children explore two different discussion topics as a group
- Session 2 – children explore two further different discussion topics as a group
- Post-test – children re-evaluate the 12 statements across three discussion topics (four statements per topic)

7.2.1.5 Preliminary participants

Nine children, aged 8- to 9-years-old (Female = 4) took part, from two Year 4 classes. No children was reported to have SEN or EAL. The school was a two-form entry from a large city in the Midlands of the UK. It was judged to be performing excellently on national scales, and is below average for both pupils from minority ethnic groups and those entitled to free school meals. Seven of the children had one parent who worked at the university. The class teacher inadvertently introduced a sampling bias by selecting those children who were most interested in science. Therefore, the majority of the sample already had some knowledge of evolution at pre-test.

7.2.1.6 Preliminary procedure

Children were split into three groups. Each group were assessed using three different organism traits and their corresponding explanations, as detailed in Table 7.2. Therefore, the pre- and post-test assessment covered nine different organisms' traits, with each child being questioned about 12 explanations for the existence of these traits. For each explanation children were presented with a Concept Cartoon depicting the discussion topic (an organisms' trait) and explanation for its existence, they were then asked to decide whether it was a good or bad explanation for how the species has changed over time. Then were asked to provide a rationale for their judgement. In terms of intervention activities, each group took part in two intervention sessions, these sessions lasted approximately 20 minutes. Each session explored two organisms' traits and their corresponding explanations. The two intervention sessions comprised of:

- Introduction – children (in groups of three) were reminded on the focus of the previous session/pre-test: that they were exploring different explanations for why species have changed over time – good or bad scientific explanations;

- Main discussion – children were presented with a Concept Cartoon, which displayed four possible explanations about a certain trait. They discussed the statements on the Concept Cartoon with teacher guidance. This ended with a summary of the commonalities and differences between the good and bad explanations;
- Evaluation – children were then read four other explanations by the researcher. Children quickly judged these new statements as good (appropriate) or bad (inappropriate), while linking back to the main discussion conclusions.

In total, the pilot investigated the suitability of 19 traits (9 at assessment and 10 during the intervention sessions). Traits from the candidate set (Table 7.2) were randomised to the assessment and interventions sessions, which were then randomised to participant group. Finally, the order of the four explanations for each trait were randomised.

7.2.1.7 Preliminary measures

Children's responses at pre- and post-test were coded on two scales:

- If their initial judgement of the explanation was correct, see Table 7.3.
- On the content of their rationale. If children correctly supported their judgement, with reference to the variation of a trait in an evolutionary explanation or to the problems associated with a want-based explanation, they scored one. All other rationales scored zero (e.g., a non-response or simply re-reading the explanation).

Table 7.3: *Coding rubric for judgement of statements*

	Participant judgement	
Statement	Good	Bad
Want-based explanation	Score = 0	Score = 1
Evolutionary explanation	Score = 1	Score = 0

7.2.1.8 Preliminary results and discussion

It should be noted that the vast majority of children's responses did comment upon trait function. However, this was usually in reference to how the trait function was appropriate, with children often stating that the trait function made sense, that they had seen the organism using the trait in the stated way. In these cases children then proceeded to analyse the Concept Cartoon explanation based on the explanation for the origin of the trait. Table 7.4 and Table 7.5 shows those traits where children rejected the explanation based on the function of the trait, where their response did not consider the origin of the trait. For example:

- [lions' manes are for menacing] is 'silly because how come something be scared of a mane, I think it makes them look nicer';
- [peacocks' feathers are for scaring predators] is 'not true cos how are they going to scare predators away with a tail, it just means they will get spotted easier';
- [zebras' stripes are for mosquito repellent] is silly because 'how can stripes get rid of mosquitoes! I can't wear a stripy top and just not get bit, that won't work'.

The results in Table 7.4 and Table 7.5 show there are only seven examples where both trait functions were not a problem. These are reindeers' antlers, giraffes' necks, rhinos' horns, camels' humps, rabbits' ears, leopards' spots and elephants' trunks.

Table 7.4: *Traits using during the assessments – did children focus on function?*

Organism	Trait	Function	Focus on function?
zebras	stripes	for confusing predators	
		for mosquito repellent	Yes
octopuses	suckers	for catching fish	
		for prying shells apart	Yes
anteaters	tongue	for eating ants	
		for cleaning	Yes
jellyfishes	stings	for catching prey	
		for fighting	Yes
horses	hoof	for running away	Yes
		for kicking	Yes
reindeers	antlers	for fighting	
		for finding food	
giraffes	neck	for eating leaves	
		for watching for danger	
peacocks	feathers	for scaring predators	
		for attraction	Yes
spiders	silk	for catching prey	
		for anchoring	Yes

Table 7.5: *Traits using during the sessions – did children focus on function?*

Organism	Trait	Function	Focus on function?
rhinos	horn	for fighting	
		for finding food	
camels	hump	for storing fat	
		for storing water	
snakes	venom	for catching prey	Yes
		for fighting	Yes
butterflies	patterns	for warning predators	Yes
		for camouflage	
flamingos	legs	for finding food	
		for keeping clean	Yes
squirrels	tail	for balancing	Yes
		for warmth	
lions	mane	for protection	
		for menacing	Yes
rabbits	ears	for hearing danger	
		for keeping cool	
leopards	spots	for camouflage (prey)	
		for camouflage (predators)	
elephants	trunk	for finding food	
		for spraying water	

For children's judgements, their average score improved from 66.6% (*SD* = 24.6) correct judgements at pre-test to 95.3% (*SD* = 8.4) correct judgements at post-test. A score increase of 28.7 percentage points (95% CL 9.25 to 48.1). A dependent *t*-test was conducted and the results showed a significant score difference, t (8) = 3.30, $p.$ = .008, with a strong effect size, d_{rm} = 2.24. Children's understanding of the elements of appropriate explanations about species change over time improved from 38.8% (*SD* = 28.8) correct at pre-test to 78.7% (*SD* = 13.2) correct rationales at post-test. A score increase of 39.9 percentage points (95% CL 16.5 to 63.0). A dependent *t*-test was conducted and the results showed a significant score difference, t (8) = 3.76, $p.$ = .003, with a strong effect size, d_{rm} = 2.48. Therefore, the results suggest that following the intervention, children's understanding of the elements of appropriate and inappropriate explanations improved considerably.

One observational finding was that the assessment procedure was overly long, at times lasting 15 - 20 minutes. While this did not appear impact upon children's performance or attention, it is a problem for assessing a full class of children within a busy school timetable. For this reason the decision was made to reduce the number of assessment Concept Cartoons from 12 to 8, in order to further limit time needed to assess each child. It is maintained that this lower number of measures is sufficient due to the considerably large effect size found in Study 2. Furthermore, the Concept Cartoon statements were shorted. This editing primarily focused on the first line of each explanation, with the aim to avoid repetition in statements. The differences between the old and new forms of explanation are shown in Table 7.6 (want-based explanations) and Table 7.7 (evolutionary explanations).

Table 7.6: *Comparison of old and new want-based explanation wording*

	Old	New
Want-based (abstracted)	X's ancestors did not used to have Y. All of them were born without Y. One day some X wanted Y. They thought that Y would be better for P. So they all grew Y to do P. The Xs with Y did P more and so survived to have more babies. Their babies also had P. Over a long time all Xs were born with Y, which are for P.	X's ancestors were all born without Y. However, some X's thought that Y would be good for P. So they started to grow Y to do P. The Xs with Y did P more and so survived to have more babies. Their babies also had Y. Over a long time all Xs were born with Y, which are for P.
Want-based (example)	Camels' ancestors did not used to have humps. All of them were born without humps. One day some camels wanted a hump. They thought that humps would be better for storing fat in. So they all grew humps for storing up fat for when there was no food. The camels with humps stored more fat and so survived to have more babies. Their babies also had humps. Over a long time all camels were born with humps, which are better for storing up fat for when there was no food.	Camels' ancestors were all born without humps. However, some camels thought that humps would be good for storing fat in. So over a number of years they all started to develop humps for storing up fat for when there was no food. The camels with humps stored more fat and so survived to have more babies. Their babies also had humps. Over a long time all camels were born with humps, which are for storing fat for when there was no food.

Table 7.7: *Comparison of old and new evolutionary explanation wording*

	Old	New
Evolutionary (abstracted)	X's ancestors had a diminished version of Y. When they were born, there was variation in Y. Some Xs had less pronounced Y, some had more pronounced Y, but most were in the middle. Those with pronounced Y were better at P. The Xs with Y did P more and so survived to have more babies. Some of their babies also had P. Over a long time all Xs were born with Y, which are for P.	When X's ancestors were born, there was variation in Y. Some Xs had less pronounced Y, some had more pronounced Y, but most were in the middle. Those with pronounced Y were better at P. The Xs with Y did P more and so survived to have more babies. Some of their babies also had Y. Over a long time all Xs were born with Y, which are for P.
Evolutionary (example)	Camels' ancestors had small humps. When they were born, there was variation in hump size. Some camels had small humps, some had larger humps, but most were in the middle. Those with large humps were better at storing fat for when there was no food. The camels with large humps stored more fat and so survived to have more babies. Some of their babies also had large humps. Over a long time all camels were born with humps, which are better for storing up fat for when there was no food.	When camels' ancestors were born, there was variation in hump size. Some camels had small humps, some had larger humps, but most were in the middle. Those with large humps were better at storing fat for when there was no food. The camels with large humps stored more fat and so survived to have more babies. Some of their babies also had large humps. Over a long time all camels were born with humps, which are for storing up fat for when there was no food.

7.2.1.9 Implications for main study

The primary aim of the preliminary study was to avoid using explanations where trait functions were so controversial as to take children's focus away from evaluating the reasons for the origin of the trait. The preliminary study resulted in seven traits (each with two trait functions) that satisfied this criteria. However, there were a number of other examples where one trait function was seen as realistic by the children but the second trait function was not, for example, lions' manes being used for menacing versus being used for protection (see Table 7.5 for further examples). It was decided that these realistic Concept Cartoon statements (ones where the trait function was believable) would be used in the plenary portion of the intervention, in what was termed a *mixed-bag*, the split of traits can be seen in section 7.2.2.5.

The mixed-bag of discussion topics included all of the organism traits' where children viewed both possible trait functions as realistic (e.g., giraffes' necks) and the organism traits were children's view only one trait function as realistic (e.g., peacocks' feathers), see Table 7.4 and Table 7.5 for full list of discussion topics. Essentially this mixed-bag contained all of the trait functions that children thought were believable, ones for which the trait could actually be used.

In the main study explanations from this mixed-bag of traits were randomly drawn during the plenary phase of the intervention, where the group quickly evaluates a number of additional explanations. This mixed-bag allowed children to use what has been discussed on a wider range of examples, to abstract over multiple examples. This model of intervention also bears more similarity to the Study 2 intervention, where children investigated a Concept Cartoon then evaluated multiple examples posed by the puppet.

The preliminary study did suggest that the intervention works, following the intervention sessions there was significant improvements in children's understanding of what were appropriate (evolutionary) versus inappropriate (want-based) explanations for species change over time. While children were particularly competent at the pre-test, scoring over 60% correct for their initial judgement and almost 40% correct for their rationales, they did not represent a typical Year 4 sample. While the class teacher did provide a range of ability (as requested) this range was in relation to their English and mathematics skill, not their science skill. In fact the children selected were those most interested in science, therefore, inadvertently biasing the sample. As a consequence the observed high pre-test scores were not expected to be repeated in the main study.

A further note on assessment, children's judgement score can be inflated if they correctly judge a want-based explanation as bad, but then go on to support this with a rationale which focuses on dislike of a trait function (e.g., they do not believe that zebras stripes' help to keep away mosquitoes). Therefore, children's rationale scores are a better measure of actual intervention impact. As children's rationales can only be correct if they focus on the origin of the trait (natural variation or organism desire), and those rationales which support their judgement by discussing the function of a trait are not counted in the measurement. Consequently, both judgement and rationale

scores were used in the main study, but the primary measure of effectiveness was children's rationale score.

7.2.2 Main study

7.2.2.1 Aim

This study investigated situations where the existence of a trait is explained with relation to functional-teleology (i.e., a purpose it performs for a species). It asked the question *do discursive activities about the appropriateness of scientific explanations develop children's understanding of appropriate (evolutionary) and inappropriate (want-based) mechanisms for the existence of organisms' traits?* In these explanations the origin of a trait (mechanism) is explained by either:

a) a want-based explanation – organisms developed the trait to serve a specific purpose;

b) an evolutionary explanation – in a population there was continuous variation of a trait.

The explanation for the prevalence of the trait in a current species is the same for both trait origins: the presence (of the pronounced version) of the trait conferred a selective-advantage to those organisms, in that environment. It is hypothesised that:

- At pre-test children in both conditions would incorrectly endorse want-based explanations for the origin of organisms' traits;
- At pre-test children would demonstrate little understanding of the elements of appropriate (evolutionary) and inappropriate (want-based) explanations for the origin of organisms' traits;
- Following the intervention, the treatment group (only) would suggest want-based explanations are inappropriate and would understand the elements of what constitute appropriate and inappropriate explanations for the origin of organisms' traits.

7.2.2.2 Design

The experiment used a pre-/post-test experimental design with a control group. A true experimental design was used, as opposed to a quasi-experimental design, as the sample consisted of only one Year 4 class. A no treatment control group was used, as there was no comparable treatment to compare the intervention against.

7.2.2.3 Participants

24 children, aged 8- to 9-years-old (109 to 118 months-old, M = 114 months-old, SD = 7.65), took part in the study (Female = 13, English as an Additional Language = 5, Special Educational Needs = 7, Free School Meals = 2). Children were from a single Year 4 class in a primary school in a large town in the East Midlands of the UK. The school was single-form entry, was judged to be performing well on national scales, and was below the national average for both pupils from minority ethnic groups and those entitled to free school meals. Children were randomly assigned to

either the treatment group (12 children, Female = 6 EAL = 3, SEN = 4, FSM = 1) or the control group (12 children, Female = 7 EAL = 2, SEN = 3, FSM = 1).

7.2.2.4 Procedure

The procedure was similar to the preliminary study, pre-test assessment began with the child being told about who the interviewer was, where they had come from and the purpose of the interviewer's visit: 'to talk and think about some different science topics'. Participants were then asked if they were happy to answer some questions about how different animals have changed over time and for the conversation to be recorded. Participants were told:

> *that animal's ancestors, those that lived a very, very long time ago, like the great-great-great-great-grandparents of modern animals, used to look rather different. For example, giraffes ancestors' used to have much shorter necks, but over time they changed and now modern giraffes have long necks. What I have here are a number of cartoons showing explanations for how species have changed, and I just want to you tell me if you think they are good or bad explanations for how species have changed over time.*

The researcher then proceeded to read out the explanations on the assessment Concept Cartoons. For each Concept Cartoon children were asked:

- to judge the explanation as good (appropriate) or bad (inappropriate) – this allowed for the assessment of children's initial reaction regarding their perceived correctness of an explanation;
- to provide a rationale for their judgement – this allowed for the assessment of children's understanding of, and skill in articulating, the elements of appropriate or inappropriate explanations.

At post-test assessment, children were merely reminded about the pre-test were they had to think about good and bad explanations for how species have changed over time. Participants took part in assessment in register order.

All the children (both intervention and control group) were assessed, before and after the intervention, using eight questions for two organisms' traits (reindeers' antlers and giraffes' necks), each with two trait functions (see Table 7.8). The order of the four explanations for a single organisms' trait was randomised for each participant, as was the order of the two organisms.

Following the pre-test children were randomised to the treatment or control group. To recap, the intervention sessions comprised of:

- Introduction: children (in groups of four) were presented with a Concept Cartoon, which displayed two want-based and two evolutionary explanations for the existence of a species'

trait (Session one = elephants long trunks, Session two = rabbits long ears), these four explanations covered two different trait functions;

- Main discussion: children discussed the statements on the Concept Cartoon with teacher guidance, after discussing all four explanations the group discussed commonalities between good and bad explanations;
- Plenary: additional explanations from the mixed-bag were read out to children who were asked to quickly judge the explanations as good (appropriate) or bad (inappropriate), in light of the prior discussion.

The children in the treatment group were randomised into groups of four for each of the two sessions. The study took place across five weeks at the end of the Summer term (June - July). Assessment took place during the first and fifth weeks, and interventions sessions were conducted during the second and third weeks. Pre- and post-test individual assessment sessions lasted approximately 12 minutes, and group intervention sessions lasted 15-20 minutes. Assessment and interventions took place in the school library, which was unused during the sessions. All assessment was audio recorded and transcribed for coding.

7.2.2.5 Materials

Table 7.8 shows the organisms' traits and corresponding trait functions used in the Concept Cartoons. This is split by assessment, intervention main discussion and intervention mixed-bag. Want-based and evolutionary explanations were wording according to the abstract explanation wordings detailed in Table 7.6 and Table 7.7. An example assessment Concept Cartoon is provided in Figure 7.1, assessment Concept Cartoons were split by sex, with female participants receiving assessment Concept Cartoons with only female characters represented, the same practice was conducted with male participants. This was done to avoid an influence of gender, that is, to avoid children judging a Concept Cartoon based on the gender of the depicted Concept Cartoon character. An example lesson Concept Cartoon is provided in Figure 7.2 along with an accompanying lesson plan in Figure 7.3. As the intervention sessions were mixed in terms of participant gender, the gender of the Concept Cartoon characters were randomised to want-based and evolutionary explanations for the intervention session Concept Cartoon prompts. The full set of Concept Cartoons are provided in Appendix 4.

Table 7.8: *Discussion topics used in the main study*

Pre-test and post-test discussion topics		
Organism	Trait	Function
reindeers	antlers	for fighting
		for finding food
giraffes	neck	for eating leaves
		for watching for danger
Intervention discussion topics (main discussion)		
Organism	Trait	Function
elephants	trunk	for finding food
		for spraying water
rabbits	ears	for hearing danger
		for keeping cool
Intervention discussion topics (mixed-bag)		
Organism	Trait	Function
camels	hump	for storing fat
		for storing water
leopards	spots	for camouflage (as prey)
		for camouflage (as predators)
rhinos	horn	for fighting
		for finding food
anteaters	tongue	for eating ants
butterflies	patterns	for camouflage
flamingos	legs	for finding food
jellyfishes	stings	for catching prey
lions	mane	for protection
octopuses	suckers	for catching fish
peacocks	feathers	for scaring predators
spiders	silk	for catching prey
squirrels	tail	for warmth

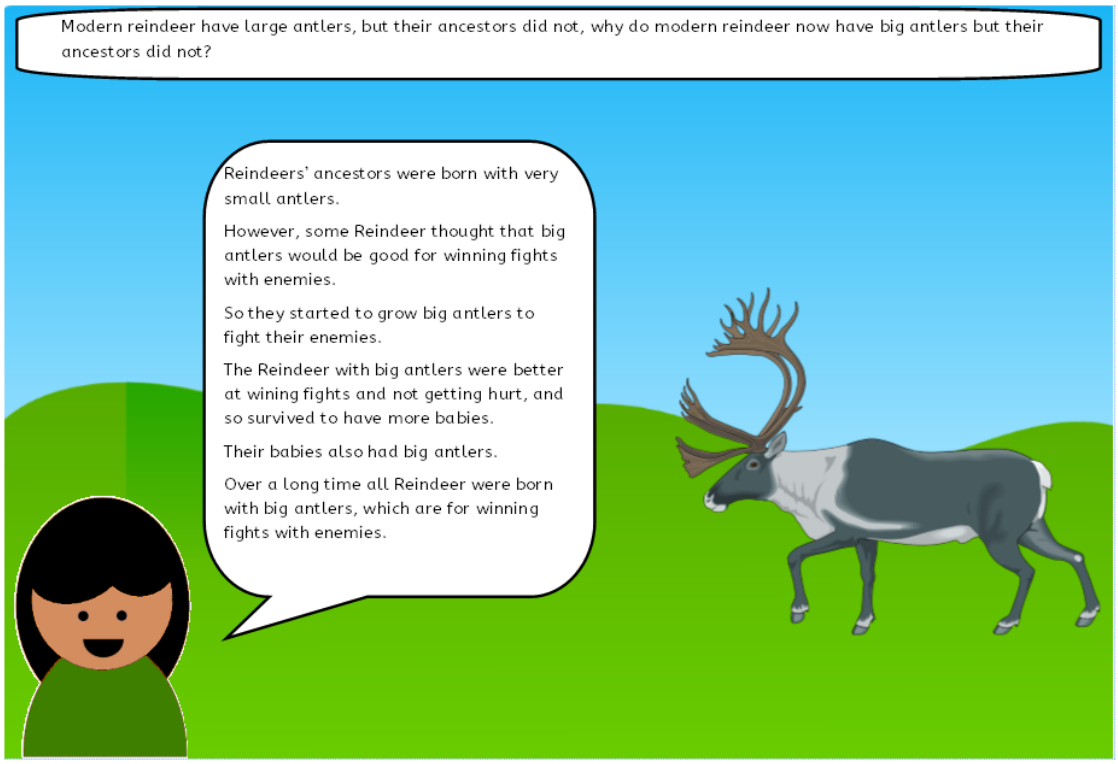


Figure 7.1. Example assessment (Female) Concept Cartoon for Study 3: Want-based explanation 1 for why reindeer have big antlers.



Figure 7.2. Intervention session 1 Concept Cartoon for Study 3: session prompt for why rabbits have long ears.

Groupings	Group __:	Date:
Session ___	Topic_____	Time:

What to do:

Introduction (3 minutes)

What were we doing last week, can anyone recap for us?
Remind chn about why I am here - talking about why things happen
What were good explanations – one that explain how X got Y
What were bad explanations – they don't really explain how X got Y
Introduce children to the 1st topic for today.
Show and explain the concepts cartoon
It shows some people thinking about why x have Y

Main discussion (10 minutes)

Continued to discuss the topic and explanations - use questions to guide the discussion. Encouraged chn to consider:

- What is the function of the topic;
- What caused it to develop;
- Do these explanations logically account for the existence of Y?

Encourage debate and a range of views from all children.
Introduce questions from box if children struggle.
Teacher to summarise children's discussion, what views were similar and different across the four statements. What conclusions can be drawn, regarding their views on a) function and b) development of trait.

Evaluation of statements (5 minutes)

Similar to the main discussion, but go through a number of the mixed bag cartoons, time permitting. Apply conclusions from main discussion
Chn to use thumbs up /down for good/bad explanation
Then quickly check their views, on development and function - can refer to the summary in previous section.

Resources: Recorder, relevant concept cartoons, A3 paper for children's ideas,

Questions for clarification

- What do you think? (to new child);
- Did everyone hear that, is there anything anyone wants to add?;
- So you're saying ……? (to add clarification to idea);
- How does this relate/differ to X's idea before? (to link ideas back);
- That reminds me of an idea I heard a while ago, that … (to add more examples);
- Who thinks this a good idea?;
- Are there any problems with this idea?;
- Take your time/ have a little think/ is there anything you want to add/ can you tell us a little bit more about X?;

Questions for probing ideas

- Can anyone think of anything else X is used for? (to search for counter examples);
- Do you think it explains how X got Y?
- If X did use Y to do P/Q would W/E be a good mechanism?
- What's the worst part (of the explanation)?
- Any other way X could get Y?
- Why might people say an animal wants Y?

Figure 7.3. Generic lesson plan for Study 3

7.2.2.6 Measures

Children's judgments were coded using the rubric in Table 7.3, their rationales for their judgements were coded using the rubric in Table 7.9. Approximately 10% of the data (40 responses) were coded by a second researcher. Cohen Kappa indicated a very good level of agreement ($K = .931$, $p < .001$).

Table 7.9: *Coding rubric for children rationales of their judgements*

Score	Explanation type	Example
1	The rationale for the judgment of a statement discusses the style of explanation used in the statement. Specifically, an evolutionary explanation is good because it offers a causal explanation to how the trait developed – through variation in the population and natural selection.	[Evolutionary explanation for giraffes necks] is good cos says there is variation in the past, and those with long necks lived long and the short necks ones died out
	For a want-based explanations, the statement is bad because it does not offer a causal explanation for how the trait developed. Instead it states that some of the species decided that the trait would be beneficial, or wanted the trait for self-serving purposes.	[Want-based explanation for reindeer antlers] is bad because it is just saying that they wanted bigger antlers and just decided to start growing them
0	Any other form of rationale, including but not limited to, a non-response; Repeating: The rationale for the judgment of an statement is a repeated version of the Concept Cartoon explanation;	[Evolutionary explanation for giraffes necks] is good because it means that the giraffes can get more food to eat
	Trait function: The rationale for the judgment of an explanation focuses solely on the function which the trait performs, this function could be endorsed or refuted; Axiomatic: The rationale for the judgment simply states the statement is true/false, no additional information is provided.	[Want-based explanation for reindeer antlers] is good cos that is a fact, everyone know that they fight with them

7.2.2.7 Ethics

The study was conducted within the ethical guidelines from the University of Nottingham (UoN, 2015) and the British Educational Research Association (BERA, 2011). After consent had been provided by the school's head teacher and Year 4 class teacher, opt-out consent forms were sent to

parents. The rationale for opt-out consent is discussed in section 3.4. In summary, the opt-out consent was only considered to be provisional consent; therefore, after being informed about the purpose of the interview, children were asked if they agree to take part. If they dissented the interview would not take place. The interview would have been stopped if a child indicated dissent through discomfort or withdrawal, although this did not happen.

7.3 Results

7.3.1 Data checking

To test the efficacy of the intervention two variables were created: judgment score and rationale score, these were participants' mean score for their eight judgements or eight rationales; each with score range of 0 – 8. These variables exist at both pre-test and post-test[4]. To test for differences between the treatment and control groups' judgements and rationales scores at pre-test, two independent *t*-tests were conducted. The independent variable was group, with two levels (control and treatment), the dependent variable was judgement score or rationale score, for the respective tests. Comparing the two groups judgment scores showed no significant difference, t (22) = 1.75, $p.$ = .094, d = .71, comparing the two groups rationale scores also showed no significant difference, t (22) = 1.11, p = .278, d = .46. Consequently, there was no significant difference between the two groups performance (control and treatment) at pre-test.

To explore the influence of discussion topic, which was shown to be a factor in levels of teleological explanation during Study 1 (section 4.3.2), Cochran's Q tests were used to explore differences in the levels of correct judgments at pre- and post-test. The maximum-corrected measure (η^2_Q) is provided as an effect size (Serlin, Carr & Marascuillo, 1982). These tests were used to assess if children were treating the eight statements about four trait functions, for two organisms as a homogeneous group or if there was variation in their responses levels between the eight statements. As there was no difference between the two experimental groups at pre-test they are collapsed and analysed together. While in Study 1 multiple chi-square goodness-of-fit tests were used to further investigate significant Cochran's Q tests (see section 4.3.2) that procedure could not be used with this dataset, as high levels of correct judgements for evolutionary explanations would have been offset by the low levels of correct judgement for want-based explanations. Therefore, comparing individual discussed topics against the discussed topic mean would have skewed the results of the chi-square test. Consequently, descriptive statistics are provided.

[4] At pre-test children's rationale score deviated from normality for both the control (W = .675, df = 12, $p.$ < .001) and treatment group (W = .465, df = 12, $p.$ < .001), see section 3.3.4 for discussion on normality of data.

Table 7.10: *Correct judgements and rationales at pre-test by explanations type*

Explanation type	Percentage of correct pre-test judgments	Percentage of correct pre-test rationales
Giraffe want-based 1	20.8%	0.0%
Giraffe want-based 2	29.1%	4.1%
Giraffe evolutionary 1	70.8%	0.0%
Giraffe evolutionary 2	75.0%	8.3%
Reindeer want-based 1	16.6%	4.1%
Reindeer want-based 2	29.1%	4.1%
Reindeer evolutionary 1	54.1%	4.1%
Reindeer evolutionary 2	75.0%	8.3%

At pre-test, Cochran's *Q* showed that the children's judgements of explanations were not consistent across the statements (*Q* (7) 42.35, *p*. < .001, η^2_Q = .25). The descriptive statistics in Table 7.10 would indicate that this was because teleological statements were often incorrectly endorsed by children. This is in line with the hypothesis. For children's pre-test rationales Cochran's *Q* showed no significant difference between the explanations, *Q* (7) 4.30, *p*. = .744, *n* = 24, η^2_Q = .02. As Table 7.10 shows this is due to almost all of children's rationales showing no understanding of appropriate or inappropriate styles of explanation. While there is some influence of explanatory type being judged (want-based or evolutionary), what these results suggest is that children's responses were not influenced by trait function. For example, if children argued against the idea that giraffes necks were used for looking out for predators (trait function 2) but accepting of the notion that giraffes necks were used for reaching leaves at the tops of trees (Trait function 1) then there should be a large difference in the two want-based judgements. As children would not endorse the want-based explanation for trait function 2 (leading to higher correct judgements) but would endorse the want-based explanation for trait function 1 (leading to lower correct judgements). There would also be a similar but opposite trend for the two evolutionary explanations. These trends are not apparent in the pre-test judgements. Essentially, the method used to select the intervention discussion topics avoided the bias associated with questioned topics, as found in Study 1.

7.3.2 Children's judgements

A mixed ANCOVA was conducted to investigate the influence of the intervention on children's judgements about the existence of organisms' traits. The within-groups independent variable was assessment, with two levels (pre-test and post-test), the between-groups independent variable was experimental group (control or treatment). The covariate was children's age in months, which was mean centred to avoid compromising the main effect (Enders & Toffig, 2007). The dependent variable was participants' judgement score.

The analysis revealed a main effect of assessment (pre-test vs. post-test) on children's judgement scores, F (1, 21) = 48.63, p. < .001, η_p^2 = .698. There was also a main effect of group (control vs. treatment), F (1, 21) = 14.30, p. = .007, η_p^2 = .296, and a significant interaction between assessment and group upon children's correct judgments, F (1, 21) = 43.52, p. < .001, η_p^2 = .615. There was no significant effect of the covariate (mean centred age in months), F (1, 21) = .30, p. = .589, η_p^2 = .014, nor a significant interaction between assessment and age, F (1, 21) = 0.13, p. = .714, η_p^2 = .007.

Multiple pairwise comparisons were conducted to further investigate these results, Bonferroni corrections were applied to control for multiple tests. The independent *t*-tests showed no significant difference between the groups pre-test scores, t (22) = 1.75, p. > .999, d = .71, however, there was a significant difference between the groups post-test scores, t (22) = 5.92, p. < .001, d = 2.42. The paired *t*-tests showed a non-significant difference in the control group's judgement score from pre- to post-test, t (11) = .60, p. > .999, d_{rm} = 0.18; however, there was a significant difference in the treatment group's judgement score from pre- to post-test, t (11) = 12.93, p. < .001, d_{rm} = 3.98.

Table 7.11: *Mean judgement scores[1] at pre- and post-test, by experimental group (Max. score = 8)*

	Control Group (n = 12)	Treatment Group (n = 12)
Mean Pre-test score (*SE*)	4.05 (0.31)	3.19 (1.13)
Mean Post-test score (*SE*)	4.34 (0.39)	7.48 (0.39)

[1] covariate (mean age in months) evaluated at: -.08

These results and the group means in Table 6.5 indicate that following the intervention the treatment group were significantly better at correctly judging statements (from 39.8% to 93.5% correct), whereas the control group showed a non-significant increase in judgment score (from 50.6% to 54.2% correct).

7.3.3 Children's rationales

A mixed ANCOVA was conducted to investigate the influence of the intervention on children's understanding of, and skill in articulating, the elements of appropriate, or inappropriate, explanations about the existence of organisms' traits. The within-groups independent variable was assessment, with two levels (pre-test and post-test), the between-groups independent variable was experimental group (control or treatment). The covariate was children's age in months, which was mean centred to avoid compromising the main effect. The dependent variable was participants' rationale score.

The analysis revealed a main effect of assessment (pre-test vs. post-test), on children's rationale scores, F (1, 21) = 85.33, p. < .001, η_p^2 = .803. There was also a main effect of group

(control vs. treatment), F (1, 21) = 25.64, p. < .001, η_p^2 = .550, and a significant interaction between assessment and group upon children's correct rationales, F (1, 21) = 53.53, p. < .001, η_p^2 = .718. There was no significant effect of the covariate (mean centred age in months), F (1, 21) = .03, p. = .856, η_p^2 = .002, nor a significant interaction between assessment and age, F (1, 21) = 0.27, p. = .603, η_p^2 = .013.

Multiple pairwise comparisons were conducted to further investigate these results, Bonferroni corrections were applied to control for multiple tests. The independent *t*-tests showed no significant difference between the groups pre-test rationale scores, t (22) = 1.11, p. > .999, d = .46, however, there was a significant difference between the groups post-test rationale scores, t (22) = 1.75, p. < .001, d = 2.83. The paired *t*-tests showed there was no significant difference in the control group's rationale scores from pre- to post-test, t (11) = 1.14, p. > .999, d_{rm} = 0.18; however, there was a significant difference in the treatment group's rationale score from pre- to post-test, t (11) = 13.21, p. < .001, d_{rm} = 3.98.

Table 7.12: *Mean rationale scores[1] at pre- and post-test, by experimental group (Max. score = 8)*

	Control group *(n* = 12)	Treatment group (*n* = 12)
Mean pre-test score* (SE)	0.45 (0.16)	0.13 (0.16)
Mean post-test score* (SE)	0.90 (0.44)	5.01 (0.44)

[1] covariate (mean age in months) evaluated at: -.08

These results and the group means in Table 6.6 show the control group did not significantly differ in their rationale scores from pre- to post-test (from 5.6% to 11.2%). However, following the intervention the treatment group had a substantial significant increase (from 1.6% to 62.6%) in providing rationales that demonstrated awareness of the elements of appropriate, or inappropriate, explanations for organisms' traits. Examples of these types of rationales include:

- [Giraffes evolutionary 1] is good 'cos they survived if they had long necks'.
- [Giraffes evolutionary 2] is good 'cos like small giraffes, they can't see much so it's hard to survive for them, but not the long necks'.
- [Giraffes want-based 1] is bad because 'it's just saying they decide to grow it. You need to put how, how, how do they grow it, how. Cos I want wings, I want a money tree as well coming out my butt, hmmm, see, see it's not happening, where are my wings, where's my money tree?'.
- [Giraffes want-based 2] is bad because 'they can't just start to grow long necks'.
- [Reindeers evolutionary 1] is good because 'like there is a variation in size, so you can just look at the size and they can just fight, the ones with the big antlers will fight but the small ones will have a bit of a disadvantage'.

- [Reindeers evolutionary 2] is good because ‘I just like that again cos it’s not saying they just started to grow but that there was variation’.
- [Reindeers want-based 1] is bad because ‘for fitting with enemies is good, cos I’ve seen that, but cos they started to grow antlers, it doesn’t make any sense’.
- [Reindeers want-based 2] is bad because ‘nothing can grow really, we can grow but antlers on reindeers, they wouldn’t just go let’s grow antlers and they would pop out’.

7.4 Discussion

7.4.1 Assessment results

Following the success of the Study 2 intervention (chapter 6), this study attempted to use the same generic intervention structure to improve:

- Children’s correct judgements of want-based (design-teleological) and evolutionary (scientific) explanations about the existence of organisms’ traits;
- Children’s understanding of the elements of appropriate, and inappropriate, explanations about the existence of organisms’ traits.

A concern raised by the results of Study 1 was that trait function may influence children’s responses. For this reason a preliminary study was conducted to find trait functions which would be immune to the bias associated with within-category variance (i.e., children’s responses to questions about the trait functions were consistent) and would feasible trait functions (i.e., the function would not be rejected immediately due to a perceived lack of viability). The preliminary results and the initial data checking (section 7.3.1) suggested that the trait functions used in this study were immune to within-category variance and were viewed by children as feasible trait functions.

With regard to the analysis of the intervention the results showed that the intervention was successful in achieving the two intertwined aims. Children can be taught to recognise that want-based explanations are inappropriate to account for the development of organisms’ traits. These two aims were measured by children’s judgement scores and their rationale scores. With regard to children’s judgement scores, the treatment group showed a considerable increase in their correct judgments about the development of organisms’ traits, from 41% correct to 94% correct. However, the control groups mean score shows little change, from 50% to 54% at pre- to post-test. The analysis indicated that the condition to which children were assigned had a significant effect upon their score improvement from pre- to post-test. Pairwise comparisons showed that the treatment group had a significant score increase with considerable effect size, whereas the control groups score increase was nonsignificant.

Considering the influence of explanatory type, at pre-test there was a significant difference in levels of judgement across the different Concept Cartoon explanations (from 17% to 75%

correct). That is, following the intervention children were more likely to correctly judge evolutionary explanations as good than to recognise that want-based explanations are incorrect. This split of high proportions of correct judgements for evolutionary explanations and low proportions of correct judgements for want-based explanations corresponds with the study's underpinning conceptual framework: that children will incorrectly endorse want-based explanations for the origin of organisms' traits. At post-test the control group maintained this tendency to claim want-based explanations are appropriate. However, the treatment group showed no significant difference in the post-test judgments, they improved in judging both evolutionary and want-based explanations. Therefore, following the intervention there was a large and significant improvement in children's judgements of explanations for the development of organisms' traits across the discussion topics assessed.

With regard to children's rationale scores, the treatment group showed a considerable increase in their understanding of the elements of appropriate, and inappropriate, explanations about the origin of organisms' traits, from 2% to 63%. However, the control group's mean scores show little change, from 5% to 12% at pre- to post-test. Analysis indicated that the condition to which children were assigned had a significant effect upon their score improvement from pre- to post-test. Pairwise comparisons showed that the treatment group had a significant score increase with considerable effect size, whereas the control group's score increase was non-significant. Therefore, following the intervention, there was a large and significant improvement in children's understanding of appropriate (evolutionary) and inappropriate (want-based) explanations about the origin of organisms' traits.

Taken together the results of the intervention's groups pre- and post-test assessment indicate, it is argued, that the intervention was successful its aims to improve:

- Children's judgements of want-based (design-teleological) and evolutionary (scientific) explanations about the existence of organisms' traits;
- Children's understanding of appropriate, and inappropriate, explanations about the existence of organisms' traits.

While these aims were achieved for both want-based and evolutionary explanations, the intervention appears to be more successful for teaching children to recognise and understand why want-based explanations are incorrect. Therefore, meeting the main aim of the research question, to develop children's understanding of why certain teleological explanations are inappropriate in scientific contexts. The results of this study show that design-teleology is not the barrier it is often claimed to be. Children can be easily taught to recognise goal-oriented design-teleological statements and criticise their use as scientific explanations. Therefore, suggesting that the barrier of teleological reasoning will be lessened when children are learning about evolution.

7.4.2 Intervention evaluation

This section discusses the actual process of delivering the intervention. First, the assessment method is explored before considering the appropriateness of the intervention sessions for use in educational settings. Chapter 6 (section 6.5.2) argued that the success and the simplicity of the Study 2 intervention was due to its theoretical underpinnings (dialogic talk and P4C) and the activities used (Concept Cartoons), as it proved children with space and time to critically reflect on, and discuss, preconceived ideas in a supportive group environment. Consequently, as Study 3 has the same theoretical underpinnings and the same activities, these aspects of the intervention are not explored here as the same conclusions apply.

Considering only the participants' judgements of teleological statements indicates that, at pre-test, children's correctly judged 24% of the teleological statements; that is, they incorrectly supported teleological explanations around 76% of the time. Comparing these results to a similar studies, shows a similar pattern of results. Kelemen (1999b) found 65% endorsement of teleological explanations, DiYanni & Kelemen (2005) found 71% teleological endorsement for organisms' traits. Therefore, the assessment method used appears to result in similar levels of teleology as indicated elsewhere in the literature, suggesting strong validity of measures.

A final consideration for assessment method is the length of assessment. As noted in section 7.2.2.4, the design of the original assessment called for twelve items. However, due to the time needed for children to respond to these questions, in the main study this was reduced to eight items. While this decrease did not affect the analysis of the efficacy of the intervention (see section 7.3.1), it would be problematic to implement in a scale up study. As recruited teachers implementing the intervention would not have the time to spend 10-15 minutes assessing each individual child. Therefore, two possible solutions are available. The first being to only assess a small number of children, who would represent the entire class. While this would solve the time-cost problem, the sample may not be representative within a single class. The second solution would be to develop a method for group assessment. This would result in each child having their progress tracked and a decrease in assessment time, although there would be problems regarding independence of observations. Consequently, group assessment methods may need developing if a scale up intervention trail was conducted.

In terms of incorporating activities into the curriculum, the literature review (section 2.8.7) has addressed this question finding that the National Curriculum (DfE, 2013) encourages children to develop their thinking skills and to be assisted in understanding scientific methods of investigation about the world. The intervention also supports the development of specific knowledge about Biology that is required to be taught in Year 2, 4 and 6. The actual intervention activities, as in Study 3, demand no lengthy intervention-specific training for a teacher to use. These activities are within the skill set of all primary teachers, the intervention sessions are similar to circle time activities, a key lesson format in the repertoire of every teacher. Therefore, once a teacher

understands the aims and purpose of the intervention it should theoretically be easy to incorporate into their teaching practice. The intervention is at its core simple and limited in terms of time resources: after only 30 - 40 minutes of group discussion children showed considerable improvements in their understanding of explanations for the existence of organisms' traits.

7.4.3 Conclusion

The aim of Study 3 was to explore if the intervention format used successfully in Study 2 could be applied to a different ontological category, that of organisms' traits. The results of Study 3, it is argued, show that the intervention was successful in the aim to improve children's understanding of why want-based explanations are inappropriate, and why evolutionary explanations are appropriate, to account for the existence of organisms' traits. Following the intervention children in the treatment group showed a considerable improvement in their rationales for why explanations are appropriate or inappropriate. Further consideration of explanatory type (evolutionary vs. want-based) suggests that the intervention was more effective at improving children's understanding of why want-based explanations are inappropriate, rather than why evolutionary explanation are appropriate. However, this is not considered problematic as the intervention is proposed to be used alongside, or prior to, formal teaching about evolution. The results of Study 3 indicate that teleology may not be the major barrier it is claimed to be, in relation to learning about evolution. After only limited discussion about styles of explanation in science, children were considerably better at disregarding want-based explanations and at understanding why these forms of explanation are inappropriate to account for the existence of organisms' traits.

8. Discussion

The main focus of this thesis has been children's use of teleological styles of explanations for the existence of natural kinds. A key underpinning idea has been the view that teleology is a collection of different explanatory styles (e.g., design-teleology or relational-teleology) rather than a method of reasoning. This notion has informed many decisions made in the thesis, especially ideas around the interplay between cultural patterns of language and children's use of teleological explanations and the impact that this has upon designing methods of instruction. This final chapter summarises the studies conducted and conclusions reached. It is structured by research questions and also by two main themes: measuring teleology and limiting teleology. The chapter concludes with a number of limitations and possible further directions for research.

8.1 Measuring teleology

8.1.1 RQ1: Does question wording influence children's explanations about the existence of natural phenomena?

The first main research question in Study 1 focused on the influence of question wording on the tendency of children to provide teleological explanations for the existence of natural phenomena (e.g., day or rainbows). Section 2.6 of the literature reviewed the data collection methods used in previous work, focusing on how prior studies have measured children's teleological explanations: the specific questions children were asked to assess their preconceptions. It was argued that a number of studies have used what can be considered leading questions, for example, the oft-cited paper by Kelemen (1999a) asked children 'what is X for?' in relation to different types of organisms, natural objects, natural phenomena and artefacts. These leading questions, it was proposed, can place demand characteristics upon children that result in them providing a certain form of response (Orne, 1962; Weber & Cook, 1972), in this case a purposeful explanation for the existence of X. This investigation had a similar rationale to McGarrigle & Donaldson's (1974) 'naughty teddy' critique of Piaget's (1954) conservation task: that the way in which the task was presented to children may have negatively impacted their performance and therefore children's knowledge and understanding may have been underestimated. Consequently, it was hypothesised that question wording would influence children's responses, with a teleologically leading question (what is X for?) resulting in more teleological responses that a neutrally worded question (why is there X?).

The results of Study 1 (section 4.3.1) confirmed this hypothesis. Question wording had a significant and large effect on the type of response children provided and this effect was maintained across the age range investigated (5- to 8-years-old). There are two clear implications from this finding, the first being that assessment of children's preconceptions must be designed in such a way as to avoid the bias associated with teleologically leading questions. Whether in practice or in

research situations children should not be accidentally led into provided purposeful rationales for the existence of natural kinds. Secondly, caution is required when interpreting the findings of prior work, which may have inadvertently used leading questions to assess children's preconceptions (see Kampourakis et al., 2012a; Keil, 1995; Kelemen, 1999a, 1999b; Polling & Evans, 2002). This is not an argument suggesting that the studies which claim that children use high levels of teleological explanation are without merit. After all, even when responding to the neutrally worded question used in Study 1 the majority of children's responses (68.2%) were still teleological (compared to 92.1% when using the leading-question format). It is, however, an argument proposing that these studies may have overestimated children's tendency to provide teleological explanations.

8.1.2 RQ2: Does questioned topic influence children's tendency to provide a teleological response?

Many studies exploring children's use of teleological explanations have investigated children's preconceptions about the existence of various things (e.g., scissors, scree or servals' spots). These topics of discussion are usually grouped together into a number of different categories that are then explored and contrasted (e.g., artefacts, natural objects or parts of organisms). Exploration of children's teleological explanations for exemplars of different categories has resulted in two competing theories about children's application of teleological explanations to categories of kinds. Children are either claimed to be selective in their application, as they use teleological explanations in the categories of organisms' traits and artefacts (see Kampourakis et al., 2012a; Keil, 1995) or promiscuous in their application, as they use teleological explanations for any exemplar regardless of category (see Kelemen, 1999c). While grouping discussion topics into categories is a common feature of previous work, there is little information regarding how these categories are defined, other than the category label. Following Chi (1992; 1997; 2008) section 2.3.2 of the literature review problematized this lack of category definition and proposed a systematic method: ontological classification. It is argued that this framework provides a systematic, easy to use method of grouping, which avoids problems with overlapping categories when comparing different research studies. However, as discussed in section 2.3.3, occasionally there may be boundary cases that do not fit neatly into a single category, an example was given on naturfacts (natural objects used for a purpose) and how these are redesigned over time to become artefacts in their own right. On the rare occasion that such cases are discussed a rationale should be provided as to the how and why they were placed in a certain category.

Section 2.6.2 of the literature review detailed that prior work (e.g., Kampourakis, 2012b; Kelemen, 1999a) has only used a small number of examples of things within an ontological category to account for an entire category. For example, Kelemen (1999a) asked children about the existence

of two examples of natural objects to assess their preconceptions for the entire category. It was argued that this limited number of examples is insufficient to provide a generalised view of an entire ontological category. Using a limited number of supposed exemplars, it was proposed, can lead to an under or overestimation of children's tendency to provide teleological explanations for a certain ontological category, unless it was proven that such exemplars are representative of a certain category (no discussion of representative exemplars has been found in the literature). For this reason the second research question in Study 1 investigated whether questioned topic influenced children's tendency to provide a teleological response.

Study 1 explored this question by asking children to explain the existence of ten different examples of natural phenomena. Analysis of the types of responses (teleological vs. scientific) children provided (see section 4.3.2) showed that teleological explanations were not used consistently across the ten questioned topics. Certain natural phenomena (e.g., waterfalls) received a significantly greater number of teleological responses than others. It is argued that the results show that there is a large response type variation across examples of natural phenomena.

This finding has two implications. Firstly, with regard to the debate around children's selective versus promiscuous use of teleological explanations, the results of Study 1 present a complicated picture suggesting both a promiscuous and selective application. The data support a promiscuous application of teleology when considering children responses at the level of ontological category. That is, the notion of promiscuous teleology would suggest that children would provide a large number of teleological rationales for all ontological categories, including the category of natural phenomena investigated in Study 1. When considering only the neutrally worded question, 68.2% of children's explanations were teleological, therefore, supporting claims that children do not only use teleological explanations for the categories of organisms' traits and artefacts (a selective application of teleology by category, see Kampourakis et al., 2012a; Keil, 1995) but do for other ontological categories: a promiscuous application (see Kelemen, 1999c). However, exploring children responses at the individual example level presents a more nuanced picture, there was large variation in the numbers of teleological responses depending upon the questioned topic. For example, for storms 32% of children's answers were teleological whereas for waterfalls 87.5% of children responses were teleological explanations. It is proposed that this variation in response type across examples of a single category does not correspond with the terminology currently used in the literature. *Selective* and *promiscuous teleology* relate to the classification level of ontological categories – not examples of these categories. Consequently, it could be argued that the results of Study 1 show a promiscuous use of teleological explanation if children's explanations are only considered at the level of ontological category. It is argued that taking this position would be a mistake, as within the category of natural phenomena there was a selective application of teleology depending upon the example of natural phenomena under discussion. Therefore, future debate surrounding the selective versus promiscuous application must

clarify whether the debate refers to the use of teleology at only the category level or if it will take into account the considerable variation in teleological responses across individual examples of an ontological category.

The second implication is that caution is needed when interpreting the findings of previous studies that have used a small number of supposed exemplars to explore children's preconceptions about an ontological category. Depending upon the examples used, prior work may have under- or overestimated the number of teleological explanations children provide within a certain category. Consequently, further exploration of children's teleological explanations needs to take into account that a small number of questioned topics may inadvertently bias the dataset. Therefore, either a) a large number of examples are needed to account for within-category variation or b) future studies must select and justify a group of representative exemplars if the focus is a category level investigation.

8.1.3 SQ1: What types of teleology do children display when they provide a teleological explanation?

Review of the current literature suggested that when exploring children's use of teleological explanations prior work usually does not distinguish between or acknowledge different forms of teleological explanation; rather teleology is treated as one form of explanation (a notable exception is the theoretical paper by ojalehto et al., 2013). Section 2.5.3 detailed the opinions surrounding the use of teleological explanations in scientific contexts. Building upon this, and in conjuncture with section 2.5.2, it was argued that there are different forms of teleological explanations which all fall under the umbrella heading of teleology. However, in scientific contexts the appropriateness of a teleological explanation depends upon both the form of teleological explanation used and the ontological category under discussion.

Pertinent to this thesis were teleological explanations for organisms' traits and natural phenomena. It was proposed that functional-teleological explanations for the existence of organisms' traits (explaining something in relation to a function that it performs) are necessary in scientific contexts, as it is difficult and unusual to discuss a part of an organism without reference to the function that that part serves. Additionally, the use of an evolutionary explanation is required to support this functional-teleological explanation if the discussion focuses on the origin of a certain trait, whereas the use of design-teleology (a want-based explanation where organisms decided to develop a trait to service a particular purpose) to account for the origin of a trait is inappropriate in scientific contexts. The proposal that a functional-teleological explanation is justified for organisms' traits and that an evolutionary, rather than a want-based, explanation is required for the origin of such traits was part of the underpinning theory for Study 3 (discussed in section 8.2.1).

For natural phenomena no such debate exists, as in scientific contexts all forms of teleological explanations are inappropriate to account for their existence. Section 2.5.2 argued that the majority of prior work has treated teleological explanations for natural phenomena (and natural objects) as design-teleology (explaining something in relation to a purpose which it serves). For example, ‘there is rain for watering the grass’ or ‘day is for waking up the birds’. However, following ojalehto et al. (2013), a new form of teleological explanation was detailed, that of relational-teleology. In this form of teleological explanation something is explained in relation to a purpose for which it can be used or an effect that it causes, and so the purpose or effect is human-centric. For example, ‘snow is for children to make snowmen' or 'night is for people to go to sleep'. The notion of relational-teleology was supported with reference to the influence of language on children’s use of teleological explanations (section 2.4.4). In short it was outlined that a child’s individual culture is constructed from and influenced by various subcultures of which children are part (as summarised by Bronfenbrenner, 1979) and that their culture affects the styles and patterns of language which a child uses to express themselves (Hoff, 2006). It was argued in certain subcultures, relational-teleological language may be used in discussions about natural kinds (e.g., a parent telling a child to go to bed because night is for people to go to sleep) without the literal implication that they actually exist for that purpose. Therefore, a child may reuse or recycle this form of explanation when discussing it in a scientific context, due to lack of understanding about appropriate and inappropriate forms of explanations in the subculture of science. While these different types of teleology are considered to be distinct abstract categories, it should be recognised that actual concrete examples of teleological explanation (e.g., children’s statements) may be better thought of as existing on a spectrum of teleology. As such certain teleological statements may inhabit a graduated area between what can be thought of as design-teleology and relation-teleology, with such cases it needs to ascertain whether they lean more towards one area of the spectrum than the other. Section 4.2.5 detailed a linguistic method used to resolve such cases where relational-teleological statements will have a human agent inserted. No prior work was found which explored children’s use of different forms of teleological explanation. Consequently, the first sub-question of Study 1 focused on the type of teleological explanations children used to account for natural phenomena.

To investigate the type of teleology children used in Study 1, children’s teleological explanations were coded into design- or relational-teleology. More specifically children actual teleological statements were coded into the abstract categories of design- or relational-teleology, where children’s answers could be thought of as existing on a spectrum of teleological statements. Answers that were vague or inhabited the graduated area between these two types of teleology were resolved by linguistic analysis to ascertain if a child had inserted a human agent into their response in place of the questioned-topic. The results provided an interesting picture, at the category level the majority of children’s teleological explanations (72.5%) were a form of relational-teleology

rather than design-teleology, the main form of teleology discussed in the literature. Although again there was an influence of questioned topic, certain ones (day, light and night) received significantly more relational responses and one topic (storms) had a larger number of design-teleological answers. This suggests that when children provide teleological explanations they may not always be expressing a literal belief that something exists *for a specific purpose* but rather discussing it in relation to something *for which it could be used*. While it is expected that there will be within category variation, this finding should be applicable to the majority of natural phenomena.

This finding has two implications: firstly, future work must take into account the type of teleology children are using to explain natural kinds in order to provide a more refined picture of children's preconceptions. Secondly, children's teleological explanations can be influenced by cultural methods of discussion around natural phenomena. That is, children participate in discussion around natural phenomena and this discussion often concerns an effect that natural phenomena have on peoples' lives or a purpose for which they can be co-opted by people (relational-teleology). For example, a child may not believe rain is for watering the grass but may use this explanation due to prior conversations where rain was discussed in relation to it watering the grass. Therefore, it is possible that relational-teleology is a natural pattern of language that may be used by children for explanations for the existence of natural phenomena because they do not understand that it is not an appropriate style of explanation in scientific contexts. Consequently, it is argued that this finding indicates that children's use of teleology may be limited by developing their understanding of appropriate forms of scientific explanation.

8.1.4 SQ2: Do children's explanations about natural phenomena result from fragmented or coherent knowledge?

There are varying perspectives on the structure of children's preconceptions. Considering these perspectives as a continuum suggests that there are two prominent positions in the literature that can be thought of as occupying the furthest reaches of either side of this continuum. That is, one side of the continuum leans towards viewing children's knowledge as having a more coherent structure with the other side leaning towards it having a more fragmented structure. (Although more central positions have been advocated, e.g., the action-based theory of conceptual growth, Howe, 1998). A more coherent viewpoint would suggest that for a certain concept children's explanatory framework, detailed by their preconceptions, would be consistently applied by individuals across contexts; therefore, demonstrating an explanatory system that displays considerable coherence (Ioannides & Vosniadou, 2002). Whereas a more fragmented stance suggests that children's preconceptions would differ depending upon the context in which they are explaining, and when asked about a certain concept a group of children may provide a wide range of contradictory responses. This demonstrates an explanatory system consisting of a collection of fragments that

combine when needed to form an idea (diSessa, 2014). In relation to teleological explanations, previous work would appear to indicate that children's preconceptions about the existence of natural kinds are the result of more coherent knowledge about the natural world. However, this coherent view has not been specifically stated in the literature. A position towards the coherent side of the continuum can be inferred from the prior studies that claim that children broadly apply teleological explanation across examples of an ontological category (e.g., natural objects or organisms) and teleology is consistent across individuals (i.e., children provide a small number of similar teleological explanations for a certain example). Section 2.5.1.2 proposed a fragmented view for children's preconceptions about natural kinds, in which teleological explanations would be more nuanced and contradictory (i.e., not stable for an example or category). It was argued that if teleology results from fragmented ideas, then when a group of children is asked to provide an explanation for a range of contexts (e.g., different examples of a category), children's responses should a) exhibit reasonable variation within children (i.e., a child would not provide similar levels of teleology across a range of examples); and b) exhibit variation across children (i.e., a group of children would provide a range of purposeful responses for any given example of a category).

Study 1 provided an opportunity to explore the structure of children's preconceptions about the existence of natural kinds and to suggest if teleological explanation is the result of more coherent or more fragmented ideas. Answering this question (SQ2) involved exploring within-child variation across contexts (i.e., how variable are children's explanations for a certain example of natural phenomena) and between-child variation across contexts (i.e., how variable are the styles of explanation children use across the natural phenomena investigated). Children's responses across the ten questioned topics were analysed, including both scientific versus teleological answers and design- versus relational-teleological explanations. It was proposed that across the range of natural phenomena, children's responses did exhibit a) reasonable variation within children (i.e., a child varied in response type across the natural phenomena); and b) exhibited variation across children (i.e., the participants provided a range of varied response types for an example of natural phenomena). Consequently, children's naïve ideas about the existence of natural phenomena are argued to be a form of fragmented knowledge. That is, they result from structures of knowledge that can be considered to have a more fragmented structure when considering that there is a continuum of positions, with the furthest reaches of this continuum indicating a more coherent or more fragmented structure of naïve ideas.

This finding had one main implication, following diSessa, that if children's preconceptions about natural phenomena are the result of a more fragmented form of knowledge then it would indicate a diminished barrier to learning. A diminished barrier to learning is proposed, not because a fragmented view suggests that teleological explanations are less problematic but, because this fragmented view provides a new method of tackling children's use of teleological explanation. It is proposed that through the reframing and reshaping of fragments of children's preconceptions,

teachers can ‘chip away’ at children’s ideas that cause them to use inappropriate teleological explanations in scientific contexts.

8.1.5 RQ3: Does the natural discourse children experience around natural phenomena influence children’s explanations for their existence?

There are a number of factors believed to influence children’s use of teleological explanations, including the influence of intentional artefacts (section 2.4.1), religion (section 2.4.2) and evolution (section 2.4.3). This thesis proposed an additional factor, the influence of language (section 2.4.4). As mentioned in section 8.1.3, the influence of language proposes that children have their own individual culture which, following Hoff (2006), is an amalgamation of the numerous subcultures of which they are part, for example, their family’s subculture or their school’s subculture. These subcultures all exert varying pressures on children and, consequently, the styles and patterns of language that children use to explain themselves. One such form of explanation is relational-teleology (section 2.5.2), the relating of natural kinds to purposes they serve for humankind, potentially with an implicit implication that they do not actually serve such a purpose. It was proposed that if children are part of subcultures that discuss natural kinds in relation to purposes they serve for humankind, then when asked to explain natural kinds, then children may draw upon this discourse to account for natural kinds with relational-teleological explanations. Therefore, children’s use of relational-teleological explanations in scientific contexts may not indicate a specific belief that something exists for a particular purpose. Instead, it may indicate the use of cultural language patterns that children do not understand are inappropriate in scientific contexts. Study 1b was designed to explore if the patterns of language children experience may influence their use of teleological explanations.

Study 1b sought to answer the question (RQ3) *does the natural discourse children experience about natural phenomena influence children’s explanations for their existence?* This research question was investigated using two component questions, (CQ1) *what specific examples of teleological and scientific explanations do children experience about natural phenomena?* and (CQ2) *what are the themes associated with natural phenomena in the discussions children experience?* Corpus linguistics was used as the method to explore these questions as it provided a systematic way to analyse the naturally-occurring discourse children experience. The results of the analysis indicated that children experience few specific explanations about the existence of natural phenomena. Therefore, the explanations children provide are not likely to be repetitions of previously heard accounts. Exploring collocates of natural phenomena, the words that appear significantly more frequently in discussion about a certain example of natural phenomena than others, presented a different explanation. Collocates of natural phenomena were grouped into a number of thematic categories, for example, *activity*, *location* or *time period*, which were used to propose three conceptual metaphors: TIME AS NATURAL PHENOMENA, NATURAL PHENOMENA ARE

ACTIVITIES and NATURAL PHENOMENA ARE A RESOURCE IN PLAY. Following conceptual metaphor theory (see section 5.1.2), it is proposed that these conceptual metaphors structure how children conceptualise natural phenomena, that is, children's experience of such metaphors could influence how they think and talk about natural phenomena. This is supported by the results of Study 1 that show the majority of children's explanations use relational-teleology, proposing that certain natural phenomena exist for the purpose of signifying a time period, supporting a certain activity or for being used in play.

This finding has one main implication, for this thesis, that children's natural discourse around natural phenomena is associated with a number of key themes. These themes manifest in children's use of relational-teleological explanations for the existence of natural phenomena. This finding supports the claims made in this thesis about the influence of language on children's use of teleological explanations. Children's teleological explanations are shaped by the natural discourse they experience, and shaped by the different subcultures of which children are a part. It could be suggested that this association between children's explanations and their discourse is obvious, however, no known link has previously been made between the children's use of teleological explanations and the discourse they experience.

8.1.6 Summary

These research questions and their answers are related to the theoretical and practical aspects of the measurement of children's use of teleological explanations. RQ1 and RQ2 show how the way in which questions are worded and the subject of the question (questioned topic) can influence children's use of teleological explanation, both associated with considerable variation in response type. Exploring the type of teleology children use (SQ1), how their preconceptions may result from fragmented knowledge (SQ2) and how the discourse they experience may influence their explanatory style (RQ3) has consequences for the development of both theory about teleology and of practical aspects of assessing children's use of teleological explanations. While a number of implications have been outlined with respect to each question, there is one common thread: the implication for instruction. It was argued that these findings show that teleology is less of a barrier to science education than previously thought. While it was shown that certain methods have a led to an overestimation of children's use of teleology, it was still the dominant style of explanation in Study 1. However, after consideration of the fragmented structure of knowledge that leads to teleological explanations and that children routinely use forms of teleology influenced by cultural patterns in language, a method of tackling teleology was proposed, one which reconceptualises the barrier that is teleology. Children's use of teleological explanation could be limited with the use of interventions that aim to reshape children's knowledge fragments, through developing children's

understanding that while individuals may use teleological formations in social situations they may not be appropriate forms of explanation for scientific contexts.

8.2 Limiting teleology

The majority of research explicitly on children's use of teleological explanations focuses on cataloguing children's preconceptions, with little work on specifically limiting children's use of teleology or on developing their understanding of this style of explanation. A major theme of this thesis was to expand this underexplored area, to limit children's endorsement of teleology through developing their understanding of it. This section discusses Studies 2 and 3, which sought to limit children's use of teleology.

8.2.1 RQ4 & RQ5: Do discursive activities about the appropriateness of scientific explanations support children in developing an understanding of appropriate and inappropriate explanations for the existence of natural kinds?

As outlined in section 8.1.6 and in the main body of the thesis (e.g., see section 2.7.2), an argument was made that children's use of teleological explanations could be limited through the use of interventions which sought to develop their understanding that certain explanations are not appropriate in scientific contexts. This was supported by two notions; firstly, that children's teleological explanations may result from fragmented knowledge about the natural world. This would indicate that their preconceptions could be altered through the gradual reframing of knowledge elements, through educators chipping away at certain aspects of an incorrect idea. Secondly, that one area which may be receptive to this reframing may be children's understanding of where teleological explanations are appropriate. Summarised in section 8.1.3 and supported by section 8.1.5 was the proposal that children's teleological explanations are influenced by cultural patterns of language. Children are members of subcultures that may discuss natural kinds in relation to purposes for which they could be used (relational-teleology), which is supported by conceptual metaphors that were found to be associated with natural phenomena. This discourse that children experience could lead them to providing non-literal teleological explanations in scientific contexts, as they do not realise that it is an inappropriate form of explanation for the scientific subculture. Consequently, reframing children's knowledge of this explanatory style may limit children's use of teleology in science education.

A number of theories and ideas explored in the literature review support the idea that a discussion-oriented intervention may be an appropriate method to support children's development in relation to the research aim. For example, the argumentative theory of reasoning (Mercier & Sperber, 2017) outlines that while individual's reasoning can lead to poorly constructed preconceptions, group discussion can support the creation of well-reasoned arguments, as it

provides individuals the opportunity to evaluate and develop the arguments of others (see section 2.4.3.2). Another key theory is the idea of dialogic teaching (e.g., see Alexander, 2006; Mercer, 2008), in such situations classroom talk is seen as the foundation of learning, where groups of children supported by a teacher work together to collaboratively construct their understanding of a concept (see section 2.8.4). These theories were influential in the development of the interventions, as was Philosophy for Children. P4C (e.g., see Smith, 2010b) is a thinking skills programme that draws heavily on dialogic teaching to support children's understanding of philosophical and abstract concepts, concepts that share similarities with teleology. In summary, the designed interventions were based on the structure of P4C sessions and were underpinned by both the argumentative theory of reasoning and dialogic teaching.

The interventions focused on two different age ranges (5- to 7-year-olds and 8- to 9-year-olds) and two different ontological categories where teleological explanations are used by children (natural phenomena and organisms' traits). As such each study has a slightly different research question. Study 2 explored (RQ4) *do discursive activities about the appropriateness of scientific explanations enable children to develop an understanding of why causal explanations are appropriate and teleological explanations are inappropriate to account for the existence of natural phenomena?* Study 3 investigated (RQ5) *where the purpose of a trait is explained with relation to functional-teleology, do discursive activities about the appropriateness of scientific explanations enable children to develop an understanding of why evolutionary explanations are appropriate and want-based explanations are inappropriate to account for the origin of organisms' traits?* While both interventions had slightly different focuses, the overarching design of intervention sessions and assessment methods was the same.

Both interventions were successful in terms of limiting children's endorsement of and developing their understanding of teleological explanations. In both studies the treatment group showed significant improvement from pre-test and considerably outperformed the control group at post-test, these significant differences were supported by very large effect sizes. These results concern both children's initial judgements about the 'correctness' of different explanatory styles and their understanding of why these explanatory styles are appropriate or inappropriate. It was argued in both chapters 6 and 7 that these results show that children's understanding of teleological styles of explanation can be reformed, and that children can develop an awareness that, while people may use teleological patterns of language in daily life, these forms of explanation are not appropriate in scientific contexts.

One main implication arising from these results is that children's use of teleological explanation is not the major barrier it is claimed to be. Children can be taught to understand where and why certain forms of teleology are inappropriate through a series of simple, short and effective interventions. The activities used in the interventions would require no specific training, the activities (Concept Cartoons, puppets and group discussion) are part of the repertoire of primary

school teachers. With regard the time needed to incorporate the intervention into classroom practice, the second version of the intervention (Study 3) involved working with children for a maximum of 40 minutes and the structure of the interventions do not necessitate these activities need to happen in one block. Therefore, the interventions are not time intensive, educators could easily incorporate elements of the intervention into the starters of science lessons. Consequently, it is maintained that the thesis was successful in its aim to develop a method that would limit children's use of teleological explanations. In line with the theoretical underpinnings of real world research (see section 3.3.1), this is a successful outcome in that it is applicable to and solves a problem existing in educational environments.

8.3 Limitations

There were a number of limitations to the studies conducted, the main limitations are provided in this section, split by study. Furthermore, there are a number of other implications that are not presented in this section, rather they are covered in section 8.4: future directions.

8.3.1 Study 1

There were three clear limitations in Study 1. Firstly, there may have been a carry-over effect, where those children who received their first set of questions using the leading question wording were primed to provide teleological answers for the second question set, using the neutrally question wording. The analysis indicated there was a nonsignificant interaction between question wording and question set order on children's responses, although this was not the only statistical measure used, partial-eta showed a small effect size, as such there may have been a minor interaction between question wording and question set order. Further consideration of the data indicates that for children who received the leading-treatment first there does appear to be a trend for this to result in lower levels of scientific responses then when they received the open-treatment first (26% scientific responses compared to 38% scientific responses when the open-treatment was used first, see Table 4.3). This possible trend would have made the hypothesis test more conservative, as a carry-over effect would cause the open-treatment data set to be more similar to the leading-treatment dataset, meaning that there would be a smaller difference between the two datasets. Consequently, children's actual level of scientific responses may have been underestimated in the full dataset. Ideally, there would have been more delay between the two treatments during data collection. However, due to the constraints of working in a school environment it was not possible to interview children twice. While the time taken to interview the children on two separate occasions would have been the same, once classroom movement time is factored in, it would have resulted in a research timetable that was untenable for the school.

Secondly, as mentioned in the Study 1 chapter, the investigation into the influence of questioned topic and exploration of type of teleology was limited by only using half the dataset.

This was due to the removal of children's responses to the leading-treatment, as it was shown to have a considerable influence on response type. Therefore, the analysis for questioned topic and type of teleology did not explore individual participants' explanations across all ten questioned topics, rather the open dataset comprised of children's rationales for five randomly assigned questioned topics. It is acknowledged that having data for all ten examples of natural phenomena would have provided a more thorough picture of children's explanations. However, it was considered an acceptable compromise, as it allowed Study 1 to be used to explore both a) the influence of questioned topic and b) the main hypothesis regarding the influence of question wording. Within the time frame for this thesis, investigation of these questions in two separate studies may not have been possible.

Thirdly, in Study 1 there was no data analysis conducted regarding students with English as an Additional Language or Special Educational Needs, as to include either of these as independent variables would have required a sample size so large that it would have been impractical for the scope of this thesis. While there is no indication in the literature that these two groups of students would act differently with regard to teleological styles of explanation, it is still considered a limitation. Research on these groups is needed as it is important for any planning undertaken in educational settings: teachers need to differentiate where, and if, appropriate. If possible, future studies could explore if these groups react differently to question prompts.

8.3.2 Study 1b

A limitation of Study 1b was the limited concordance investigation. The corpus developed contained very little teleological or scientific concordances; therefore, limited inferences could be made from the results of the first part of the corpus investigation. The development of larger corpora may provide the evidence to better assess the specific examples of teleological and scientific discussion which children experience, or would support the claim that children do not experience explanations about the existence of natural phenomena.

Regarding the second part of the corpus study (collocate investigation and related conceptual metaphor), the derived conceptual metaphors from collocate categories are open to interpretation. It was not the intention to claim that these are the only three possible metaphors for natural phenomena. Rather they are the three that emerged from this study and, consequently, different corpora investigations may produce different metaphors depending upon research focus. Further investigation of conceptual metaphors for natural phenomena will support or refute the conclusions proposed in Study 1b.

8.3.3 Study 2

There was one main limitation of Study 2, the timetabling of assessment. The post-tests took place on the last two days of term during free play activities, for both the Year 1 and 2 cohorts. This

unavoidable scheduling would not have provided a desirable working environment, instead one that contained multiple distractions, possibly diverting participants' attention from the task. This may have resulted in both the treatment and control groups underperforming due to the scheduling of the post-tests.

8.3.4 Study 3

A possible limitation of Study 3 relates to treatment bleed effects. For children's rationales (i.e., their accounts for why they judged an explanation as correct or incorrect) it was hypothesised that the pre- and post-test scores would remain the same for the control group, this was not observed. The descriptive statistics for and analysis of the control group's post-test rationales (see section 7.3.3) indicated that correct rationales were influenced by explanatory type. There were a larger number of correct rationales for want-based explanations than for evolutionary explanations (although not enough to show a significant improvement from pre- to post-test). It is proposed that these results are explained through a bleed effect (Shadish, Cook & Campbell, 2001) where children from the control group overheard discussions about the intervention, or discussed the intervention with children in the treatment group. Essentially, children may have discussed how the sessions involved talking about 'how animals wanted to change and how silly that is'. This is supported by observations by the researcher who overheard participants in the control group talking with their friends in the treatment group and asking what was happening in the intervention sessions. This bleed effect between treatment conditions is unfortunate but did not appear to effect the main study results: children in the control group showed no significant overall improvement in their rationales from pre- to post-test. Furthermore, any bleed effect would have made the analysis more conservative as the no treatment control group could have been partially influenced by the intervention.

8.4 Future directions

While a large number of possible research directions could be proposed from the results and limitations of this thesis, this section details two major future directions. The first focuses on ontological category and the results of Study 1. It is acknowledged that only investigating examples of natural phenomena is a limitation of Study 1, thus further research is needed to confirm if the results are applicable to different ontological categories. This is in relation to all four research questions: the influence of question wording and questioned topic, the type of teleology children use and the form of children's preconceptions (coherent vs. fragmented). Exploring these ideas in relation to other ontological categories is relatively simple for examples of natural objects, as the study design could remain the same apart from the change in questioned topics. Investigating types of teleological explanations in relation to organisms' traits would require the inclusion of different types of teleology, which were not explored in Study 1, specifically religious-teleology and functional-teleology. Functional-teleology would provide a challenge because, as argued in section

2.5.3.2, functional explanations for the existence of organism' traits are justified. Therefore, what would be required is a number of follow-up questions targeting children's ideas about the origin of such traits, to ascertain if they use want-based or evolutionary explanations. This future direction could a) provide support for or against the results of Study 1; and b) improve the current understanding around children's teleological explanations for the existence of natural kinds.

The second future direction involves further development and testing of the interventions used in Studies 2 and 3. A shortcoming of the assessment method used in these studies, in terms of tracking conceptual change in children, is the lack of a delayed post-test for retention, as interventions may only show short term effects (Newby, 2010, Shadish, Cook & Campbell, 2002). While this would have been beneficial, the constraints of the research environments meant a delayed post-test was impractical. Should future iterations of the interventions be conducted then early recruitment of schools is desirable to allow sufficient time to develop a research timetable that meets the needs of both the schools and the research design (Rochelle et al., 2014).

The results of Studies 2 and 3 relate to the efficacy of the interventions, in that they were shown to work better than a no treatment control. While it may appear a self-evidential claim that the interventions are better than no input, it is actually a demonstration of proof of concept, proof that children can be taught to understanding teleological forms of explanation. What is now needed is a test of relative efficacy. The activities used in the interventions may be the most effective activities for developing children's understanding or they may be the least effective activities. Future studies are required which compare the interventions against others that seek to develop similar conceptual knowledge. As there is no similar interventions focusing on limiting children's use of teleology, a comparable intervention for an active control group may consist of directly teaching children the correct scientific explanations to see if that is a more efficient method for limiting children's use of teleology.

Future work looking at further development and testing of the interventions, through improved assessment and active control groups, would a) provide support for/against the findings of this thesis; and b) contribute to the theoretical and practical knowledge about limiting children's use of teleological explanations in scientific contexts.

8.5 Final remarks

This thesis has taken the position that teleology is a collection of a number of different styles of explanation, related by their collective focus on *purpose* or *goal-directed behaviour* as rationales for why something exists in the natural world. Furthermore, the 'correctness' of teleological styles of explanation depends, in part, upon the context in which they are used. Taking this stance, four studies have been conducted on children's use of teleological explanations in scientific contexts. These studies have shown that language is both a problem and a solution to children's use of

teleological explanations. In terms of a contribution to knowledge this thesis shows that children's use of teleological explanations is more nuanced than currently indicated and can be limited with a series of simple, short discursive lessons.

In light of these studies it is argued that there is a need to reconceptualise how children's use of teleological explanations is assessed, as there are a number of factors that indicate that teleology is more complex than previously suggested. Furthermore, through the use of discussion-based activities, children can develop their understanding of teleology and their awareness of the context in which these forms of explanation should and should not be used.

9. References

Abelson, R. P. (1995). *Statistics as principled argument*. Hove: Psychology Press.

Adey, P., & Shayer, M. (2002). Cognitive acceleration comes of age. In M. Shayer & P. Adey (Eds.), *Learning intelligence: Cognitive acceleration across the curriculum from 5 to 15 years* (pp. 1-17). Buckingham: Open University Press.

Adey, P., Robertson, A., & Venville, G. (2002). Effects of a cognitive acceleration programme on year 1 pupils. *British Journal of Educational Psychology, 72*(1), 1-25. doi:10.1348/000709902158748

Alexander, R. (2006). *Towards dialogical teaching: Rethinking classroom talk* (3rd ed.). York: Dialogos.

Allen, C. (2009). Teleological notions in biology. *The Stanford Encyclopedia of Philosophy*. Winter 2009. Retrieved from http://plato.stanford.edu/archives/win2009/entries/teleology-biology/

Allen, M. (2014). *Misconceptions in primary science*. London: Open University Press.

Allen, W. L., Cuthill, I. C., Scott-Samuel, N. E., & Baddeley, R. (2010). Why the leopard got its spots: Relating pattern development to ecology in felids. *Proceedings of the Royal Society B: Biological Sciences*. doi:10.1098/rspb.2010.1734

Amin, T. G. (2009). Conceptual metaphor meets conceptual change. *Human Development, 52*, 165-197.

Amin, T., Smith, C., & Wiser, M. (2014). Student conceptions and conceptual change: Three overlapping phases of research. In N. G. Lederman & S. K. Abell (Eds.), *Handbook of research on science education* (2nd ed., pp. 57-81). New York: Routledge.

Amundson, R., & Lauder, G. V. (1994). Function without purpose. *Biology and Philosophy, 9*(4), 443-469.

Anderson-Inman, L., & Ditson, L. (1999). Computer-based concept mapping. *Learning and leading with technology, 26*(6-13).

Andriessen, J., & Baker, M. (2014). Arguing to learn. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 439-460). Cambridge: Cambridge University Press.

Anthony, L. (2014). *AntConc (3.4.4w)* [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from http://www.antlab.sci.waseda.ac.jp/

Athey, C. (1991). *Extending thought in young children: A parent-teacher partnership*. London: Paul Chapman Publishing.

Ayala, F. J. (1970). Teleological explanations in evolutionary biology. *Philosophy of Science, 37*(1), 1-15.

Banerjee, K., & Bloom, P. (2014). Why did this happen to me? Religious believers' and non-believers' teleological reasoning about life events. *Cognition, 133*, 277-303.

Barnes, M. E., Evans, E. M., Hazel, A., Brownell, E. R., & Nesse, R. (2017). Teleological reasoning, not acceptance of evolution, impacts students' ability to learn natural selection. *Evolution: Education and Outreach, 10*(7), 2-12. doi:10.1186/s12052-017-0070-6

Basalla, G. (1988) *The evolution of technology*. Cambridge: Cambridge University Press.

Berenstain, S., & Berenstain, J. (1997). *The Berenstain bears' big book of science and nature*. New York: Random House Inc.

Benton, T., & Craib, I. (2011). *Philosophy of social science* (2 ed.). Hampshire: Palgrave Macmillan.

BERA (2011). *Ethical guidelines for educational researchers*. London: British Educational Research Association.

Berger, L. R. (1980). The Winnicott squiggle game: A vehicle for communicating with the school-aged child. *Paediatrics, 66*(6), 921-924.

Biddle, C., & Schafft, K. A. (2014). Axiology and anomaly in the practice of mixed methods work: Pragmatism, valuation, and the transformative paradigm. *Journal of Mixed Methods Research*. doi:10.1177/1558689814533157

Bjorklund, D. F., Cassel, W. S., Bjorklund, B. R., Brown, R. D., Park, C. L., Ernst, K., & Owen, F. A. (2000). Social demand characteristics in children's and adults' eyewitness memory and suggestibility: The effect of different interviewers on free recall and recognition. *Applied Cognitive Psychology, 14*, 421-433.

Bloom, P., & Weisberg, S. (2007). Childhood origins of adult resistance to science. *Science, 316*, 996-997.

BNC (2015). British national corpus. Retrieved from http://www.natcorp.ox.ac.uk/

Boudry, M., Blancke, S., & Pigliucci, M. (2015). What makes weird beliefs thrive? The epidemiology of pseudoscience. *Philosophical Psychology, 28*(8), 1177-1198. doi:10.1080/09515089.2014.971946

Bronfenbrenner, U. (1979). *The ecology of human development*. Cambridge, MA: Harvard University Press.

Brown, D. E. (2010). *Students' conceptions - coherent or fragmented? And what difference does it make?* Paper presented at the National Association of Research in Science Teaching, Philadephia, PA.

Bruner, J. S. (1966). *Towards a theory of instruction*. Harvard University Press: London.

Bruner, J. S. (1996). *The culture of education*. London: Harvard University Press.

Bruner, J. S., & Haste, H. (1987). Introduction. In J. Bruner & H. Haste (Eds.), *Making sense: The child's construction of the world*. London: Routledge.

Bryman, A. (2012). *Social research methods*. Oxford: Open University Press.

Bytheway, B. & Johnson, J. (2002). 'Doing diary-based research', in Jamieson, A. & Victor, C.R. (eds). *Researching Ageing and Later Life: the Practice of Social Gerontology*, Oxford: Blackwells Publishing, 155-74.

Callanan, M. A., & Oakes, L. M. (1992). Preschoolers' questions and parents' explanations: Causal thinking in everyday action. *Cognitive Development, 7*(1), 213-233.

Cao, T. Y. (2004). Ontology and scientific explanations. In J. Cornwell (Ed.), *Explanations: Styles of explanation in science* (pp. 173-196). Oxford: Oxford University Press.

Carey, S. (1985). *Conceptual change in childhood*. London: The MIT Press.

Carey, S. (2000). Science education as conceptual change. *Journal of Applied Developmental Psychology, 21*(1), 13-19.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63*(4), i-174. doi:10.2307/1166214

Casler, K., & Kelemen, D. (2008). Developmental continuity in teleo-functional explanation: Reasoning about nature among Romanian Romani adults. *Journal of Cognition and Development, 9*(3), 340-362. doi:10.1080/15248370802248556

Cefai, C., Ferrario, E., Cavioni, V., Carter, A., & Grech, T. (2014). Circle time for social and emotional learning in primary school. *Pastoral Care in Education, 32*(2), 116-130. doi:10.1080/02643944.2013.861506

Chen, Y.-C., Hand, B., & Norton-Meier, L. (2017). Teacher roles of questioning in early elementary science classrooms: A framework promoting student cognitive complexities in argumentation. *Research in Science Education, 47*(2), 373-405. doi:10.1007/s11165-015-9506-6

Chi, M. T. H. (1992). Conceptual change within and across ontological categories: Examples from learning and discovery in science. In R. Giere (Ed.), *Cognitive models of science: Minnesota studies in the philosophy of science* (pp. 129-186). Minneapolis: University of Minnesota Press.

Chi, M. T. H. (1997). Creativity: Shifting across ontological categories flexibly. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 209-234). Washington, D.C.: American Psychological Association.

Chi, M. T. H. (2008). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In S. Vosniadou (Ed.), *Handbook of research on conceptual change*. Hillsdale, NJ: Erlbaum.

Chi, M. T. H., Slotta, J. D., & Leeuw, N. D. (1994). From things to processes: A theory of conceptual change for learning science concepts. *Learning and Instruction, 4*, 27-43.

Chin, C., & Osborne, J. (2008). Students' questions: A potential resource for teaching and learning science. *Studies in Science Education, 44*, 1-39.

Chin, C., & Teou, L. Y. (2009). Using concept cartoons in formative assessment: Scaffolding students' argumentation. *International Journal of Science Education, 31*(10), 1307-1332.

Christensen, P., & James, A. (Eds.). (2000). *Research with children: Perspectives and practices*. London: Falmer Press.

Christidou, V., Koulaidis, V., & Christidis, T. (1997). Children's use of metaphors in relation to their mental models: The case of the ozone layer and its depletion. *Research in Science Education, 27*(4), 541-552.

Claidière, N., Trouche, E., & Mercier, H. (2017). Argumentation and the diffusion of counter-intuitive beliefs. *Journal of Experimental Psychology: General, 146*(7), 1052-1066.

Clark, A. (2006a). Language, embodiment, and the cognitive niche. *Trends in Cognitive Sciences, 10*(2), 370-374.

Clark, D. B. (2006b). Longitudinal conceptual change in students' understanding of thermal equilibrium: An examination of the process of conceptual restructuring. *Cognition and Instruction, 24*(4), 467-563.

Clark, D. B. (2011). Comparison of students' knowledge structure coherence and understanding of force in the Philippines, turkey, china, Mexico, and the united states. *Journal of the Learning Sciences, 20*(2), 207-261. doi:10.1080/10508406.2010.508028

Clough, E. E., & Driver, R. (1985). Secondary students' conceptions of the conduction of heat: Bringing together scientific and personal views. *Physics Education, 20*(4), 176-186.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). London: Routledge.

Coley, J. D., Arenson, M., Xu, Y., & Tanner, K. D. (2017). Intuitive biological thought: Developmental changes and effects of biology education in late adolescence. *Cognitive Psychology, 92*, 1-21. doi:https://doi.org/10.1016/j.cogpsych.2016.11.001

Conroy, H., & Harcourt, D. (2009). Informed agreement to participate: Beginning the partnership with children in research. *Early Child Development and Care, 179*(2), 157-165.

Coolican, H. (2009). *Research methods and statistics in psychology* (5th ed.). London: Hodder Education.

Corriveau, K. H., Chen, E. E., & Harris, P. L. (2015). Judgements about fact and fiction by children from religious and nonreligious backgrounds. *Cognitive Science, 39*, 353-382.

Crowell, A., & Kuhn, D. (2014). Developing dialogic argumentation skills: A 3-year intervention study. *Journal of Cognition and Development, 15*(2), 363-381. doi:10.1080/15248372.2012.725187

Crook, C., & Garrett, D. (2005). The positivist paradigm in contemporary social science research. In B. Somekh & C. Lewin (Eds.), *Research methods in the social sciences* (pp. 207-214). London: Sage.

Curtis, W., & Pettigrew, A. (2009). *Learning in contemporary culture*. Exeter: Learning Matters Ltd.

Davies, J. (2010). Anthropomorphism in science. *EMBO Reports, 11*(10), 721-721. doi:10.1038/embor.2010.143

Davis, P. R., Horn, M. S., & Sherin, B. L. (2013). The right kind of wrong: A "knowledge in pieces" approach to science learning in museums. *Curator: The Museum Journal, 56*(1), 31-46. doi:10.1111/cura.12005

Dawes, L. (2004). Research report: Talk and learning in classroom science. *International Journal of Science Education, 26*(6), 677-695.

Dawkins, R. (2007). *The god delusion*. London: Bantam Press.

DCSF (2008). *Practice guidance for the early years foundations stage*. Nottingham: DCSF.

Deignan, A. (2005). *Metaphor and corpus linguistics*. Amsterdam: John Benjamins Publishing.

Dennett, D. C. (1989). *The intentional stance*. Massachusetts: The MIT Press.

DfE. (2013). *The national curriculum in England: Key stages 1 and 2 framework document*. HMSO Retrieved from https://www.gov.uk/government/collections/national-curriculum.

DfE (2014) *Statutory Framework for the Early Years Foundation Stage: setting the standards for learning, development and care for children from birth to five*. Retrieved from https://www.gov.uk/government/publications/early-years-foundation-stage-framework--2

DfE. (2016). *Educational excellence everywhere*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/508447/Educational_Excellence_Everywhere.pdf.

DfEE. (1999). *The national curriculum in England: Key stages 1 and 2*. London: HMSO.

Di-Poï, N., & Milinkovitch, M. C. (2016). The anatomical placode in reptile scale morphogenesis indicates shared ancestry among skin appendages in amniotes. *Science Advances, 2*(6). doi:10.1126/sciadv.1600708

diSessa, A. A. (2013). A bird's eye view of "pieces" vs. "Coherence" controversy. In S. Vosniadou (Ed.), *Handbook of conceptual change research* (2nd ed., pp. 31-48). New York: Routledge.

diSessa, A. A. (2014). A history of conceptual change research: Threads and fault lines. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (2nd ed., pp. 88-108). Cambridge: Cambridge University Press.

diSessa, A. A., Gillespie, N. M., & Esterly, J. B. (2004). Coherence versus fragmentation in the development of the concept of force. *Cognitive Science, 28*(6), 843-900. doi:10.1207/s15516709cog2806_1

DiYanni, C., & Kelemen, D. (2005). Time to get a new mountain? The role of function in children's conceptions of natural kinds. *Cognition, 97*(3), 327-335. doi:10.1016/j.cognition.2004.10.002

Dockett, S., Einarsdóttir, J., & Perry, B. (2009). Researching with children: Ethical tensions. *Journal of Early Childhood Research, 7*(3), 283-298.

Dockett, S., Einarsdóttir, J., & Perry, B. (2012). Young children's decision about research participation: Opting out. *International Journal of Early Years Education, 20*(3), 244-256.

Donaldson, M. (1987). *Children's minds*. London: Fontana Press.

Driver, R. (2008) *The pupil as scientist?*. Milton Keynes: Open Universtiy Press.

Driver, R., Asoko, H., Leach, J., Mortimer, E., & Scott, P. (1994). Constructing scientific knowledge in the classroom. *Educational Researcher, 23*(7), 5-12.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education, 84*(3), 287-312.

Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (2007). *Taking science to school: Learning and teaching science in grades k-8*. Washington: The National Academies Press.

Duit, R., & Treagust, D. F. (2003). Conceptual change: A powerful framework for improving science teaching and learning. *International Journal of Science Education, 25*(6), 671-688.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*(2), 121-138.

ESRC (2010). The research ethics guidebook: A guide for social scientists. Retrieved from http://www.ethicsguidebook.ac.uk/Opt-in-and-opt-out-sampling-94

ESRC (2016) *The UK Data Service*. Retrieved from https://www.ukdataservice.ac.uk/

Evans, E. M. (2000). Beyond scopes: Why creationism is here to stay. In K. S. Rosengren, C. N. Johnson, & P. L. Harris (Eds.), *Imagining the impossible: Magical, scientific, and religious thinking in children* (pp. 305-333). Cambridge: Cambridge University Press.

Evans, E. M. (2013). Conceptual change and evolutionary biology: Taking a developmental perspective. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (2 ed., pp. 220-239). New York: Routledge.

Evans, E. M. (2018). Bridging the gap: From intuitive to scientific reasoning -the case of evolution. In K. Rutten, S. Blancke, & R. Soetaert (Eds.), *Perspectives on science and culture*. West Lafayette, IN: Purdue University Press.

Evans, V., & Green, M. (2006). *Cognitive linguistics: An introduction*. Edinburgh: Edinburgh University Press.

Fender, J. G., & Crowley, K. (2007). How parental explanation changes what children learn from everyday scientific thinking. *Journal of Applied Developmental Psychology, 28*(3), 189-210.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4 ed.). London: SAGE Publications Ltd.

Field, A., & Hole, G. (2003). *How to design and report experiments*. London: SAGE.

Fielding, N. G., & Fielding, J. L. (2000). Resistance and adaptation to criminal identity: Using secondary analysis to evaluate classic studies of crime and deviance. *Sociology, 34*(4), 671-689.

Flewitt, R. (2005). Conducting research with young children: Some ethical considerations. *Early Child Development and Care, 175*(6), 553-565.

Frankish, K. (2010). Dual-process and dual-system theories of reasoning. *Philosophy Compass, 5*(10), 914-926. doi:10.1111/j.1747-9991.2010.00330.x

Fisher, R. (2007). Dialogic teaching: Developing thinking and metacognition through philosophical discussion. *Early Child Development and Care, 177*(6-7), 615-631. doi:10.1080/03004430701378985

Freeman, M., & Mathison, S. (2009). *Researching children's experiences*. London: Guilford Press.

Friedler, Y., Zohar, A., & Tamir, P. (1993). The effect of age and of learning on the ability to distinguish between anthropomorphic and teleological explanations. *International Journal of Science Education, 15*(4), 439-443. doi:10.1080/0950069930150407

Galli, L., M., G., & Meinardi, E., N. (2011). The role of teleological thinking in learning the Darwinian model of evolution. *Evolution: Education & Outreach, 4*, 145-152.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40*(2), 177.

Gelman, S. A., & Legare, C. H. (2011). Concepts and folk theories. *Annual Review of Anthropology, 40*(1), 379-398. doi:10.1146/annurev-anthro-081309-145822

Gelman, S. A., & Rhodes, M. (2012). "Two-thousand years of stasis": How psychological essentialism impedes evolutionary understanding. In K. S. Rosengren, S. Brem, E. M. Evans, & G. Sinatra (Eds.), *Evolution challenges: Integrating research and practice in teaching and learning about evolution* (pp. 3-22). Cambridge: Oxford University Press.

Gibbs, R. (1994). *The poetics of mind*. Cambridge: Cambridge University Press.

Gilbert, J. K., Watts, D. M., & Osborne, R. J. (1982). Students' conceptions of ideas in mechanics. *Physics Education, 17*(2), 62-66.

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research, 42*(3), 237-288. doi:10.2307/1169991

Gregory, T. R. (2009). Understanding natural selection: Essential concepts and common misconceptions. *Evolution: Education and Outreach, 2*(2), 156-175. doi:10.1007/s12052-009-0128-1

Goldacre, B. (2008). *Bad science*. London: Fourth Estate.

Goldacre, B. (2015). *I think you'll find it's a bit more complicated than that*. London: Fourth Estate.

Gorard, S., Siddiqui, N., & See, B. H. (2015). *Philosophy for children: Evaluation report and executive summary*. Millbank: Education Endowment Foundation.

Gould, S. J. (1992). *Bully for Brontosauraus*. London: Penguin.

Gouran, D. S. (2012). A response to Hugo Mercier and dan Sperber's "why do humans reason? Arguments for an argumentative theory". *Argumentation and Advocacy, 48*(3), 186-188. doi:10.1080/00028533.2012.11821767

Grasset, L. (2016). *How the zebra got its stripes: Tales from the weird and wonderful world of evolution*. London: Profile Books Ltd.

Greany, T., Barnes, I., Mostafa, T., Pensiero, N., & Swensson, C. (2016). *Trends in maths and science study (timss): National report for England*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/572850/TIMSS_2015_England_Report_FINAL_for_govuk_-_reformatted.pdf.

Greenfield, D. B., Jirout, J., Dominguez, X., Greenberg, A., Maier, M., & Fuccillo, J. (2009). Science in the preschool classroom: A programmatic research agenda to improve science readiness. *Early Education and Development, 20*(2), 238-264. doi:10.1080/10409280802595441

Greif, M. L., Nelson, D. G. K., Keil, F. C., & Gutierrez, F. (2006). What do children want to know about animals and artifacts? Domain-specific requests for information. *Psychological Science, 17*(6), 455-459. doi:10.1111/j.1467-9280.2006.01727.x

Gries, S. Th. (2006) 'Introduction', in Gries, S. Th. and Stefanowitsch, A. (eds) *Corpora in cognitive linguistics: corpus-based approaches to syntax and lexis*, Mouton de Gruyter: New York, pp. 1-19.

Gries, S. Th. (2009) 'What is corpus linguistics?', *Language and Linguistics Compass*, 3(5), pp.1225-1241.

Guastello, E. F., Beasley, T. M., & Sinatra, R. C. (2000). Concept mapping effects on science content comprehension of low-achieving inner-city seventh graders. *Remedial and special education, 21*(6), 356-364.

Gupta, A., Hammer, D., & Redish, E. F. (2010). The case for dynamic models of learners' ontologies in physics. *Journal of the Learning Sciences, 19*(4), 285-321.

Grugeon, E., Hubbard, L., Smith, C., & Dawes, L. (2012). *Teaching speaking and listening in the primary school*. Routledge: Oxon.

Hallam, S., Ireson, J., & Davies, J. (2004). Primary pupils' experiences of different types of grouping in school. *British Educational Research Journal, 30*(4), 515-533. doi:10.1080/0141192042000237211

Hanke, D. (2004). Teleology: The explanation that bedevils biology. In J. Cornwell (Ed.), *Explanations: Styles of explanation in science*. Oxford: Oxford University Press.

Hanna, E., & Meltzoff, A. N. (1993). Peer limitation by toddlers in laboratory, home and day-care contexts: Implications for social learning and memory. *Developmental Psychology, 29*(4), 701-710.

Harlen, W. (2006). *Teaching, learning and assessing science 5-12* (4th ed.). London: Sage.

Harlen, W., & Qualter, A. (2014). *The teaching of science in primary schools* (6th ed.). Oxon: Routledge.

Harwell, M. R., Rubinstein, E. N., Hayes, W. S., & Olds, C. C. (1992). Summarizing monte carlo results in methodological research: The one- and two-factor fixed effects ANOVA cases. *journal of Educational Statistics, 17*(4), 315-339.

Hatano, G., & Inagaki, K. (1994). Young children's naive theory of biology. *Cognition, 50*(1), 171-188. doi:https://doi.org/10.1016/0010-0277(94)90027-2

Hatch, J. A. (2002). *Doing qualitative research in educational settings*. Albany State University of New York Press.

Henderson, J. B., MacPherson, A., Osborne, J., & Wild, A. (2015). Beyond construction: Five arguments for the role and value of critique in learning science. *International Journal of Science Education, 37*(10), 1668-1697.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher, 30*(3), 141-158. doi:10.1119/1.2343497

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review, 26*, 55-88.

Hogan, D. (2005). Researching 'the child' in developmental psychology. In S. Greene & D. Hogan (Eds.), *Researching children's experience: Approaches and methods* (pp. 22-41). London: Sage Publications.

Hood, B. (2009). *Supersense: From superstition to religion - the brain science of belief*. London: Constable.

Howe, C. (1998) *Conceptual structure in childhood and adolescence: the case of everyday physics.* Oxon: Routledge.

Howe, C. & Abedin, M. (2013). Classroom dialogue: A systematic review across four decades of research. *Cambridge Journal of Education*, 43, 325-356.

Howe, C., Devine, A., & Tavares, T. J. (2013). Supporting conceptual change in school science: A possible role for tacit understanding. *International Journal of Science Education, 35*, 864-883.

Howe, C., Rodgers, C. & Tolmie, A. (1990). Physics in the primary school: Peer interaction and the understanding of floating and sinking. *European Journal of Psychology of Education, 5*, 459-475.

Inagaki, K., & Hatano, G. (1987). Young children's spontaneous personification as analogy. *Child Development, 58*(4), 1013-1020. doi:10.2307/1130542

Ioannides, C., & Vosniadou, S. (2002). The changing meanings of force. *Cognitive science quarterly, 2*, 5-61.

Jakobson, B., & Wickman, P. (2007). Transformation through language use: Children's spontaneous metaphors in elementary school science. *Science and education, 16*(267-289).

Jewett, E., & Kuhn, D. (2016). Social science as a tool in developing scientific thinking skills in underserved, low-achieving urban students. *Journal of Experimental Child Psychology, 143,* 154-161. doi:10.1016/j.jecp.2015.10.019

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher, 33*(7), 14-26. doi:10.3102/0013189x033007014

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research, 1*(2), 112-133. doi:10.1177/1558689806298224

Kambouri, M. (2016). Investigating early years teachers' understanding and response to children's preconceptions. *European Early Childhood Education Research Journal, 24*(6), 907-927. doi:10.1080/1350293X.2014.970857

Kambouri, M., & Briggs, M. (2013). *The teaching of science in the early years: Do early years teachers identify children's preconceptions.* Paper presented at the ESERA 2013 Conference: Science Education Research For Evidence-based Teaching and Coherence in Learning, Nicosia, Cyprus.

Kampourakis, K. (2014). *Understanding evolution*. Cambridge: Cambridge University Press.

Kampourakis, K., Palaiokrassa, E., Papadopoulou, M., Pavlidi, V., & Argyropoulou, M. (2012a). Children's intuitive teleology: Shifting the focus of evolution education research. *Evolution: Education & Outreach, 5*, 279-291.

Kampourakis, K., Pavlidi, V., Papadopoulou, M., & Palaiokrassa, E. (2012b). Children's teleological intuitions: What kind of explanations do 7–8 year olds give for the features of organisms, artifacts and natural objects? *Research in Science Education, 42*(4), 651-671. doi:10.1007/s11165-011-9219-4

Kampourakis, K., & Zogza, V. (2009). Preliminary evolutionary explanations: A basic framework for conceptual change and explanatory coherence in evolution. *Science & Education, 18*(10), 1313-1340. doi:10.1007/s11191-008-9171-5

Karlsson, P., & Bergmark, A. (2015). Compared with what? An analysis of control-group types in Cochrane and Campbell reviews of psychosocial treatment efficacy with substance use disorders. *Addiction, 110*, 420-428.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. London: The MIT Press.

Keil, F. C. (1992). The origins of an autonomous biology In M. R. Gunnar & M. Maratsos (Eds.), *Minnesota symposium on child psychology* (pp. 103-138). Hilldale, N.J.: Earlbaum.

Keil, F. C. (1994). The birth and nurturance of concepts by domains: The origins of concepts of living things. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 234-254). Cambridge: Cambridge University Press.

Keil, F. C. (1995). The growth of causal understandings of natural kinds. In D. Sperber, D. Premack, & A. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 235-262). Oxford: Clarendon Press.

Kelemen, D. (1999a). Why are rocks pointy? Children's preference for teleological explanations of the natural world. *Developmental Psychology, 35*(6), 1440-1452.

Kelemen, D. (1999b). The scope of teleological thinking in preschool children. *Cognition, 70*(3), 241-272.

Kelemen, D. (1999c). Function, goals and intention: Children's teleological reasoning about objects. *Trends in Cognitive Sciences, 3*(12), 461-468.

Kelemen, D. (2003). British and American children's preferences for teleo-functional explanations of the natural world. *Cognition, 88*(2), 201-221. doi:10.1016/s0010-0277(03)00024-6

Kelemen, D. (2012). Teleological minds: How natural intuitions about agency and purpose influence learning about evolution. In K. Rosengren, S. K. Brem, E. M. Evans, & G. M. Sinatra (Eds.), *Evolution challenges: Integrating research and practice in teaching and learning about evolution* (pp. 66-93). Oxford: Oxford University Press.

Kelemen, D., Callanan, M. A., Casler, K., & Perez-Granados, D. R. (2005). Why things happen: Teleological explanations in parent-child conversations. *Developmental Psychology, 41*(1), 251-264.

Kelemen, D., & DiYanni, C. (2005). Intuitions about origins: Purpose and intelligent design in children's reasoning about nature. *Journal of Cognition and Development, 6*(1), 3-31. doi:10.1207/s15327647jcd0601_2

Kelemen, D., Emmons, N. A., Schillaci, R. S., & Ganea, P. A. (2014). Young children can be taught basic natural selection using a picture-storybook intervention. *Psychological Science, 25*(4), 893-902. doi:10.1177/0956797613516009

Kelemen, D., & Rosset, E. (2009). The human function compunction: Teleological explanations in adults. *Cognition, 111*, 138-143.

Kelemen, D., Seston, R., & Saint Georges, L. (2012). The designing mind: Children's reasoning about intended function and artefact structure. *Journal of Cognition and Development, 13*(4), 439-453. doi:10.1080/15248372.2011.608200

Keogh, B., & Naylor, S. (2012). *Concept cartoons: What have we learnt?* Paper presented at the Fibonacci project European conference, Leicester, UK.

King, G., Keohane, R. O., & Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research*. Chichester: Princeton University Press.

Kipling, R. (1908). *Just so stories*. London: Macmillian.

Klahr, D., & Chen, Z. (2003). Overcoming the positive-capture strategy in young children: Learning about indeterminacy. *Child Development, 75*(5), 1275-1296.

Klahr, D., Zimmerman, C., & Jirout, J. (2011). Educational interventions to advance children's scientific thinking. *Science, 19*(333), 971-975.

Korman, J., Voiklis, J., & Malle, B. F. (2015). The social life of cognition. *Cognition, 135*, 30-65.

Kövecses, Z. (2002). *Metaphor: A practical introduction*. Cambridge: Cambridge university press.

Kruskal, W. H. (1960). Some remarks on wild observations. *Technometrics, 2*, 1-3.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96*(4), 674-689. doi:10.1037//0033-295x.96.4.674

Kuhn, D., Ramsey, S., & Arvidsson, T. S. (2015). Developing multivariable thinkers. *Cognitive Development, 35*, 92-110. doi:10.1016/j.cogdev.2014.11.003

Lakoff, G., Espenson, J., & Schwartz, A. (1991). *Master metaphor list*. Cognitive Linguistics Group. University of California. Retrieved from http://araw.mede.uic.edu/~alansz/metaphor/METAPHORLIST.pdf

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. London: The Univeristy of Chicago Press.

Layton, D. (1981). The schooling of science in England. In R. MacLeod & P. Collins (Eds.), *The parliament of science*. Northwood: Science Reviews Ltd.

Leach, J., Driver, R., Scott, P., & Wood-Robinson, C. (1992). *Progression in understanding of ecology concepts by pupils aged 5 to 16*. CLIS (Children's Learning in Science Research Group, Centre for Studies in Science and Mathematics Education), Leeds: University of Leeds.

Lehner, R., & Schauble, L. (2006). Scientific thinking and science literacy. In W. Damon, R. Lehner, K. A. Renninger, & I. E. Sigel (Eds.), *Handbook of child psychology: Child psychology in practice* (6 ed., Vol. 4, pp. 153-196). Hoboken: Wiley.

Lindeman, M., Svedholm-Hakkinen, A. M., & Lipsanen, J. (2015). Ontological confusions but not mentalizing abilities predict religious belief, paranormal belief, and belief in supernatural purpose. *Cognition, 134*, 63-79.

Lipton, P. (1991). *Inference to the best explanation* (2 ed.). London: Routledge.

Lipton, P. (2004). What good is explanation? In J. Cornwell (Ed.), *Explanations: Styles of explanation in science* (pp. 1-22). Oxford: Oxford University Press.

Lovell, K. (1961). *The growth of basic mathematical and scientific concepts in children*. London: University of London Press.

Loxley, P., Dawes, L., Nicholls, L., & Dore, B. (2010). *Teaching primary science: Promoting enjoyment and developing understanding*. Harlow: Pearson Education Limited.

Lundy, L., & McEvoy, L. (2012). Childhood, the united nations convention of the rights of the child, and research: What constitutes a 'rights-based' approach? In M. Freeman (Ed.), *Law and childhood studies: Current legal issues* (pp. 75-91). Oxford: Oxford University Press.

Mack, L. (2010). The philosophical underpinnings of educational research. *Polyglossia, 19*(10), 5-11.

MacNaughton, G. (2008). *Shaping early childhood* (2 ed.). Berkshire: Open University Press.

MacWhinney, Brian. (2000). *The CHILDES Project: Tools for analyzing talk*. (3 ed.). Mahwah: Lawrence Erlbaum.

Markowsky, J. K. (1975). Why anthropomorphism in children's literature? *Elementary English, 52*(4), 460-466.

McCloskey, M. (1983). Naive theories of motion. In D. Gentner & A. Stevens (Eds.), *Mental models*. Hillsdale, NJ: Lawrence Erlbaum.

McEnery, T. (2006). *Swearing in English: Bad language, purity and power from 1586 to the present* Abington: Routledge.

McEnery, T., Xiao, R. & Tono, Y. (2006) *Corpus-based language studies: an advanced resource book*. Oxon: Routledge

McGarrigle, J., & Donaldson, M. (1974). Conservation accidents. *Cognition, 3*, 341-350.

Mercer, N. (2002). Developing dialogues. In G. Wells & G. Claxton (Eds.), *Learning for life in the 21st century* (pp. 141-153). Oxford: Blackwell Publishing Ltd.

Mercer, N., Dawes, L., & Staarman, J. K. (2009). Dialogic teaching in the primary science classroom. *Language and Education, 23*(4), 1-17.

Mercer, N., Dawes, L., Wegerif, R., & Sams, C. (2004). Reasoning as a scientist: Ways of helping children to use language to learn science. *British Educational Research Journal, 30*(3), 359-377. doi:10.1080/0141192041000168968 9

Mercer, N., & Howe, C. (2012). Explaining the dialogic processes of teaching and learning: The value and potential of sociocultural theory. *Learning, Culture and Social Interaction, 1*(1), 12-21. doi:https://doi.org/10.1016/j.lcsi.2012.03.001

Mercer, N., & Littleton, K. (2007). *Dialogue and the development of children's thinking: A sociocultural approach*. London: Routledge.

Mercier, H. (2013). Using evolutionary thinking to cut across disciplines. In T. R. Zentrall & P. H. Crowley (Eds.), *Comparative decision making*. Oxford: Oxford Scholarship Online.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences, 34*, 57-111.

Mercier, H., & Sperber, D. (2017). *The enigma of reason: A new theory of human understanding*. London: Penguin.

Merzyn, G. (2005). A comparison of some linguistic variables in fifteen science texts. *Research in Science Education in Europe*, 308.

Millar, R., & Osborne, J. (1998). *Beyond 2000: Science education for the future*. London: King's College London.

Michl, J. (2002) 'On seeing design as redesign: an exploration of a neglected problem in design education', *Scandinavian Journal of Design History*, 12, p. 7-23.

Mishler, E. G. (1979). Meaning in context: Is there any other kind? *Harvard Educational Review, 49*(1), 1-19.

Morris, M., Merritt, M., Fairclough, S., Birrell, N., & Howitt, C. (2007). Trialling concept cartoons in early childhood teaching and learning of science. *Teaching Science, 53*(2), 42-45.

Moshman, D., & Geil, M. (1998). Collaborative reasoning: Evidence for collective rationality. *Thinking and Reasoning, 4*(3), 231-248.

Muijs, D. (2011). *Doing quantitative research in education with spss*. London: Sage Publications.

Murphy, K. R., & Alexander, P. A. (2013). Situating text, talk, and transfer in conceptual change: Concluding thoughts. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (2 ed., pp. 601-621). Oxon: Routledge.

Naylor, S., & Keogh, B. (1999). Concept cartoons, teaching and learning in science: An evaluation. *International Journal of Science Education, 21*(4), 431-446.

Naylor, S., & Keogh, B. (2000). *Concept cartoons in science education*. Sandbach: Millgate House Publishers.

Naylor, S., Keogh, B., Downing, B., Maloney, J., & Simon, S. (2007). The puppets project: Using puppets to promote engagement and talk in science. In R. Pintó & D. Couso (Eds.), *Contributions from science education research* (pp. 289-296). Dordrecht: Springer Netherlands.

Newby, P. (2010). *Research methods for education*. Essex: Pearson Education Limited.

Nickerson, R. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*(2), 175-220.

Nitz, S., Ainsworth, S. E., Nerdel, C., & Prechtl, H. (2014). Do student perceptions of teaching predict the development of representational competence and biological knowledge? *Learning and Instruction, 31*, 13-22.

Novak, J. D., & Cañas, A. J. (2006). The theory underlying concept maps and how to construct them. *Florida Institute for Human and Machine Cognition, 1*, 2006-2001.

Novick, L. R., Pickering, J., MacDonald, T., Diamond, J., Ainsworth, S., Aquino, A. E., Scott, M. (2014). Depicting the tree of life in museums: Guiding principles from psychological research. *Evolution: Education and Outreach, 7*(1), 25. doi:10.1186/s12052-014-0025-0

O'Kane, C. (2000). The development of participatory techniques: Facilitating children's views about decisions which affect them. In P. Christensen & A. James (Eds.), *Research with children: Perspectives and practices* (2 ed., pp. 125-155). Oxon: Routledge.

ojalehto, B., Waxman, S. R., & Medin, D. L. (2013). Teleological reasoning about nature: Intentional design or relational perspectives? *Trends in Cognitive Sciences, 17*(4), 166-171.

Oliver, M., & Venville, G. (2015). Cognitive acceleration through science education: The case for thinking through science. In R. Wegerif, L. Li, & J. Kaufman (Eds.), *The Routledge international handbook of research on teaching thinking* (pp. 378-388). Oxon: Routledge.

Oliver, M., & Venville, G. (2016). Bringing CASE in from the cold: The teaching and learning of thinking. *Research in Science Education*, 1-18. doi:10.1007/s11165-015-9489-3

Olson, D. R. (2003). *Psychological theory and educational reform: How school remakes mind and society*. Cambridge: Cambridge University Press.

Opfer, J. E., & Gelman, S. A. (2001). Children's and adults' models for predicting teleological action: The development of a biology-based model. *Child Development, 72*(5), 1367-1381.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17*, 776-783.

Osborne, J., Collins, S., Ratcliffe, M., Millar, R., & Duschl, R. (2003). What "ideas-about-science" should be taught in school science?: A delphi study of the expert community. *Journal of research in science teaching, 40*(7), 692-720.

Özdemir, G., & Clark, D. B. (2007). An overview of conceptual change theories. *Eurasia Journal of Mathematics, Science and Technology Education, 3*(4), 251-261.

Peacock, G., Wright, D., Johnsey, R., & Sharp, J. (2011). *Primary science: Knowledge and understanding* (5th edn. ed.). Exeter: Learning Matters.

Pearson, M., & Somekh, B. (2003). Concept-mapping as a research tool: A study of primary children's representations of information and communication technologies (ict). *Education and information technologies, 8*(1), 5-22.

Pennebaker, J. W. (2011). *The secret life of pronouns*. London: Bloomsbury Press.

Piaget, J. (1929). *The child's conception of the world*. London: Routledge & Kegan Paul Ltd.

Piaget, J. (1952). *The child's conception of number*. london: Routledge and Kegan Paul.

Pihlaja, S. (2013). 'It's all red ink':The interpretation of biblical metaphor among evangelical Christian youtube users.

Polling, D. A., & Evans, E. M. (2002). Why do birds of a feather flock together? Developmental change in the use of multiple explanations: Intention, teleology and essentialism. *British Journal of Developmental Psychology, 20*(1), 89-112.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education, 66*(2), 211-227.

Potter, G. (2000). *The philosophy of social science: New perspectives*. London: Routledge.

Pring, R. (2004). *The philosophy of educational research* London: Continnuum.

Ribeiro, M. G. L., Larentis, A. L., Caldas, L. A., Garcia, T. C., Terra, L. L., Herbst, M. H., & Almeida, R. V. (2015). On the debate about teleology in biology: The notion of "teleological obstacle". *Historia Ciencias Saude-Manguinhos, 22*(4). doi:10.1590/s0104-59702015005000003

Robson, C. (2011). *Real world research* (3rd ed.). Sussex: John Wiley & Sons Ltd.

Roschelle, J., Feng, M., Gallager, H. A., Murphy, R., Harris, C., Kamdar, D., & Trinidad, G. (2014). *Recruiting participants for large-scale random assignment experiments in school settings*. Menlo Park, CA: SRI International.

Roth, K. (2014). Elementary science teaching. In N. G. Lederman & S. K. Abell (Eds.), *Handbook research on science education* (Vol. 2, pp. 361-364).

Rottman, J., Zhu, L., Wang, W., Seston Schillaci, R., Clark, K. J., & Kelemen, D. (2017). Cultural influences on the teleological stance: Evidence from China. *Religion, Brain & Behavior, 7*(1), 17-26. doi:10.1080/2153599X.2015.1118402

Runciman, D. (2017). How climate scepticism turned into something more dangerous. *The Guardian*. Retrieved from https://www.theguardian.com/environment/2017/jul/07/climate-change-denial-scepticism-cynicism-politics

Ruse, M. (1989). Teleology in biology: Is it a cause for concern? *TREE, 4*(2), 51-54.

Ruse, M. (2000). Teleology: Yesterday, today, and tomorrow? *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 31*(1), 213-232.

Ruse, M. (2005). Darwinism and mechanism: Metaphor in science. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Science, 36*(2), 285-302.

Russ, R. S., Coffey, J. E., Hammer, D., & Hutchison, P. (2008). Making classroom assessment more accountable to scientific reasoning: A case for attending to mechanistic thinking. *Science Education, 93*(5), 875-891.

Samarapungavan, A., & Wiers, R. W. (1997). Children's thoughts on the origin of species: A study of explanatory coherence. *Cognitive Science, 21*(2), 147-177. doi:https://doi.org/10.1016/S0364-0213(99)80021-4

Sampson, G. R. (2007). There is no language instinct. *Ilha Do Desterro, 52*(8), 35-63.

Sampson, G. R., & Babarczy, A. (2013). *Grammar without grammaticality: Growth and limits of grammatical precision*. Boston: De Grugter.

Sánchez Tapia, I., Gelman, S. A., Hollander, M. A., Manczak, E. M., Mannheim, B., & Escalante, C. (2016). Development of teleological explanations in Peruvian Quechua-speaking and u.S. English-speaking preschoolers and adults. *Child Development, 87*(3), 747-758. doi:10.1111/cdev.12497

Sawyer, R. K. (2014). Introduction: The new science of learning. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 1-20). Cambridge: Cambridge University Press.

Schallert, D., Kleiman, G., & Rubin, A. (1977). *Analysis of differences between written and oral language*. Urbana, IL: Center for Reading Study.

Schwandt, T. A. (2000). Three epistemological stances for qualitative inquiry: Interpretivism, hermeneutics, and social constructionism. In N. K. Denzin & Y. S. Lincoln (Eds.), *The handbook of qualitative research* (2 ed., pp. 189-215). London: Sage Publications.

Scott, J. (2008). Children as respondents: The challenge for quantitative methods. In P. Christensen & A. James (Eds.), *Research with children: Perspectives and practice* (2 ed.). London: Routledge.

Seleznyov, S., & Adhami, M. (2015). *Let's think: Coming of age for cognitive acceleration*. Paper presented at the LSRI seminar, Nottingham.

Serlin, R. C., Carr, J., & Marascuilo, L. A. (1982). A measure of association for selected nonparametric procedures. *Psychological Bulletin, 92*(3), 786-790.

Sharp, J., Peacock, G., Johnsey, R., Simon, S., Smith, R., Corss, A., & Harris, D. (2014). *Primary science: Teaching theory and practice* (7 ed.). London: Learning Matters.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs: For generalized causal inference*. Belmont: Wadsworth.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52*(3/4), 591-611. doi:10.2307/2333709

Shayer, M. (2003). Not just Piaget; not just Vygotsky, and certainly not Vygotsky as alternative to Piaget. *Learning and Instruction, 13*(5), 465-485. doi:http://dx.doi.org/10.1016/S0959-4752(03)00092-6

Shayer, M., & Adhami, M. (2007). Fostering cognitive development through the context of mathematics: Results of the came project. *Educational Studies in Mathematics, 64*(3), 265-291.

Shermer, M. (2002). *Why people believe weird things*. New York: Saint Martin's Press.

Siegler, R. S. (1998). *Children's thinking* (3rd ed.). London: Prentice-Hall International.

Simon, S., Naylor, S., Keogh, B., Maloney, J., & Downing, B. (2008). Puppets promoting engagement and talk in science. *International Journal of Science Education, 30*(9), 1229-1248.

Singer, M. A., & Nievod, A. (2003). New age therapies. In S. O. Lilienfeld, S. J. Lynn, & J. M. Lhor (Eds.), *Science and pseudoscience in clinical psychology* (pp. 176-204). New York: Guilford Press.

Smith, E. (2010a). Is there a crisis in school science education in the UK? *Educational Review, 62*(2), 189-202.

Smith, J. (2010b). *Talk, thinking and philosophy in the primary classroom*. Exeter: Learning Matters Ltd.

Smith, J. P., diSessa, A. A., & Roschelle, J. (1993). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences, 3*(2), 115-163.

Smith, P. (2001). *Talking classrooms: Shaping children's learning through oral language instruction*. Newark: International Reading Association.

Smith, P. K., Cowie, H., & Blades, M. (2003). *Understanding children's development* (4 ed.). London: Blackwell Publishers.

Smith, P. K., Cowie, H., & Blades, M. (2015). *Understanding children's development* (6 ed.). Chichester: Wiley.

Southerland, S. A., Abrams, E., Cummins, C. L., & Anzelmo, J. (2001). Understanding students' explanations of biological phenomena: Conceptual frameworks or p-prims? *Science Education, 85*(4), 328-348. doi:10.1002/sce.1013

Stake, R. E. (2010). *Qualitative research: Studying how things work*. London: The Guilford Press.

Stubbs, M. (1995) Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of Language, 2* (1), 1-24.

Sutherland, P. (2000). *Cognitive development today: Piaget and his critics*. London: Sage Publications.

Switek, B. (2014). Getting to the root of fur. *National Geographic*.

Taber, K. S. (2014). Alternative conceptions/frameworks/misconceptions. In R. Gunstone (Ed.), *Encyclopedia of science education* (pp. 1-5). Dordrecht: Springer Netherlands.

Taggart, G., Ridley, K., Rudd, P., & Benefield, P. (2005). *Thinking skills in the early years: A literature review*. Retrieved from Slough:

Talanquer, V. (2007). Explanations and teleology in chemistry education. *International Journal of Science Education, 29*(7), 853-870.

Tamir, P., & Zohar, A. (1991). Anthropomorphism and teleology in reasoning about biological phenomena. *Science Education, 75*(1), 57-67. doi:10.1002/sce.3730750106

Tao, Y., Oliver, M., & Venville, G. J. (2011). Chinese and Australian year 3 children's conceptual understanding of science: A multiple comparative case study. *International Journal of Science Education, 34*(6), 879-901.

Tao, Y., Oliver, M., & Venville, G. J. (2013). A comparison of approaches to the teaching and learning of science in Chinese and Australian elementary classrooms: Cultural and socioeconomic complexities. *Journal of research in science teaching, 50*(1), 33-61.

Thomas, K. (1984). *Man and the natural world: Changing attitudes in England 1500-1800*. London: Penguin Group.

Thomaz, M. F., Malaquis, I. M., Valente, M. C., & Antunes, M. J. (1995). An attempt to overcome alternative conceptions related to heat and temperature. *Physics Education, 30*(1), 19-26.

Thulin, S., & Pramling, N. (2009). Anthropomorphically speaking: On communication between teachers and children in early childhood biology education. *International Journal of Early Years Education, 17*(2), 137-150.

Tomasello, M. (1999). *The cultural origins of human cognition*. United States of America: First Harvard University Press.

Treagust, D. F., Chittleborough, G. D., & Mamiala, L. T. (2002). Students' understanding of the role of scientific models in learning science. *International Journal of Science Education, 24*, 357-368.

Treagust, D. F., & Tsui, C. (2014). General instructional methods and strategies. In N. G. Lederman & S. K. Abell (Eds.), *Handbook of research on science education* (Vol. II, pp. 303-320). Oxon: Routledge.

Trickey, S., & Topping, K. J. (2004). 'Philosophy for children': A systematic review. *Research Papers in Education, 19*(3), 365-380.

Tryphon, A., & Voneche, J. (1996). *Piaget - Vygotsky: The social genesis of thought*. New York: Psychology Press.

Tu, T. (2006). Preschool science environment: What is available in a preschool classroom? *Early Childhood Education Journal, 33*(4), 245-251.

Tunnicliffe, S. D., & Reiss, M. J. (1999). Building a model of the environment: How do children see animals? *Journal of Biological Education, 33*(3), 142-148. doi:10.1080/00219266.1999.9655654

United Nations (1989). *United Nations convention on the rights of the child (uncrc)*. Geneva: United Nations.

UoN (2015). School of education: Research ethics. Retrieved from https://www.nottingham.ac.uk/educationstudentintranet/researchethics/index.aspx

UREC. (no date). The appropriateness of opt in and opt out recruitment methodologies. Retrieved from www.sheffield.ac.uk/polopoly_fs/1.165632!/file/AGN-Opt-In-Opt-Out.pdf

Urquiza-Haas, E. G., & Kotrschal, K. (2015). The mind behind anthropomorphic thinking: Attribution of mental states to other species. *Animal Behaviour, 109*, 167-176. doi:https://doi.org/10.1016/j.anbehav.2015.08.011

Venville, G., Adey, P., Larkin, S., Robertson, A., & Fulham, H. (2003). Fostering thinking through science in the early years of schooling. *International Journal of Science Education, 25*(11), 1313-1331.

Venville, G., & Oliver, M. (2015). The impact of a cognitive acceleration programme in science on students in an academically selective high school. *Thinking Skills and Creativity, 15*, 48-60. doi:10.1016/j.tsc.2014.11.004

Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology, 24*, 535-585.

Vosniadou, S., & Skopeliti, I. (2017). Is it the earth that turns or the sun that goes behind the mountains?: Students' misconceptions about the day/night cycle after reading a science text. *International Journal of Science Education, 39*(15), 2027-2051.

Vygotsky, L. (1978). *Mind in society* (M. Cole, V. John-Stenier, M. Scribner, & E. Souberman Eds.). London: Harvard University Press.

Vygotsky, L. (1986). *Thought and language* (A. Kozulin Ed.). Cambridge: The MIT Press.

Weber, S., & Cook, T. D. (1972). Subject effects in laboratory research: An examination of subject roles, demand characteristics, and valid inference. *Psychological Bulletin, 77*(4), 273-295.

Wegerif, R., Mercer, N., & Dawes, L. (1999). From social interaction to individual reasoning: An empirical investigation of a possible socio-cultural model of cognitive development. *Learning and Instruction, 9*(6), 493-516.

Weinstein, R. S. (1976). Reading group membership in first grade: Teacher behaviors and pupil experience over time. *Journal of Educational Psychology, 68*, 103-116.

Wellington, J. (2004). *Educational research: Contemporary issues and practical approaches*. Oxford: Bloomsbury Publishing.

Wellington, J., & Osborne, J. (2001). *Language and literacy in science education*. Buckingham: Open University Press.

Whorf, B. L. (1956). *Language, thought and reality: Selected writings of benjamin lee whorf* (J. B. Carroll Ed.). Cambridge: The MIT Press.

Wicken, J. S. (1981). Causal explanations in classical and statistical thermodynamics. *Philosophy of Science, 48*(1), 65-77.

Wild, K., Kilgarriff, A., & Tugwell, D. (2013). The oxford children's corpus: Using a children's corpus in lexicography. *International Journal of Lexicography, 26*(2), 190-218.

Wilkinson, A. (1971). *The foundations of language: Talking and reading in young children*. London: Oxford Universtiy Press.

Willard, A. K., & Norenzayan, A. (2013). Cognitive biases explain religious belief, paranormal belief, and belief in life's purpose. *Cognition, 129*, 379-391.

Wilson, S., & Maclean, R. (2011). *Research methods and data analysis for psychology*. Berkshire: McGraw-Hill Education.

Winnicott, D. W. (1971). *Therapeutic consultations in child psychiatry*. London: The Hogarth Press.

Wood-Robinson, C. (1995). Children's biological ideas: Knowledge about ecology, inheritance, and evolution. In S. M. Glynn & R. Duit (Eds.), *Learning science in the schools: Research reforming practice*. New Jersey: Lawrence Erlbaum Associates.

Woodfield, A. (1976). *Teleology*. Cambridge: Cambridge University Press.

Woodhead, M., & Faulkner, D. (2000). Subjects, objects or participants?: Dilemmas of psychological research with children. In P. Christensen & A. James (Eds.), *Research with children: Perspectives and practicies* (2 ed., pp. 10-39). Oxon: Routledge.

Wynne, C. D. L. (2007). What are animals? Why anthropomorphism is still not a scientific approach to behavior. *Comparative Cognition & Behavior Reviews, 2*, 125-135.

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review, 27*, 172-223.

Zimmerman, D. W. (1994). A note on the influence of outliers on parametric and nonparametric tests. *The Journal of General Psychology, 121*(4), 391-401.

Zohar, A., & Ginossar, S. (1998). Lifting the taboo regarding teleological and anthropomorphism in biology education - heretical suggestions. *Science Education, 82*, 679-697.

10. Appendices

Appendix 1: Combined information sheet and opt-out consent letter used in Study 1

Appendix 2: Full collocate list split by collocate categories and keyword

Appendix 3: Concept cartoons used in Study 2

Appendix 4: Concept cartoons used in Study 3

Appendix 1: Example information sheet and opt-out consent letter (Study 1)

Information Sheet for Parents: How does question wording affect children's scientific answers?

My name is Jonathan Halls and I am a PhD student at the University of Nottingham, in the Learning Sciences Research Institute, School of Education. My research investigates how young children learn science and what can be done to improve their knowledge.

The research seeks to challenge traditionally held assumptions by some academics that young children cannot learn scientific concepts. I take the belief that children are more capable than they are given credit for and, if taught correctly, they can understand scientific concepts. Furthermore, I maintain that the view that children cannot understand scientific theories has arisen from asking them the wrong questions, questions which lead children into providing a non-scientific answer.

This part of the project, which I hope your child will be involved in, examines how different question wordings affects children's responses. To be involved your child would need to partake in a relaxed 10-15 minute interview, which would be audio-recorded. The interview would not seek to challenge your child's beliefs, merely elicit their views about different science topics. They would be free to stop the interview or to not answer any question.

The proposed research project has been cleared by the University of Nottingham Ethics committee and follows the ethical protocols outlined by the British Educational Research Association. This means that your child's identity will be anonymised from the outset and all materials relating to your child would be confidential and only available to my supervisor and myself. You would have the right to withdraw your child from the study at any point before, during or after the interview. The results of the project will be provided to the school and if you require I can provide an electronic copy. Additionally, the results of the study may be published online.

If you have any questions or require more information about the project please contact me using the following contact details.

If you are happy for your child to be involved please complete and return the attached consent form

Jonathan Halls
Learning Science Research Institute,
School of Education,
University of Nottingham
ttxjh76@nottingham.ac.uk
0115 846 7672

Consent Form

Project title: How does question wording affect children's scientific answers?
Researcher's name: …..…Jonathan Halls……
Supervisor's name: Professor Shaaron Ainsworth and Professor Mary Oliver……

- I have read the Participant Information Sheet and the nature and purpose of the research project has been explained to me. I understand and agree to my child taking part;
- I understand the purpose of the research project and my child's involvement in it;
- I understand that I may withdraw my child from the research project at any stage and that this will not affect their status now or in the future;
- I understand that while information gained during the study may be published, my child will not be identified and their personal results will remain confidential;
- I understand that my child will be audio taped during the interview;
- I understand that audio data will be stored electronically and transcripts of the interviews will be stored as both electronic and as hard copies. Access to these materials will be limited to the researcher and the supervisor;
- I understand that I may contact the researcher or supervisor if I require further information about the research, and that I may contact the Research Ethics Coordinator of the School of Education, University of Nottingham, if I wish to make a complaint relating to my child's or my involvement in the research.

Signed …… (Child's Parent/Guardian)

Print name ……..……**Date** ……..

Print Child's name ……

Contact details
Researcher: ttxjh76@nottingham.ac.uk; 0115 846 7672
Supervisors: Shaaron.ainsworth@nottingham.ac.uk; 0115 846 767
Mary.oliver@nottingham.ac.uk; 0115 748 4791
Research Ethics Coordinator: educationresearchethics@nottingham.ac.uk

Appendix 2: Full collocate list split by collocate categories and keyword

Collocates for day

Category	Tokens	Type	Collocates (freq.)
Time period	2304	84	Punishment (5), Backward (4), Flying (3), Forget (3), Today (309), Other (259), Next (146), Another (119), After (92), Tomorrow (79), Last (78), Night (50), Sunday (49), Before (44), Yesterday (42), Remember (37), Christmas (35), Whole (34), Week (33), Birthday (31), Saturday (30), Monday (25), Friday (24), Half (24), Times (22), Part (21), Morning (21), Wednesday (19), Mothers (18), Tuesday (18), Fathers (17), Thursday (16), Early (14), Most (14), Valentines (13), Hour (12), Year (11), Single (9), Years (9), Different (9), Valentine (8), Summer (8), Forgot (7), Life (7), Past (7), Special (7), Cards (7), Tues (6), Boxing (6), Month (6), Holidays (6), Afternoon (6), Weeks (6), Clock (6), Saint (5), April (5), Forest (5), Easter (5), God (5), Labor (4), Birth (4), Sports (4), Weekend (4), Summers (3), Allowance (3), Sport (3), Someday (3), Throughout (3), Calendar (3), Autumn (3), Everyday (3), Decided (3), Extra (3), Loft (7), Hours (12), Break (7), Wait (21), Party (11), Married (3), Born (3), Trashy (3), Card (4), Moving (3), Pay (3), Staff (9)
Activity	1018	67	Brush (11), School (139), Tea (136), Outside (39), Bed (37), Sleep (36), Work (32), Works (12), Sleeps (11), Club (11), Walked (10), Wake (9), Swimming (9), Running (9), Run (9), Food (9), Found (9), Walking (8), Hospital (7), Working (7), Brought (6), Breakfast (6), Books (6), Cartoon (5), Slept (5), Climbed (5), Holiday (5), Asleep (5), Write (5), Spanked (4), Planted (4), Concert (4), Involved (4), Museum (4), Pencils (4), Service (4), Foundation (4), Cleaning (4), Dream (4), Feed (4), Nap (4), Colouring (4), Group (4), Zoo (4),

			Football (4), Flowers (4), Pictures (4), Attendance (3), Register (3), Breaks (3), Bored (3), Trips (3), Woods (3), Candy (3), Training (3), Roller (3), Worked (3), Trip (3), Library (3), Woke (3), Picnic (3), Swim (3), Homework (3), Job (3), Washing (3), Egg (6), Numbers (3), Pass (3), Ruined (5), Farm (4) Of which are Play (Tokens = 172 (3.3%), Types =18 (5.8%)) Play (50), Dough (4), Outdoor (3), Played (13), Toys (10), Pretend (8), Toy (7), Skating (5), Games (4), Cars (4), Plays (3), Fly (8), Older (3), Fun (15), Street (9), Kite (3), Cricket (3), Playing (20)
Descriptive	592	56	Long (101), Sleepy (69), Lovely (32), Hot (20), Wear (17), Bad (17), Windy (15), Raining (14), Warm (14), Great (14), Hard (13), Wet (12), Mad (12), Tired (11), Wonderful (10), Pretty (10), Short (7), Awful (6), Brilliant (6), Upset (5), Gorgeous (5), Rough (4), Wearing (4), Difficult (4), Muddy (3), Wore (3), Noticed (3), Bright (3), Longer (3), Tiny (3), Fruit (3), Pyjamas (3), Point (3), Puddles (3), Dress (4), Real (14), Quiet (4), Sunny (47), Snowy (8), Weather (5), Round (14), Cold (9), Beautiful (17), Lucky (3), Rainy (11), Tiring (4)
Names	522	59	John (65), Mary (62), Darling (34), Children (30), Other (18), Dad (18), Father (17), Dee (14), Mummy (14), Penny (12), Peter (9), Cathy (9), Teacher (9), Dada (9), Nobody (8), Rob (8), Jeannine (8), Kathryn (7), Kid (7), Spot (7), Michael (7), Honey (7), Mister (6), Kim (6), Rabbit (6), David (6), Albert (5), Adam (5), Joey (5), Kids (5), Alice (4), Rachel (4), Parents (4), Giraffe (4), Bunny (4), Graeme (4), Auntie (4), Estelle (3), Kylie (3), Deidre (3), Prince (3), Maggie (3), Ethan (3), Ronald (3), Bobbi (3), Leanne (3), Asher (3), Keisha (3), Patrick (3), Woman (3), Noddy (3), Cows (3), Missus (3), King (3), Thomas (3), Ann (3), Neighbour (3), Jason (3), Girls (8)

NP	465	6	Day (414), Days (17), Sun (17), Rain (10), Wind (3), Snow (4)
Entertain-ment	189	24	Wolf (6), Act (3), Hungry (66), Sat (7), Scared (7), Dopey (4), Blew (4), Thinny (3), Fatty (3), Tortoise (3), Sell (3), Beanstalk (3), Cupboard (3), Bears (3), Live (11), Phoenix (3), Watch (19), Land (6), Tellytubbies (3), Television (3), Red (15), Hammers (4), Song (4), Dame (3)
Location	116	12	Camp (9), Home (29), House (29), Park (6), Teletubbyland (4), Barcelona (3), Nursery (9), Town (4), Preschool (4), Beach (3), Care (8), Class (8)

Collocates for Night

Category	Tokens	Types	Collocates (freq.)
Time period	1532	27	Last (643), Time (230), Saturday (83), Friday (60), Day (50), Other (48), Late (41), Middle (41), Tomorrow, (39), Tonight (35), Before (32), Happened (32), Morning (31), Thursday (25), Clock (23), Tuesday (20), Wednesday (16), Sunday (15), Remember (12), Christmas (11), During (11), Early (9), Hours (7), Monday (7), Afternoon (4), Summer (4), Daytime (3)
NP	1474	9	Night (1404), Nights (19), Moon (12), Sun (11), Stars (10), Rain (7), Raining (4), Star (4), Storm (3)
Activity	232	86	Ate (21), Dinner (17), Party (17), Work (14), Sick (11), Date (10), Shift (9), Juice (8), Works (8), Bath (7), Walk (7), Circus (5), Running (5), Walks (5), Burger (4), Dishes (4), Drawer (4), Fireworks (4), Leaving (4), Pizza (4), Poorly (4), Reading (4), Roll (4), Sing (4), Spilled (4), Spinach (4), Stories (4), Classes (3), Cookies (3), Meat (3), Poured (3), Rice (3), Sauce (3), Supper (3), Travels (3), Volleyball (3), Walked (3), Watermelon (3), Working (3)
			Of which are Sleep (Tokens = 681 (11.8%), Types = 39 (15.1%))
			Sleep (148), Up (144), Bed (86), Sleeping (33), Room (25), Slept (24), Wake (19), Dream (16), Dreamed (14), Asleep (13), Wear (13), Wore (13), Nappie (10), Nappies (8), Woke (8), Bedtime (7), Dreamt (7), Pillow (7), Sleeps (7), Floated (6), Rest (6), Sleepy (6), Snakes (6), Bright (5), Cot (5), Fell (5), Shirt (5), Terrible (5), Accident (3), Awake (3), Closes (3), Den (3), Diaper (3), Doddy (3), Dreams (3), Falls (3), Nap (3), Snoring (3)

			Of which are Play (Tokens = 49 (0.9%), Types = 8 (3.1%)) Playing (10), Cowboys (9), Fun (8), Game (6), Floor (5), Played (4), Stickle (4), Bricks (3)
Names	868	91	Boys (9), Cathy (9), Mother (9), Paddington (9), Kat (8), Mummy (8), Panda (8), Sweetie (8), Tigger (8), Abe (7), Christopher (7), Maisie (7), Maisy (7), Nanny (7), Pilchard (7), Turtle (7), Bill (6), Chicken (6), Family (6), Honey (6), Kate (6), Lady (6), Naomi (6), Pig (6), Guys (5), Jan (5), Kids (5), Adam (4), Babe (4), Carl (4), Chris (4), David (4), Donkey (4), Granny (4), Jessica (4), Lenny (4), Maggie (4), Max (4), Mike (4), Pooh (4), Poohbear (4), Ann (3), Anybody (3), Benjamin (3), Brum (3), Dolly (3), Eric (3), Father (3), Frog (3), Hen (3), Husband (3), Joey (3), Kayoed (3), Kim (3), Mam (3), Meg (3), Mickey (3), Pia (3), Polly (3), Poo (3), Son (3), Spot (3), Sweet (3), Tony (3), Owl (7), Animals (6), Owls (6), Flies (3), Mosquitoes (3), Swat (3)
Greeting	590	10	Good (496), Bye (30), Kiss (20), Byebye (11), Whispers (11), Great (6), Everyone (5), Shh (5), Goodbye (3), Kisses (3)
Entertainment	172	18	Story (29), Read (15), Grew (14), Beanstalk (9), Makers (8), King (5), Poisonous (5), Telling (5), Adventure (4), Bashful (4), Shoemaker (3), Watch (22), Episode (18), TV (14), Watched (7), Film (4), News (3), Television (3)
Location	81	6	Stay (32), Home (19), Spend (14), Stayed (9), Spent (4), Dublin (3)
Description	71	12	Light (12), Tired (12), Cold (10), Dark (7), Brilliant (6), Black (4), Bump (4), Wet (4), Busy (3), Dry (3), Lights (3), Sky (3)

Collocates for Rain

Category	Tokens	Types	Collocates (freq.)
NP	405	15	Rain (221), Raining (60), Cloud (21), Drops (20), Sun (17), Clouds (13), Rainy (9), Snow (9), Water (9), Snowing (7), Weather (5), Rainbow (4), Wind (4), Drop (3), Rains (3)
Descriptive	242	23	Stopped (48), Started (22), Hard (18), Little (16), Big (15), Pouring (15), Green (14), Cold (13), Sunny (11), Dried (9), Nice (8), Shower (8), Dripping (6), Heavy (5), Thunder (5), Yellow (5), Dark (4), Light (4), Really (4), Cats (3), Quite (3), Shape (3), Terrible (3)
Time period	170	18	Today (44), Day (25), Tea (13), Night (12), Time (11), Morning (9), Tonight (9), Through (8), Before (6), Home (6), Yesterday (6), Afternoon (3), Friday (3), Late (3), Later (3), Remember (3), Summer (3), Tomorrow (3)
Activity	135	21	Wet (20), Starting (4), Umbrella (11), Quick (9), Car (7), Bike (6), Pants (6), Coat (5), Cover (5), Plants (5), Shoes (5), Bothering (4), Wear (4), Apple (3), Boots (3), Flowers (3), Grow (3), Indoors (3), Of which are Play (Tokens = 29 (2.2%), Types = 4 (3.3%)) Play (15), Pretend (7), Stepped (4), Happy (3)
Name	101	13	Fraser (23), Darling (20), Mummy (13), Dear (10), Eleanor (6), Tallulah (5), Cathy (4), Mummy (4), Ste (4), Helen (3), Joey (3), Mark (3), Vicky (3)
Entertainment	82	14	Spider (11), Washed (11), Singing (10), Incy (8), Sing (8), Spout (5), Dance (4), Door (4), God (4), Poor (4), Wincy (4), Lemon (3), Old (3), Swing (3)

Location	81	7	Outside (51), Inside (11), House (5), Park (5), Window (5), Hole (4)
Instruction	67	4	Look (55), Hear (5), Looking (4), Watch (3)
Posturing	58	5	Think (34), Coming (7), Might (7), Thought (6), Hope (4)

Collocates for Rainbows

Category	Tokens	Types	Collocates (freq.)
Descriptive	145	14	Fat (66), Big (26), Colors (12), Beautiful (6), Blue (4), Good (4), Orange (4), Sometimes (4), Wow (4), Color (3), Nice (3), Red (3), Trail (3), Yellow (3)
NP	58	1	Rainbow (58)
Name	39	9	Eleanor (6), Ric (6), Dada (5), Emma (5), Darling (4), Girl (4), Fay (3), Mum (3), Nic (3)
Activity	34	6	Make (7), Making (7), Watching (7), Painting (6), Makes (4), Finished (3)
Entertainment	32	6	TV (12), Song (5), Watch (5), Bears (4), Book (3), Program (3)
Instruction	16	1	Look (16)

Collocates for Snow

Category	Tokens	Types	Collocates (freq.)
Entertainment	376	19	White (207), Bear (59), Seven (12), Dwarfs (11), Mister (10), Crying (7), Witch (7), Beard (5), Books (5), Fleece (5), Queen (4), Dog (3), Furry (3), Prince (3), Dashing (3), Horse (3), Watch (18), Speeder (7), Watched (4)
Descriptive	367	52	White (61), Snowing (33), Gone (14), Little (12), Water (12), Cold (11), Ground (10), Rocks (10), Beautiful (9), Covered (8), Melt (8), Through (8), Drop (7), Fun (7), Melts (7), Wow (7), Best (6), Falls (6), Happened (6), Hard (6), Melted (6), Tree (6), Colors (5), Everything (5), Falling (5), Inches (5), Melting (5), Color (4), Feet (4), Happy (4), Hate (4), Mountain (4), Snowy (4), Started (4), Cracks (3), Deep (3), Finished (3), Flakes (3), Frosty (3), Garden (3), Happens (3), Hurt (3), Love (3), Pile (3), Plenty (3), Pretty (3), Sitting (3), Stuck (3), Trees (3), Wet (3), Window (3), Yellow (3)
Activity	334	42	Stay (23), Tuned (17), Pants (12), Suit (10), Watching (10), Picture (8), Boots (6), Draw (6), Walking (6), Inside (5), Stand (5), Coat (4), Pocket (4), Standing (4), Jacket (3), Stayed (3) Of which are Play (Tokens = 208 (13.6%), Types = 26 (15.4%)) Ball (25), Man (23), Play (21), Make (18), Made (14), Snowman (12), Spot (11), Built (10), Pretend (7), Snowballs (6), Eyes (5), Jumped (5), Making (5), Minute (5), Playing (5), Box (4), Ready (4), Run (4), Kick (3), Lady (3), Men (3), Nuts (3), Played (3), Ride (3), Throw (3), Throwing (3)

NP	226	8	Snow (191), Rain (9), Raining (7), Storm (5), Snowed (4), Sun (4), Ice (3), Snowflakes (3)
Time period	108	16	Time (22), Today (15), Winter (14), Christmas (10), Day (5), Remember (5), Season (5), Summer (5), Tonight (4), Wintertime (4), Yesterday (4), Autumn (3), Friday (3), January (3), Last (3), Morning (3)
Names	94	14	Eleanor (19), Darling (17), Mom (9), Mommy (8), Dad (6), Fay (6), Marky (5), Kay (4), Mike (4), Steve (4), Birds (3), Gabriella (3), Greg (3), Teddy (3)
Posturing	93	11	Think (26), Gonna (24), Wonder (13), Never (6), Around (5), Keep (4), Guess (3), Reading (3), Somebody (3), Supposed (3), Wish (3)
Location	58	7	Outside (23), House (16), Place (5), Head (4), Newcastle (4), Chicago (3), Everywhere (3)

Collocates for Storms

Category	Tokens	Types	Collocates (freq.)
NP	17	2	Storm (12), Snow (5)
Time period	14	4	Clock (4), Coming (4), Night (3), Nine (3)
Descriptive	6	2	Scared (3), Squeals (3)

Collocates for Waterfalls

Category	Tokens	Types	Collocates (freq.)
Action	12	2	Under (9), Bar (3)

Collocates for Waves

Category	Tokens	Types	Collocates (freq.)
Descriptive	48	10	Big (14), Hit (7), Pool (6), Wrong (5), Waves (4), Huge (3), Sound (3), Crash (3), Water (3)
Activity	29	3	Wave (22), Jumping (4), Over (3)
NP	16	1	Tidal (16)

Appendix 3: Concept Cartoons used in Study 2





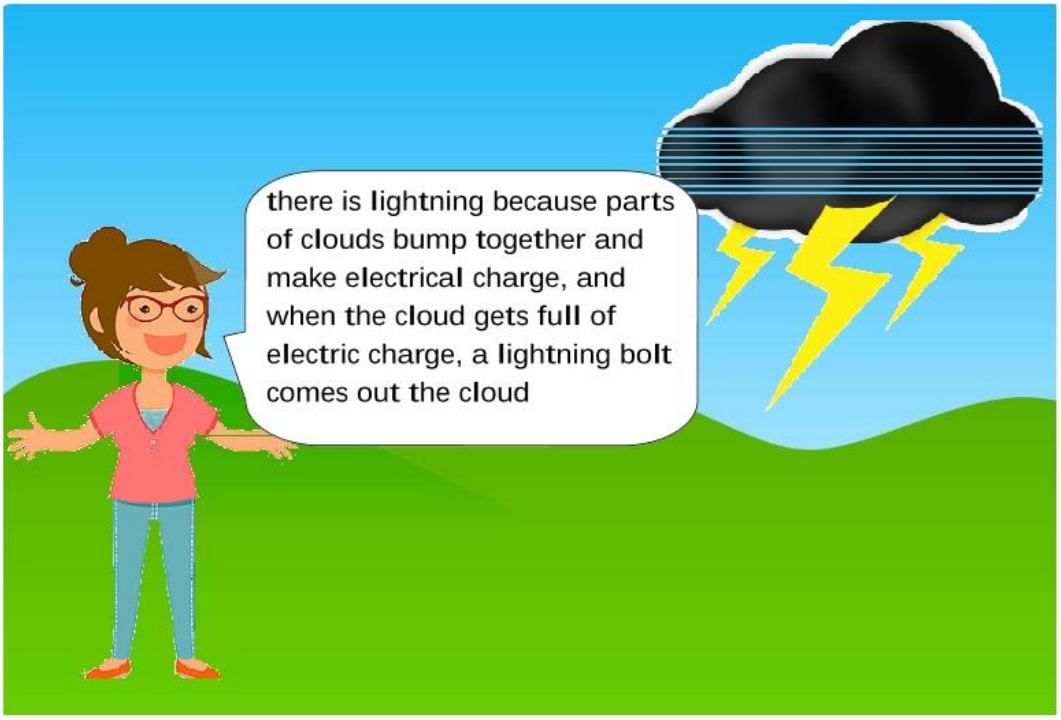

there is day because the Earth is spinning around and when the bit we are on faces the sun, then we have daytime


there is night because the sun goes down and then the sky turns dark


there is night time so people know when to go to sleep


there is lightning for lighting up a dark, stormy sky


there is lightning so people can use it to power electric appliances, like the TV or lights


there is lightning because parts of clouds bump together and make electrical charge, and when the cloud gets full of electric charge, a lightning bolt comes out the cloud


there is snow for telling some animals to hibernate over winter


there is snow so children can make snowmen

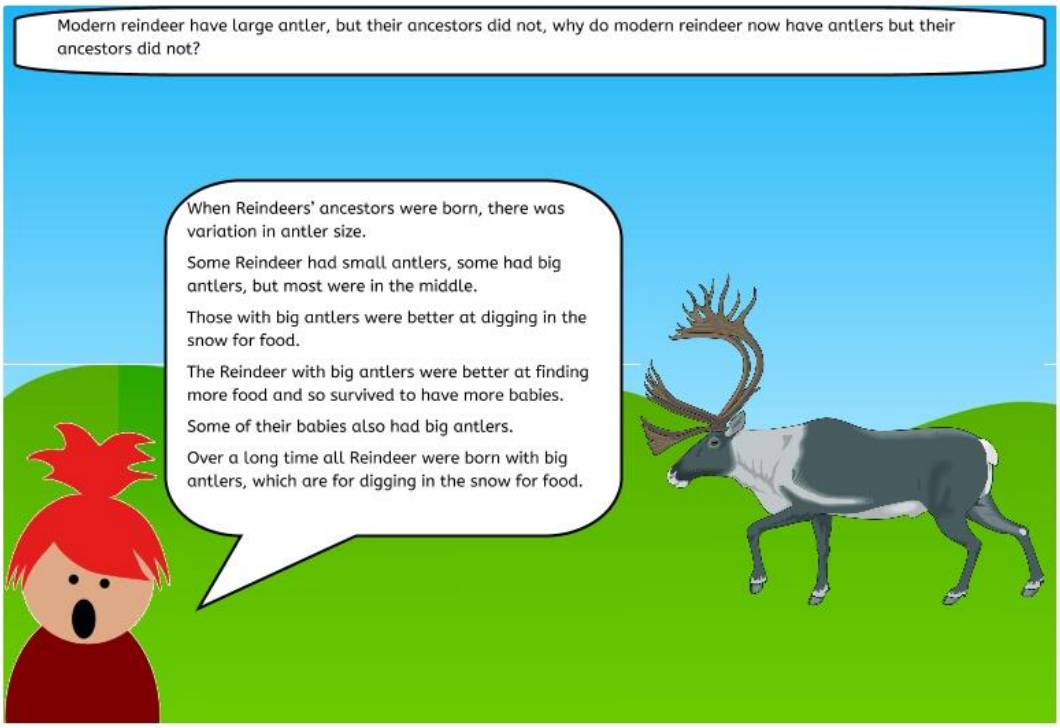

there is snow because it is really cold and the rain drops freeze and turn into snowflakes

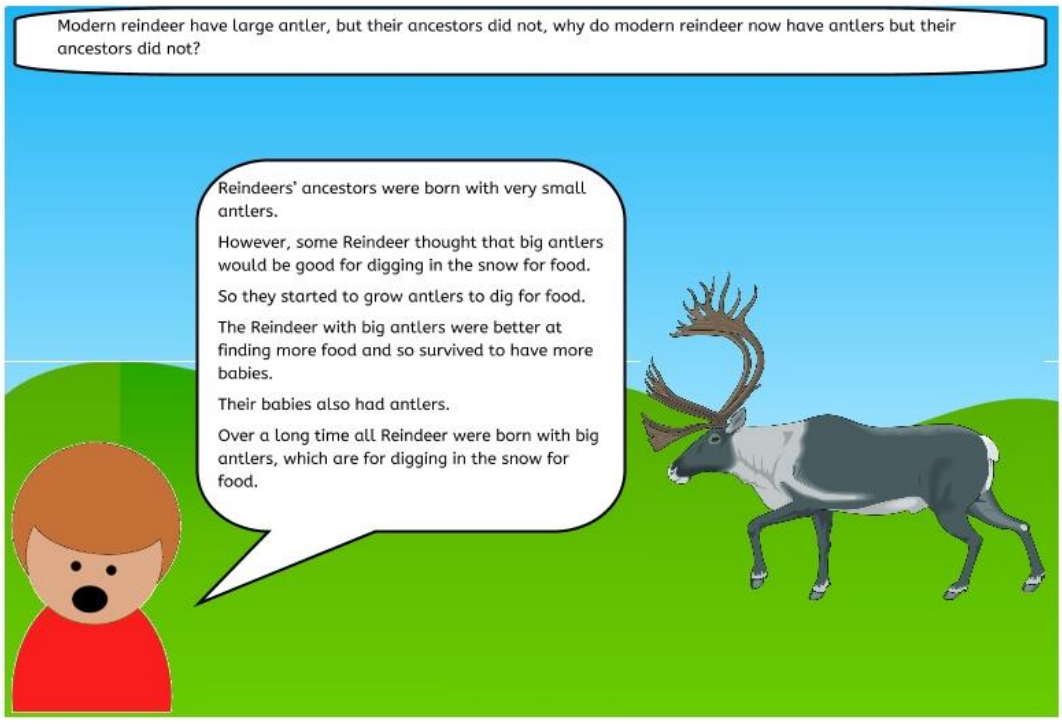

there is wind for blowing the leaves off trees

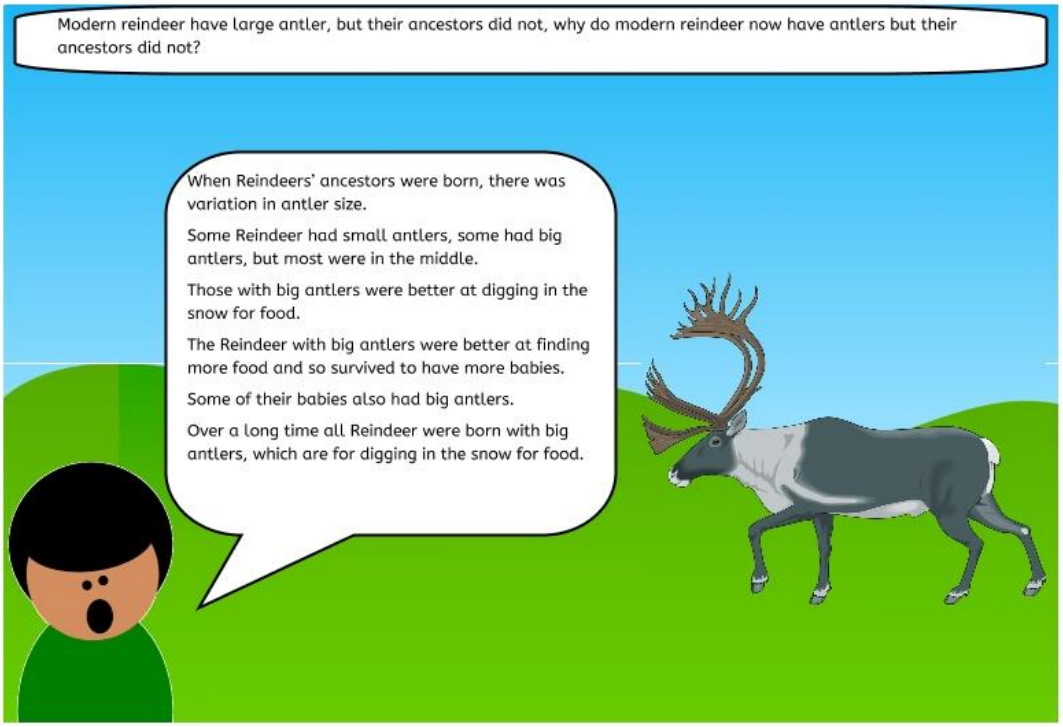

there is wind so children can fly kites

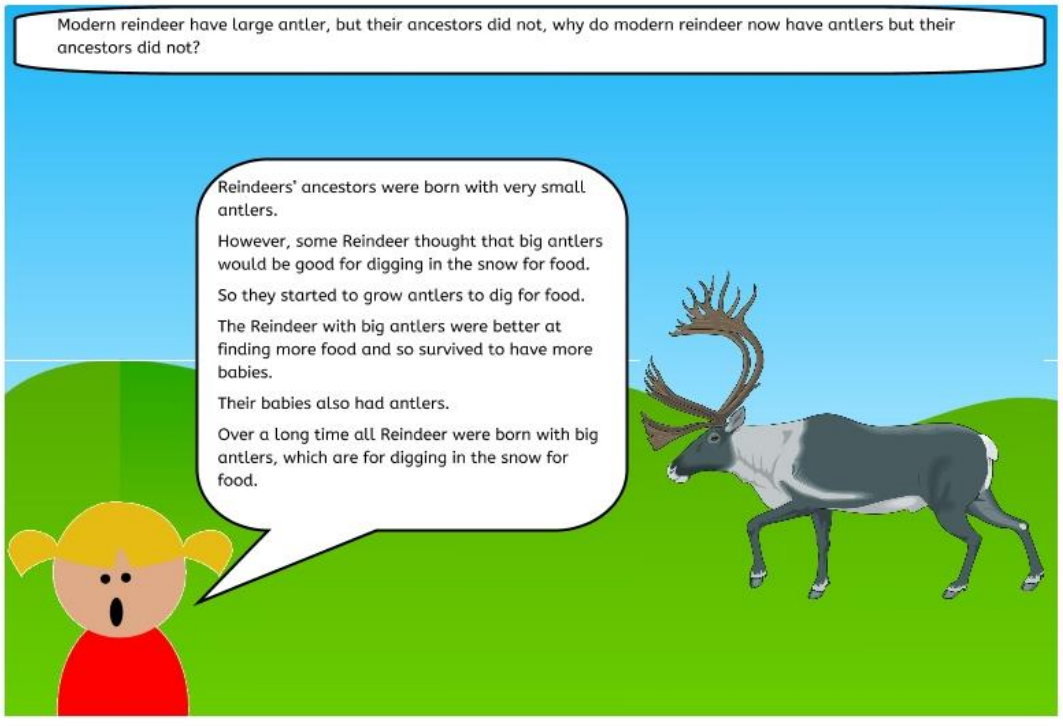

there is wind because the air is moving, in the sky hot air moves up and cold air rushes into the space left

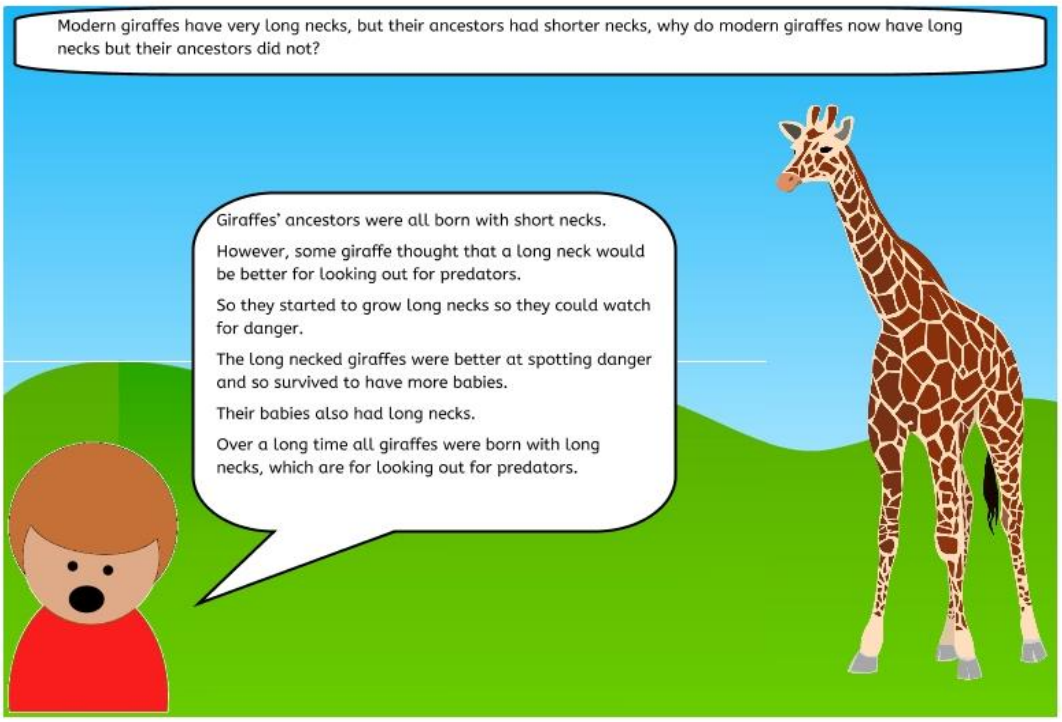
Appendix 4: Concept Cartoons used in Study 3

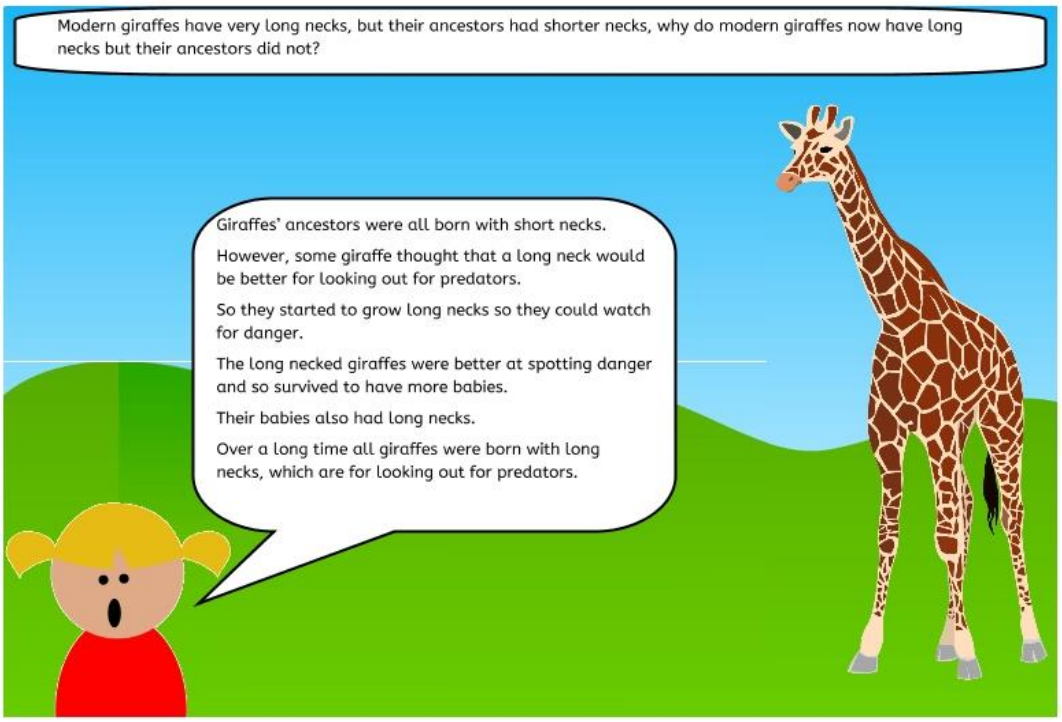


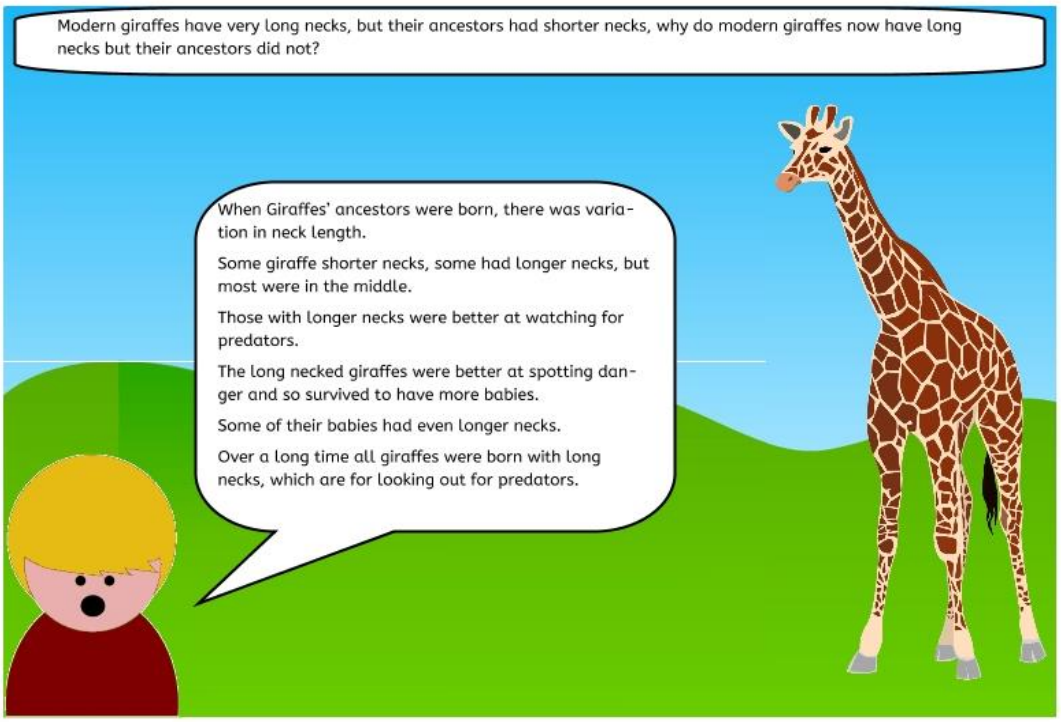


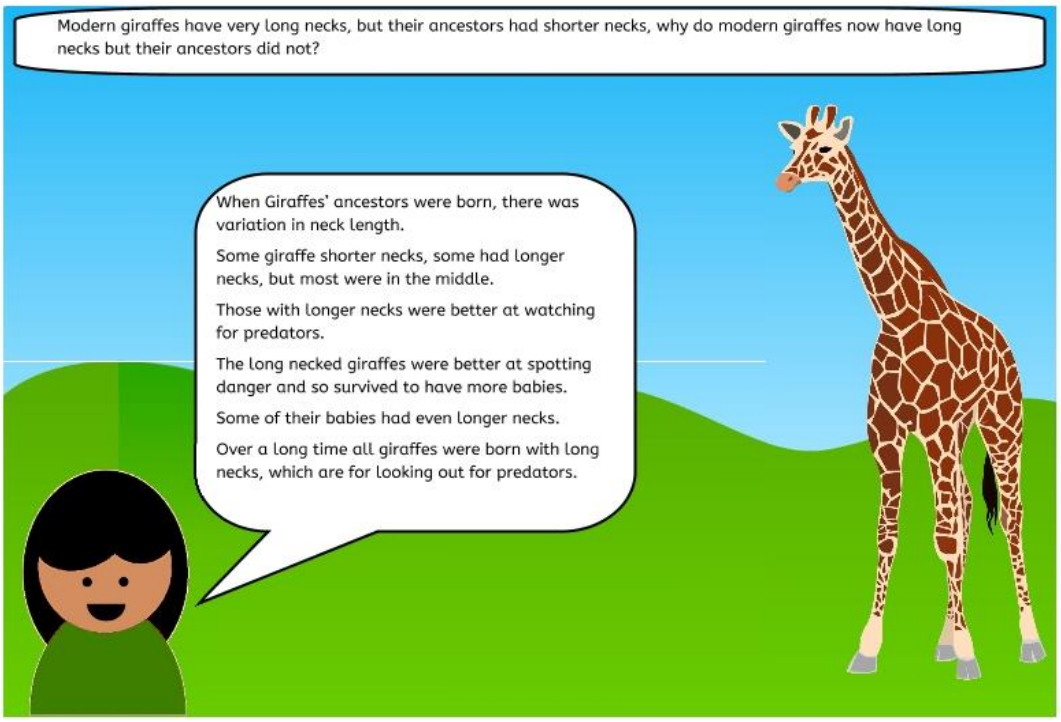

Modern reindeer have large antler, but their ancestors did not, why do modern reindeer now have antlers but their ancestors did not?
When Reindeers' ancestors were born, there was variation in antler size.
Some Reindeer had small antlers, some had big antlers, but most were in the middle.
Those with big antlers were better at digging in the snow for food.
The Reindeer with big antlers were better at finding more food and so survived to have more babies.
Some of their babies also had big antlers.
Over a long time all Reindeer were born with big antlers, which are for digging in the snow for food.


Modern reindeer have large antler, but their ancestors did not, why do modern reindeer now have antlers but their ancestors did not?
Reindeers' ancestors were born with very small antlers.
However, some Reindeer thought that big antlers would be good for digging in the snow for food.
So they started to grow antlers to dig for food.
The Reindeer with big antlers were better at finding more food and so survived to have more babies.
Their babies also had antlers.
Over a long time all Reindeer were born with big antlers, which are for digging in the snow for food.


Modern giraffes have very long necks, but their ancestors had shorter necks, why do modern giraffes now have long necks but their ancestors did not?
Giraffes' ancestors were all born with short necks.
However, some giraffe thought that a long neck would be better for looking out for predators.
So they started to grow long necks so they could watch for danger.
The long necked giraffes were better at spotting danger and so survived to have more babies.
Their babies also had long necks.
Over a long time all giraffes were born with long necks, which are for looking out for predators.


Modern giraffes have very long necks, but their ancestors had shorter necks, why do modern giraffes now have long necks but their ancestors did not?
Giraffes' ancestors were all born with short necks.
However, some giraffe thought that a long neck would be better for looking out for predators.
So they started to grow long necks so they could watch for danger.
The long necked giraffes were better at spotting danger and so survived to have more babies.
Their babies also had long necks.
Over a long time all giraffes were born with long necks, which are for looking out for predators.


Modern giraffes have very long necks, but their ancestors had shorter necks, why do modern giraffes now have long necks but their ancestors did not?
When Giraffes' ancestors were born, there was variation in neck length.
Some giraffe shorter necks, some had longer necks, but most were in the middle.
Those with longer necks were better at watching for predators.
The long necked giraffes were better at spotting danger and so survived to have more babies.
Some of their babies had even longer necks.
Over a long time all giraffes were born with long necks, which are for looking out for predators.


Modern giraffes have very long necks, but their ancestors had shorter necks, why do modern giraffes now have long necks but their ancestors did not?
When Giraffes' ancestors were born, there was variation in neck length.
Some giraffe shorter necks, some had longer necks, but most were in the middle.
Those with longer necks were better at watching for predators.
The long necked giraffes were better at spotting danger and so survived to have more babies.
Some of their babies had even longer necks.
Over a long time all giraffes were born with long necks, which are for looking out for predators.