KEY WORDS FOR LEARNING SCIENCE: THE STRUCTURE OF STUDENTS' SIGN NETWORKS IN MECHANICS

Richard David Taylor

Kellogg College



A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2023

Contents

CONTENTS **1**

ABSTRACT **6**

1 INTRODUCTION **7**

1.1 Context and Aims of the Study **7**

1.2 Thesis Outline **9**

1.2.1 Literature Review 9

1.2.2 Research Questions 15

1.2.3 Sign Network Methodology 16

1.2.4 Sign Network Analysis 18

1.2.5 Interview Methodology 19

1.3 Summary **20**

2 LITERATURE REVIEW **21**

2.1 The Role of Conceptual Change in Learning Science **21**

2.2 Diversity and Lack of Consensus in Conceptual Change Models **22**

2.3 What is a Concept? **25**

2.3.1 The Role of Concepts in Thought 25

2.3.2 Psychological Theories of Concepts 26

2.3.3 Scientific Concepts 32

2.4 What is Conceptual Change? **41**

2.4.1 The Innate Conceptual Repertoire 41

2.4.2 Changes in the Conceptual System 42

2.4.3 The Mechanism of Conceptual Change 43

2.4.4 Summary 47

2.5 Carey's Theory of the Origin of Concepts **48**

2.5.1 Core concepts 48

2.5.2 Conceptual Discontinuity in Conceptual Development 50

2.5.3 Quinian Bootstrapping 56

2.6 Sign Networks 61
2.6.1 Network Science 61
2.6.2 Sign Network Studies 72
2.6.3 Summary of Sign Network Studies 82

2.7 Summary, Contribution to Knowledge and Research Questions 84

3 MEASUREMENT AND CONSTRUCTION OF INDIVIDUAL SIGN (WORD) NETWORKS 86

3.1 Introduction 86

3.2 Selecting and Defining the Scientific Domain of Knowledge 86

3.3 Measurement and construction of students' sign networks 88
3.3.1 School Setting 91
3.3.2 Participants 92
3.3.3 Snowball Sampling Procedure for Measuring Network Vertices (Signs) and Edges (Sign Connections) 112
3.3.4 Student Sign Network Construction 122
3.3.5 Summary 134

4 RESULTS OF THE SIGN (WORD) NETWORK ANALYSIS 137

4.1 Introduction 137

4.2 Physicists' and Psychologists' Understanding of Mechanics and Vocabulary 137
4.2.1 The Force Concept Inventory 138
4.2.2 Vocabulary Tests 146
4.2.3 Differences in the Conceptual Understanding of Mechanics and Vocabulary Knowledge in Physics and Psychology Cohorts 151

4.3 Sign Network Analysis 154
4.3.1 Introduction to the Sign Network Analysis 154
4.3.2 Basic Network Properties 157
4.3.3 Network Vertex Degree Distributions 163
4.3.4 Hub Words in the k_{in} Distributions 185
4.3.5 Summary of Sign Network Analysis 189

5 METHODOLOGY TO INVESTIGATE THE FUNCTION OF HUB-WORDS 191

5.1 Introduction 191
5.1.1 Participants 193
5.1.2 Procedure 193
5.1.3 Materials 200
5.1.4 Interview Transcription 206

5.2 Data Analysis 207
5.2.1 Rationale for Content and Lexicogrammatical Analysis 207
5.2.2 Lexicogrammatical Analysis of Hub-Word Use 209

5.3 Summary 221

6 RESULTS OF THE INTERVIEW ANALYSIS TO INVESTIGATE THE FUNCTION OF HUB-WORDS 222

6.1 Frequency of Hub-Word Use in Interviews 222

6.2 The Function of Hub-Word in Descriptions and Explanations 231
6.2.1 Force 231
6.2.2 Mass 242

6.3 Holistic Analysis of Explanations 254
6.3.1 Comparing Participants' Interview Responses with the FCI scores and Expected Explanations 255
6.3.2 Participants' Descriptions of Motion 256

6.4 Summary 260

7 DISCUSSION 261

7.1 Claim 1: The Physicists and Psychologists have a Different Conceptual Understanding of Mechanics 263

7.2 Claim 2: The Networks have Similar Macrostructure but there are Differences in the Microstructure 265
7.2.1 The Networks are Approximately the Same Size 267
7.2.2 The Networks are Sparse and have a Low Average Degree 268
7.2.3 The Networks have Short Average Path Lengths 268
7.2.4 The Significance of High Clustering Coefficients 268
7.2.5 Comparison of Sign Networks with Concept Networks 270
7.2.6 Summary 271
7.2.7 Implications for Teaching 272

7.3 Claim 3: The Sign Networks Contain Hub-Words 274

7.3.1 In-Degree Distributions 274
7.3.2 Out-Degree Distributions 275
7.3.3 Summary 277
7.3.4 Implications for Teaching 278

7.4 Claim 4: The Sign Networks Have Shared Hub-Words 279
7.4.1 Placeholder Structure Precedes Full Conceptual Understanding 280
7.4.2 Summary 283
7.4.3 Implications for teaching 284
7.4.4 How do Words Become Hubs in a Sign Network? 286

7.5 Claim 5: The Sign Networks Contain Some Different Hub-Words 293
7.5.1 Implications for Teaching 295

7.6 Limitations and Future Work 296
7.6.1 Further Evidence for Quinian Bootstrapping 296
7.6.2 Small Sample Sizes and Using Data Lost Through Attrition 297
7.6.3 Additional Measures of Network structure 297
7.6.4 Communities in the Sign Networks 298
7.6.5 Quinian Bootstrapping in Groups of Student with Different Language Profiles 300

7.7 Summary 301

8 CONCLUSION 302

8.1 Summary of Claims 303
8.1.1 Summary and Contribution to Knowledge 306

8.2 Summary of Implications for Teaching 307

9 REFERENCES 309

10 APPENDICES 321

10.1 CUREC Approval 321
10.1.1 CUREC 1A Application Form 322
10.1.2 Headteacher Consent Letter 332
10.1.3 Participant Consent Form 334
10.1.4 Participant Information Sheet 335

10.2 Merged Word Tables 338

10.3 Compound Word and Phrase Tables 361

10.4 Strict Test Criteria Tables for the Sign Network Degree Distributions 367

10.5 Hub Word Frequency Tables 369

10.6 Agree 'Ideal' Explanations of the Motion of the Objects in the Demonstrations with Force Concept Inventory (FCI) Cross References 372

10.6.1 Film Canister Rocket 372

10.6.2 Falling Cupcake Case 374

10.7 Demonstration coverage of the conceptual dimensions and sub-dimensions identified in the Force Conceptual Inventory (FCI). 375

10.8 Interview Questioning Protocol 380

10.8.1 Falling cupcake case 380

10.8.2 Film Canister Rocket 383

10.9 Additional Tables from the Interview Analysis 386

Abstract

Learning science requires conceptual change. Hence, understanding this process is important to develop successful methods of science teaching. Susan Carey's Theory of the Origin of Concepts (TOOC) is one of the most complete theories of conceptual change. TOOC identifies Quinian bootstrapping as a mechanism that drives conceptual development, with a network of connections between signs (words) forming a placeholder sign structure, which then guides the development of concepts. The nature of placeholder sign structure within scientific domains of knowledge is not known.

This research has examined sign networks in mechanics formed by 6 A-level Physics students learning mechanics and has contrasted these networks with the networks of 5 A-level Psychology students (all aged 16 to 18). Considered word associations and snowball sampling were used to measure individual directed sign networks. The structure of these networks was characterised using basic network parameters and by fitting the degree distributions with different models (e.g., Poisson distribution or power-law) to determine whether the networks contained highly connected hub-words. The Force Conceptual Inventory (FCI) and interviews were used to assess participants' conceptual understanding of mechanics. The interview transcripts were analysed using content and lexicogrammatical analysis.

The FCI scores and the interview analysis showed the physicists and psychologists have a different conceptual understanding of mechanics, but that the macrostructure of all the placeholder sign networks is similar. The networks are sparse, have short average path lengths and low average degree, and high clustering. However, the connections (microstructure) between signs are different in the physics and psychology networks. These structural features suggest conceptual development is guided by a small number of connections (sparseness and low average degree) between neighbouring signs (clusters), which are closely connected to other parts of the network (short average path length).

All the sign networks contain hub-words, with, *force*, *energy*, *time*, and *particle* occurring frequently in all networks, but *mass* is only found in the physics networks. The presence of hub-words suggests the clusters of signs are connected via these hub-words and that the sign structure 'grows', or is organised, around hub-words. Furthermore, the interview analysis of the use of *force* in explanations and the FCI scores suggest this structural feature is present in the psychology networks before these participants have a full conceptual understanding of mechanics (i.e., force is recognised as a cause of motion, rather than acceleration). This is consistent with TOOC account of Quinian bootstrapping driving conceptual development and has implications for the pedagogical practice in physics and beyond.

1 Introduction

1.1 Context and Aims of the Study

An important part of a child's scientific education is learning the ideas and theories that will allow them to understand and engage with our modern scientific and technologically driven society (Osborne & Dillon, 2008). However, children often find these theories notoriously difficult to learn (see Chi, 2005; Treagust & Duit, 2008). Extensive research over many decades suggests this is because children's understanding of the basic concepts (e.g., *force*, *acceleration*), which are used to build these theories, differ from a scientist's understanding of these concepts (e.g., Driver, Guesne, & Tiberghien, 1985). Hence, learning science is not simply a process of acquiring new knowledge using well-understood existing concepts, but a process that requires changes in the way a child defines and categorises the world by changing their understanding of basic concepts (Amin & Levrini, 2017a; Treagust & Duit, 2008). Therefore, evidence based science teaching should be informed by knowledge of conceptual change, and particularly what drives this process (Driver, Squires, Rushworth, & Wood-Robinson, 1994).

The process of conceptual change has been extensively researched (Amin & Levrini, 2017a; Treagust & Duit, 2008) and extensively debated (Amin & Levrini, 2017a; diSessa & Sherin, 1998; Rusanen, 2014; Rusanen & Pöyhönen, 2013; Sherin, 2017b). Despite this, there is no widely accepted consensus on what constitutes conceptual change, what changes during conceptual change, and which mechanisms drive conceptual change (Amin & Levrini, 2017a). The lack of agreed basic assumptions (e.g., what defines a concept), criteria for identifying the phenomenon (e.g., the timescale that conceptual change takes place over), and clear methodological approaches has limited the progress in this field (Amin & Levrini, 2017a; diSessa & Sherin, 1998; Rusanen, 2014; Rusanen & Pöyhönen, 2013). Hence, Amin and Levrini (2017b) and Sherin (2017a) suggest a minimal shared theoretical language needs to be adopted to allow models of conceptual change to be developed to enable more precise model construction, theory building and evaluation.

Sherin (2017a) suggest this shared theoretical language can be based on commonalities in conceptual change models that exist when these models are viewed at high levels of abstraction. At these high levels, most models of concepts can be viewed as systems of conceptual elements (Rusanen & Pöyhönen, 2013; Sherin, 2017a). Hence, at an abstract level conceptual change

becomes a reordering of these systems of elements (Rusanen, 2014; Sherin, 2017a). Therefore, Sherin (2017a) identifies the following shared theoretical language:

1. There is a system of elements described at multiple levels.
2. This system exhibits some variability in behaviour depending on context.
3. Changes in the system are often driven by some sort of external influence.
4. Changes can be slow and incremental.
5. These changes might result in some sort of 'hybrid'.

This language (specifically points 1 and 3) is used to compare three common views of conceptual change: the coordination class view (diSessa & Sherin, 1998), the framework or intuitive theory view (Vosniadou & Skopeliti, 2014), and the ontological shift view (Chi & Slotta, 1993; Chi, Slotta, & De Leeuw, 1994). Conceptual change within the intuitive theory view is explained by Susan Carey's Theory of the Origin of Concepts (TOC) (Carey, 2009). The TOOC emerges from this comparison as the most complete theory of concepts and conceptual change (Amin, 2017).

Language plays a crucial role in driving conceptual change in TOOC through the process of Quinian bootstrapping. This process suggests that a child needs to form a network of connections between words (or signs) to begin to form an understanding of a concept. This network of connections then guides the development of the concept. However, little is known about the structure of the word or sign networks associated with learning scientific concepts. Hence, the aim this research was to examine the structure of the sign networks formed by A-level Physics students learning mechanics and has contrasted these networks with A-level Psychology students, who have a more basic understanding of this fundamental topic in Physics. The thesis outline below sets out in greater detail the rationale for investigating these sign networks within the context of current literature, the methodology used to measure and analyse these networks, and the overall structure of this thesis.

1.2 Thesis Outline

1.2.1 Literature Review

The literature review begins by outlining the debate around what a scientific concept is, and why the lack of an agreed definition or model of a scientific concept has led to issues with finding a mechanism for conceptual change. This debate is introduced by exploring models or views of concepts that are prominent within cognitive psychology. Within this field, a concept is loosely defined a 'unit' of thought (Carey, 2009). This definition can be better understood through the Representational Theory of Mind (RTM) (A. Clark, 2014).

The RTM identifies reasoning as a computational process on sentence-like structures in a mental 'language of thought' (A. Clark, 2014), which is sometimes called mentalese (Pinker, 2009). In mentalese, concepts are the equivalent of words (A. Clark, 2014). Like words, concepts can develop meaning by referring to categories or 'classes of objects in the world' (Murphy, 2002, p. 1). Extensive research in cognitive psychology has addressed what the content of these mental representation of categories might be and how this content is formed (Murphy, 2002). The main theories of concepts in this field are the prototype theory (Hampton, 1979; Medin & Smith, 1984; Rosch, 1975a), exemplar theory (Medin & Schaffer, 1978), and theory-theory or knowledge view (Murphy & Medin, 1985). These theories are discussed to provide a context for the following discussion of scientific concepts.

Murphy (2002) stresses that there is no accepted view that one of these theories is correct or complete. Indeed, he argues that it is likely that a combination of the ideas from these theories will be needed to explain concepts. For instance, he suggests that an integrated theory of concepts is likely be a hybrid between a prototype and a knowledge view. (Keil, 1989) adds some structure to Murphy's hybrid view by suggesting that prototype and knowledge conceptual elements might form concepts in an associative network. This picture of a concept is consistent with a dual factor view of concepts (Block, 1986; Block & Campell, 1987), where a concept contains both broad and narrow content. Narrow content is determined by the intrinsic properties of an object. In contrast, broad content refers to features of an object that depend on how it interacts with its environment or wider knowledge.

There are obvious parallels in views of concepts proposed by Murphy (2002) and Keil (1989) and the theoretical language proposed by Sherin (2017a). Despite these parallels, Rusanen and Pöyhönen (2013) argue that there is often not enough recognition of when researchers are

referring what they call psychological or scientific concepts. Typically, psychological concepts have been developed to explain mental representation of objects. Hence, within mentalese, these concepts are equivalent to nouns. The prototype/knowledge associative network proposed by Keil (1989) has not been explicitly designed to explain concepts, like scientific concepts, which might be more like verbs (i.e., actions which cause events). diSessa and Sherin (1998) point out this distinction, and develop their own theory of scientific concepts, which they call coordination classes, that involves systems of elements called p-prims and knowledge elements such as physical formula (e.g., F=ma). P-prims can be most closely identified with narrow content, whilst knowledge elements are associated with broad content.

diSessa and Sherin (1998) also reject the assumption that concepts are 'unitary mental structures' (p1170) and define coordination classes as more diffuse structures that drawn upon many different parts of knowledge. However, other theories of scientific concepts preserve the idea of concepts as units of thought and connect more directly with the theories that have been developed in cognitive science. For instance, Carey (1985b, 2009) and Vosniadou and Skopeliti (2014) have developed theories of scientific concepts based on the theory-theory view of concepts. In this framework or intuitive theory view, concepts are defined within the context of mental theories. Again, the conceptual elements in the intuitive theory views form a network or a system, as they are part of a mental theory, and contain narrow and broad content (Carey, 2000).

Finally, the ontological view of scientific concepts follows the assumption that concepts are mental representations of categories (Chi & Roscoe, 2002), which is how psychological concepts are defined (Murphy, 2002). However, these categories are extended from just including matter to processes and mental states (Chi et al., 1994). The ontological view is less specific about what the conceptual elements might be, although, Chi (2005) suggest a concepts ontological category is determined by 'plausible attributes' (p161). For instance, it is plausible for an object to have a colour as an attribute but not an event such as a football match. Again, it is likely these attributes have both narrow and broad content.

This thesis does not attempt to resolve the difficulties in defining scientific concepts. The purpose of discussing this debate is to explore differences in proposed models, and in doing so emphasise why the lack of agreement in the nature of scientific concepts has made it difficult to agree a mechanism for conceptual change. However, this discussion also highlights that, despite the lack of an agreed definition, these models fit with a broad definition of 'a system of conceptual elements with some elements with narrow content and some elements with broad

content', which is consistent with the shared theoretical language suggested by Sherin (2017a). This thesis adopts this pragmatic and minimal definition of a concept. s

Once a general definition of a concept is adopted, Carey (2015) argues that examining the process of conceptual acquisition provides a powerful analytical tool for developing and distinguishing between different views of concepts. She identifies three principal questions that any account of conceptual development must address. These are:

1. What is the innate conceptual repertoire?
2. How do any initial conceptual representations differ from a descendent conceptual system?
3. What is (are) the mechanism(s) that achieves the transformation of the initial into the descendent final state?

TOOC, and the intuitive view of concepts it incorporates, provides a more complete answer to questions one and three when compared with the coordination and ontological view of concepts.

The innate conceptual repertoire is the basic stock of conceptual representations that any subsequent concepts are built from (Carey, 2009). These concepts are innate in the sense that they are not the output of any learning process (Spelke, 2000; Spelke & Kinzler, 2007). In TOOC, the innate conceptual repertoire consists of core concepts of *object*, *number*, *agent*, and *cause* (Carey, 2009, 2011). Carey (2009) presents extensive empirical evidence for the existence of these concepts. In the coordination class view, the basic conceptual elements are p-prims. However, diSessa (1993) acknowledges that these concepts elements are learnt and not innate. Moreover, Henderson, Langbeheim, and Chi (2017) points out that p-prims are probably made up from simpler conceptual elements (e.g., core concepts). The innate concepts that are the foundation of any concepts in the ontological view are not specified. Therefore, TOOC provides the most complete account of the fundamental concepts that form any subsequent concepts.

TOOC also provides a more complete account of the mechanism that drives conceptual change. At an abstract level, and within the shared theoretical language suggest by Sherin (2017a), conceptual change becomes a reordering of a system of elements (Rusanen, 2014; Sherin, 2017a). Hence, any mechanism for conceptual change drives this reordering of elements. Rusanen (2014) points out that this mechanism is often assumed to be an internal ('inside the head') mechanism that makes the system 'better' in some way, either in terms of its

ability to solve problems (Posner, Strike, Hewson, & Gertzog, 1982; Strike & Posner, 1992), utility (Ohlsson, 2009a, 2009b) or coherence (diSessa, 1993; diSessa, Gillespie, & Esterly, 2004; Thagard, 2000). However, Rusanen (2014) also highlights problem solving ability, utility or coherence are normative states for the system, which specify end goals for any system, rather than specific mechanisms to achieve these goals. Both the coordination class and ontological view of conceptual changes fail to define a precise mechanism to reorder any system of conceptual elements to achieve the normative goals identified by Rusanen (2014).

In contrast, TOOC identifies language (and more generally symbol systems) as a mechanism that organises any learning processes that are required to a make a conceptual system 'better', and hence addresses the concerns raised by Rusanen (2014). However, in identifying language as playing a crucial role in conceptual change, TOOC also addresses a fundamental problem with any model of conceptual development that relies on comparison between the efficacy of thinking, using old and new conceptual understanding. This problem was originally articulated by Fodor (1975) and concerns hypothesis testing as a mechanism for conceptual change. As an advocate of the representational theory of mind, Fodor's argument assumes that thoughts are expressed by a syntactic arrangement of concepts. Therefore, an individual cannot express a thought (e.g., an explanation of a phenomenon or solution to a problem) if they do not possess the concepts that make up that thought. Hence, learning a new concept cannot involve a comparison between thoughts that use the old concept and a thought that uses the new concept because the new concept has not yet been learnt!

TOOC resolves Fodor's paradox by arguing that conceptual change is driven by a learning process called Quinian Bootstrapping, rather than hypothesis testing. The Quinian bootstrapping process begins when an individual connects a mental representation with a newly coined or learned external symbol or sign, usually a word. Initially, the mental representation associated with the new sign is far from a fully developed concept. In this sense, the sign is initially a 'placeholder', whose meaning is defined by its set of connections or interrelations between other signs. This set of interrelations in the placeholder, or sign, structure guide modelling activities (e.g., analogy monitoring and construction, thought experiments and inductive inference) that allow an individual to develop a full understanding of the concept the placeholder sign represents. Hence, conceptual development begins, and is guided by, the connections between signs (usually words). This is the key theoretical idea that runs throughout this thesis.

Carey (2009) provides a detailed account of how Quinian bootstrapping leads to the development of the conceptual representation natural number (i.e., one, two, three, four, etc… or n+1). This account is based upon Carey's extensive empirical knowledge of the core concept of number, but also her knowledge of the placeholder sign structure that guides this development (i.e., a child's initial ability to recite '*one*, *two*, *three*, *four*, etc…' before they understand the meaning of this sequence). Carey uses this knowledge of the placeholder sign structure to define stages in the development of the concept natural number and to marshal empirical evidence to support these stages. Hence, knowledge of placeholder sign structure is crucial in developing this precise and detailed account of the development of this concept.

Carey also provides an account of how Quinian bootstrapping might lead to the development of concepts with intuitive or scientific theories. However, her account of role of Quinian bootstrapping in these examples is less detailed and is based upon circumstantial, rather than direct evidence. For instance, the studies by Smith (2007) and Wiser and Amin (2007) that have analysed the efficacy of teaching approaches that are based on Quinian bootstrapping are cited by Carey as evidence for her theory. However, these studies do not directly measure the sign structure for individual or groups as they attempt to learn scientific concepts. Hence, the detailed knowledge of the placeholder sign structure that guides the development of concepts within intuitive theories is not known. The previous paragraph argues that Carey (2009) uses this knowledge of placeholder sign structure to develop her detailed account of the acquisition of the concept of natural number. Therefore, knowledge of the placeholder sign structure that forms when individual learn intuitive theories could allow more detailed accounts of how Quinian bootstrapping might drive the development of concepts within these theories.

There are some empirical studies that do measure what could be called a sign structure (e.g., Cachapuz & Maskill, 1989; Slotte & Lonka, 1999), which are captured using word associations. The findings of these studies support the general idea that students develop a sign structure when learning scientific concepts. However, these studies do not use the analysis tools that have been developed as part of modern network science (Barabási & Pósfai, 2016; Newman, 2018) to analyse sign structure rigorously and systematically. Hence, they provide a largely qualitative insight into the role any sign structure might have in learning concepts.

In contrast, empirical studies by Koponen and Pehkonen (2009) and Koponen and Nousiainen (2018b) have used analytical techniques informed by network science and graph theory to quantitatively investigate the structure and growth of concept networks. These studies have demonstrated the advantages of using quantitative measures to systematically compare the

networks of novices and experts and identify differences in their structure. Moreover, they have shown how knowledge of measures of network structure (e.g., high clustering coefficients) can then be used to develop and constrain models that explain the growth of these networks. However, these concept networks are constructed using constraints on types of vertices that are allowed and how edges between these vertices can be formed. In these networks, vertices can be concepts or quantities and the edges that connect these vertices are either experimental procedures (e.g., operational definition of concept or laws or demonstrations) or modelling procedures (e.g., simple definitions or logical deductions). By applying these constraints, the concept networks represent scientifically acceptable knowledge, rather than subjective beliefs.

However, TOOC specifically argues that a sign is initially associated with placeholder values before these placeholder values develop into fully formed concepts. Hence, a sign network is developed before than the fully formed concept network it will ultimately be associated with. Furthermore, this initial sign structure guides the development of the procedural rules that are the hallmark of a fully developed concept network. Hence, the TOOC predicts that a learner will recognise that there is a connection between signs before they can fully explain why they are connected. Therefore, less stringent criteria (i.e., word associations) for forming edges are more appropriate to investigate sign, rather than concept, networks. In this thesis, the strength of the criteria for defining vertices and edges guides whether a network is more likely to represent sign or underlying concepts (i.e., whether the network is a sign or concept network).

There are several studies that have applied less stringent criteria, in the form of free word associations, to create both aggregate (De Deyne & Storms, 2008a; Steyvers & Tenenbaum, 2005) and individual sign networks (Morais, Olsson, & Schooler, 2013). These studies have also used techniques from network science and graph theory to analyse the networks they create. There is some evidence that these sign networks have a scale-free structure (Steyvers & Tenenbaum, 2005). This type of structure is characterised by most signs (or words) only connected to a limited number of other signs (i.e., one or two). However, a few signs are very highly connected, accounting for most of the connection in the network (Barabási & Pósfai, 2016; Newman, 2018). These very highly connected vertices – or hubs – are referred to as hub-words. Steyvers and Tenenbaum (2005) develop a model of network growth based on preferential attachment (Barabási & Albert, 1999; Dorogovtsev & Mendes, 2002) that explains the growth and structure of these sign networks they measure, and the presence of hub-words in these networks. This model incorporates Carey's ideas about conceptual development through the development of sign structure. But furthermore, the Steyvers and Tenenbaum

(2005) model implies a sign network 'grows' around highly connected hub-words, which they argue are likely to be associated with key concepts within a domain of knowledge. Hence, it is likely that individuals with different levels of conceptual understanding in a domain of knowledge would have different hub-words.

Describing the nature of any placeholder sign structure that is involved in the development of concepts within intuitive theories is the focus of this thesis. Hopefully, the more detailed knowledge of this placeholder sign structure will enable a more precise detailed account of the Quinian bootstrapping process to be developed. Moreover, this knowledge might highlight aspect of the process that are supported by empirical evidence. Hence, this thesis broadly aims to fulfil the criteria defined by Amin and Levrini (2017b) of developing more a more detailed account of an aspects of conceptual change and that can be subjected to empirical evaluation.

This aim is achieved by investigating the nature of any similarities and differences in the placeholder sign structure of individuals with difference levels of expertise in mechanics. Furthermore, the possibility that sign networks grow around hub-words, and hence novices and experts have different hub-words, is a key idea that is tested when the similarities and differences between the sign network for different cohorts are investigated. Furthermore, this study also tests the assumption that any hub-words are associated with key concepts by analysis how these words are used when individuals give explanations of phenomena in a domain of knowledge.

1.2.2 Research Questions

The aims of this study are formalised in these research questions.

1. What are the sign networks of individuals within a scientific domain of knowledge?
2. What are the similarities and differences between the networks for individuals who have different levels of scientific understanding (novices and experts)?
    a. Are hub-words responsible for most connections in the networks?
    b. If so, are the hub-words the same for different individuals?
3. Do these individuals use hub-words from their networks when they reason about a scientific experiment or demonstration? If so, how?

These research questions were addressed using a research design that was split into two phases. During the first phase, sign networks were mapped and analysed to answer research questions 1 and 2. This phase of the methodology is described in Chapter 3. During the second phase,

interviews with participants explored how individuals used any hub-words in their semantic networks when reasoning about scientific demonstrations or experiments, addressing research question 3. This phase of the methodology is described in Chapter 5. The two phases of the methodology are outlined in greater more detail in the following sections.

1.2.3 Sign Network Methodology

The individual sign networks were made using a snowball sampling method, which has been adapted from an original method developed by Morais et al. (2013). The Morais et al. (2013) method uses free word association to measure sign networks for entire lexicons. Each participant in the Morais et al. (2013) study was initially presented with a set of 'seed' or stimulus words and asked to list words they associated with these words. The response words then became the stimulus words in the next iteration of word association collection. Hence, a participant's sign network was sampled through the initial set of seed word 'snowballing' to a much larger sample of the words in that participant's vocabulary.

It is important to stress that the Morais et al. (2013) method samples an individual sign network, rather than capturing the complete network (Newman, 2018). However, Morais et al. (2013) use simulations to demonstrate that their sampled networks accurately measure network parameters if a certain number of word-association iterations are exceeded. Unfortunately, the timeframes needed to collect this number of iterations were impractical for a study that uses school age participants and takes place in a school environment. The sign network chapter explains how this method was adapted and why it was deemed to accurately capture individual sign networks for the participants in this study. The results of a pilot study are used to justify this methodology.

This methodology also adopts a slightly different procedure to collect word associations from the Morais et al. (2013) study. In the Morais et al. (2013) study, participants are encouraged to list any word that comes to mind in response to a stimulus word in a one minute timeframe. In contrast, this study asks participants to list response words that 'help them understand' a stimulus word in an unlimited timeframe. This change from free to considered responses is designed to encourage students to list signs (response words) that may be an important part of the placeholder structure around a given sign (stimulus word) but without the stringent conditions for linking words require in concept maps (Koponen & Pehkonen, 2009) or the very loose conditions defined by free word associations (Morais et al., 2013).

Directed networks for each participant were constructed using the measured word associations. An edge is draw between signs (vertex) if they are word-associations. The direction of the edge shows whether a sign is a response or a stimulus word, with edges pointing away from stimulus words. There are two issues arising with the construction of the directed network that are addressed:

1. Finishing of the network.

   In the snowball sampling method, the response words in an iteration become the stimulus words in next iteration. However, this process cannot carry on indefinitely due to time constraints. Hence, there is a last iteration of response words that do not become stimulus words. When a directed network is constructed, these words will have no edges leaving them (i.e., stimulus to response). The procedure for how this issue is dealt with, and the possible implications for the structure of the networks, are outlined.

2. Possible sign merges, and compound words and phrases

   a. *Word merges*

      Participants often respond to stimulus words using different grammatical forms of words. For instance, force-pull or push-pulled. When constructing the network, it was decided to merge these words. The rationale for this decision is outlined.

   b. *Compound words*

      Participants were explicitly asked not to use compound words or phrases as a response to stimulus words. Despite this, these types of response were regularly occurred. Hence, each of these responses required a decision about whether to split up the compound word or phrase into multiple vertices or preserve the response as a single vertex. The procedure to ensure these decisions were made systematically and were underpinned by an appropriate rationale is described.

In addition to the capture of the word associations and constructions of the networks, this methodology outlines the selection and recruitment of participants, the school setting, and the rationale behind choosing mechanics as specific domain of knowledge. Ethical issues are also anticipated and addressed.

1.2.4 Sign Network Analysis

The sign network analysis addresses research questions one (*What are the sign networks of individuals within a scientific domain of knowledge?)* and two (*What are the similarities and differences between the networks for individuals who have different levels of scientific understanding (novices and experts)?),* which is further broken down into two sections, which are addressed by different parts of the analysis.

The first part of the analysis identifies similarities and differences in the physics and psychology cohorts' conceptual understanding of mechanics and vocabulary. Participants' conceptual understanding of mechanics is assessed using the force concept inventory (FCI) (Hestenes, Wells, & Swackhamer, 1992). The design of the FCI is discussed and it is stressed that the FCI is a valid and reliable test of whether participants can solve mechanics problems in a way that is consistent with an 'expert' physicists (Savinainen & Scott, 2002). There is some debate about whether the FCI is a valid measure of the conceptual content individuals use to arrive at these solutions (Eaton & Willoughby, 2018; Huffman & Heller, 1995). This debate highlights the importance of using interviews to further investigate how the participants build their explanations in mechanics.

The second part of the analysis identifies similarities and differences in the sign networks for the physics and psychology cohorts. Again, this analysis was carried out in several phases. The first phase used standard measures (e.g., average shortest path length, clustering coefficient) to characterise the basic structure of the network (Newman, 2018).

The second phase of the analysis examined the in and out degree distributions for the networks. This analysis applied the strict test developed by Broido and Clauset (2019) to establish whether these network have a scale-free structure, which would be consistent with previous analysis of sign networks and consistent with the Steyvers and Tenenbaum (2005) model, which explain sign network growth through preferential attachment to hub-words. The definition of a scale-free structure adopted by Broido and Clauset (2019) has been questioned by Voitalov, van der Hoorn, van der Hofstad, and Krioukov (2019). However, the strict test still defines an accessible and robust statistical procedure for determined whether the degree distributions of networks are heavy-tailed, and hence contain hub-words (Newman, 2018). Arguably, the presence of hub-words in the placeholder sign networks is a more interesting finding that whether the networks have a scale-free structure because this suggests the networks could be growing around particular signs (words).

The final phase of this analysis investigates any hub-words that are present in the networks. It aims to establish whether any hub-words are shared by both the physicists and psychologists by looking at how frequently hub-words occur in these two different groups of participants.

1.2.5 Interview Methodology

The second phase of this research aims to answer research question 3, which is concerned with how individuals use the words, and especially any hub-words, when they are asked to give scientific explanations of how objects move. To address research question 3, semi-structured interviews were used (Cohen, Manion, & Morrison, 2018; diSessa, 2007; Ginsburg, 1997). During an interview, a participant was shown a demonstration. For example, a cup-cake case falling from 1m, first on its own and then with paperclips added to it. He or she were then asked to describe and then explain what happened in the demonstration. The tasks and the questions associated with each task were carefully chosen to align the conceptual dimensions identified in the FCI. Each of the interviews were audio-recorded and then transcribed, to investigate how any hub-words are used in each participant's descriptions and explanations. This investigation involved a content analysis followed by a lexicogrammatical (LG) analysis using systemic functional linguistics (SFL) (Halliday & Matthiessen, 2004). This methodology is similar to previous methodologies that has also been used to examine how individuals explain physical phenomena (e.g., Seah, Clarke, & Hart, 2011).

The methodology for this phase of the research was again tested in a pilot-study. The results of the pilot study established the efficacy of interviews as a method of data collection. Moreover, these results demonstrate the ability of content and then LG analysis to identify the use of hub-words in participants' explanation and descriptions. The initial content analysis identified when a participant is giving a description or an explanation. The LG analysis provided a robust method for identifying the function of each participant's hub-words.

1.3 Summary

The central aim of this thesis is to examine the placeholder sign structure that forms when individual learn intuitive theories. The methodology outlined in the previous sections outlines how this thesis intends to achieve this aim. Primarily, the sign network analysis addresses research questions 1 and 2, which are concerned with the similarities and differences in the placeholder sign networks for the physics and psychology cohorts. The interview analysis addresses research question 3, which asks if and how hub-words are used in participants' explanations of how objects' move. Hence, the interview analysis probes participants conceptual understanding of mechanics. However, the design of the second phase of the analysis is influenced by the results of the sign network analysis. Hence, the results of this analysis are presented in Chapter 4, which precedes Chapter 5 that describes the interview analysis. Chapter 6 then presents the results of this interview analysis.

The findings from both analyses are brought together in the discussion (Chapter 7). The discussion demonstrates how knowledge of the placeholder sign structure that forms when individual learn intuitive theories can be used to develop a more detailed and precise account of the Quinian bootstrapping process that, within the TOOC, drives conceptual change. Hence, this thesis contributes to the more precise model construction, theory building, and empirical evaluation approach to conceptual change advocated for by Amin and Levrini (2017a).

2 Literature Review

2.1 The Role of Conceptual Change in Learning Science

Children are curious about the natural world. They ask a wide variety of scientific questions (Baram-Tsabari, Sethi, Bry, & Yarden, 2006) and seek answers to these questions from a number of different sources, including parents (Dabney, Tai, & Scott, 2016; Perez-Granados & Callanan, 1997) and television programmes (Swirski, Baram-Tsabari, & Yarden, 2018). They also survive and think in a world that requires some knowledge of natural phenomena. For instance, children have to make decisions about whether objects are alive or not (Carey, 1985b), and are therefore going to behave in particular ways, or about how an object might move (diSessa, 1993). As the above examples suggest, it is likely children develop their understanding of these phenomena from their own experience of the natural world and everyday culture.

So, children arrive in science classrooms, not as blank slates or empty vessels, but with a wealth of ideas about natural phenomena (Driver et al., 1985; Driver et al., 1994). Unfortunately, these ideas are often at odds with accepted scientific explanations (theories). Hence, an important aim of science education is to reconcile children's everyday explanations with scientific theories (e.g., Driver et al. (1994); Department of Education (2015)), but this is often not achieved. Children's everyday explanations are robust and longlisting despite repeated teaching of the accepted scientific theories (see Chi, 2005; Treagust & Duit, 2008).

It is generally accepted that the challenges children face when learning scientific explanations stems from the fact that children's understanding of the fundamental concepts that underpin these explanation are different (Rusanen & Pöyhönen, 2013). Hence, learning science is not simply a process of acquiring new knowledge, but a process that requires changes in individuals', '*ontological commitments, inferential relationships and standards of explanations*' (Rusanen, 2014, p. 1414). In other words, children must change their understanding of existing concepts or learn new concepts to learn scientific theories; this learning process is frequently called 'conceptual change'.

2.2 Diversity and Lack of Consensus in Conceptual Change Models

Learning science requires conceptual change. Hence, a major focus for science education research is trying to understand the process of conceptual change and the factors that might affect or drive this process (see Amin and Levrini (2017a), Treagust and Duit (2008), and Duit and Treagust (2003) for summaries). However, conceptual change is also a focus of research in cognitive psychology (see Chapter 10 in Murphy (2002)) and philosophy of science (Kuhn, 1996). Many researchers have fused ideas from these disciplines to create explanations of conceptual changes (Rusanen & Pöyhönen, 2013). For instance, a seminal account of how children change their understanding of scientific concepts by Strike and Posner (1992) was strongly influenced by Thomas Kuhn's historical account of the development of scientific theories (Kuhn, 1996). However, the diversity of theoretical perspectives and phenomena of interest in these research fields has led to a lack of consensus about the basic assumptions of conceptual change research and appropriate methodologies (Rusanen, 2014; Rusanen & Pöyhönen, 2013; Sherin, 2017b). For instance, research by Carey (1985b) focuses on how children develop biological concepts (e.g. living thing) over a number of years (ages 4 to 10). In contrast, Levrini and DiSessa (2008) examine how students (age 18 to 19) develop their understanding of the concept of proper-time over the course of a single classroom episode. Furthermore, these two examples also illustrate what diSessa and Sherin (1998) identify as a lack of stated and agreed upon definition of what a concept is in many accounts of conceptual change. 'Living thing' can be identified as a subgroup of 'objects', which is consistent with the common psychological definition of a concept as a '*mental representation of a category (class of objects in the world)*' (Murphy, 2002). However, it does not seem plausible that 'proper-time' is a class of object (diSessa & Sherin, 1998). Rather, proper-time is more consistent with what Rusanen and Pöyhönen (2013) call a scientific concept.

Whilst it is not the case in the research by Carey (1985b) and Levrini and DiSessa (2008), Sherin (2017a) argues that the differences in underlying assumptions, target phenomena or timescales in conceptual change models are often not explicitly acknowledged and discussed, or are simply not treated as differences. Therefore, comparing conceptual change models is difficult because it is not clear whether the underlying assumptions in different models are coherent and whether they are trying to explain the same learning phenomena (Sherin, 2017a). Hence, Rusanen (2014) argue it might be more appropriate to think of conceptual change as an umbrella term that refers to many different models or accounts of conceptual change. Moreover, there is so

much diversity in what proposed models of conceptual change try to explain, she suggests a 'grand theory' of conceptual change might not be possible. In contrast, an account of conceptual change may be a composite requiring many different explanations for different phenomena.

Sherin (2017a) acknowledges the diversity in conceptual change models and the arguments for splitting this field of research. However, he suggests that communalities in conceptual change models can be identified if these models are viewed at a high level of abstraction. Importantly, identifying these communalities allows some shared theoretical language to be developed for comparison purposes. This shared theoretical language has allowed some level of consensus about what needs to happen during conceptual change and, therefore, what any mechanism for conceptual change needs to explain. The shared theoretical language identified by Sherin (2017a, p. 74) that applies to most conceptual change models is listed below:

1. There is a system of elements described at multiple levels.
2. This system exhibits some variability in behaviour depending on context.
3. Changes in the system are often driven by some sort of external influence.
4. Changes can be slow and incremental.
5. These changes might result in some sort of 'hybrid'.

When comparing models of conceptual change, Rusanen and Pöyhönen (2013) also emphasise that it is important to recognise whether a model explains changes in what they call psychological or scientific concepts. They broadly identify psychological concepts as concepts that are the focus of research in cognitive psychology and scientific concepts as those that are the focus of study in science education or the philosophy of science. The next sections in this literature review use the commonalities in conceptual change models identified by Sherin (2017a) and his common theoretical language, as well as the distinction between psychological and scientific concepts, to explore the key theoretical ideas in this thesis.

Section 2.3 discusses models of both psychological and scientific concepts to explore current thinking about what a concept is. Section 2.4 builds on this discussion to examine what conceptual change is. Section 2.5 presents Susan Carey's Theory of the Origin of Concepts (TOOC) (Carey, 2009) as the most complete theoretical framework for understanding conceptual change, and specifically Quinian bootstrapping, as a mechanism that drives conceptual development. Section 2.6 discusses empirical studies of sign networks, which

provide further insights into this Quinian bootstrapping process. The contribution this thesis makes to knowledge is identified through the discussion of these network studies.

2.3 What is a Concept?

Concepts play a fundamental role in theories of the mind. They have been referred to as *'units'* (Carey, 2009) of thought and *'metal glue'* (Murphy, 2002). As basic components of our thought, concepts are involved in '*identifying objects and events as being in a certain category, drawing inferences about novel entities, and communicating about them.'* (Murphy, 2002, p. 3) The goal of the psychology of concepts is to understand how concepts achieve these functions. To do this, the role concepts play in thought needs to be specified, as described in the next section.

2.3.1 The Role of Concepts in Thought

What does it mean for a concept to be a 'unit of thought' (Carey, 2009) or 'mental glue' (Murphy, 2002)?. This section aims to address this question by outlining the role of concepts within a more general theory of cognition. This theory is the computational theory of mind (CTM) (Braisby & Gellatly, 2012; A. Clark, 2014).

The CTM is the closest thing psychology has to a general theory because it has been extremely successful in explaining how a brain, composed of nothing but ordinary matter, can be intelligent (Braisby & Gellatly, 2012; A. Clark, 2014; Pinker, 2009). Or as Pinker (2009) crudely, but effectively, phrases it; how can meat think? The CTM frames cognition (functions that the brain performs) as an information processing problem (A. Clark, 2014; Pinker, 2009). Advances in computer science have demonstrated that information processing problems can be effectively solved by algorithms operating with formal logic on universal Turing machines (computers) (A. Clark, 2014). It is developments in these three areas that have enabled cognitive scientists to develop a convincing, if not thoroughly developed, theory of how a mind, made of ordinary matter, rather than a spiritual substance, can think intelligently. Put more simply, cognitive scientists have realised that complex thought can be explained by repeated processes (algorithms) that use symbols and operations between symbols (formal logic) on physical computers (Turing machines), which are made of either silicon chips or neurons (A. Clark, 2014).

In a silicon computer, the algorithms and logical operations that define how information is processes, or how a computer thinks, are encoding using programme languages such as C (Dasgupta, 2016). The programme language of the mind is a mental language called mentalese (Pinker, 2009). Like programming and natural languages, mentalese consists of symbols, that represent meaning, and syntax that define the operations that combine these symbols into

statements or expression that express more ideas or thoughts. This version of the CTM is known as the language of thought hypothesis (LOT) (Cain, 2002). The LOT hypothesis is not the only version of the CTM. For instance, the CTM can also be implemented through connectionism, parallel distributed processing, and neutral networks (A. Clark, 2014). However, the LOT is the most intuitive model for understanding how concepts function in thought.

Within the LOT, concepts are the mental symbols, or the mental words (Rescorla, 2019), which are roughly equivalent to a lexical word (Margolis & Laurence, 2022). Hence, in the mentalese statement, *force causes acceleration*, force, acceleration, and cause are all concepts. Therefore, concepts are the primitives or 'units' of thought that used to express more complex notions. The general theories of concepts that are discussed in the next section have focused on how these mental words are represented or structured in the mind.

2.3.2 Psychological Theories of Concepts

The majority of research on concepts in cognitive psychology has focused on mental representation of categories (classes of object) (Murphy, 2002). The following discussion outlines the empirical evidence that has been collected in this research and the main theories that have been proposed to explain this evidence. In doing so, this section explores the general principles that any theories of concepts are likely to need to include.

2.3.2.1 The Classical View of Concepts

The classical view predates cognitive psychology and draws upon many philosophical arguments. In this view, an object is a member of a category if it satisfies necessary and sufficient criteria associated with that category (Margolis & Laurence, 2022; Murphy, 2002). For instance, in the ubiquitous example of a bachelor, a person is a member of this category if they meet the criteria of being male and unmarried. However, the classical view has been widely discredited using theoretical arguments and robust empirical evidence. For example, Wittgenstein famously used the example of a game to demonstrate that finding necessary and sufficient criteria for most categories of object is notoriously difficult (Grayling, 2001; McGinn, 2013). Empirically, several studies have demonstrated effects associated with how individuals approach categorisation that the classical view finds very difficult to explain. These effects are category fuzziness (Hampton, 1979; McCloskey & Glucksberg, 1978) and typicality (Rips, Shoben, & Smith, 1973; Rosch, 1975b; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976).

Category fuzziness captures the idea that there are examples of objects that do not fit neatly into categories with necessary and sufficient criteria. For example, would an unmarried man who has lived with his partner, and their three children, in a stable relationship for twenty years be classified as a bachelor? Or is an olive a fruit? A series of studies by Hampton (1979) demonstrated that these examples are the norm, rather than the exception, with most of his participants very rarely rating examples of items as clear members of a category. In a similar study, found McCloskey and Glucksberg (1978) that participants frequently changed their mind about category membership.

Typicality effects were first discovered by Rosch (1975b) and demonstrate that for most categories, there are better (typical members) and worst (atypical members) of categories. For instance, a cod or trout is a good example of a fish, but a seahorse is not. Like category fuzziness, these typicality effects cannot be easily explained by the classical view (Rosch, 1975b; Rosch et al., 1976).

2.3.2.2 The Prototype and Exemplar Views

Prototype and exemplar views are a response to the failure of the classical view to explain category fuzziness and typicality effects (Murphy, 2002). In the prototype view, a concept is a mental representation that summarises a category using a list of features that members of that category usually have (Hampton, 1979; Rosch, 1975b). However, each of the features are weighted to indicate their importance in determining category membership. For example, the features 'has feathers' or 'can fly' would have high weights for the category, *bird*. However, the colours, brown, red, grey, might have lower weights.

A decision on whether an object is a member of a category is made by summing the weight of the features present in the object and then subtracting the weight of any features that are absent. If this number is above a certain threshold, called the categorisation criterion, then the object is a member of the category. Hence, similarity, as measured by the comparison of cumulative value from the feature weight calculation and the categorisation criterion, is used to judge category membership, rather than the presence of definition features, as is the case in classical view. This allows the prototype view to explain category typicality and fuzziness effects. Typicality effects are explained because the 'best' typical category members have more features and, therefore, a high weight comparison score. However, 'worst' atypical category members are not excluded from a category because, although their weight comparison score is lower, it is still above the categorisation criterion. Usually, items will have weight comparison scores

that are close to the category criterion value, rather much higher than it. Moreover, the same item might have weight comparison scores that are close to the criterion value for more than one category. These aspects of the prototype view explain category fuzziness.

The exemplar view rejects the assumption made by the prototype view that a concept is some form of summary or average representation of all members of a category (Murphy, 2002). In contrast, a concept in the exemplar view is the collection, or set, of all an individual's memories of objects that are members of a category (Medin & Schaffer, 1978). For example, an individual's concept of dog is all the dogs that that person remembers. An individual makes a categorisation decision by matching a new object to the best fitting object stored in their memory. If the best fit memory is in their collection of dog memories, then the new object is a dog. Despite the differences in how concepts are represented, like the prototype view, the exemplar view determines category memberships by similarity, rather than the presence of defining features, as is the case with the classical view. Hence, the exemplar view can also explain category fuzziness and typicality effects.

2.3.2.3 Summary of the Prototype and Exemplar Views

The studies that have informed the development of prototype and exemplar views usually involve categories that have been deliberated selected to divorce any categorisation decision from outside knowledge (Murphy, 2002). For instance, participants have been presented with '*geometric shapes, alphanumeric strings, patches of colour, dot patterns, and schematic faces*' (Murphy, 2002, p. 141) to categorise. Murphy (2002) states that there are two main reasons for this. First, researchers feel it is important to present concepts that are unknown to participants. Second, by using simple stimuli, the studies using these categories will be able to investigate general learning principles that apply across different domains of knowledge. Hence, prototype and exemplar views of concepts explain mental representations of categories that can be formed based solely on experience (i.e., interaction with stimuli from the world, rather than explicitly taught knowledge). However, it is very likely that knowledge does play a role in the structure of psychological concepts. This role is discussed in the next section.

2.3.2.4 The Knowledge View of Concepts

The prototype and exemplar views were developed as a response to the failure of the classical view. Similarly, the knowledge or theory view was developed to account for the limitations of the prototype and exemplar views (Murphy, 2002). Murphy and Medin (1985) use the idea of conceptual coherence to demonstrate these limitations. They define conceptual coherence as a

collection of objects that form a useful group. For example, it is useful to form a category of chair because, as Murphy (2002) points out, when an individual encounters a new object that is a member of this group, they do not have to study it from scratch. They know whether this new example of a chair is dead or alive, what its function is, whether it will hurt them or help them.

Murphy and Medin (1985) address the question of what makes a category coherent (in other words useful). An obvious and powerful answer to this question is that objects, events, or entities form a conceptual coherent group if they are similar to each other. As Murphy and Medin (1985, p. 291) state, '*similarity may be the glue that makes a category learnable and useful*'. Indeed, the prototype and exemplar view of concepts are both based on similarity (see section 2.3.2.2). However, they argue that similarity alone cannot explain conceptual coherence, and they use flaws in the prototype and exemplar theories to illustrate their argument. Their critique of prototype view focuses on how the features that are used to define the summary representation of a category are selected. For example, in the case of robin, the features 'can fly', 'has feathers', and has a 'red breast', might be highly weighted as they are common features of robins and, hence, specify how robins are similar. However, so are the features, 'weighs under 1kg', 'can be found in Europe and the United States', 'lives outside'. Murphy and Medin (1985) argue that features like 'can fly', 'have a red breast' are selected as salient and given high weights, whilst 'weighs under 1kg' are not, through knowledge of robins and birds more generally.

Evidence that knowledge affects concepts also comes from studies of conceptual development. For instance, in series of studies, Carey (1985b) investigated children's understanding of the concepts of a living thing and an animal and how this understanding evolved as they aged. She identified major changes in children's knowledge of biology and demonstrated that these changes affect how they classify a living thing or an animal. In a similar study, Keil (1989) asked children to categorise animals that had undergone some sort of transformative process that altered their superficial characteristics (e.g., appearance). For instance, a racoon that had been dyed to look like skunk. Keil (1989) found that younger children (age 5) with relatively little biological knowledge, were likely to reclassify the animal (i.e., racoon to skunk), whilst 7-year-olds were less likely to accept the change, and finally most 9-year-olds did not accept that reclassification was possible. In a series of related studies, Keil (1989) demonstrates that younger children tended to categorise animals based on characteristic features (e.g., colour and pattern of fur), in a way that was consistent with the prototype view. However, older children relied on what Keil (1989) called defining featured based on their biological knowledge (e.g.,

whether an the object lays eggs). Keil (1989) argues that the results of his study demonstrate that children move beyond just classifying objects based on similarity (i.e., looking like a skunk or a racoon), towards using their knowledge (i.e., animals do not change species when their fur is dyed) to determine classification.

The studies of Murphy and Medin (1985), Carey (1985b), and Keil (1989) demonstrate that concepts are '*part and parcel of your general knowledge of the world*' (Murphy, 2002, p. 60). This view is expressed more precisely by arguing that concepts are defined by their role in our mental theories about the world (Murphy, 2002). For example, the concept of baby would be defined in relation to knowledge about reproduction (Carey, 1985b, 1988a, 2000). Hence, in the knowledge view, a concept is defined by its relationship to other concepts in a mental theory. Hence, the knowledge view is sometimes referred to, rather awkwardly, as the theory-theory (Murphy, 2002). These mental theories or intuitive theories are not equivalent to scientific theories (Murphy, 2002. They can be naïve and in some cases contradictory (Vosniadou & Skopeliti, 2014). However, like scientific theories, they are can be used to make predictions, interpret evidence, and provide explanations of phenomena in their domain (Weiskopf, 2023).

2.3.2.5 Which Theory 'Wins'?

Murphy (2002) ends his comprehensive review of the different theories of psychological concepts by playfully asking whether there is a clear winner between the prototype, exemplar, and knowledge (or theory theory) views. His answer is no. In justifying this response, Murphy (2002) makes several points which are salient for the argument in this thesis and to the shared theoretical language introduced by Sherin (2017a).

First, at present, there is no integrated theory of concepts that can explain all the empirical evidence. The prototype and exemplar views perform well at explaining category fuzziness and typicality effects when concepts are formed by recognising similarity in features that can be identified from basic stimuli from experience. However, Murphy (2002) lists a number of issues that have been identified with exemplar theory (e.g., an inability to explain hierarchical structure between concepts or induction) that suggest the prototype view may be a better model in these cases. The knowledge view addresses some limitations of the prototype view by explaining how features might be selected or prioritised and how knowledge of a domain is incorporated into conceptual understanding. Hence, Murphy (2002) suggests a prototype plus knowledge is likely to develop an integrated picture of a concept.

This picture of a concept is consistent with a dual factor view of concepts (Block, 1986; Block & Campell, 1987), where a concept contains both broad and narrow content. Narrow content is determined by the intrinsic properties of an object. For example, in the studies carried out by Keil (1989), whether an animal looks like a skunk or a racoon. In contrast, broad content refers to features of an object that depend on its environment. For instance, in the example of Keil's racoon, classifying a racoon as a mammal could involve observing a racoon feed a smaller version of itself milk. Hence, this classification decision involves broad content as it requires knowledge that cannot be gained from simply observing this object (i.e., a single racoon) in isolation from its environment. However, this decision may also involve explicit cultural knowledge (e.g., theories of scientific classification). Therefore, in this example, broad content is loosely associated with the knowledge view part of a concept and narrow content with the prototype view part. However, in the concept of water, which Putnam (1975) originally used to illustrate the differences between narrow and broad content, the intrinsic, and therefore narrow content of water, is the knowledge that its chemical formula is H_2O, rather than any knowledge of how it interacts with its environment (e.g., falls from the sky), which defines its broad content (Brown, 2022). Hence, in this example it is the narrow content that is associated with the knowledge view and the broad content that is more closely aligned with the prototype view.

The racoon and water examples illustrate that the boundaries between narrow and broad content are not always obvious or well defined. Indeed, the notion of distinguishing narrow and broad content is controversial (Brown, 2022). However, Block (1986) suggests that narrow content can be identified using the notion of conceptual role. Conceptual role is defined by a concept's function within thought. Hence, in the racoon example, the concept of racoon functions as an example of a living thing but as an object at a more basic level. Block argues this basic level primarily defines a concept's narrow content (Carey, 2011).

Murphy (2002) describes the integrated prototype plus knowledge picture of concept as messy and this is illustrated by this discussion. However, Keil (1989) provide some structure to this picture by proposing that the prototype and knowledge aspects of a concept exist in a associative network (see Figure 2.1). A concept is defined by a cluster of association between different conceptual elements. This associative network view of a concept highlights that whilst concepts are viewed as 'units' or of thought in the Language of Though hypothesis, they do have 'sub-atomic' structure and behave in a way that is influenced by their environment (i.e., their connection other concepts and sub-conceptual elements). This view of concepts clearly

resonates with the view of concepts as a network of elements, described at different levels, and also allow for the possible 'hybrids' as concepts develop (Sherin, 2017a).

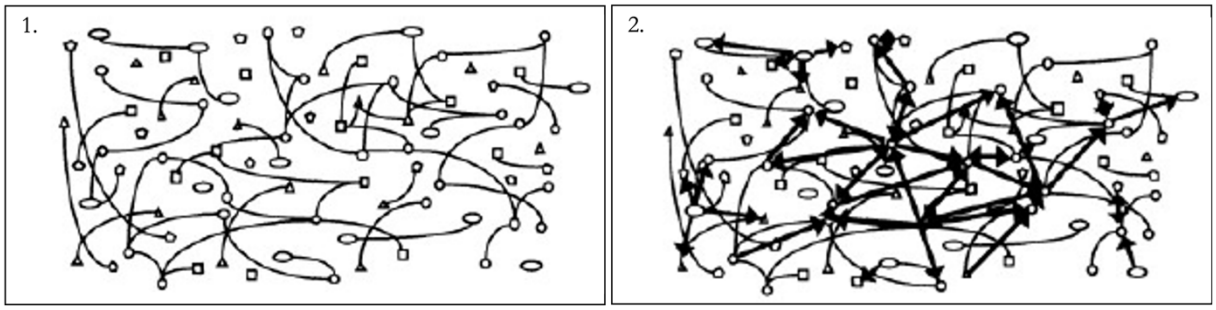


Figure 2.1 A concept as an associative network. The shapes or vertices in the network represent conceptual elements, which are either sensory-perceptual or have some conceptual content. Initially (1) elements are connected according to narrow (prototype like) content. But then, as the concept develops, connections are also made according to broad (knowledge like) content.

The second salient point builds upon Murphy's argument that there is no integrated theory of concepts and as such the prototype, exemplar, and knowledge views have developed to explain specific functions. In the case of the psychological concepts these views explain categorisation. Murphy stresses that it is likely that not all concepts will share this function. Hence, different types of concepts might require very different structures to psychological concepts. This possibility is discussed in the next section which introduces three different views of scientific concepts by primarily comparing the different functions in thought that scientific and psychological concepts might perform.

2.3.3 Scientific Concepts

The prototype view, exemplary view and knowledge views have been explicitly designed to explain psychological concepts, which are typically associated with mental representation of objects. Hence, within mentalese, these concepts are equivalent to nouns. However, concepts like force, and energy, which are typically viewed as scientific concepts (see diSessa (1993), diSessa and Sherin (1998), and Ioannides and Vosniadou (2002)) might function more like verbs (i.e., actions which cause events). This section explores three views of concepts that have been developed to explain these scientific concepts, which are often cited in reviews of conceptual change in science education due to their prevalence in this area of research (e.g., Amin & Levrini, 2017a). The selection of three views emphasises that the aim of this discussion is not to give an exhaustive review of all views of scientific concepts and associated models of conceptual change. Rather, the goal is to highlight significant differences in these views, which have limited progress in developing a consensus around the nature of conceptual change, but

also to stress that there is some commonality in these views at abstract level, which justifies the description of scientific concepts as a system of elements given by Sherin (2017a).

2.3.3.1 Coordination Classes View of Scientific Concepts

diSessa and Sherin (1998) have developed a model of scientific concepts, which they call coordination classes. Like the approach taken by cognitive psychology (described in the previous section), diSessa and Sherin (1998) define concepts as mental representations and then set out to describe the structure of these representations. However, they do not see the function of scientific concepts as being the same as the psychological concepts that have been describe in the previous section, which represent categories. They illustrate the difference between these two types of concepts through performance. Coordination classes are used by individuals to 'see' information in the world. For psychological concepts, this would involve 'seeing' that an object is a member of a category (e.g., a bird). However, scientific concepts often do not involve this type of decision. For instance, determining an object's velocity involves a judgement on a continuum (i.e., is the velocity high or low, greater than another object), rather than simply a yes or no decision about membership.

diSessa and Sherin (1998) refine their vague performance specification of 'seeing' values of quantities by adding the criteria of integration and invariance. Integration refers to the idea that to arrive a value for a conceptual quantity, multiple observations might have to be coordinated. For example, the distance an object travels in a certain time, and in what direction, are observations that need to be noted and combined to determine velocity. However, to 'see' velocity in more than one situation, observations need to be made and combined in a consistent way. This criterion is invariance.

The primary mental structures that achieve integration and invariance are read-out strategies and a causal net. Read-out strategies select the information that needs to be noted in a particular situation to 'see' the concept in question (e.g., distance travelled in certain direction and time for velocity). Causal net combines or 'processes' the information from the read-out strategies to determine the value of concept. For example, the magnitude of velocity is the distance travelled per unit time, or the magnitude of the force need to cause' acceleration is given by F=ma (diSessa & Sherin, 1998, p. 1176). As these two examples illustrate, knowledge of equations that define physical laws and quantities play a pivotal role in forming an expert causal net. Moreover, the nature of the casual net strongly influences the readout strategies that are needed to develop a coordination class. In less developed coordination classes, the causal net

is heavily influenced by individual's intuitive sense of mechanism, which represents their ideas about cause (diSessa, 1993). In novices, the sense of mechanism (and consequently the causal net) is dominated by phenomenological primitives or p-prims (diSessa, 1993). P-prims capture individuals' explanations of cause that be gained from sensory and perceptual experience only, rather than explicit knowledge. Two examples of p-prims are:

1. **Ohm's p-prim**

   An agent acts through a resistance to produce a result. For example, a box can be made to move on a surface but will move slower if is pushed on a rough, rather than a smooth surface.

2. **Force as a mover**

   An impetus acting in a burst on an object will cause that object to change direction and/or speed in the direction of the impetus. For example, hitting a ball with a bat.

These two examples illustrate that there are aspects of p-prims that could be defined as narrow content. For instance, in the force as a mover example, force is defined as a cause. Similarly, Ohm's p-prims includes element that are defined as agents.

As coordination classes develop, p-prims are integrated with the knowledge elements of the causal net. For instance, Ohm's p-prim and Force as a mover are reinterpreted and coordinated as F=ma. Hence, the narrow content of p-prims (e.g., force as a cause) is interpreted in terms of broad content (e.g., force as a cause of acceleration).

The above discussion illustrates how the conceptual elements (i.e., causal net, readout strategies, p-prims) that make up a coordination class are all closely related and interaction with each other. Hence, diSessa (2002) refers to coordination classes as '*complex knowledge systems*' (p30) or a conceptual ecology. Moreover, this knowledge system is widely distributed and cannot be seen as a 'unitary mental structure (concept node), or even a small number of mental structure' (diSessa & Sherin, 1998, p. 1170). This is a break from assumption of concepts as a 'unit of thought' that operates as a mental word in a mentalese language of thought model that is prevalent in cognitive psychology. In contrast, the views of scientific concepts that are discussed in the following sections (2.3.3.2 and 2.3.3.3) adhere more closely to the view of concepts as 'mental words' and, therefore, align more closely with notion of psychological concept.

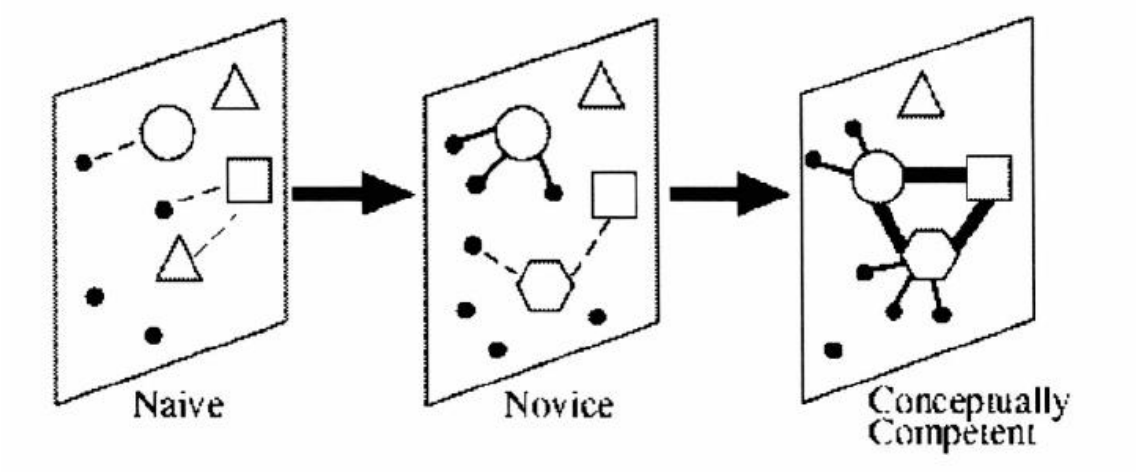


Figure 2.2 Coordination classes as a 'complex system of knowledge elements' (diSessa & Sherin, 1998, p. 1172). Taken from diSessa (2002, p. 31).

2.3.3.2 Framework or Intuitive Theory View of Scientific Concepts

Coordination classes are motivated by a deliberate break from many of the assumptions of the psychological theories of concepts discussed in section 2.3.2. In contrast, framework or intuitive theory[1] view of scientific concepts are an extension of the knowledge view in section 2.3.2.4, which were original developed by Carey (1985b) to explain the categories of living thing. In this view, a concept view is defined by its role in an associated intuitive theory of knowledge. This role may involve conceptual elements that connect a concept to entities in the world (reference) but also elements that are defined by how this concept functions in thought (Carey, 2009, p. 5). These functions include:

1. Representing causal and explanatory knowledge.
2. Making inferences and predictions.
3. Selecting features that support categorisation decisions.

Hence, intuitive theories play a role in how individuals recognise the intrinsic properties of an object but also how a concept interacts with wider knowledge. Hence, the intuitive theory view of scientific concepts is consistent with the view of psychological concepts outlined in section 2.3.2 that proposes a dual theory of conception content, which incorporates both broad and narrow content (Block, 1986; Block & Campell, 1987).

[1] Note that the terms mental, framework, and intuitive theories are used in the literature to refer to the same view of scientific concepts. For instance, Ioannides and Vosniadou (2002) use 'framework theory' but Carey (2009) use 'intuitive theory'. However, both authors are referring to the same idea. Moreover, 'scientific theory' can be used to refer to a mental theory that aligns with an accepted scientific theory. In this thesis, the term 'intuitive theory' is adopted to be consistent with the use of this term in Carey's Theory of the Origin of Concepts (Carey, 2009).

To illustrate this view of scientific concepts, Figure 2.3 shows how the concept of 'baby' could be embedded in an intuitive theory that defines this concept. The baby example is adapted from empirical observations made by Carey (1985a, 1985b, 1988a).

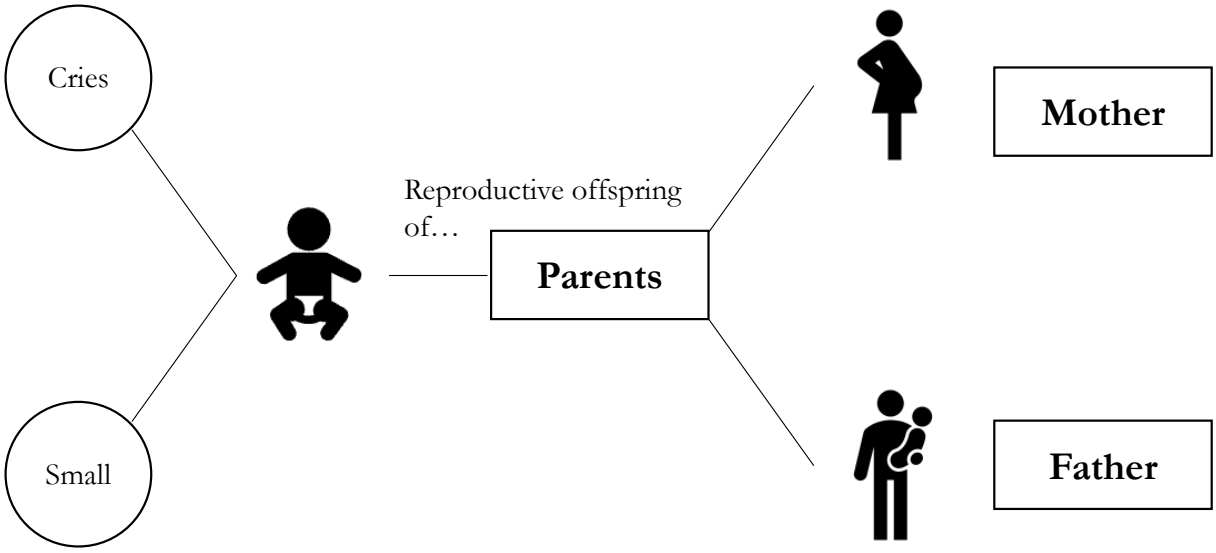


Figure 2.3 A hypothetical concept of a baby adapted from empirical observations made by Carey (1985a, 1985b, 1988a). The narrow content associated with the concept is on the left-hand side of the diagram (circles)). The broad content is on the right-hand side. The rectangular outlines indicate possible concepts embedded within the intuitive theory build from these constituent concepts.

Figure 2.3 highlights several important features of the intuitive theory view. First, the target concept exists within an associative network or system of concepts that defines the intuitive theory. Within this system, some conceptual elements are associated with narrow content (e.g., small, cries), whilst others are associated with broad content (e.g., reproductive offspring). This is analogous to an associative network view of a concept proposed by Keil (1989) (see Figure 2.1). Second, within each system there are some associations that define the core conceptual role of the concept, whereas other associations are more peripheral (Carey, 2000). These core associations or properties can change as a concept develops. This defines conceptual change in the intuitive view, rather than knowledge acquisition (Carey, 2000, 2011). Finally, concepts can have different roles within theories that are not limited to referring to categories of objects). Hence, the intuitive theory view of concepts can extend from psychological concepts to scientific concepts.

2.3.3.3 Ontological View of Scientific Concepts

In contrast to the intuitive theory and coordination approach, the ontological view of scientific concepts adheres to the assumption that concepts are mental representations of categories, which is the psychological view (Chi & Roscoe, 2002). However, these categories are extended from just including matter to processes and mental states (Chi et al., 1994). In this view, children

have robust misconceptions because they incorrectly categorise scientific concepts. For example, electricity is seen as a substance, a category of matter, rather than a process (Chi & Roscoe, 2002). Therefore, conceptual change is explained in terms of ontological shifts (Chi et al., 1994), where a child changes their fundamental belief about what a type of thing a concept is by moving that concept from one ontological category to another. Hence, in this view, the core or defining features of a scientific concept are the conceptual elements that determine its ontological category. Chi (2005, p161) calls these conceptual elements a concept's ontological attributes which could be associated with narrow content. The ontological shifts required by Chi et al. (1994) could be brought about by changes in broad content. For instance, explicit knowledge about electricity that would help to redefine this concept as a process, rather than a substance.

Ontological categories are distinguished by 'mutually exclusive sets of plausible attributes' (Chi, 2005, p. 161). For instance, it is plausible for an object to have a colour as an attribute but not an event like a football match is not coloured. The ontological view is vague about how these ontological attributes might be represented in concepts. However, Chi and Roscoe (2002) seem to suggest they could be predicate statements (i.e., x is a…) that play a significant role in determining a category's ontological attribute. For instance, the statement 'the dog is brown', as opposed to the 'the dog is an hour long', suggests the item dog is an object, rather than a process. These kinds of statement are consistent with a prototype view and an intuitive view of concepts.

2.3.3.4 Summary

Murphy (2002) stresses that there is no integrated theory of concepts and, hence, there is likely to be a need for bespoke models of concepts fitted to their function (i.e., determining category membership) (see section 2.3.2.5). This has been taken on board by the theories of scientific concepts discussed in the previous sections, especially diSessa and Sherin (1998), and has led to three different approaches, which have been outlined in the previous sections. As was the case with psychological concepts (see section 2.3.2), the shared theoretical language suggested by Sherin (2017a) can be used to compare these views of scientific concepts.

First, it is clear that each model fits into Sherin's description of a 'system of elements, described at different levels' (Sherin, 2017a). In a coordination class, the read-out strategies (element 1, higher level) interact with the causal net (element 2, higher level), which is in term made up of p-prims (element 1, lower level, broad content) and other knowledge elements, such as

equations, (element 2, lower level, narrow content) which are also connected. In the intuitive theory view of concepts, a target concept is embedded in system of other concepts (element 1, higher level), which articulates the intuitive theory that defines part of conceptual role of the target concept (narrow content). There are also conceptual elements (element 2, lower level) that determine how the concepts refers to entities within the world (broad content). In an ontological view, conceptual elements might include attributes that establish a concept's ontological category.

Second, there are differences in how the system of elements is organised in each of the models and the elements operate in thought. The ontological view implies a tree-like structure (see Chi (1992)). In contrast, the coordination classes and intuitive theory views do not require a top-down structure, but rather a more web-like structure (see Figure 2.2 and Figure 2.3). In thought, the concepts described by the intuitive theory and ontological views are atomic in the sense they can operate as 'mental words' in a mentalise language of thought. However, diSessa and Sherin (1998) do not accept this view and argue that coordination classes 'are somewhat fuzzy' (p1170) and do not fit the idea of a 'unitary mental structure' (p1170).

Third, despite all being systems of elements, this discussion highlights that the nature of the elements in each system is a major difference between these views. However, despite these differences, all three models include elements that are learnt through experience and interaction with the world (broad content), and elements that are associated with more intrinsic features (narrow content). Hence, all models are dual content views of concepts. The view of scientific concepts as a system of both broad and narrow based elements generally connects these models to the model of psychology concepts developed by Keil (1989), which was also an associative network of experiential (broad content) and knowledge (narrow content) based elements. This general commonality is important as it highlights that there is some agreement between the concepts described by cognitive psychology and science education research. Hence, the theoretical insights, empirical results, and techniques into researching the nature of concepts and conceptual development that have been discovered by cognitive psychology can be applied to enhance and our understanding of scientific concepts, as Rusanen and Pöyhönen (2013) suggest, as long as the distinction between the two types of concept are recognised. In this thesis, there are several instances when ideas from the study of psychological concepts have been applied to enable further research into scientific concepts. These include, applying Fodor's theoretical arguments for radical concept nativism (Cain, 2002) to models of conceptual change in science education research, and using Carey's theory of the origin of concepts (Carey, 2009),

which draws upon both theoretical and empirical research in cognitive science to explain conceptual development,

It is clear from this discussion that there are significant differences in the three models of scientific concepts. This lack of agreement is problematic for conceptual change models because it is difficult to agree what needs to change in conceptual development when there not a clear definition of what a scientific concept is. diSessa and Sherin (1998) forcefully make this argument. Therefore it is understandable that diSessa et al. (2004) have written challenging the intuitive or framework view of scientific concepts by questioning whether children's naïve scientific ideas can be viewed as coherent theories. In reply, Vosniadou (2017) has questioned diSessa's criticisms. Whilst this debate has taken place, both the coordination class view, intuitive view, and ontological view of concepts have all been used in empirical studies to analyse conceptual change (e.g., Brookes and Etkina (2009); Levrini and DiSessa (2008); Parnafes (2007); Vosniadou and Skopeliti (2014)). In many cases, the data from these studies can be interpreted in terms of each view, with no clear advantage to using one view over another. This ongoing debate about the nature of scientific concepts has limited progress in research on the nature of conceptual change and the mechanisms responsible (Amin & Levrini, 2017a).

However, this discussion also highlights the argument made by Sherin (2017a) and Rusanen (2014) that despite the lack of an agreed definition, there are commonalities in models of scientific concepts, if these models are viewed at high levels of abstraction. At these high levels, a broad definition of a scientific concept might be:

> *A system or network of elements, with elements at different levels. Some of these elements are associated with intrinsic properties of the of the concept, which define the concepts core function (narrow content). Other elements are associated with how the concept interacts with its environment (broad content). Elements associated with narrow and broad content can be influenced by explicit knowledge.*

This is a minimalist and pragmatic definition. However, as Amin and Levrini (2017b) argue conceptual change research has perhaps reached a point when a '*minimal shared theoretical vocabulary'* (p2) is needed to make it possible to make precise models of conceptual change and to compare different theories. The next section demonstrates that this minimal and broad definition is useful to investigating conceptual change and enables progress to be made in terms

of understanding this process and selecting which of the three views of scientific concept presented here give the most complete account of conceptual change.

2.4 What is Conceptual Change?

diSessa and Sherin (1998) strongly argue that a precise definition of a concept is needed to make progress in researching mechanism for conceptual change. They argue that this definition is needed to specify what actually changes in conceptual change or development. In their argument, they cite Carey's research on conceptual development (specifically Carey (1988b)) as an example of research that lacks this precise definition and, therefore, is limited in what it shows about conceptual change. However, to a certain extent, Carey (2015) reverses this argument. She acknowledges any theory of conceptual change needs a general assumption about what a concept is. However, she also argues that once this general definition is in place, this allows conceptual acquisition or change to be studied in a robust way. Moreover, researching conceptual development leads to improvements in understanding of what a concept is (for example, see the conclusion of Carey (2009)). This position is adopted in this thesis and this section demonstrates how the broad definition of a concept as a system of elements (including some with broad content and other with narrow content) can be used in an approach to studying conceptual acquisition that can lead to a greater understanding of conceptual change, and particularly the mechanism of conceptual change. The approach to conceptual acquisition that is used in this section adopts the three principles questions that Carey (2015, p. 15) argues any theory of conceptual development need to address. These are:

1. What is the innate conceptual repertoire?
2. How do any initial conceptual representations differ from a descendent conceptual system?
3. What is (are) the mechanism(s) that achieves the transformation of the initial into the descendent final state?

Each of these questions are discussed in the following sections and it is highlighted where the answers to these question limit or promote a particular view of scientific concepts.

2.4.1 The Innate Conceptual Repertoire

The innate concepts repertoire refers to the initial stock of concepts that are available to individuals to build additional concepts during any learning process (Spelke, 2000; Spelke, Breinlinger, Macomber, & Jacobson, 1992; Spelke & Kinzler, 2007). Hence, they must exist prior to any learning and in this sense, they are innate or unlearnt (Carey, 2009). In the broad definition of a scientific concept as a system of elements, they are the initial elements or vertices in the network.

In coordination classes, the innate conceptual repertoire would appear to consist of p-prims. diSessa (1993) identifies p-prims as existing somewhere between sensory and perceptual representations and named concepts (i.e., knowledge) in a cognitive hierarchy. However, diSessa (1993) is clear that p-prims are learnt from experience. If p-prims are learnt, then they are not the not part of an innate conceptual repertoire. Therefore, p-prims themselves must be formed from more primitive representations (Henderson et al., 2017). This is clear when the content of the examples of p-prims given in section 2.3.3.1 are analysed. For instance, Ohm's p-prims states; '*an agent or causal impetus acts through resistance or interference to produce a result*' (diSessa, 1993, p. 217), and Force as a mover states; '*A directed impetus acts in a burst on an object resulting in displacement and/or speed in the same direction'* (diSessa, 1993, p. 217). Clearly, both p-prims are framed in terms of agents, causes, and objects.

Carey (2009) identifies each of these ideas as a core concept. Core concepts form the initial conceptual repertoire in Carey's theory of the origin of concept (Carey, 2009). Carey presents extensive empirical evidence for the existence of innate core concepts (see section 2.5.1 for a more detailed account of this evidence). Carey (2009) also explains how core concepts could form part of intuitive theories. Hence, at a very fundamental level, the coordination classes view of scientific concepts give a less complete picture of scientific concepts, when compared with the intuitive theory view because it does not specify innate conceptual repertoire, and therefore the initial starting point for conceptual development. This is also true of the ontological view of scientific concepts, which does not comment on what the initial conceptual repertoire might be.

2.4.2 Changes in the Conceptual System

The broad definition of a scientific concept adopted in this thesis, and described is section 2.4.4, is a system of element with narrow and broad content. Hence, at this abstract level, conceptual change is a change in these systems of elements (Rusanen, 2014; Sherin, 2017a). This change can happen in several ways, which have been identified and characterised by Lappi (2007).

First, the system of elements can be reorganised by changing the way these elements are connected. This could be by joining elements that are initially isolated to form a more connected and organised whole. This is consistent with the view of conceptual change proposed by diSessa and Sherin (1998) in the coordination class view. Alternatively, elements can shift from one part of the network to another. This is broadly consistent with the ontological view of conceptual change proposed by Chi et al. (1994).

Second, if a concept is defined as a system of elements, there will be a set of elements that belong to this system and others that do not. This set of elements could change in two ways. Two or more sets of elements that form distinct systems, and therefore different concepts, could merge to form a single conceptual system. This is conceptual *coalescence* Carey (1978); (Carey, 1988b). Alternatively, a set of elements that once formed a single system, and hence a single concept, could split to from two sets and two separate concepts. This is conceptual *differentiation* (Carey, 1978, 1985b). In this case, it could be that some elements are shared between the two final systems of elements. Coalescence and differentiation are both examples of reinterpreting the system of conceptual elements.

Finally, new vertices, representing new conceptual elements, could appear in the conceptual system that could be used to form new systems of elements, i.e. a new concept. This final possibility is the most problematic for models of conceptual change because it implies that new representational resources are being built from elements that are not part of the initial conceptual repertoire. This causes serious issues for some types of learning process (i.e., hypothesis testing) that have been suggested as mechanisms for conceptual change. These issues are discussed in detail in the next section.

2.4.3 The Mechanism of Conceptual Change

If conceptual change is defined as changes in a system of conceptual elements, then the mechanism for conceptual change is whatever coordinates or drive these changes. Rusanen (2014) argues that in many models of conceptual change this is often an internal 'inside the head' mechanism that makes the system 'better' in some way, either in terms of being more fruitful, intelligible and plausible (Posner et al., 1982; Strike & Posner, 1992), have greater utility (Ohlsson, 2009a, 2009b) or more coherence (diSessa, 1993; diSessa et al., 2004; Thagard, 2000).

Posner et al. (1982) define the terms fruitful, intelligible, and plausible quite loosely. Intelligible implies any new arrangement of a conceptual system must make sense to a learner, plausible implies that it must be believable, and fruitful implies that the new system must be more productive in terms of solving problems compared with the previous system. Ohlsson (2009a, 2009b) defines utility in terms of the usefulness of the system. In their view, a more useful system requires less cognitive load and will lead to more efficient, faster cognition and more satisfactory end states (Rusanen, 2014). Finally, diSessa (1993) define coherence in terms of the connectedness of the system or coherence.

All three of these examples involves a comparison between the old and the new conceptual system and an evaluation of which system is 'better' in the terms that they have defined (i.e., utility, coherence, fruitfulness). However, comparison and evaluation as a mechanism or catalyst for conceptual change suffers from two major issues.

First, Rusanen (2014) points out that greater utility, coherence, plausibly, or fruitfulness are all normative goals for a conceptual system. Hence, they are not in themselves mechanisms for conceptual change. These goals might specify end states but not how the system gets to this state. Rusanen (2014) argues that currently there is no plausible internal 'in the head' mechanism that organises learning processes to arrive at these normative goals.

Carey (2009) and many sociocultural theories of learning (e.g., Vygotsky (1986); Wertsch (1991)) avoid this problem by citing external influences, such as language, as organising the reordering of conceptual systems. However, Carey (2009) specifically turns to language to address a second major issue with any model of conceptual development that relies on comparison between the efficacy of thinking, using old and new conceptual understanding. This issue was originally articulated by Fodor (1975) in support of a strong version of his conceptual continuity thesis and radical conceptual nativism. The conceptual continuity thesis states that it is not possible to develop concepts that cannot be expressed in terms of existing representational resources (i.e., existing concepts). Hence, all concepts are innate, which includes concepts like house, elephant, proton, and carburettor (Cain, 2002, p. 69). This is radical concept nativism.

As the name suggests, radical concept nativism is seen as a very strong, and perhaps extreme, claim and, as such, it has few supporters (Laurence & Margolis, 2002). However, it has proven difficult to identify what is wrong with Fodor's argument and examining the premises of the argument is relevant here as they indicate which of the views of scientific concepts presented in section 2.3.3 might be most useful in understanding conceptual change.

Fodor's argument of radical conceptual nativism is based on three premises (Rey, 2014). The first premise states that all learning mechanisms are essentially hypothesis testing (Rey, 2014). In other words, all learning involves a comparison between whether one idea or another best explains an experiences. Figure 2.4, illustrates Fodor's example of how hypothesis testing might work when learning a concept (Cain, 2002).

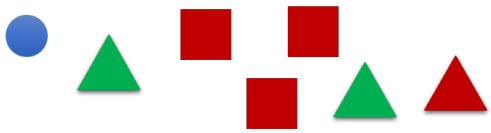
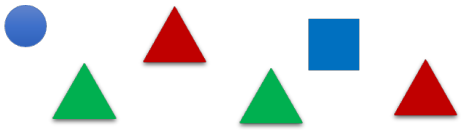
Evidence		Hypothesis
There are three glurgs here.		A glurg is a triangle or a square.
There are three glurgs here.		A glurg is a triangle.
There is two glurgs here.		A glurg is a red triangle.
There are two glurgs here.		A glurg is a green triangle.

Figure 2.4 An example of learning a concept by comparing a hypothesis with experiential evidence (hypothesis testing) adapted from an example used by Cain (2002).

A more scientific example comes from Taylor (2012), who found that 15-year-old students often see acceleration occurring when an object 'getting faster', rather than 'a change in velocity' (see Figure 2.5).

Evidence **Examples and teacher input about whether acceleration is taking place.**		**Hypothesis 1** **An object accelerates if it gets faster.**	**Hypothesis 2** **An object accelerates if its velocity changes.**
A ball is held stationary 1m above the floor.	There is no acceleration.	✓	✓
A ball is dropped from 1m and its speed increases from 0.2 m/s to 0.5 m/s.	There is acceleration.	✓	✓
The balls is now falling at constant speed of 0.5 m/s.	There is no acceleration.	✓	✓
The ball hits the desk and comes to a halt. It's speed decreases from 0.5 m/s to 0 m/s.	There is acceleration.	✗	✓
A car is travelling west at 5 m/s. It travels around a bend at a constant speed. After the bend it is travelling south.	There is acceleration.	✗	✓

Figure 2.5 An example of learning the scientific concept of accelerations might be approached hypothesis testing. Taken from exerts of classroom dialogue identified by Taylor (2012).

The problem Fodor identifies with hypothesis testing as a mechanism of concept acquisition is summarised by Carey (2009): '*[you] cannot learn what you cannot represent*' (p18). The example of

learning the *glurg* concept illustrates how this observation leads to a paradox. In this case, to be able to think '*a glurg is a green triangle*' in mentalese, an individual must already be able to represent '*green triangle*'. Hence, they already have this concept. Similarly, to think, '*ahh, acceleration must mean getting faster or slower*', you must already be able to represent this concept. This argument can be applied to all mechanisms of conceptual change that involve comparison between a new and old conceptual system (i.e., hypothesis testing).

Cain (2002) points out that an obvious solution to Fodor's argument for radical concept nativism is to propose that the concept green triangle is formed by combining green and triangle, which are themselves innate concepts. This is the basis of Fodor's second premise, which states:

> *'One can only learn new concepts by creating and confirming hypothesis formulated in terms of logical connectives from antecedentally available primitive concepts'.* (Rey, 2014, p. 111)

Hence, conceptual elements that are already present in a conceptual system can be used as 'bricks' to 'build' new concepts to make comparisons with older concepts. In a broad definition of a concept as a system of elements this would be consistent with reorganisation or reinterpretation, but not with the introduction of new vertex elements.

However, Fodor argues that most concepts are not composite. This is the third premise of his argument (Rey, 2014). Hence, new concepts cannot be formed from reordering or reinterpretation but only by invention of new conceptual elements, and therefore new representational resources. This is inconsistent with the first two premises of Fodor's argument. Therefore, all concepts must be native (i.e., innate) (Cain, 2002).

Carey's theory of conceptual development is a rebuttal of Fodor's argument for radical concept nativism and conceptual continuity and does so by refuting the first and second premise in Fodor's argument in three ways. First, Carey presents extensive and robust empirical evidence for the nature of the representations in the innate conceptual repertoire. Second, she uses this knowledge of the innate conceptual resources to demonstrates that in conceptual development, individuals create representational resources (subsequent concepts) more powerful than their antecedental representations (prior concepts). In other words, new elements can be added to any system of elements and then used to build more sophisticated concepts. Third, Carey's theory of conceptual development explains in detail how language, as opposed to hypothesis testing, can organise specific learning processes to develop new concepts from the concepts in the innate conceptual repertoire. As has already been noted, Carey emphasis on language's

organising role in conceptual development also addresses Rusannen's concern about the lack of a mechanism to organise learning processes to make a conceptual system 'better'. Hence, Carey's theory of conceptual change, and the intuitive theory view of scientific concepts it adopts, provides a complete account of both scientific concepts and conceptual change (Amin, 2017).

In contrast, the models of conceptual change implied by the coordination class and ontological view of scientific concepts are less complete for two main reasons. First, both views do not specify a precise mechanism that organises learning processes that reorganise the systems of conceptual elements (Rusanen, 2014). Although, diSessa and Sherin (1998) imply a role in symbolic language as part of this mechanism by identify equations (e.g., F=ma) as structures that coordinate p-prims. Second, the models of conceptual development implied by the coordination class and ontological views do not convincingly specify the innate conceptual repertoire. Hence, they fail to completely show how more powerful representation resources (e.g., p-prims), which form the basic elements in their conceptual system, can be created from less powerful representations (e.g., sensory-perceptual representations). Therefore, it is not clear how they overcome the issues with conceptual learning, specifically associated with hypothesis testing, that Fodor's first and second premises highlight.

2.4.4 Summary

This section has adopted a pragmatic definition of a scientific concept in terms of a system of elements with broad and narrow content. This parsimonious definition reflects the minimal shared theoretical language (Sherin, 2017a) that can be applied to describe three different views of scientific concepts. In adopting this definition, this section has demonstrated that it is possible to begin to compare models of conceptual changes and develop more precision in how they describe conceptual change. This has not been an exhaustive comparison of all theories of conceptual change to definitively rule out certain theories. However, this comparison has shown that Carey's theory of conceptual change, and the intuitive view of scientific concepts it incorporates, gives a more comprehensive answer to the questions, 'what is a concept?', and 'what is conceptual change?', when compared with the coordination class or ontological views of scientific concepts. Therefore, Carey's theory of the origin of concepts is adopted as the theoretical framework in this thesis. Carey's theory is explained in detail in the next section because of its importance.

2.5 Carey's Theory of the Origin of Concepts

Susan Carey's Theory of the Origin of Concepts (TOOC) (Carey, 2009, 2011) is a comprehensive account of conceptual development (Amin, 2017). The TOOC aims to resolve the paradox Fodor identifies concerning learning via hypothesis testing, and in doing so highlights the fundamental role of language as one plausible mechanism that drives conceptual development. Carey outlines the TOOC by addressing the principle three questions Carey (2015) suggests any theory of conceptual development needs to answer. These questions have already been introduced in section 2.4, and are: what is the nature of innate conceptual repertoire, what are the changes that take place during conceptual change, and what is the mechanism for change? Carey's response to these questions forms the three theses of the TOOC which are described below.

2.5.1 Core concepts

The innate conceptual repertoire represents the original building blocks or elements for any more complex representations. Hence, these building blocks can be identified by breaking concepts down into their most fundamental components, either through examining behaviour in empirical studies or through theoretical arguments, or a combination of both (Carey, 2009; Spelke & Kinzler, 2007). Historically, sensory, and perceptual representations have been assumed to be this fundamental level because they are our most basic response to stimuli (Berkeley, 1910; Locke, 1975; Piaget, 1954; Quine, 1960). However, Carey rejects this assumption and argues that the innate repertoire contains conceptual as well as sensory/perceptual representations. She calls these innate conceptual representations core concepts.

Carey identifies two logically independent and empirically distinct properties of concepts that suggest they have conceptual content, as opposed to simply sensory/perceptive content. First, they cannot be reduced to what Carey calls spatiotemporal or sensory vocabulary. Second, they have a rich conceptual role (Carey, 2009, 2011) in mental processes, which is primarily distinguished by Carey from the role of sensory/perceptual representation by inference. Inference is a hallmark of conceptual thinking because, as Carey states sensory/perceptions '*represent the here and now, and almost nothing follows from the fact that something is red, whereas rich inferences are licensed by identifying something as an agent*' (p9).

Carey presents extensive empirical evidence from several studies that demonstrate that very young children (2 to 5 months of age) possess mental representations that cannot be explained by sensory/perceptual representations and, hence, require conceptual content. Carey uses this evidence to establish that core cognition includes representations of *object*, *number*, *agent*, *goal*, and *cause*. This evidence is outlined in four chapters of the TOOC. Hence, it is too extensive to be described in detail here. Instead, a single example is presented here to illustrate Carey's argument that sensory or spatiotemporal stimuli (e.g., location, paths of motion, shapes, or colours) alone cannot account for the ability of very children to perform certain tasks, which require some sort of conceptual representation that includes the ability to infer. The very young age of the children (approximately 2 months of age) is taken as evidence that these conceptual representations are innate.

Spelke, Kestenbaum, Simons, and Wein (1995) use the fact that even very young children will express signs of surprise (i.e., looking for longer at a scene) to demonstrate that infants (1 or 2 months of age) have a concept of object. Their experiment is shown in Figure 2.6. The key question in both cases is: how many objects are involved in the event? For A, the answer is two objects. If only a single object were involved, this object would have to dematerialize behind the right-hand screen and rematerialize behind left-hand screen. The path a single object would have to take would be spatio-temporally discontinuous. In contrast, event B can only be explained with a spatio-temporally continuous path of a single object between the two screens. Spelke et al. (1995) demonstrated that infants consistently recognised the unexpected outcomes for both A and B through longer looking times. Hence, they concluded that these infants could establish representation of objects using spatio-temporal discontinuity. In other words, very young infants must be able to represent objects when they are hidden (no spatio-temporal input) and use knowledge associated with that representation (i.e., that objects do not appear and disappear) to infer that the unexpected results in both cases are odd. Therefore, they have a core concept of object.

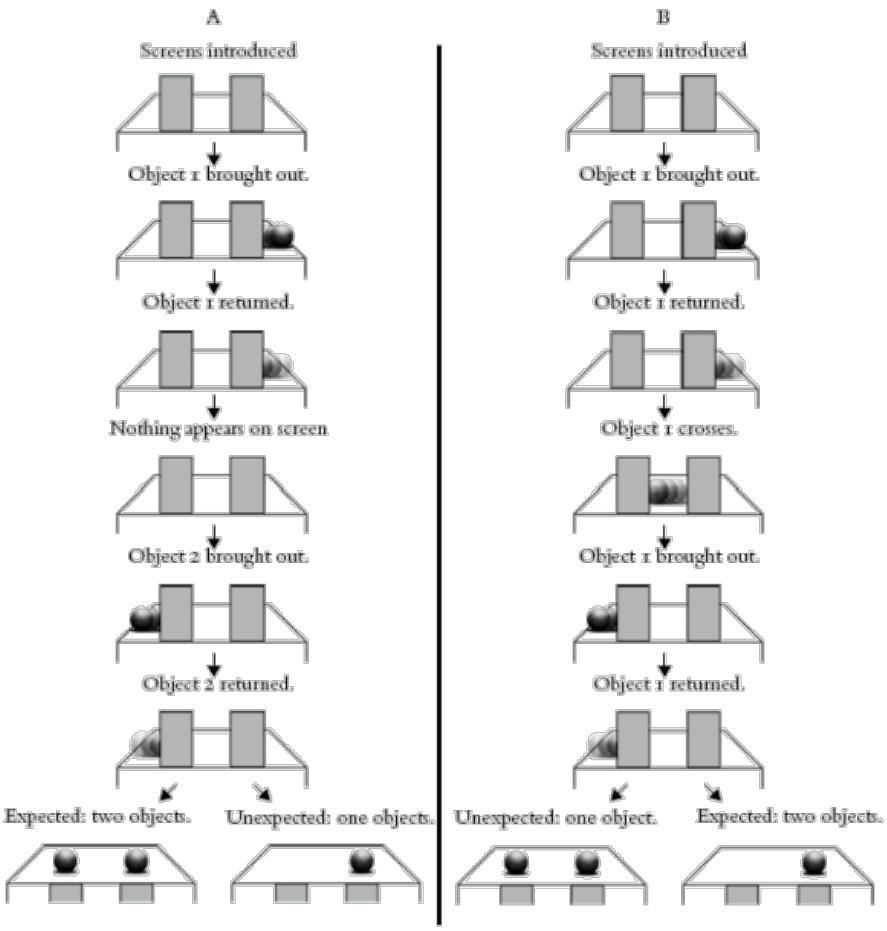


Figure 2.6 Diagram of the experimental procedure used by Spelke et al. (1995) to demonstrate that infants have a core concept of object. The figure is reproduced from Carey (2009, p. 43). Note the typo in condition B – 'Expected: two objects' should read 'Expected: two objects' and 'Unexpected: one object' should read 'Unexpected: two objects.'

Despite having conceptual content, core concepts differ from the concepts that are defined by intuitive theories. These concepts are language-like representations and make use of symbols to express complex theories. In contrast, core concepts are iconic, rather than symbolic, representations (Carey, 2009).

2.5.2 Conceptual Discontinuity in Conceptual Development

Carey (2009) second thesis rejects Fodor's continuity thesis (2nd premise) and argues that human beings (alone amongst animals) can create representational resources that are more powerful and fundamentally different to sensory/perceptual representations and core concepts. Carey arrives at this thesis through her extensive characterisation of the innate conceptual repertoire and several case studies of conceptual development. Through this analysis, she identifies several incidences of conceptual discontinuities, when more advanced conceptual systems are incommensurate with antecedent systems. Carey argues (empirically and

theoretically) that discontinuities can occur at two levels of abstraction, defined by whether the discontinuity occurs at a domain general or domain specific level of knowledge.

First, at a domain general level, the structure of core cognition is inconsistent with more sophisticated concepts (e.g., planet or germ, see Carey (2011, p. 118)), which are defined within intuitive theories (Carey, 2009). These concepts are represented through language-like structure (i.e., mentalese), and are not iconic, like core concepts. Furthermore, concepts within intuitive theories can develop and change as the intuitive theory that defines them develops. In contrast, core concepts are fixed throughout an individual's lifetime. Finally, core concepts exist in isolation from other concepts, whereas concepts in intuitive theories are defined in terms of how the relationships they form with other concepts in these theories.

The differences above are at a domain general level because they apply to conceptual differences across different domains of knowledge. However, discontinuity can also exist in domain specific levels. These discontinuities are identified by comparing prior (CS1) and subsequent conceptual systems (CS2). Four examples of these conceptual discontinuities are described below. The first and second examples, natural and rational number, illustrate when the CS2 is more representationally powerful than CS1. The third and four examples, taken from thermodynamics and theories of matter, demonstrate when how a subsequent CS2 can be built from concepts in CS1 that are incommensurate with the final concepts that are embedded in CS2. These are both examples of discontinuities that involve scientific concepts.

2.5.2.1 Discontinuity in Natural and Rational Number

The conceptual system that initially represents number is core cognition (CS1). Core cognition has two systems for representing number: parallel individuation of small set of entities in a working memory model, and analogue magnitude representation of number (Carey, 2009, 2011). Both ways of representing number are iconic and are summarised in Figure 2.7. Neither parallel individuation nor analogy magnitude representation can represent a series of infinite positive integers (i.e., 1, 2, 3, etc… or n+1). Parallel individual has an upper limit of sets of 3 or 4. Analogue magnitude representations are not limited in this way. However, they only represent approximate cardinal values. They are unable to represent exactly one because they are arbitrary magnitude and rely on comparison to represent higher or lower numbers. Hence, they cannot represent n+1. In contrast, the CS2 represents number as verbal numerals embedded in a count routine. Therefore, CS2 can represent an infinite series of positive integers

and, therefore, has greater representation power than both representational systems (parallel individuation and analogue magnitude) for number in CS1.

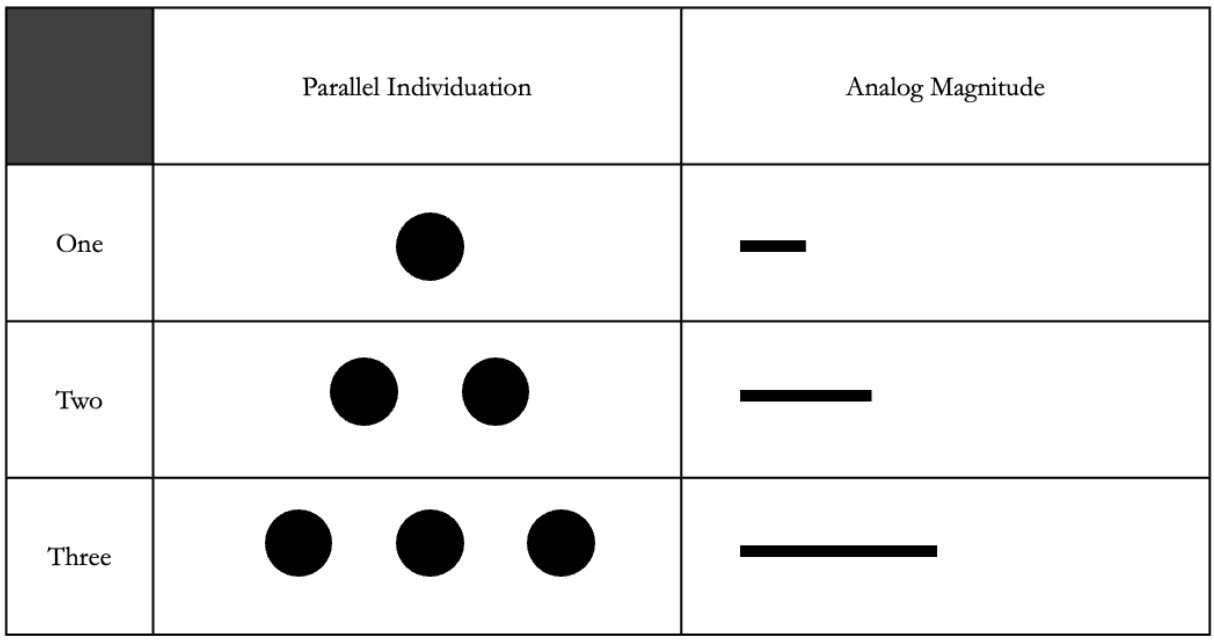


Figure 2.7 The iconic representation to represent numbers up to three used by core number cognition (Carey, 2009).

Rational number is another example within mathematics that demonstrates a conceptual discontinuity where the later conceptual system (CS2) has more representation power then the former system (CS1). In this example, CS1 is the positive integer series that was CS2 in the previous example. So, in this system a 'number' is 1, 2, 3, etc… In CS2, the rational number system, a 'number' is '*any point on a number line that can be expressed by x/y, where x and y are positive integers*' (Carey, 2009, p. 119). Hence, CS2 requires division on integers to be represented. Division is not represented in CS1, which is expressed entirely in terms of addition, subtraction, and multiplication of integers. Hence, the transition from natural number (CS1) to rational number (CS2) is another example of conceptual discontinuity because CS2 requires additional conceptual resources with increased representational capacity.

2.5.2.2 Discontinuity in Intuitive Theories

Discontinuity at a local level exists when different intuitive theories are constructed with concepts that are incommensurable (Carey, 2009). These intuitive theories are articulated by the concepts they contain, which are also associated with word or signs. Hence, each conceptual system that expresses an intuitive theory has an associated language system. Through a comparison of language systems, Carey argues it is possible, using a method originally developed

by Kitcher (1978, 1988), to identify when concepts in different conceptual systems are incommensurate.

Carey suggests concepts are incommensurate when there is a '*huh'* moment when an individual with one conceptual system encounters another incompatible conceptual system. This '*huh'* moment exists because the first language system (LS1) cannot be translated into the second language system (LS2) without what Carey calls a translator's gloss. This gloss is required because the LS1 system does not map onto the same conceptual system as the LS2 system. This could be for a several reasons. For example, concepts could have been added, deleted, differentiated, or coalesced. Carey illustrates this argument with an example of from the historic development of thermodynamics and an example from children's development of their theory of matter. These examples are described in the following sections.

2.5.2.2.1 Discontinuity in Thermodynamics Concepts

Black's Theory of Thermodynamics was the first theory to include the terms *heat* and *temperature* that mapped onto two distinctly different concepts (CS2 in this example) that represent two different physical entities (Carey, 2009). Heat is an extensive quantity that is associated with the amount of energy transferred between thermodynamics systems. Hence, the amount of heat in two cups of water is the sum of the individual amounts of heat in each cup. Whereas temperature is an intensive quantity that is defined as a property of a body that determines whether there will be a net flow of heat into or out of the body from a neighbouring body or region. Therefore, heat and temperature are related. However, unlike heat, when two cups of water with difference temperatures are added together, the temperature of the mixture is not the sum of the individual temperatures.

Through a historic analysis of the Academy of Florence theories of thermodynamics, Wiser and Carey (1983) demonstrate that term *heat* is used in conceptual systems (CS1 in this example) that existed prior to Black's thermodynamics. However, the term *temperature* is not. Moreover, the term heat is often used ways that is consistent with aspects of temperature, as an intensive value, as well as with a CS2 interpretation of heat, as a cause of change. However, heat is used as a single variable in explaining mechanical phenomena. This undifferentiated concept of heat/temperature can only play a coherent role in a theory of thermodynamics because these theories also include the concept of degree of cold, which is absent from modern theories of thermodynamics.

This historic example illustrates how language can identify incommensurability in theories. First, the term *heat* in LS1 is associated with an undifferentiated concept of temperature/heat in CS1. This undifferentiated concept becomes two separate concepts in CS2 associated with two distinct physical entities, and the separate terms of *heat* and *temperature*. Second, the term *cold* in LS1 is associated with the concept 'degree of cold' in CS1. This concept is absent in CS2, and the term *cold* is now associated the concept of 'absence of heat'.

What has been described in the previous paragraph is the translator's gloss that demonstrates that a discontinuity exists in the concepts that form CS1 and CS2. This discontinuity exists because there is not a simple mapping between LS1 and a shared conceptual system that can also easily be mapped onto LS2. Or put another way, an LS1 term cannot simply be swapped for an LS2 term. To understand LS2, a new conceptual system, CS2, is required. Hence, CS1 and CS2 contain concepts that are incommensurate.

Carey goes on to present empirical evidence that these discontinuities in theories exist in conceptual development as well as development of scientific theories. These discontinuities occur as children develop an adult theory of matter.

2.5.2.2.2 Discontinuity in Theory of Matter Concepts

Carey (2009) identifies Piaget's well-known observation that preschool children are 'non-conservers' as an example of a conceptual discontinuity associated with children's and adults' theory of matter. Piaget and Inhelder (1974) observed that if a preschool child is given two balls made of the same amount of clay, they will agree that the balls weigh the same and take up the same space. However, if one of the balls is flattened into a disc, they will often claim the ball has more clay, takes up more space or weighs more. Hence, they are non-conservers.

Piaget explains this phenomenon by arguing that children are incapable of the same level of logical thinking as adults, but also that concepts in children's and adults' intuitive theories of matter are incommensurate. The idea that children and adults have significant differences in their ability to think logically has largely been discredited (see Carey (1985a, 1985b) for reviews of the criticisms of Piaget's stages of logical development). However, Carey agrees with Piaget that children's inability to conserve is due to children having a different conceptual system, CS1, associated with matter compared with adults, CS2. Piaget argued CS1 is characterised with an undifferentiated concept of size/weight and the concept of density was absent. However, Carey presents evidence from her own empirical studies that dispute this conclusion (Smith, Carey, &

Wiser, 1985). Instead, Carey argues that rather than density being absent from CS1, density is undifferentiated from weight.

Like 'degree of heat', Carey suggests that the undifferentiated concept of weight/density operates in CS1 as a 'degree of heaviness'. Again, understanding what is meant by 'degree of heaviness' is the translator's gloss that indicates conceptual discontinuity. Carey argues this concept is a relative quantity which is associated with an object, rather than a substance that is neither systematically extensive nor intensive. In contrast, in CS2 weight is an extensive quantity that is associated with substance. Density is also associated with substances but is intensive.

Children use degree of heaviness to judge how objects feel. For instance, a bag of rice feels heavy relative to a grain of rice. In this example, degree of heaviness corresponds with weight. However, this is not always the case. For example, heaviness can be personal in the sense that one person might find a book harder to lift than another. Furthermore, the undifferentiated weight/density or degree of heaviness concepts allows children to make apparently nonsensical judgements about objects. For instance, a popcorn kennel becomes lighter when it is popped or an object will eventually reach a point when it has negligible weight in everyday terms if it is chopped up into small enough pieces (Carey, 1991).

Using degree of heaviness as a relative quantity associated with objects also leads to children to make decisions about the nature of entities that are incommensurate with adults' decisions. For example, although even very young children can distinguish between real and unreal entities (based on perception), they do not fully distinguish between real material and immaterial entities (Carey, 1991; Smith, 2007; Smith, Maclin, Grosslight, & Davis, 1997). For instance, heat is seen as material, whilst powders, liquids and gases can be classed as immaterial. Furthermore, typically, children refer to air as made of 'nothing'. These examples demonstrates that children do not necessarily see weight as being a necessary property of matter.

Carey summaries the differences between a child (CS1) and adult's (CS2) intuitive theory of matter as:

- **Differentiation**

  CS1: undifferentiated concept of weight/density

  CS2: differentiated concept of weight (extensive, additive) and density (intensive, weight per unit volume).

- **Coalescence**

  CS1: liquids, gases seen as not matter, whilst other entities (i.e., heat) are seen as matter.

  CS2: Liquids and gases are seen as matter.

- **Change in conceptual role (conceptual core)**

  CS1: matter/substance decided by perceptual access and some causal interaction. Weight is not necessarily a property of all matter.

  CS2: Having weight and occupying spaces is a necessary property of matter.

This summary illustrates how several changes in individual concepts needs to occur in a coordinated way to enable a child to make the transition between CS1 and CS2. This is an important feature of the model of conceptual change the TOOC promotes. The next section describes the Quinian bootstrapping mechanism Carey suggests coordinates these changes and enables the conceptual discontinuities that have been described in this section to be overcome.

2.5.3 Quinian Bootstrapping

Carey's thorough examination of conceptual discontinuity demonstrates that conceptual development involves jumps in individuals' ability to represent the world around them by inventing new concepts. These new concepts are either more representationally powerful (i.e., language based rather than iconic, which enables infinite number representation rather than just up to four) or they play a different conceptual role in the theories that define them, which make them incommensurate. Hence, learning involves not just knowledge acquisition but conceptual change. As Carey points out, this presents a major issue for any learning mechanism that is based on a comparison between CS1 and CS2.

> *Characterising the learning mechanisms that underlie conceptual change is a formidable explanatory challenge. Many classes of learning mechanisms that underlie the major share of knowledge acquisition consist of selection or concatenation over an existing conceptual base. These include hypothesis testing, parameter setting, correlation detection, and many others. Additional learning mechanisms, of some other sort, must be implicated in conceptual change, for the conceptual system that is their output cannot be expressed in the conceptual base that is their input.*
>
> Carey (2009, p. 365)

Fodor reacts to this problem by arguing for conceptual continuity (as has been explained in section 2.4.3). However, the empirical evidence described section 2.5.2 has largely disproved this conclusion. So, this leaves Carey with the problem of finding an alternative learning mechanism for creating new conceptual resources that does not involve hypothesis testing. Carey's alternative learning mechanism and solution to this problem is Quinian bootstrapping, which is the third thesis in the TOOC.

The Quinian bootstrapping process begins when an individual connects a mental representation with a newly coined or learned external symbol or sign, usually a word. Initially, the mental representation associated with the new sign is far from a fully developed concept. In this sense, the sign is initially a 'placeholder', whose meaning is defined by its set of connections or interrelations between other signs. This set of interrelations in the placeholder, or sign, structure guide modelling activities (e.g., analogy monitoring and construction, thought experiments and inductive inference) that allow an individual to develop a full understanding of the concept the placeholder sign represents. Hence, conceptual development begins, and is coordinated, by the connections between signs (usually words). This is the key theoretical idea that runs throughout this thesis. The two examples below illustrate how the process of Quinian bootstrapping operates in childhood conceptual development, forming the concept of natural number and weight and density in an intuitive theory of matter. These are two of the conceptual discontinuities identified by Carey and described in sections 2.5.2.1 and 2.5.2.2.

2.5.3.1.1 Quinian Bootstrapping for Form the Concept of Natural Number

A common experience for anyone who works with or is familiar with young children is teaching them to count by reciting the count list (one, two, three, four, etc…). Carey (2009) identifies this count list as the first phase in bootstrapping an understanding of natural number. It is the required placeholder structure. Initially, the meaning of *one*, *two*, *three* is limited to the relationship between these signs. However, as the bootstrapping process proceeds, a child will begin to map the meaning of one, two, three and four onto the parallel individuation representations of these numbers, which is the core concept of number, through analogy. Once this has taken place, the order of one, two, three in the placeholder structure allows a child to identify that one, two or two, three involves the additional of 'one more'. Now the child can make an 'inductive leap' and infer that five, six, or ten, eleven, also indicates 'one more'. This is how a child learns to represent natural number (CS2), the more powerful representation resources, from core cognition (CS1). Importantly, the placeholder structure has organised the initial conceptual resources (i.e., the core concept of number in terms of parallel individuation)

and the learning processes (analogy, comparison, induction) available to the child to enable them to form a new conceptual representation.

2.5.3.1.2 Quinian Bootstrapping to Form Concepts of Density and Weight in an Intuitive Theory of Matter

Quinian bootstrapping is also the mechanism that drives changes in children's intuitive theories of matter. Developing differentiated concepts of weight and density is central to the development of children's intuitive theory of matter to an adult theory of matter (see section 2.5.2.2.2). Carey uses the teaching sequence developed by Smith (2007) to identify possible stages in the Quinian bootstrapping that achieves this differentiation between weight and density, and also coordinates the required simultaneous changes in conceptual resources outlined at the end of section 2.5.2.2.2 (e.g., liquids and gases are viewed as material).

Carey argues the differentiating of weight and density begins with the introduction of the formula $D=W/V$ [2]acting as a placeholder structure. This structure introduces a new mental sign of density and signals that this sign is different from weight. Carey argues this initial placeholder structure organises several modelling processes, thought experiments, and analogical mapping process in the teaching sequence that enable the full conceptual content for these weight and density signs to develop. In the teaching sequence, Carey also suggests a dot per box diagram is a placeholder structure that organises learning activities to emphasise weight as an extensive quantity and density as an intensive quantity.

2.5.3.1.3 Summary of the Quinian Bootstrapping Examples and gap in knowledge

In both examples above, Carey presents Quinian bootstrapping as a plausible mechanism for conceptual development that does not rely on hypothesis testing and, therefore, able to resolve the issues Fodor raises with conceptual development based solely on this type of learning. Hence, a precise understanding of the bootstrapping process is of great interest in conceptual change research as it could be a key process that drives conceptual change. This knowledge could be part of the basic knowledge needed to create detailed models of conceptual change, that can be empirical tested, which are currently not defined (Amin & Levrini, 2017b; Sherin, 2017a).

In the example of the development of the concept of natural number, the bootstrapping process is well specified and substantiated with empirical evidence. For instance, Carey first

[2] Carey uses mass and weight interchangeably in this example. Hence, D=W/V, rather than D=M/V.

demonstrates that core cognition that uses parallel individuation can only represent number up to four. She then presents evidence that children can memorise the placeholder structure and, importantly, that they have this structure before they are able to represent natural number. Therefore, placeholder structure exists before a full conceptual understanding, as the TOOC predicts. This detailed account of the bootstrapping process and the empirical evidence Carey presents to support this account relies on knowledge of core cognition, the antecedent conceptual system, but also the precise identification of the placeholder structure (i.e., the count sequence) that organises the learning processes.

The bootstrapping mechanism for the transition from CS1 to CS2 in children's theory of matter is not presented in the same detail as the bootstrapping of rational number and with the same level of empirical evidence to support the process. The studies by Smith (2007) and Wiser and Amin (2007) are cited by Carey as evidence for her theory. However, both these studies present evidence that teaching sequence and activities that are developed to promote Quinian bootstrapping are more successful in enabling children to develop an adult theory of matter. Hence, the evidence from these studies is circumstantial.

Several empirical studies not cited by Carey also provide evidence that language development plays an important role in conceptual development. For instance, Brookes and Etkina (2009) used linguistic analysis to demonstrate that experimenting with the meaning of force plays a role in the development of both historic theories of mechanics and students understanding of this topic. Similarly, Mortimer and Scott (2003) observed that experienced and highly skilled science teachers used a variety of pedagogical methods (i.e., types of questioning and patterns of student/teaching interaction) to develop their students' understanding of key scientific terminology for a given topic (e.g., rusting or particle). However, none of these studies directly measure any placeholder sign structure for individuals or groups of students as they attempt to learn scientific concepts or compare sign structure for individuals with different conceptual understanding. Moreover, they do not attempt to determine whether this placeholder sign structure precedes conceptual understanding, which would be direct evidence to support the TOOC in the case of intuitive theories. Hence, again, they provide circumstantial rather than direct evidence for the Quinian bootstrapping of intuitive theories.

In contrast, Carey's carefully described example of the development of the concept of natural number demonstrates that a clear picture of placeholder sign structure (i.e., the count sequence) provides a framework to study the Quinian bootstrapping process, and conceptual development. This suggests the bootstrapping mechanism that develops concept in intuitive

theories could be described with more precision if more was known about the placeholder structures that organises the required learning processes. Moreover, knowledge of this structure could guide the search for evidence that would support this view of conceptual development, as was the case with knowledge of the count sequence in the bootstrapping of natural number. The next section looks at studies that have examined sign structure and discusses how knowledge of this sign structure might inform an understanding of how Quinian bootstrapping leads to the development of intuitive theories in scientific domains of knowledge.

2.6 Sign Networks

This section discusses empirical studies that have investigated placeholder or sign structure which could be associated with the process of Quinian bootstrapping and, therefore, the mechanism for conceptual development. Typically, these studies have used networks to as a tool to map this placeholder structure. In this thesis, these networks are referred to as sign networks. However, these networks have also been referred to as semantic networks or concept maps. A distinction between these types of networks will be discussed in section 2.6.2.2.

The aim of this discussion is to highlight some of the key findings from these studies that indicate what a placeholder structure might look like. However, these findings emerge from analyses that draw heavily upon the theoretical developments in network science or graph theory. Hence, the next section introduces network science and the key ideas from this discipline that enable the findings and implications of the empirical studies discussed in sections 2.6.2.1 , 2.6.2.2, and 2.6.2.3 to be understood.

2.6.1 Network Science

2.6.1.1 Network Construction

Networks, or graphs as they are often known in mathematical literature, are constructed with vertices (nodes) and edges (links) between these vertices (Barabási & Pósfai, 2016; Caldarelli, 2007; Newman, 2018). For continuity and clarity, network will continue to be used rather than graph.

2.6.1.1.1 Vertices and Edges as Basic Units of Analysis in a Network

The vertex and edges (also known as nodes and connections) represent the basic units of analysis in any network and, therefore, determine the level of analysis that can be achieved with the network. For instance, a social network could be constructed to show collaboration between teachers in a school using teachers as vertices and connecting these vertices if teachers have shared resources or lesson plans. This network could be used to analyse the structure of individual collaborations within a school community. However, teachers could be collated, or merged, according to which department they work in (e.g., Science, Mathematics, English) and these departments could form the vertices in a network, again with an edge between the departments if teachers have collaborated. In this case, this network could only be used to analyse the structure of collaborations between school departments, but not individual teachers.

Hence, merging teachers into departments, and using these departments as vertices, creates a less detailed network, which restricts the level of analysis possible with the network.

However, it is not simply the case that vertices should always be defined to achieve the finest grain possible (e.g., teachers rather than department). Rather, any decision about how to define vertices, and indeed edges between these vertices, should reflect the aim of the study in question (Newman, 2018). For example, most studies that analyse transport networks use geographical locations as vertices and routes between these locations as edges (e.g., Bagler, 2008). However, Sen et al. (2003) in a study of the Indian rail network argue that what matter most to passengers is whether they have to change trains during their journey, rather than the number of stops a train makes. Hence, they draw edges between two locations (vertices) only if there is a direct train between these locations. Therefore, in their analysis of this network, the distance between two vertices (i.e., the number of edges between these vertices) represents the number of changes of train required to make the journey, rather than the number of stops. Hence, the network design has enabled a much more useful analysis given the aim of the study (Newman, 2018).

In a sign network, individual words are usually the vertices. Edges between words can be constructed in several ways. For instance, Cachapuz and Maskill (1987) draw an edge between words if the relatedness coefficient (Garskof & Houston, 1963) between two words is greater than a certain threshold. In contrast, Morais et al. (2013) draw edges between stimulus and response words in a word-association task. Furthermore, some studies ask participants to use their knowledge of a particular area of domain to specify vertices and the edges between these vertices. For instance, when creating their networks, Koponen and Nousiainen (2018b) require participants to identify vertices as concepts, laws, models, or experiments, and only draw an edge between these vertices if they were able to identify the action or procedure that connects them. Therefore, these decisions reflect the depth of conceptual understanding these networks aim to measure. This depth will become important when considering whether studies are, in fact, measuring a placeholder sign structure. This distinction is discussed in greater depth in sections 2.6.2 and 3.3.3.2.

2.6.1.1.2 Directed and Undirected Networks

The technique of drawing edges between stimulus and response word used by Morais et al. (2013) also demonstrates how a network can be directed or undirected. For instance, a directed network would show a distinction between an in-going edge, which shows a word is a response to a stimulus word, and an out-going edge, which shows that a word is a stimulus word

connecting to a response word. An undirected network would not make this distinction. Figure 2.8 illustrates this difference between a directed and undirected sign network.

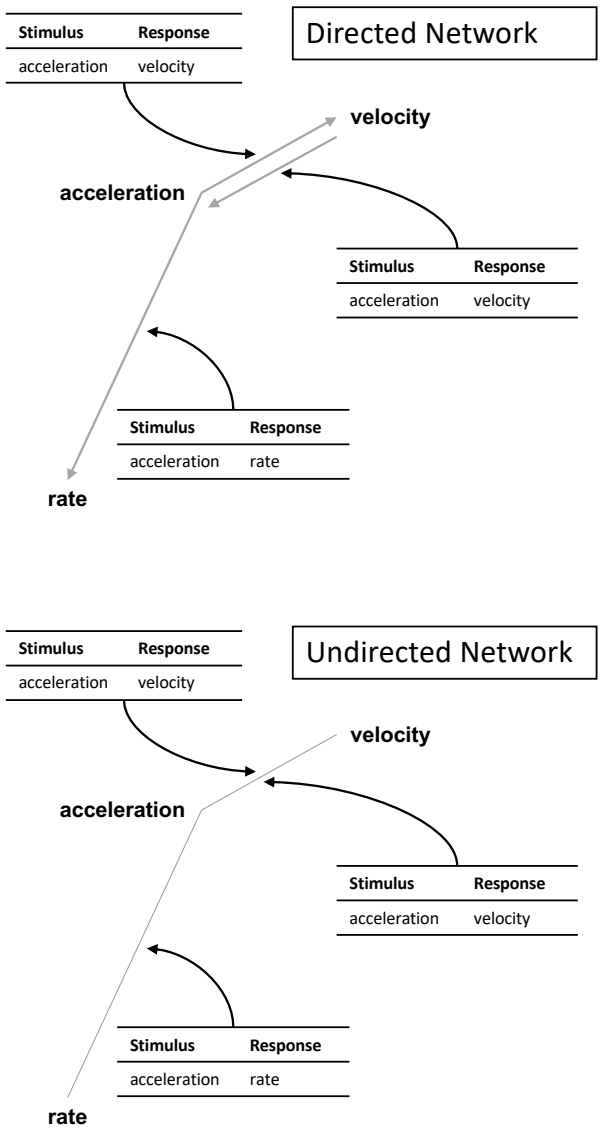


Figure 2.8 The difference between a directed and undirected sign network.

A network can be constructed from a complete knowledge of the vertices and edges that exist in that network. For instance, in social network for a school, all the students and teachers (the vertices) would be known, as would the connections between students in the same class and

the class teacher (the edges). However, this is not always the case. For example, in the recent COVID-19 pandemic, it was very important to identify the social contacts for an infected individual to create a network of an individuals who may have been infected, but it is clearly not always possible to identify all social contacts, especially if the infected person has travelled by train or visited a busy location. Hence, the social network created for this person represents a sample of the complete network. There are several techniques that can be used to sample networks. Snowball sampling is a common technique, which again, is familiar from most people's experience of the COVID pandemic because this was the technique used to identify individual social contacts. In this example, the sampling begins with the infected person being asked to name their social contacts. These contacts become the initial vertices and edges in the sample network. However, these sampled vertices are then asked to name their social contacts, which become new vertices and edges, and so on. Hence, the sample snowballs from the original vertices. Snowball sampling is used in this thesis. It is an effective technique for finding vertices and edges in a network but it does have limitations (Newman, 2018). These are discussed in section 3.3.

2.6.1.2 Basic Network Parameters, Key Theoretical Ideas and Measures

This section introduces the basic parameters and key theoretical ideas that are used to describe and analyse a network and, hence, form the foundations for any quantitative study of networks (see sections 2.6.1.2.1, 2.6.1.2.2, and 2.6.1.2.3). Moreover, these parameters and ideas have been developed into measures that capture interesting features of network structure (e.g., centrality and clustering). These measures are important for making sense of large networks, which can be bewildering to visually analyse, and for moving beyond a qualitative description of a network (Newman, 2018). They are described in sections 2.6.1.2.4 and 2.6.1.2.5.

2.6.1.2.1 Network Order, Size, Density and Sparseness

Once a network has been constructed, there are two basic parameters that quantitatively describe a network (Barabási & Pósfai, 2016; Caldarelli, 2007; Newman, 2018). These are:

- Network order (n) is the number of vertices in a network.
- Network size (m) is the number of edges in a network.

In a complete network, every vertex is connected to every other vertex. In this case, the network size is a maximum. For a directed network with n vertices, the maximum number of possible edges can be calculated by (Newman, 2018):

$$m_{max} = n(n-1)$$

The connectance or density (ρ) of a network is defined as the fraction of the maximum possible edges that are actually present in the network (Newman, 2018).

$$\rho = \frac{m}{n(n-1)} = \frac{m}{m_{max}}$$

A network is said to be dense if ρ remains non-zero as the network gets larger and n increases (Newman, 2018). In contrast, a network is sparse if $\rho \to 0$ in the limit of large n. However, as Newman (2018) points out, this definition only applies if n can tend to ∞. Hence, in practice, a network is usually said to be sparse if ρ is small, which implies that most of the possible edges in a network are not present.

2.6.1.2.2 Degree

The *degree* of network vertex (k) is one of the simplest, yet useful and widely used network concepts (Newman, 2018). Degree plays a central role in this thesis. In an undirected network, the degree of a vertex is the number of edges that connect to that vertex. In a directed network, a distinction can be made between the number of incoming edges (in-degree, k_{in}) and the number of outgoing edges (out-degree, k_{out}) that connect to a vertex.

The mean vertex degree for an undirected network is given by (Barabási & Pósfai, 2016):

$$< \mathrm{k} >= \frac{1}{n}\sum_{i=1}^{n} k_i$$

2.6.1.2.3 Paths and Distances in Networks

In any physical system, the distance between items of interest is a key parameter. In a network, distance is measured by *path length* (Barabási & Pósfai, 2016), which are in turn described *paths* and *walks* (Newman, 2018). A walk in a network is a move between any sequence of vertices that are connected by an edge. In a walk, it is possible to revisit a node by moving along an edge more than once. Walks that do not retrace themselves are called *paths*. The length of a path or *path length* is the number of edges that must be crossed between two vertices and, hence, separate the vertices.

- The shortest path (d) between any two vertices is the path that connects these vertices with the fewest number of edges (Barabási & Pósfai, 2016).

- The diameter (D) of the network is the maximum shortest path (Barabási & Pósfai, 2016).
- The average path length (L) is the average distance between all pair of vertices in a network. The average path length in a directed network is given by (Barabási & Pósfai, 2016):

$$L = \frac{1}{n(n-1)} \sum_{\substack{i,j=1,n \\ i \neq j}} d_{i,j}$$

Shortest paths and path lengths are key ideas that define several network phenomena (Newman, 2018). For instance, low shortest path lengths characterise the small-world effect that is present in many networks (Newman, 2018). The small world effect was first made famous by the experiments of the psychologist Stanley Milgram, who found that residents of the USA selected at random were able to forward a letter to a named target person using on average six intermediaries (Milgram, 1967). Hence, it is a small world! Similarly, Albert, Jeong, and Barabási (1999) found that any two webpages on the world-wide-web could be reached through on average nineteen other webpages.

There does not necessarily have to be a path that exists between all vertices in network. Hence, networks can contain two or more separate parts that are disconnected from each other. These parts are known as components (Newman, 2018).

2.6.1.2.4 Network Centrality

The basic parameters and key theoretical ideas described in the previous sections have been used to define several measures that capture the salient features of network structures. One of these features is centrality. Centrality aims to identify the most important vertices in a network (Newman, 2018). However, there are different definitions of importance, and this has led to several difference measures of centrality. In this section, two measures of centrality are introduced.

- Degree centrality is a simple, but effective, measure of centrality. It defines the most important vertices in the network as the most connected vertices. Hence, it is measured by the vertex degree.

- Betweenness centrality draws upon the idea of paths. It defines important vertices as vertices that lie on many paths between other vertices. Figure 2.9 illustrates the notion

of betweenness centrality and its usefulness as contrasting measure of centrality to degree centrality. In this figure, vertex A has a high betweenness centrality as it sits on a path between groups 1 and 2, but a low degree (k=2). Figure 2.9 also highlights the crucial role betweenness centrality can have in identifying vertices that are crucial for connecting different groups or communities within networks.

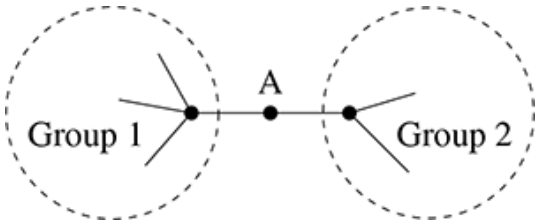


Figure 2.9 The notion of betweenness centrality in a network. Vertex A has a low degree but is 'central' in connecting group 1 and group 2 vertices in the network. Taken from Newman (2018, p. 175).

2.6.1.2.5 Groups of Vertices

Most networks naturally split into groups (Newman, 2018). For instance, a social network within a school will divide into groups of friends, classes, teachers within a department and so on. Therefore, the definition and identification of groups within networks has been a major focus of network theory. This section introduces the idea of clique and then uses this to define an important measures of network grouping called the clustering coefficient.

A clique is present in an undirected network if every member of a set of vertices is connected by an edge to every other (Newman, 2018). Figure 2.10 show a clique with three vertices. The clique with three vertices defines a perfect transitive relationship. In this case, if vertex A is connected to vertex B, and vertex B is connected to vertex C, then vertex A is also connected to vertex C. Another way to define this relationship is in terms of a path of length two between ABC (Figure 2.10). This path forms a clique, and defines a perfect transitive relationship, if a third edge exists between A and C. The path is then called a closed path. The clustering coefficient is used to define the level of transitivity in a network and is given by:

$$C = \frac{number\ of\ closed\ paths\ of\ length\ 2}{number\ of\ paths\ of\ length\ 2}$$

C = 1 implies perfect transitivity and a network of cliques. In contrast, C = 0 implies no closed paths of length 2.

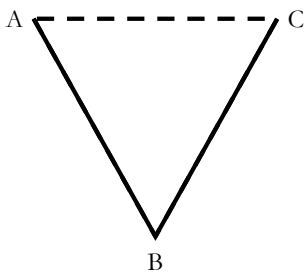


Figure 2.10 A clique in a network. The path between ABC is said to be closed if a third edge between A and C is present. Taken from Newman (2018, p. 183).

The clustering coefficient can be interpreted as a measure of the probability any two randomly chosen vertices in the network will share a connection with a third neighbouring vertex (Watts & Strogatz, 1998). For instance, in a social network, which typically have clustering coefficient of 0.2, this implies that two friends are more likely to have a shared friendship (or connection) with a third person than if they did not know each other (Fortunato, 2010).

The clustering coefficient for a vertex i with degree k_i is defined as:

$$C_i = \frac{2E_i}{k_i(k_i - 1)}$$

E_i represent the number of edges between the k_i neighbours of vertex i (Barabási & Pósfai, 2016, p. 64). The average clustering coefficient over all N nodes is given by:

$$\langle C \rangle = \frac{1}{N}\sum_{i=1}^{N} C_i$$

2.6.1.3 Models of Network Structure

The measures described in the previous section reduce the complexity of a network to numerical values that can be easily understood and compared (Newman, 2018). However, these measures also provide reference points for models of network structure to explain (Barabási & Pósfai, 2016; Newman, 2018). These models provide a more detailed picture of network structures and often incorporate theories of network development (Barabási & Albert, 1999; Barabási & Pósfai, 2016; Caldarelli, 2007; Newman, 2018) . Hence, they facilitate a greater understanding of a network. This section explores and contrasts two important models of network structure and growth.

Random graph models are one type of simplest but most important models of network formation (Barabási, 2002; Barabási & Pósfai, 2016). As the name suggests, the edges (or connections) between vertices are made randomly in this type of model (Barabási & Pósfai, 2016; Newman, 2018). Random graph models are of particular interest in network analysis, which Barabási and Pósfai (2016) point out can be confusing given that it is unlikely that the connection between vertices in real networks are formed randomly. For instance, it does not seem plausible that a word (or sign) randomly connects to any other word in an individual's vocabulary. However, random models are key tools in network analysis because they provide a way of distinguishing between properties of a network that can arise by chance, versus other properties that require more explanation (Barabási & Pósfai, 2016). For instance, the short paths lengths that characterise many real networks, including social networks (Milgram, 1967) and sign networks (e.g., Cancho & Sole, 2001), are an example of a network property that can arise from random connections, and no further explanation is required (Barabási & Pósfai, 2016). However, this is not always the case. For instance, random models predict clustering coefficients that are lower than those that are measured in many real networks. Hence, the clustering coefficient measures a property of networks that is a 'signature' of greater order in the network structure that requires a deeper explanation than simply randomly connecting vertices (Barabási & Pósfai, 2016). The idea of a 'signature' feature in network structure that requires explanation is a key idea in this thesis.

The shape of the vertex degree distribution is another key 'signature' property that can indicate greater order in a network structure (Barabási & Pósfai, 2016). The vertex degree distribution is defined as the fraction of vertices in a network with degree k (Barabási & Pósfai, 2016; Caldarelli, 2007). In the limit of large n, the degree-distribution is given by the expression:

$$P(k) = \frac{n_k}{n}$$

where n is the total number of vertices and n_k is the number of vertices with degree k (Barabási & Pósfai, 2016). P(k) can be understood as the probability that a randomly chosen vertex in the network will have a degree, k (Newman, 2018). In a network with a random structure, the degree distribution will follow a Poisson distribution (Barabási & Pósfai, 2016). In contrast, the degree distribution for many real networks seem to follow a power-law, which suggests they have a scale-free structure (e.g., Broder et al., 2000).

The shapes and particularly the different tails of the Poisson and Power-law distributions indicate a key difference between a network with a random and scale-free structure (Barabási & Pósfai, 2016) (see Figure 2.11). A Poisson distribution has a clear peak, which corresponds to the average number of edges per vertex (average degree) and a narrow range of possible values for the number of edges per vertex. The average degree acts as the 'scale' of the random network. In contrast, the power-law distribution indicates most vertices only have a few edges. However, it is possible to find a small number of vertices with a very large number of edges. Hence, the degree distributions are right-skewed and are referred to as 'heavy-tailed'. The vertices in the tail of these distributions are known as hubs. There is also a large range for the average degree. This implies that if a vertex is chosen at random, the selected vertex degree could be small or very large, in contrast with a random network, where the degree of the selected vertex is likely to have a small range centred around the average degree or scale of the network. Hence, a network that has a power-law degree distribution does not have a characteristic scale – it is scale-free.

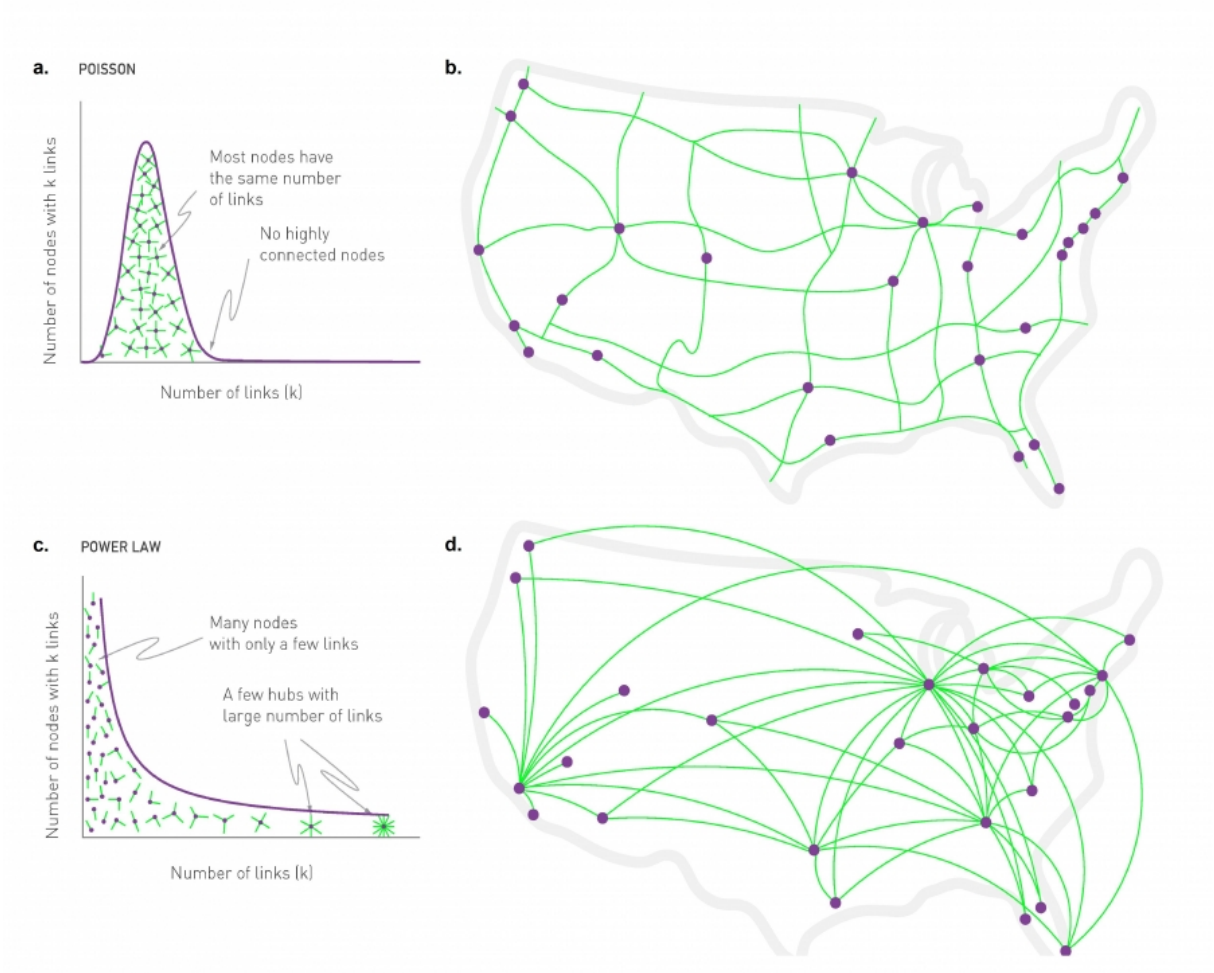


Figure 2.11 The difference between a random and a scale-free network structure illustrated by comparing road and air connection in the USA (taken from (Barabási & Pósfai, 2016))

Like a high clustering coefficient, a scale-free structure, and particularly the presence of hubs, is a 'signature' property of a network that requires greater explanation than random connections between vertices (Barabási & Albert, 1999; Barabási & Pósfai, 2016). Put simply, an explanation is required as to why some vertices become very highly connected (i.e., become hubs), whilst most do not (Barabási, 2002). Most models that provide this explanation utilise two assumptions that are absent from random models: network growth and preferential attachment (Barabási & Albert, 1999; Barabási & Pósfai, 2016). Random models assume a fixed number of vertices (Barabási & Pósfai, 2016; Newman, 2010). In contrast, models that explain scale-free structures begin with a small number of vertices and 'grow' by adding new vertices to the network (Barabási & Albert, 1999; Barabási & Pósfai, 2016). Moreover, when new vertices are added they are more likely to connect to existing vertices with higher numbers of edges, rather than with equal probability (i.e., at random). Hence, they preferentially attach to vertices with a higher degree.

This discussion has so far focused on random models of networks or models that are specifically designed to generate scale-free structure. This focus reflects the level of interest that has been directed towards establishing whether real networks have a scale-free structure because of the interesting properties of these networks (i.e., their fractal geometry) (Caldarelli, 2007). However, framing a discussion of network models in this way risks giving the impression of a false dichotomy, which is not supported by empirical studies of networks. These evidence from these studies show that the claim that most networks have a scale-free structure is not conclusive (Broido & Clauset, 2019) and is controversial (Voitalov et al., 2019). However, the claim that the degree distributions of most real networks are heavy tailed, and therefore contain hubs, is widely accepted (Newman, 2018).

There are several models that can explain degree distributions that contain hubs. For instance, the configuration model, which specifies the degree for each vertex in a network, but then allows these vertices to connect randomly (Newman, 2018). Or a vertex copying model (Kleinberg, Kumar, Raghavan, Rajagopalan, & Tomkins, 1999), where when new vertex enters a network, a sub-section of the connections between vertices in the existing network are 'copied' onto the new vertex. These examples illustrate that the true dichotomy in network models is whether a network property can be explained by purely random connections between vertices, or whether additional mechanisms or assumptions are needed (i.e., preferential attachment, or specifying the degree for the vertices). Hence, network models are an important analytical tool that allow ideas about network growth to be tested and compared with the empirical properties of networks.

2.6.2 Sign Network Studies

Several empirical studies have investigated what could be said to be sign networks. However, there are important differences in how these studies measure, construct and analyse networks that affects what aspects of the Quinian bootstrapping process, and particularly the placeholder structure, these studies investigate. These differences can be grouped according to the following criteria:

1. Target domain of knowledge.
2. Criteria for drawing edges between vertices.
3. Individual or group networks.
4. Analytical techniques.

Figure 2.12 shows network studies according to where they sit on a continuum between creating networks in a scientific or general domain of knowledge, and their criteria for selecting vertices and drawing edges between them. This second axis represents whether the network could be called a sign (loose condition for edges) or a concept network (stringent conditions for edges). This distinction defines the level of declarative knowledge that is required (see discussion in section 2.6.2.2). Figure 2.12 show the groups of studies that are discussed in each section below, but also identifies the gap in the research, which defines part of this thesis' contribution to knowledge.

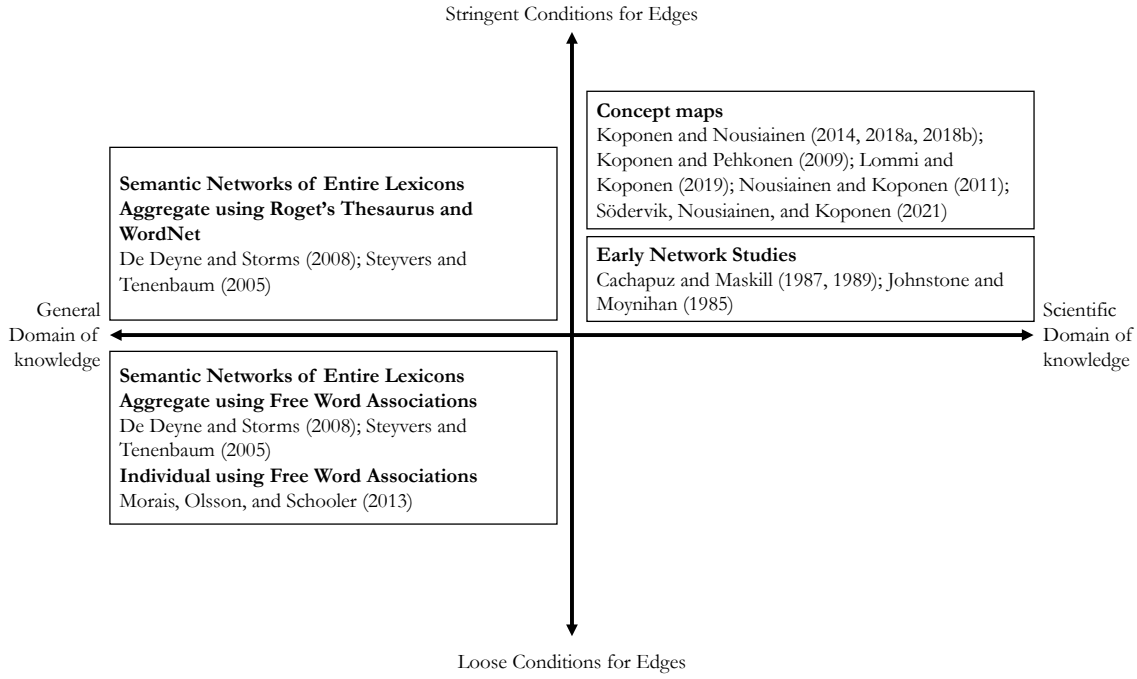


Figure 2.12 Previous studies of potential sign networks plotted according to the whether they target a general or scientific domain of knowledge and whether they apply stringent or loose condition to draw edges between vertices. Note the lack of studies that target a scientific domain of knowledge with looser conditions for drawing edges.

2.6.2.1 Early Empirical Studies of 'Sign' Networks

Cachapuz and Maskill (1987); Hovardas and Korfiatis (2006) and Nakiboğlu (2008) investigated the changes in the word-association given by students of various ages (from secondary school to university level) to stimulus words before and after instruction in different domains of scientific knowledge (e.g., collision theory, Le Chatelier's Principle). These studies all used free-association tasks that asked students to response to a stimulus word with the first words that come to mind rather than a considered response. Students were asked to give more than one word as a response (typically 10 words).

Analysis of the word-association in each of these studies was performed at a group level and all these analyses report changes in the words that are associated with stimulus words after instruction. Furthermore, Cachapuz and Maskill (1987) used a post-instruction performance test to split students studying a collision theory topic into high, middle and low attainment groups. The students in the higher attaining group made more new associations than the students in the middle and lower groups. Furthermore, these higher attaining students were more likely to identify a particular word (collision) with several of the stimulus words. Similarly, Hovardas and Korfiatis (2006) found that many of the new word-associations made by university students came from the ecology domain of scientific knowledge, which was being studied, after instruction. Moreover, the frequency of response words from other domains of knowledge decreased.

Cachapuz and Maskill (1987) and Nakiboğlu (2008) also investigated how the global structure of the connections between words changes by constructing semantic networks for groups of students before and after instruction. These sign networks were constructed using a relatedness coefficient (Garskof & Houston, 1963). This relatedness coefficient measures the similarity in meaning between two words by assuming that words with same meaning have identical word-associations. For instance, if a participant responded to *force* with *mass* and *acceleration* and *acceleration* with *mass* and *force*, the related coefficient would be one. A lower relatedness coefficient indicates that some associations are not shared.

In their networks, Cachapuz and Maskill (1987) and Nakiboğlu (2008) connect words that have an average relatedness coefficient above certain threshold values. Hence, a connection between two words in the network indicates that these words have a high number of shared word-associations within the group. The sign networks in both studies show 'islands' of connected words before tuition. These islands are connected after tuition. However, Cachapuz and Maskill (1987) find these islands remain in the sign networks they construct for their group of low-attainment students. Hence, these networks indicate that more words share the same word-associations after tuition and in higher attaining groups.

In a related study, Johnstone and Moynihan (1985) also used the relatedness coefficient to construct semantic networks for groups of 'poor' and 'good'[3] students (14-15 years old) studying chemistry. These networks for poor and good students also show differences in the

[3] This use of 'poor' and 'good' to describe students has been retained despite the biased nature of these terms to remain consistent with Johnstone and Moynihan (1985) original descriptions.

connections between words and in the total number of connections deemed to be correct by the researchers. Again, this supports the claim that individual associations between words and the overall structure of the sign network created by these associations may affect ability to think scientifically.

The findings of these studies support the general idea that students develop a sign structure when learning scientific concepts. However, what these studies say about the sign structure is limited for two reasons.

First, the relatedness coefficient is quite an unusual measure to use to draw edges between signs, as it connects words that share a high number of associations, rather than showing the structure of these associations. This limitation is recognised by Nakiboğlu (2008) who uses the frequencies of word association to define edges. In her study, an edge is formed between a response and stimulus word if the response word has been given more than 11 times by participants in the group. However, this is still quite a stringent criterion for forming edges and consequently these sign networks could appear less connected that they are.

Second, these studies do not use the analytical tools that have been developed as part of modern network science (Barabási & Pósfai, 2016; Newman, 2018) to analyse sign structure rigorously and systematically. Hence, they provide a largely qualitative insight into the role any sign structure might have in learning concepts. This qualitative analysis lacks a robust comparison of similarities and differences between networks and does not enable the development of any models of network growth that could provide further insight into sign network structure. This is not a limitation of the studies discussed in the next sections.

2.6.2.2 Scientific Sign Networks Analysed with Network Science Techniques

In contrast, empirical studies by Koponen and Pehkonen (2009), Koponen and Nousiainen (2014), Koponen and Nousiainen (2018a), Koponen and Nousiainen (2018b), and Lommi and Koponen (2019) have used analytical techniques informed by network science and graph theory to quantitatively investigate the structure and growth of individual sign (concept) networks for pre-service Physics teachers (studying Physics at university level) in specific scientific domains of knowledge (i.e., electromagnetism and the history of science). These studies have highlighted several salient features of these networks.

First, they are not hierarchical but web-like, which is inconsistent with the ontological view of scientific concepts (see section 2.3.3.4 and Chi and Slotta (1993)). Second, they have high

clustering coefficients, which indicate that if two signs or concepts are connected, they are more likely to share a connection with a third concept (as was outlined in 2.6.1.2.5). Koponen and Nousiainen (2018b) point that that these cliques and transitive patterns are a feature of several models of knowledge construction and processing (Goldwater & Schalk, 2016; Halford, Wilson, & Phillips, 2010; Kemp & Tenenbaum, 2008, 2009). Third, measures of betweenness centrality have identified key concepts in the network that lie on many paths between other concepts in the network (Koponen & Nousiainen, 2014). For this reason, these concepts play an important role in the flow of information in the network and, therefore, how other concepts relate to each other (Koponen & Nousiainen, 2018a, 2018b). Furthermore, Koponen and Nousiainen (2018a) found that individual's networks did not possess all the key concepts that were present in a group network created by collated the 12 individual networks in their study of electromagnetism. Hence, individuals do not necessarily possess all concepts required for complete knowledge of this domain. The degree distribution of the networks peaked at small values of degree (~3) but also contained some vertices with larger degrees (up to ~12 to 14) (Koponen & Nousiainen, 2018b). Finally, the networks are modular, with modules associated with phenomenally distinct but connected areas (e.g., electricity, magnetism, and electromagnetic induction).

The features of the networks identified by Koponen and Nousiainen (2018b), such as the high clustering coefficients and vertices with high degrees (i.e., hubs), are 'signature' features that indicate that the connections in the network are not being made randomly (see section 2.6.1.3). Hence, there are likely to be underlying processes that determine how these connections are made.

Koponen and Nousiainen (2018b) have developed a model of network growth that is able to replicate the structure of their measured individual networks and, therefore, explain the 'signature' features in the networks. This model is underpinned by three basics assumptions Koponen and Nousiainen (2018b) make about the learning process. These are:

1. A new concept entering the network will make connections that are typically within three to seven genealogical steps from a small number of ancestral concepts (from two to four).
2. Concepts are grouped according to phenomenological meaning (i.e., concepts related to electricity typically connect with other concepts related to electricity, and similarly for concepts related to magnetism). The number of genealogical steps determines whether

a concept is connected (three to seven steps) within or outside a module (more than ten).

3. Four basic linking motifs are used to connect a new concept to the network.

These basic assumptions are incorporated into their model of network growth by using a discrete gamma distribution to determine the number of genealogical steps when a new concept is introduced into the network. A probabilistic process is also used to select a linking motif for any new concept entering the network. Hence, this model is an example of a configuration model (Koponen & Nousiainen, 2018b).

Koponen and Nousiainen (2018b) compare the simulated networks created by their model with their measured networks. This comparison shows that their model can successfully replicate the structural features of the measured networks. Hence, Koponen and Nousiainen (2018b) demonstrate that models of network growth can provide a means of operationalising assumptions about the learning processes involved in conceptual development and, therefore, enabling these assumptions to be tested against empirical data (i.e., measured networks).

The studies described in this section have demonstrated the advantages of using quantitative measures to systematically compare the networks of novice and experts and identify differences in their structure, and how knowledge of these measures (e.g., high clustering coefficients) can then be used to develop and constrain models that explain the growth of these networks. These models can then inform our understanding of the mechanism of conceptual development. However, these concept networks are constructed using stringent constraints on types of vertices that are allowed and how edges between these vertices can be formed. In these networks, vertices can be concepts or quantities and the edges that connect these vertices are either experimental procedures (e.g., operational definition of concept or laws or demonstrations) or modelling procedures (e.g., simple definitions or logical deductions). By applying these constraints, the concept networks represent scientifically acceptable knowledge, rather than subjective beliefs.

However, TOOC specifically argues that a sign is initially associated with placeholder values before these placeholder values develop into fully formed concepts. Hence, a sign network is likely to be initially more developed than the fully formed concept network it will ultimately be associated with. Furthermore, this initial sign structure guides the development of the procedural rules that are the hallmark of a fully developed concept network. Hence, the TOOC predicts that a learner will recognise that there is a connection between signs before they can

fully explain why they are connected. Therefore, less stringent criteria (i.e., word associations) for forming edges are more appropriate to investigate sign, rather than concept, networks.

In this thesis, the strength of the criteria for defining vertices and edges guides whether a network is more likely to represent sign or underlying concepts (i.e., whether the network is a sign or concept network). The next section presents the findings from studies that have applied less stringent criteria to form connection between words. Hence, these studies are more likely to represent placeholder sign networks. However, these studies have focused on entire vocabularies, and therefore several domains of knowledge, rather than just a scientific domain.

2.6.2.3 Sign Networks for Complete Vocabularies

The studies in the previous section have investigated networks in scientific domains of knowledge. This section presents the findings of studies that have made sign networks for a general domain of knowledge and, therefore, the networks capture sign structure for entire lexicons. These studies either use free word-associations to define the edges between signs (De Deyne & Storms, 2008a; Morais et al., 2013; Steyvers & Tenenbaum, 2005) or considered connections between words in Roget's Thesaurus or Wordnet (Steyvers & Tenenbaum, 2005). Hence, these studies can be placed at both ends of the continuum between measuring placeholder (free word associations) and concept structure (wordnet, thesaurus).

In the word association networks, Steyvers and Tenenbaum (2005) and De Deyne and Storms (2008a) use English (Nelson, McEvoy, & Schreiber, 1999) and Dutch (De Deyne & Storms, 2008b) word associations norms respectively to create sign networks. Hence, these are group networks. However, Morais et al. (2013) argue these group or aggregate networks could distort aspects of the sign network structure. Hence, they develop a snowball sampling technique, which still uses word associations to identify vertices and edges, but measures sign networks for individuals. The networks Steyvers and Tenenbaum (2005) create using Roget's Thesaurus and Wordnet are group networks because the thesaurus and Wordnet are collaborative projects that aim to capture a population's or cultural understanding of language.

All these studies use techniques from network science and graph theory to analyse the sign networks they measure. The findings from these analyses are broadly similar. The networks are sparse, with densities between approximately 0.001 and 0.02, which indicates that on most vertices are only connected to a small fraction of other vertices. The networks are also characterised by short average path lengths ($3 \lesssim L \lesssim 7$). However, these average path lengths are consistent with the average path lengths that would be expected for random networks with

same order (number of vertices). In contrast, the clustering coefficients are higher ($0.1 \lesssim C \lesssim 0.3$) than would be expected for random networks of the same order ($0.001 \lesssim C \lesssim 0.005$). Hence, the clustering coefficients of these networks cannot be explained by a random network model.

Steyvers and Tenenbaum (2005) and De Deyne and Storms (2008a) claim the degree distributions for their networks (word association, thesaurus, and Wordnet) are characterised by power-laws and, hence, have a scale-free structure. This type of structure is characterised by most signs (or words) only connected to a limited number of other signs (i.e., one or two) (see section 2.6.1.3). However, a few signs are very highly connected, accounting for most of the connection in the network (Barabási & Pósfai, 2016; Newman, 2018). These very highly connected vertices – or hubs – are referred to as hub-words. Both Steyvers and Tenenbaum (2005) and De Deyne and Storms (2008a) suggest these hub-words are likely to important in some way to conceptual understanding.

Morais et al. (2013) criticises the analyses of Steyvers and Tenenbaum (2005) and De Deyne and Storms (2008a) for not creating individual sign networks but also for not using a statistical method (e.g., Clauset, Shalizi, and Newman (2009)) that is rigorous enough to robustly claim that the degree distributions are a power-law. Especially, as they do not try to use other distributions (e.g., log-normal, exponential, or power-law with exponential cut-off) to model the degree distribution data and compare these fits with the power-law fit. In their analysis they find a power-law with an exponential cut-off is the best model of the degree distributions of their individual networks. They also reanalyse the word association group networks initially created by Steyvers and Tenenbaum (2005) and De Deyne and Storms (2008a) and also find these are better modelled with an power-law, rather than with an exponential cut-off. Hence, the findings of the analysis of the degree distributions of the individual and group networks are consistent in the analysis carried out by Morais et al. (2013).

The analysis carried out Morais et al. (2013) raises doubts over whether the sign networks have a scale-free structure. However, both the power-law model and power-law model with exponential cut-off are heavy tailed distributions. Hence, the sign networks contain hub-words. But, in the case of the power law with exponential cut-off, this distribution is truncated. Therefore, there may be some mechanism limited the maximum number of connections hub-words can achieve (Morais et al., 2013). A model that can possibly explain the structure and growth of sign networks for whole vocabularies is explored in the next section.

2.6.2.3.1 A Model of Sign Network for Entire Vocabularies

High clustering coefficients (greater than those expected in a random network of the same order) and heavy-tailed degree distributions, which indicate the presence of hub-word, are signature features that suggest connections in these sign networks are not made randomly. In response to these observations, Steyvers and Tenenbaum (2005) have developed a model that explains the growth and structure of these sign networks, and the presence of hub-words in these networks. Steyvers and Tenenbaum (2005) originally designed their model to explains the scale-free structure they believed characterised their networks. Hence, the model specifically predicts a power-law degree distribution. The study by Morais et al. (2013) has raised questions about whether the degree distributions of sign network for entire lexicons are power-laws. If Morais et al. (2013) are correct, then the Steyvers and Tenenbaum (2005) model will need to be modified. However, this model is still presented here for two reasons. First, this model operationalises many of the assumptions about Quinian bootstrapping built into TOOC. Hence, it is an important analytical tool for studying the development of placeholder sign structure. Second, even if the degree distributions of sign networks are more appropriately modelled with alternative heavy tailed distributions (e.g., power-law with exponential cut-off, exponential, or log-normal distribution) rather than a power-law, Morais et al. (2013) suggest that these alternative distributions can still be modelled by building on, rather than rejecting, some of the basics assumption of the Steyvers and Tenenbaum (2005) model. Hence, this model is a starting point for trying to understand the formation of sign network structure in whole vocabularies.

The Steyvers and Tenenbaum (2005) model is based on growth and preferential attachment (Barabási & Albert, 1999; Dorogovtsev & Mendes, 2002). The sign network grows by adding new vertices (or word/sign) to the network. Each new sign has m edges it can connect to existing signs, where m is the initial number of signs in the network. A new sign initially forms an edge with any existing sign. However, the probability of connecting to an existing sign is proportional to the degree of the signs. Hence, a new sign preferentially attaches to existing signs. The remaining edges $(m-1)$ are formed with signs in the neighbourhood of the sign that was initially connected to the new sign. A model parameter called the utility determines the likelihood of a connection between each of the signs in this neighbourhood. In their model, Steyvers and Tenenbaum (2005) set the utility to 1 (i.e., all signs in a neighbourhood are equally likely to be connected) or equal to each word usage frequency.

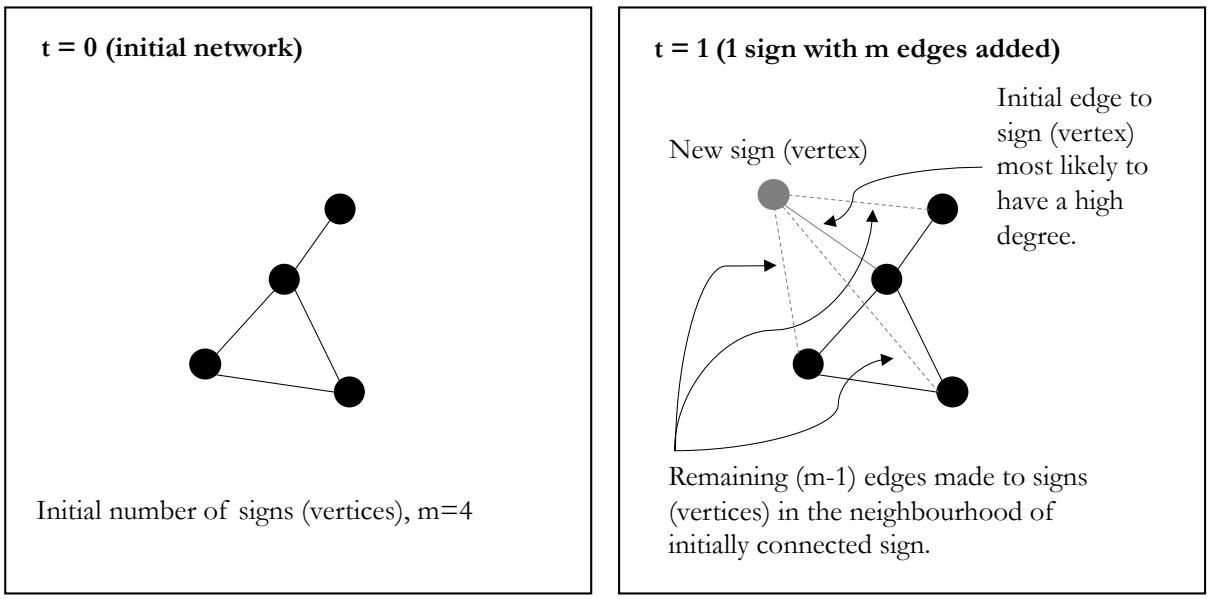


Figure 2.13 The preferential attachment model of sign network growth developed by Steyvers and Tenenbaum (2005).

Steyvers and Tenenbaum (2005) state that their model is underpinned by ideas about language and conceptual development through differentiation (e.g., Carey (1978, 1985b); E. V. Clark (1993)). This is illustrated by Figure 2.13, which shows that each time a new sign is added to the network, it differentiates the existing sign it initially connects to by changing the pattern of connections between this sign and its neighbours. More broadly, this model incorporates some of the key assumptions in Carey's theory of the origin of concepts (Carey, 2009). First, that conceptual development involves adding new signs to a sign network, which are initially placeholder values that do not link to a complete or fully understood concept. Second, the addition of these new signs changes the structure of that sign network, which then guides the development of new concepts. Moreover, each new sign connects to several existing signs, which incorporates Carey's notion of conceptual changes in initiate theories involve coordinated and simultaneous changes in several concepts (see the density example in section 2.5.2.2.2). However, the Steyvers and Tenenbaum (2005) model also suggests an interesting possibility by using preferential attachment. Namely, that the scale-free structure of the sign network is explained by the network 'growing' around key signs (or words), which become hub-words. Hence, it is possible that individuals with different conceptual understanding of mechanics (i.e., physicist and psychologists) might have different hub words because of the different levels of development in their sign networks. This idea is broadly consistent with the findings of Koponen and Nousiainen (2018b) that suggest concepts form in clusters, but there key concepts that are crucial for connecting different parts of the network. Furthermore, not

all learners in Koponen and Nousiainen (2018b) study included these concepts in their networks.

The insight provided by applying Carey's ideas through the Steyvers and Tenenbaum (2005) model informs a key aspect of this analysis. The Steyvers and Tenenbaum (2005) model specifies how the sign network needed by Carey's theory of the origin of concept might grow. A scale-free structure, and hub words, are a 'signatures' of this model version of sign network growth. Hence, as well as using key network measures to look at similarities and differences in sign networks, this the study in this thesis also aims to identify whether the sign networks in the physics and psychology groups have a scale-free structure and, hence, whether they contain hub works. However, the findings Morais et al. (2013) study questions whether the structure of large sign network is scale-free by demonstrating that several other heavy-tailed distributions can fit the degree distributions. But these alternative heavy-tailed distributions also indicate the presence of hub-words. Therefore, perhaps the most salient insight in terms of the process of Quinian bootstrapping from the analysis of sign networks made for entire vocabularies and the Steyvers and Tenenbaum (2005) model is that placeholders sign structure might grow around hub-words, rather than the networks have a scale-free structure. However, determining whether a sign network does have a scale-free structure remains relevant as this potentially constrains any model that aims to explain the formation of the network. For example, a mechanism to constrain the number of connections a hub-word can make would need to be added to the Steyvers and Tenenbaum (2005) model for this model to explain a degree distribution that follows a power-law with an exponential cut-off (Morais et al., 2013).

2.6.3 Summary of Sign Network Studies

The empirical studies discussed in this section demonstrate the efficacy of networks and network analysis to map placeholder sign structure and robustly identify similarities and differences between this structure in novice and experts. Moreover, this quantitative analysis, using accepted measures of network structure, enables models to be developed for network growth that explain the structure of these networks. These models can operationalise many of the assumption made about the nature and mechanism for conceptual change (e.g., differentiation). Hence, robust analysis of sign (placeholder) structure and comparison with models of network growth may provide greater clarity about how Quinian bootstrapping might function as a mechanism for conceptual change. For instance, the possibility that placeholder sign network grows around certain hub-words.

However, an important distinction has been made in these studies between concept and sign networks. The presence of hub-words and clustering in sign network has been found in studies that have focused on entire lexicons and, therefore, general domains of knowledge. These clusters of vertices and hub-words have been found in networks for scientific domains of knowledge (e.g., electromagnetism). However, the networks created in these domains have been concept, rather than sign maps. Hence, this study aims to create sign, rather than concept maps, in a scientific domain of knowledge to examine the placeholder structure that might guide concept development in intuitive theories for these domains, and to see whether the Steyvers and Tenenbaum (2005) model can be applied to these networks to gain insight into how they develop.

2.7 Summary, Contribution to Knowledge and Research Questions

This literature review began by stressing that learning science requires conceptual change. Hence, knowledge of the process of conceptual change, and perhaps more most importantly the mechanism that drives this change, should inform any effective pedagogy in science education. However, progress towards understanding concept change in science education has been limited by a lack of agreement about the basics ideas that underpin this area of research, for example, the nature of a scientific concept. This lack of progress has recently led to calls for some form of shared theoretical language to be adopted so different views of concepts and accounts of conceptual change can be compared. A simple, yet pragmatic, definition of a scientific concept was adopted in this review, based on the shared language introduced by Sherin (2017a), to achieve this goal. This definition was used to demonstrate that Susan Carey's TOOC gives the most complete account of conceptual development. Hence, the TOOC has been adopted as the theoretical framework for this thesis.

In TOOC, Quinian bootstrapping is the mechanism that drives conceptual change. In this process, the first phase in any conceptual change is the development of placeholder sign structure. This sign structure then guides the learning process that are required for conceptual development. Carey uses the development of the concept of natural number as evidence for this bootstrapping process. In this example, the placeholder sign structure is well known (i.e., the count sequence). Carey uses knowledge of this placeholder sign structure to gain a greater understanding of the bootstrapping process and to provide evidence for this process. Hence, she uses knowledge of the placeholder structure as an analytical tool to develop a precise and evidenced description of the development of concept of natural number. This level of descriptive precision allows aspects of the process to be empirically tested, which Amin and Levrini (2017b) identify as lacking from current accounts of the development of scientific concepts.

The placeholder sign structure which could guide the development of scientific concepts within intuitive theories is not well known and has not been studied directly. Hence, a focus of this thesis is to measure and analyse placeholder sign structure associated with the development of concepts within an intuitive theory. As was demonstrated by Carey's example of natural number, this knowledge of placeholder sign structure may enable a more precise account of conceptual development and provide opportunities to empirically test aspect of this account.

For instance, whether the development of placeholder sign structure precedes full conceptual understanding, which is a consequence of the Quinian bootstrapping process.

Some studies have measured and analysed either conceptual structure in a scientific domain of knowledge, or what is more likely to be associated with placeholder sign structure for entire lexicons. These studies have demonstrated the efficacy of using network analysis to quantitatively characterise conceptual or sign structure, and how models of network growth can provide insights in terms of how any sign structure might be formed. For instance, aggregate (group) and individual sign networks for entire lexicon contain highly connected hub-words. The model of network growth developed by Steyvers and Tenenbaum (2005) suggests sign networks grow around these hub-words.

This study is different from the studies of concept networks in scientific domains of knowledge or placeholder sign network for entire lexicons because it measures and analyses placeholder sign network in a scientific domain of knowledge. Hence, it attempts to explore the role sign structure might play in conceptual development in intuitive theories. This role is investigated by analysing the similarities and difference between the sign networks of individuals with difference levels of expertise in a scientific domain of knowledge. Furthermore, the possibility that sign networks grow around hub-words, and hence novices and experts have different hub-words, is a key idea that is tested when the similarities and differences between the sign network for different cohorts are investigated. Moreover, this study also tests the assumption that any hub-words are used by participants when they explain physical phenomena. Hence, this study investigates how placeholder sign structure could relate to conceptual understanding. Hopefully this investigation will provide insights into how Quinian bootstrapping might drive conceptual development in intuitive theories and provide evidence for this process.

These aims are formalised in the following research questions.

1. What are the sign networks of individuals within a scientific domain of knowledge?
2. What are the similarities and differences between the networks for individuals who have different levels of scientific understanding (novices and experts)?
    a. Are hub-words responsible for most connections in the networks?
    b. If so, are the hub-words the same for different individuals?
3. Do these individuals use hub-words from their networks when they reason about a scientific experiment or demonstration? If so, how?

3 Measurement and Construction of Individual Sign (Word) Networks

3.1 Introduction

The methodology in this thesis explains how this study will address the research questions stated above in section 2.4.4. To address these questions, a research design was adopted that split data collection and analysis into two phases. During the first phase, sign networks were measured and analysed for individual students (research questions 1 to 2). Then, during the second phase, students were interviewed to explore how individuals use any hub-words in their semantic networks when reasoning about scientific demonstrations or experiments (research question 3). This methodology chapter describes the first phase of the research design, including the methods used to collect the data and construct both types of sign networks (section 3.3). However, this chapter begins by justifying the selection of a scientific domain of knowledge and explains how the limits of this domain were defined.

3.2 Selecting and Defining the Scientific Domain of Knowledge

Mechanics was chosen as the scientific domain of knowledge of interest partially for pragmatic reasons of timing and convenience. The student participants of this study were taught mechanics during the first autumn term of a two-year A-level course (Year 12). Hence, data collection could take place in the spring term of this first year. This was before the students sat mock examinations in the summer of the first year and spent considerable time working on any university and job applications in the autumn term of their second year (Year 13). Therefore, using this data collection window, ensured the bureaucratic burden of this data collection occurred at a time when it was most likely to minimise any impact on their attainment (see section 3.3.2.4 for more detail).

In addition to these pragmatic reasons, mechanics was selected because extensive physics education research has examined students' understanding of forces and motion concepts (e.g. Abrahams, Homer, Sharpe, & Zhou, 2015; diSessa, 1993; diSessa et al., 2004; diSessa & Sherin, 1998; Howe, Taylor-Tavares, & Devine, 2016; Ioannides & Vosniadou, 2002). It is widely accepted that this research has established that children's understanding of mechanics prior to formal scientific teaching is not compatible, or incommensurable, with the accepted theory of

Newtonian mechanics. For instance, McCloskey (1983) suggests that many students have an 'impetus theory' of mechanics, which suggests that an object is given an internal force, or impetus, that gradually dissipate when it is set in motion by another agent. This is inconsistent with the concept of a non-contact force in Newtonian mechanics, which stresses that forces only act between an agent and an object when they are in contact with each other. Carey (2009) views incommensurability as a test for when learning in a domain of knowledge requires conceptual change, rather than just knowledge enrichment. Hence, learning mechanics is likely to involve conceptual change. Furthermore, Chi, Feltovich, and Glaser (1981) find that expert physicists categorise mechanics problems according to whether they can be solved using an approach that involves forces or energy. In contrast, novices categorise the same problems using superficial features of the problem (e.g., whether the problems involved rotation, a spring, or an inclined plane). The ability to identify a category of objects, or in this case a class of phenomena, is generally thought to be determined by conceptual understanding (Murphy, 2002). Hence, this study also suggests novices' conceptual understanding is different from experts, and therefore learning mechanics requires conceptual change. Finally, Hestenes et al. (1992) have developed the Force Concept Inventory, which is accepted as a standard test of whether an individual has a Newtonian conceptual understanding of mechanics. This standard test is used in this study as part of the methodology to address research question 2.

As well as establishing that mechanics is likely to involve conceptual change, previous research has also used mechanics as a testing area to examine theories of conceptual change. For instance, diSessa (2002) applied his theory of conceptual change, which involves primitive knowledge elements called p-prims clustering together to form more complex structures called coordination classes, to explain *Force* as a concept (or in this case as an example of a coordination class). Similarly, Parnafes (2007) uses coordination classes to explain development of 'fast'. In contrast, Ioannides and Vosniadou (2002) apply an intuitive theory approach to conceptual change, which is heavily influenced by the 'theory' theory of conceptual development (Carey, 1985a; Gopnik et al., 2004), to explain the development of *Force*. Moreover, Brookes and Etkina (2009) use a linguistic framework to chart the historic development of the meaning of the term force. Finally, Taylor (2012) found that 15- to 16-year-old students' ability to explain certain phenomena in mechanics may be associated with the connections they make between certain words (e.g. *force* and *resultant*) (also see (Taylor & Hillier, 2018)). The depth and breadth of the research in mechanics, from both a conceptual change and linguistic perspective, provides a rich context for interpreting any findings of this study and offers an opportunity to contribute to the understanding of conceptual change in this important domain of knowledge.

As well as being extensively studied in science education and conceptual change research, mechanics is also the foundation of many physical theories (e.g. statistical mechanics, which explains macroscopic physical properties such as temperature and pressure, (Kittel & Kroemer, 1997)). Hence, learning mechanics is an essential part of learning physics (diSessa & Sherin, 1998). This provides further justification for selecting mechanics as the scientific domain of knowledge. However, the ubiquitous presence of mechanics in physics makes it difficult, if not impossible, to define clear boundaries for a mechanics domain of knowledge. As a pragmatic solution to this problem, the mechanics domain of knowledge in this study was defined by topics that are included in the mechanics section of the curriculum (set out in the exam board specification, (AQA, 2017)) studied by the A-level Physics students, who were participants in this study.

3.3 Measurement and construction of students' sign networks

This section describes the methods to assess research question 1. It begins by describing the snowballing sampling methods developed by Morais et al. (2013) and how this method was adapted to measure vertices and edges in individual students' sign networks. The Morais et al. (2013) method began by selecting ten words (sign vertices) at random from 4,279 words listed as acquired before the age of nine in either of three linguistic databases (the MRC psycholinguistic database (Coltheart, 1981), the norms of Bird, Franklin, and Howard (2001), and the norms of Stadthagen-Gonzalez and Davis (2006)). These words were the initial stimulus or seed words in the Morais et al. (2013) study. In several testing sessions, Morais et al. (2013) gave their participants one minute to list any words that came to mind when they were shown a seed word. Hence, free word-associations were used to identify new sign vertices (words) and edges between signs (connections between words) in a network. The response words (the free word associations) became the stimulus words in the next iteration of the data collection procedure. Hence, the sample of measured words and connections (vertices and edges in the network) 'snowballed' from the initial seed words.

Several studies have used also used word associations to measure sign networks. For instance, Steyvers and Tenenbaum (2005) used word-association norms for 6000 English speaking participants (Nelson et al., 1999) and De Deyne and Storms (2008a) used word-associations norms for Dutch speakers (De Deyne & Storms, 2008b) to create semantic networks. However, as Morais et al. (2013) points out, by using data sets that aggregate the responses of many people, these studies measure sign networks for groups of individuals (e.g. English or Dutch speakers)

rather than individuals. This is important to recognise because it is possible these aggregate networks do not accurately reflect the structure of individual networks. For instance, Morais et al. (2013) highlights unpublished simulation work that demonstrate that a power-law degree distribution can be generated by averaging multiple individual degree distributions, which are not power-laws. Similarly, aggregate networks can distort other parameters (e.g. the average degree k, and the average path length L) which also measure the connectivity in a network (Morais et al., 2013).

The snowball sampling method designed by Morais et al. (2013) measures a semantic network for an individual and, hence, avoids the issues parameter distortion when using aggregate networks as a proxy for individual networks. However, Newman (2018) stresses any snowball sampling method will always *sample* a network rather than capture the entire (or true) network. Hence, any network parameters measured using a sampled network will always be estimates of the true values. Moreover, it is likely that this method will generate bias network samples. For example, in the limit of a large number of iterations, snowball sampling samples vertices (words in the case of a sign network) with a probability proportional to their eigenvector centrality (Newman, 2018). Therefore, a sample sign network measured using snowball sampling is likely to contain a larger proportion of words with high eigenvector centrality than words with low eigenvector centrality. These bias sample networks could lead to inaccurate estimates of parameters that measure network structure (e.g., degree distribution, average shortest path length, clustering coefficient), just as aggregate networks also fail to give accurate estimates of these parameters when they are assumed to represent averages of individual networks.

To test whether their snowball sampling method was able to accurately estimate the average vertex degree k, average shortest path length L and clustering coefficient C, Morais et al. (2013) simulated a sign network with a known structure, and then used their method to sample this network several times. For a small number of iterations (e.g., 2 or 3, when on average only 0.0005% to 0.0078% of the simulated network vertices had been sampled), the average vertex degree k was overestimated, whilst the average shortest path length L and the clustering coefficient C were underestimated using the sampled sign networks. However, by the 6th iteration, when approximately 66% of the vertices had been sampled, the estimated values of all these parameters were in close agreement with the actual values for the simulated network. Hence, Morais et al. (2013) concluded that their method gave accurate estimates of network parameters for individual sign network when a participant completed six or more iterations.

To achieve the required number of iterations to give accurate estimates of k, L and C, the total time spent in data collection by the participants in the Morais et al. (2013) study, ranged from 32 to 61 hours, in a series of one-hour sessions that occurred daily over the course of seven weeks. Hence, sampling these word-associations to create an accurate individual sign network for an entire lexicon was time consuming and involved a high level of commitment over a long period of time (daily one-hour sessions for 7 weeks). This is clearly not feasible in this study, which involves secondary school students (age 11 to 18). However, the goal of the Morais et al. (2013) study was to measure sign networks for the entire lexicon for each individual participant, by capturing their word-associations for a large proportion of the words in their vocabulary. But, to address research questions 1 and 2, a complete map of an individual's sign network is not required. Instead, this study aims to measure the sign networks for individuals in a specific domain of knowledge (mechanics) and identify similarities and difference between these networks.

The population of words in a specific domain of knowledge will be lower that the population of words in an entire lexicon. For example, in this study, the mechanics domain of knowledge was pragmatically defined (in section 3.2) to include the topics taught to A-level students in the mechanics section of the curriculum. This section of the curriculum is described in the mechanics chapters of the accompanying A-level Physics textbook (Breithaupt, 2015). These chapters include 1572 unique words. In contrast, estimates of the number of words known by individuals range between 45,000 to 80,000 (Bloom, 2000). The simulations carried out by Morais et al. (2013) demonstrated that sample networks created using word associations for approximately 33,000 out of lexicon that included 50,000 words (approximately 66%) could be used to accurately estimate the network parameters. Hence, a much smaller number words, approximately 1000 (~66% of 1572), could capture enough word associations to sample individual sign networks that would give accurate estimates of network parameters and would, therefore, allow valid comparisons between these networks.

This section of this methodology chapter argues that this is the case. It outlines an adapted version of the Morais et al. (2013) snowball sampling method, which collects a smaller sample of free word associations, and can therefore measure representative samples of domain-specific sign networks for individuals, within the logistical constraints imposed by working with students in a secondary school environment. Hence, this adapted method addresses research question 1. The arguments made for this method are based upon the finding of a pilot study, carried out between March and June 2017, to establish proof of concept of the adapted version of the

Morais et al. (2013) method. The changes to this method, as well as the general context for the study (e.g. school setting), are described in sections 3.3.1 to 3.3.4. The sign networks created in this pilot study were analysed using similar techniques to those used by Steyvers and Tenenbaum (2005), De Deyne and Storms (2008a) and Morais et al. (2013). For brevity and to maintain a coherent narrative, this analysis is described in detail in sign networks results chapter (Chapter 4). The results of the pilot study are presented in the following sections to justify the methodological choices that enable this technique to be implemented in a secondary school setting and to demonstrate the efficacy of the method to measure sign networks in this context. For each part of the method, the procedure used in the pilot study is described, then the relevant results from the pilot study are discussed, with emphasis on the implications for this part of the main study method. The main study procedures (e.g., participant selection, snowball sampling procedure) are then described.

3.3.1 School Setting

3.3.1.1 Rationale for Electing the Target School

The target school is a rural, non-selective, state-funded secondary school in the UK, where the DPhil researcher is employed part-time as a physics teacher. Hence, there are several pragmatic reasons for selecting this school. First, as a member of the school community, the researcher has been able to disseminate the findings of previous research, which have been used to inform pedagogy (Taylor, 2012; Taylor & Hillier, 2018). Hence, there is a willingness to embrace and cooperate with research as the beneficial impact of research had been demonstrated. Second, the data collection process for the main study took place over several months (see section 3.3.3.2). The extended nature of this process was anticipated by the pilot study. Therefore, it was considered a significant advantage to have close contact with participants over this long period of time to encourage continued participation and to clarify any queries that arose. Finally, there were advantages and disadvantages for recruitment of student participants in being a member of the school community. These are described in depth in section 3.3.2.3.

3.3.1.2 General Background for the School Setting

The school is larger than the average secondary school in the UK, with 1769 students aged 11 to 18 attending (Ofsted, 2017). Nearly all the students in the school population are White British (>90%) and the proportion of students who speak English as an additional language is well below the national average for UK state-funded schools (<13%) (Ofsted, 2012). The proportion of students identified as having special educational needs or who are known to be

eligible for a pupil premium payment is also below the national average. In the UK, students are eligible for pupil premium funding if they have received Free School Meals (FSM) in the last six years, have ever been a Looked After Child (LAC), or have parents who have served in the Armed Forces within the last six years (DfE, 2018). A student is entitled to FSM if their parents' income is less than £16,190 or they receive certain state benefits, including unemployment benefit (DfE, 2018). Hence, pupil premium is a crude measure of socioeconomic status, and the low proportion of students eligible for a pupil premium payment at the target school indicates that most students come from higher socioeconomic groups.

The school is current rated 'Good' by the UK Office for Standards in Education, Children's Services and Skills (Ofsted) (Ofsted, 2017). This rating broadly indicates that students make good progress in their learning because of effective teaching. Good progress is defined by Ofsted as students attaining grades in public examinations at the ages of 16 and 18 that are judged to be acceptable given their attainment in previous examinations before they entered the school at age 11. Acceptable progress is determined by comparing against the national average for progress. Hence, most students within the target school make progress in line with the national average. Ofsted assess the efficacy of teaching by observing a sample of lessons from a range of subjects within the school. Despite the overall good rating for the school, an earlier inspection judged that the progress of students in science was not yet good and there was some variability in the quality of science teaching (Ofsted, 2012).

The target school was selected for pragmatic reasons (see above). However, the general context of the setting suggests participants are drawn from a typical UK school (67% of schools are rated good by Ofsted (DfE, 2020)) in terms of standard UK measures of teaching quality and progress. Hence, this study involves participants that receive an education that is representative of UK secondary education. However, the cultural background of the students is not particularly diverse, and English language as a first language is almost ubiquitous (only 3% of students speak English as an additional language). Similarly, most students come from higher socioeconomic groups. Therefore, most of the participants in this study are likely to represent more affluent White British society.

3.3.2 Participants

All students at the target school (1769 at present) study mechanics at some points in their school career. It is clearly not feasible to include all these students in the study. Hence, smaller groups of 16- to 18-year-old students studying either A-level Physics or Psychology were targeted as

participants. The rationale for selecting these groups of students, initially for the pilot study and then for then for the main study, is outlined in sections 3.3.2.1 and 3.3.2.2. Section 3.3.2.3 describes the recruitment procedures and ethical considerations for both the pilot and main study.

3.3.2.1 Pilot Study Participants

The participants in the pilot study were four Year 12 students (all age 16 years old). Two of the students were studying A-level Physics and two of the students were studying A-level Psychology, and not A-level Physics. These groups, or cohorts, of students were selected as participants for two main reasons. First, to form two cohorts of students that were likely to produce different sign networks and, hence, facilitate comparisons between the networks. Second, these two groups represented stakeholders who could potentially gain from this research. Third, as older students, it was assumed they would engage with the research in a way that was more likely to generate valid data. These reasons are described in more detail in the sections below.

3.3.2.1.1 Using A-level Physics and A-level Psychology Students to Form Two Contrasting Cohorts of Participants

In the UK it is compulsory for all students to study a science curriculum that includes several physics topics (including mechanics) until the age of 16, when most students are examined on their science knowledge and are awarded a General Certificate in Secondary Education (GCSE) in science. In the target school, students can either opt to study for a Combined Science GCSE (equivalent to two separate GCSEs) or separate Biology, Chemistry and Physics GCSEs. The content in these GCSEs is set by the awarding body, which in the target school is AQA. The mechanics sections of the separate Physics and Combined Science GCSE include the same topics (e.g. speed, acceleration, Newton's Law) (AQA, 2015a). However, the Physics GCSE also includes moments, gears and levers (AQA, 2015b).

After their GCSE examinations, students in the target school can choose to continue to study physics, as part of a Physics A-level, if they achieve a 6 for their GCSE Science grade. The A-level Physics curriculum builds upon many of the ideas introduced at GCSE, but a number of new theories are introduced to students (e.g. the photon theory of light), which can challenge and extend students' current understanding of familiar vocabulary (see AQA (2017) and Breithaupt (2015)). For instance, students encounter the idea of a *force* causing interactions between particles (e.g., the weak force causing a neutrino to interact with a neutron and change

into proton). This phenomenon extends the idea of a *force* as a causal agent but cannot easily be reconciled with a *force* being simply a *push* or a *pull*. Hence, it is likely that A-level Physics students will need to form new sign relationships because of their A-level tuition.

It is unlikely that psychology students would have the same sign relationships, given that they will not have been exposed to the esoteric physical theories introduced at A-level. However, the entry requirements to study A-level Psychology at the target school are a grade 5 in GCSE Combined Science or in their separate science GCSEs. Therefore, it is reasonable to assume that psychology students would have some familiarity with the initial seed words used in the study (e.g. *force*, *acceleration* – see section 3.3.3.1 for full details of seed words and the selection process for these words) because the topics that introduce these terms are part of the GCSE science curriculum. Therefore, it is likely they would have some sign structure around these seed words. However, this structure would not have been enhanced by A-level tuition. Hence, using physicist and psychologist cohorts is an opportunity to measure sign networks for individuals, with different level of exposures to teaching in the mechanics domain of knowledge, and to see any differences between these networks can be identified, which could be due to differences in understanding of this domain. Therefore, the pilot study tested whether the proposed methodology was likely to answer research questions 2.

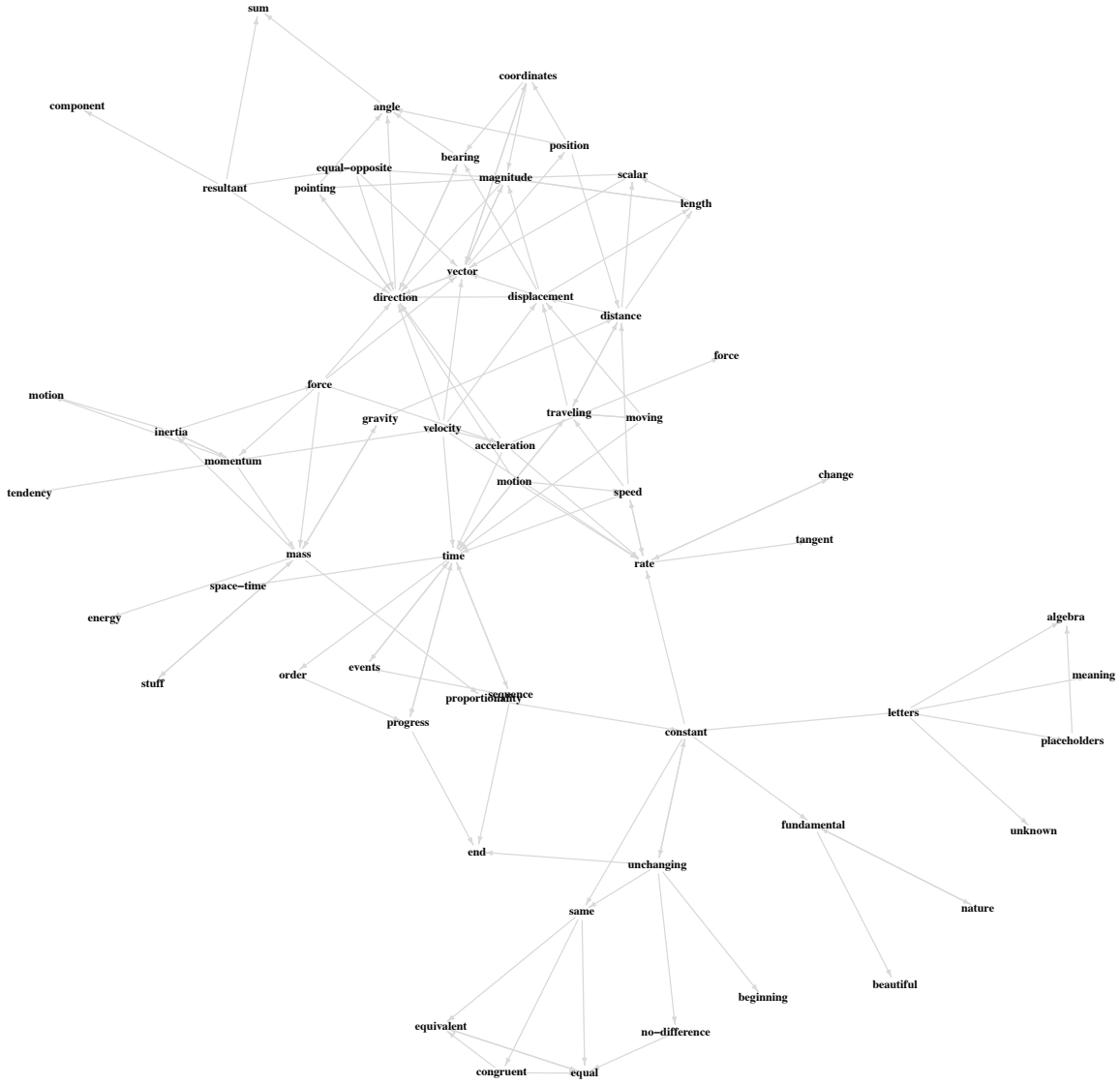


Figure 3.1 Sign network for participant 3 (physicist) in the pilot study.



Figure 3.2 Sign network for participant 2 (psychologist) in the pilot study.

Figure 3.1 and Figure 3.2 show a physicist's and psychologist's sign networks respectively. Careful visual inspection[4], of these networks shows that these members of the physics cohort and the psychology cohort, with different levels of tuition in mechanics, make different connections to key words (e.g., *force*, *mass*, *acceleration*). Moreover, the psychology network does not contain some connections between key words that might be expected (e.g., participant 1: *force* | *acceleration*)[5]. This finding is consistent with the finding of other studies (e.g., Cachapuz

[4] Note that these networks are difficult to clearly visualise and, therefore, it is difficult to qualitatively identify differences in the network. Hence, a quantitative analyse is used to compare sign networks. This is described in detail in Chapter 3.

[5] Note in the following section (and throughout this thesis) the notation PARTICIPANT: *stimulus* | *response* is used to indicate a connection between a stimulus and response word for a given participant.

& Maskill, 1987; Cachapuz & Maskill, 1989; Maskill & Cachapuz, 1989; Nakiboğlu, 2008; Van Zele, Lenaerts, & Wieme, 2004) that have measured and analysed sign networks for both groups and individuals, which suggest that an ability to reason scientifically (and therefore use scientific concepts) is associated with certain connections between words. However, these studies also suggest that greater complexity in the structure of a sign network is also associated with a more sophisticated understanding of different scientific domains of knowledge (e.g., chemical reaction kinetics (Cachapuz & Maskill, 1987)). Strikingly, some studies identify 'islands' (Cachapuz & Maskill, 1987; Nakiboğlu, 2008) or 'clusters' (Van Zele et al., 2004) of words in sign networks, which become connected after tuition but remain isolated in the networks for groups of lower attaining students (Cachapuz & Maskill, 1987).

The results of the pilot study are not entirely consistent with the findings of these studies. There are no isolated islands of words in the sign networks. Moreover, whilst the sign networks of the participants do have different numbers of words (e.g., $n_{participant\ 2} = 83$ and $n_{participant\ 4} = 132$) and connection between words (e.g., $n_{participant\ 2} = 241$ and $m_{participant\ 4} = 494$), they have similar diameters (e.g. $D_{participant\ 2} = 9$ and $D_{participant\ 4} = 9$), average shortest path lengths (e.g. $L_{participant\ 2} = 4.33$ and $L_{participant\ 4} = 4.02$), and clustering coefficients (e.g. $C_{participant\ 2} = 0.27$ and $C_{participant\ 4} = 0.22$). This would suggest the structures of sign networks for novice and experts could be more similar than has been suggested in previous studies. This discrepancy in the findings of this pilot study and previous studies can be explained by examining the differences between the method used by these studies and this pilot study to identify vertices (words or signs) and edges (connection between words or signs) in sign networks. These differences are discussed below and are used to highlight the strengths of the snowball sampling method and to justify the use of physicists and psychologists as contrasting groups.

Table 3.1 Comparison of the descriptive statistics for the semantic networks analysed in this study and those analysed by (Morais et al., 2013). Note, n is the network order (number of vertices), m is the network size (number of edges in the network), <k> is the mean vertex degree, D is the network diameter, L is the average distance or average shortest path length, C is the clustering coefficient and α is the scaling parameter, which characterises the power-law model fit to the vertex degree probability distribution.

Study	Participant	n	m	<k>	D	L	L_{random}	C	C_{random}	α
Pilot study	1	64	120	1.88	11	3.67	5.32	0.33	0.10	2.91
	2	83	3403	2.90	9	4.33	4.04	0.27	0.07	3.55
	3	59	1711	2.25	8	3.33	4.20	0.28	0.08	2.51
	4	132	494	3.74	9	4.02	3.70	0.22	0.05	5.03
Morais et al. (2013)	1	9,429		2.28	-	7.03	6.66	0.10	0.002	5.01
	2	2,303		2.30	-	6.24	5.57	0.20	0.007	3.23
	3	5,100		2.12	-	7.05	6.51	0.18	0.003	4.11
	4	1,358		2.73	-	5.66	4.75	0.32	0.001	2.75
	5	9,129		2.96	-	5.65	5.29	0.18	0.003	3.98
	6	3,239		2.40	-	6.78	6.20	0.24	0.004	3.22

Van Zele et al. (2004) measures sign networks for individuals by providing their participants with an initial set of words and asking them to use these works to make a sign network to represent their understand of a specified domain of knowledge (in this case, the atom). Participants were able to add words to create their sign map. However, essentially this method establishes whether a participant recognises connections between a given set of words. Hence, isolated clusters of words in a sign network measured using this method, indicates that an individual cannot connect the words in these cluster with words that have been provided.

In contrast to the method used in the Van Zele et al. (2004) study, the snowball-sampling method allows participants to choose any word that help them understand a stimulus words (see section 3.3.3). Response words then become stimulus words in the next iteration (Morais et al., 2013). This iteration, or recruiting of new stimulus words, allows the snowball-sampling technique to freely explore the structure of a sign network. Therefore, isolated clusters of words are present in the sign networks measured by the pilot study because the participants are free to choose the words they see as helping them to understand a stimulus word.

The studies by Cachapuz and Maskill (1987) and Nakiboğlu (2008) also begin with seed words as stimulus words and ask participant to list response words they associate with these stimulus words. However, these studies only ask participants to respond to the initial seed words. The response words do not become stimulus words in a second iteration. In other words, only one iteration is carried out. Hence, the parts of a sign network these studies explore can be limited to the closest word neighbours of the stimulus or seed words. Consequently, potential

connections between words, and therefore structure, in a sign network may not be measured. One illustration of this effect is given in Figure 3.3.

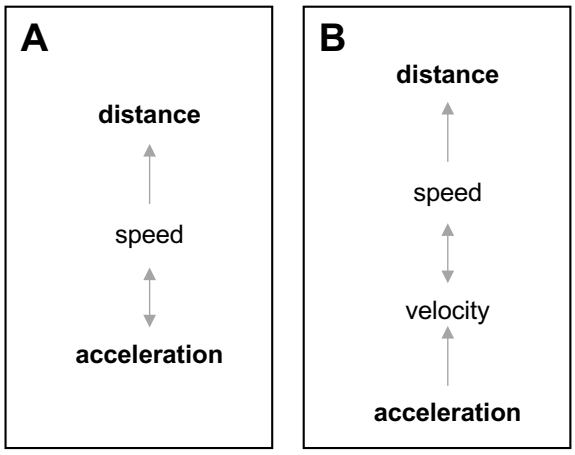


Figure 3.3 Illustration of how using only one iteration of word associations can miss connections between words and, therefore, structure in a sign network. Stimulus words are shown in bold. In example A, distance and acceleration are connected because speed is a response to both these stimulus words. However, in example B, the connection between distance and acceleration is missed because these words are connected via two words, which are not original stimulus words. This connection would be measured using the snowball sampling method because velocity and speed would be stimulus words in the next iteration.

Previous studies generally apply more stringent conditions for drawing edges and, in doing so, their analyses ignore many connections between words. For instance, Johnstone and Moynihan (1985), Cachapuz and Maskill (1987) and Nakiboğlu (2008) use the relatedness coefficient to draw edges between words, or vertices, in their semantic networks. The relatedness coefficient (Garskof & Houston, 1963) is a measure of the degree to which two cue words share the same response words. The relatedness coefficient is 1 if both cue words have the same response words and 0 if no response words are shared. Any value between 0 and 1 indicates some response words are shared. Johnstone and Moynihan (1985), Cachapuz and Maskill (1987) and Nakiboğlu (2008) only draw an edge between two words if the relatedness coefficient for these two words is above a certain arbitrary threshold value. Hence, edges are only drawn between words that share a certain number of response words. Therefore, not all the connections between the cue words and response words are recorded in the network.

This stringent condition for drawing edges and the lost information regarding all connection between any cue and response words will make any semantic networks appear to be less connected. For instance, the 'islands' of words in Cachapuz and Maskill (1987) and Nakiboğlu (2008) networks do not necessarily show that words in different islands lack any connection. They only show that words in different isolated islands do not share enough word associations for their relatedness coefficient to rise above the threshold required to connect them.

Nakiboğlu (2008) recognised some of the limitations of the relatedness coefficient technique and, in response, also created a semantic network using the frequency of responses to cue words within a group of participants. An edge was drawn between a cue and response word if the frequency of this response was above a threshold value. However, again, this stringent technique for drawing edges will underestimate the connectivity in any semantic network by discarding any infrequent responses.

In contrast, the Morais et al. (2013) and the pilot study, draw an edge in a sign network between any stimulus and response word (see section 3.3.4.1 for a more detailed description of network construction). Hence, connections between words are not lost by imposing stringent and arbitrary conditions for drawing edges between words. As a results, the sign networks measured by Morais et al. (2013) and this pilot study do not contain isolated 'islands' of words. The pilot study sign networks show that salient words within a mechanics domain of knowledge sit within a complex network of other words for both physicists and psychologists. However, in the psychologists' networks these words are connected to words that are more likely to be found in everyday vocabulary (e.g., participant 1: *acceleration* | *fast*) and some words that are likely to be important in a mechanics domain of knowledge (e.g., participant 2: *mass*) are not found in these networks. Hence, the complex structure of the psychologists' mechanics sign network might be present because the snowball sampling method is measuring the structure of a participant's wider lexicon beyond the mechanics domain of knowledge. Salient words in particular domain of knowledge might have appeared isolated in previous studies because the methods used to measure the networks in these studies have not sampled any possible connections to a wider general vocabulary (Van Zele et al., 2004) or have used methods to construct a network that have excluded these connections (Cachapuz & Maskill, 1987; Nakiboğlu, 2008).

In addition to the similarities in the general descriptive statistics of the networks, the in-degree distributions of the sign networks for both cohorts can be fitted with a power-law (see Table 3.2 and Figure 3.4). This suggests all the sign networks could have a scale-free structure. Hence, highly connected hub-words are responsible for most of the connections in these networks and for other salient structural properties (e.g., the mean shortest path length, L). Perhaps significantly, different sign networks have different hub-words (see Table 3.3). Moreover, some of the physicist's (participant 4) hub-words are consistent with recognisable physical concepts. Both of these findings are consistent with a model of sign network growth developed by Steyvers and Tenenbaum (2005). In this model, a new word entering a sign network is more likely to connect to a word with more existing connections and it neighbours. This preferential

attachment to words with more connections gives rise to hub-words and a power-law degree distribution. Hence, the results of the pilot study seem to suggest that physicist's sign networks are developing because new words entering these networks, as the mechanics topic is learnt, are preferentially connecting to highly connected hub-words, which are associated with physical concepts (e.g., *force, mass*), and their neighbours. In this way, the sign network in mechanics grows in complexity and size around hub-words from the sign structure that may already exists in an individual's general lexicon. This is consistent with Carey's theory, which suggests conceptual development begins with significant sign structure being developed around an initial placeholder sign (in this case, the hub-words *force*, *mass*, and *acceleration*), which holds particular importance.

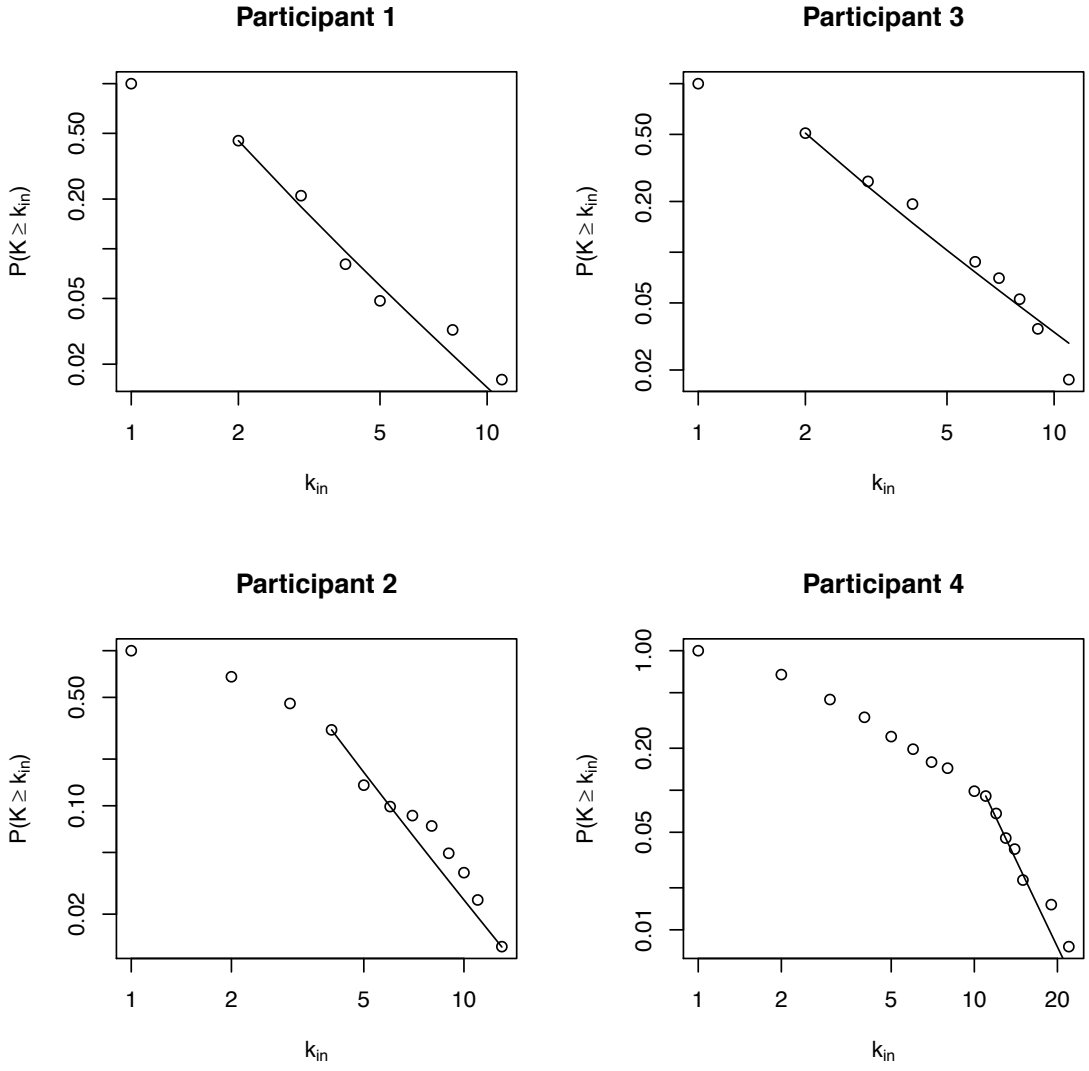


Figure 3.4 The cumulative vertex in-degree distribution for each participant. The distributions are plotted with log-log axis and solid lines that represent the best-fitting power-law model. Note the power-law models are fitted above a cut-off value, k.

*Table 3.2 The parameters and Kolmogorov-Smirnov statistics (KS) for a power-law model fit to the vertex in-degree and out-degree distributions for each participant. Note * indicates that a power-law model is a good fit to the distribution. Note, α is the scaling parameter for each power-law fit and k_{min} is the cut-off value for k in each fit.*

Distribution	Participant	α	95% confidence limits		k_{min}	95% confidence limits		KS	p
			α_{lower}	α_{upper}		$k_{min\ lower}$	$k_{min\ upper}$		
$P(k_{in})$	1	2.91	2.07	5.46	2	1	3	0.067	0.29*
	2	3.55	2.24	9.06	4	2	9	0.094	0.14*
	3	2.51	1.88	7.02	2	1	7	0.086	0.15*
	4	5.03	1.96	8.57	11	2	12	0.064	0.84*

Table 3.3 The words in each participant's network with the highest number of incoming edges. These are the highly connected hub-words. Note k_{in} is the in-degree for each vertex (word).

Participant 1		Participant 2		Participant 3		Participant 4	
Word	k_{in}	Word	k_{in}	Word	k_{in}	Word	k_{in}
speed	11	time	13	direction	11	force	22
weight	8	acceleration	11	time	9	mass	19
time	5	force	10	vector	8	acceleration	15

What is outlined above is an initial and speculative discussion regarding the growth of sign network based on based on small number of networks. However, whilst based on the limited results of the pilot study, this discussion demonstrates the efficacy of the snowball sampling method to freely explore the sign structure of words of interest. Moreover, using this method with contrasting cohorts of physicist and psychologist, with different levels of tuition in mechanics, has shown this method can address research question 2. The finding based on these comparisons of sign networks for different groups has already led to findings that illustrate a different view of how sign structure potential develops when learning science. This view is broadly consistent with both the model of sign network growth developed by Steyvers and Tenenbaum (2005) and the Quinian bootstrapping process Carey (2009). Hence, this snowball sampling method is consistent with the aim of this thesis to map placeholder sign structure to gain a greater understanding of the Quinian bootstrapping process in the TOOC, which drives the development of concepts in intuitive theories.

3.3.2.1.2 Using A-level Physics and A-level Psychology Students as Stakeholders with Potential to Benefit from the Research

As well as generating sign networks that might give insightful comparisons, both the physicist and psychologist cohorts also represent stakeholders who have the potential to benefit from this research. For the physicists, in terms of possibly identifying misconceptions they might

have and any improvements to pedagogy that arise from this study, which could improve Physics instruction. For the psychologists, through their experience participating in a study investigating learning, which they could then use to strengthen their university applications and drawn upon in any discussions around the use of human participants in psychological experiments. These benefits were shared were given as possible reasons for students to become participants during an introductory meeting to outline the aims of the study (pilot and main). During this meeting, 16 students expressed an interested in taking part in the study. Out of these 16 students, four students (two physicists and two psychologists) were selected as participants for purely logistical reasons: they were able to attend testing times that suited themselves and the researcher. When they were asked about their reasons for participating, all four students cited either wanting to help improve science teaching or a desire to take part in a psychological study, either to gain experience of this or to add to their university application.

3.3.2.1.3 Using Older A-level Students (Age 16 and above) as Participants

Targeting A-level students as participants for the pilot study was also justified by their age and the greater flexibility this gives them in terms of their school day, which includes study periods when they are not expected to attend lessons. These study periods allowed the pilot study participants to attend data collection sessions without missing lessons. Students below this age do not have this non-contact time. Therefore, their participation in this study would involve absence from face-to-face teaching time. This would not be approved by the target school and would create ethical issues because removal of these students from lessons may have a detrimental effect on their attainment. However, it was noted that expecting students to attend testing sessions during their study periods may also have some impact on their attainment. This ethical issue is addressed in more detail in section 3.3.2.4, which describes the mechanisms put in place to ensure participants could opt-out of the study if the data collection process was becoming too burdensome. However, for the pilot study, data collection was limited to four study periods, which was deemed an acceptable commitment for each participant in the CUREC approval for the study (see Appendix 10.1).

The pilot study demonstrated that it was possible to collect the necessary data in the proposed four study periods. Although, this study did also highlight that a slightly longer data collection period would be needed in the main study (see section 3.3.3.2). When asked, all the participants felt comfortable with this time commitment, did not feel burdened by the process, and indicated that they would be willing to commit some more time to data collection. However, the pilot study also showed it was necessary to have considerable flexibility and personalisation in when

the data could be collected for each student. Hence, the pilot justified the choice of A-level, with the study periods to do this, rather than younger students.

Using more mature students (16 years old or older) was also an advantage in terms of addressing research question 5, which requires participants to honestly express their thinking about several physical phenomena (e.g., the motion of falling objects). It was felt that these more mature students would be more likely to understand the necessary shift in nature of the relationship between themselves and the DPhil researcher (who is also a teacher at the target school) required to allow a transparent exploration of their ideas to take place in an interview. This issue is discussed in more detail in the interview methodology chapter (Chapter 5). However, the findings from these interviews supported the assumption that most older students would be willing to openly share their thinking with the DPhil research acting as a researcher, rather than a teacher.

3.3.2.2 Main Study Participants

The selection of target students was guided by three methodological choices, which have been described in the previous section. First, to choose contrasting cohorts (A-level Physics and Psychology students) to generate different sign networks to compare, and address research questions 2 and 3. Second, to prioritise students who represented stakeholders who were likely to gain from this research. Third, to use older students (over the age of 16) who have the necessary non-teaching time to allow data collection to take place, in terms of time comment but also flexibility. Also, these older students would have a better understanding of the shift in the nature of the relationship from teacher-student to researcher-participant that would enable the transparent sharing of ideas in interviews required to address research question 3. These choices were justified by the findings of the pilot study (see section 3.3.2.1). Hence, A-level Physics and Psychology students were also targeted to participate in the main study with the same reasons for selecting these student groups.

After introductory meetings, 34 students expressed a desire to participate in the study. These students were invited to follow-up meetings to give them further information about the study and the opportunity to formally give their consent to become participants. After these meetings, 23 students (13 physicists, 12 male and 1 female, and 10 psychologists, all female) returned permission slips and became participants. These students were sent the initial online data collection surveys. However, several of these students did not return the initial online surveys (3 participants) or did not complete the data collection process (9 participants). Hence,

final numbers of participants became 11: 6 physicists (5 male and 1 female) and 5 psychologists (5 female). The recruitment process for these students and ethical considerations, including reasons why a significant number of participants (12) were allowed to drop out of the study as part of a procedure to mitigate a potential ethical issue, are discussed in the following sections (3.3.2.3 and 3.3.2.4).

There is an obvious gender imbalance in the sample (mostly male in the physics cohort and all female in the psychology cohort). Unfortunately, this gender imbalance reflects the proportions of male and female students studying physics and psychology at the target school (i.e., the A-level physics cohort consists of 26 male and 2 female students). Hence, both the physics and psychology groups are in a sense representative of the cohorts studying these subjects at the target school. Therefore, the findings of this study should reflect the wider school population who are learning physics.

3.3.2.3 Recruitment Procedures for Participants

3.3.2.3.1 Access to Target Students

The DPhil researcher is also a teacher at the target school and is therefore a member of the target school community. Whilst this introduces some potential ethical issues (BERA, 2018), which are addressed in section 3.3.2.4, there were advantages in terms of access to participants because of the relationship between the researcher and the target school.

First, this study developed from the DPhil student's research undertaken as part of the Masters in Teaching and Learning programme (Taylor, 2012). The findings of this research were fed back to the target school and have resulted in changes in pedagogy, which are seen as beneficial. For instance, there is now a greater focus on developing writing and awareness of the meaning of language within the science department. Hence, the headteacher of the target school recognised the potential value of this research in terms of possible impact on teaching and learning within the school. Therefore, the headteacher was supportive of further research and was willing to grant permission for this study to take place within the target school (see Appendix 10.1.2).

Second, the personal relationship the DPhil researcher had with many students – having taught many of the students or interacted with them in a pastoral role – was an advantage in terms of recruiting participants. For instance, 22 students (16 physics students and 6 psychology students) were willing to take part in the pilot study. This number of students was more than

was required for the pilot. Many of the students who did not have the opportunity to take part in the pilot study have indicated they would have been willing to take part in next phase of this study. Similarly, 34 students (15 physics students and 17 psychology students) expressed an interest in taking part in the main study. The method used to recruit these students and any other students is discussed in the next section 3.3.2.4.

However, given the position of the researcher as a teacher within the school, it may be that these students felt obliged to express an interest in participating. The ethical issues this unequal power-relation creates are identified in section 3.3.2.4, and the measures used to mitigate for these issues are described Furthermore, this method of recruitment has been sanctioned as part of the CUREC approval for this pilot study (see Appendix 10.1) and this CUREC approval has been extended to cover the main study.

3.3.2.3.2 Method of Recruitment

The method of recruitment was the same for both the pilot and main study. Initially the consent of the target school was obtained through the headteacher. Then Year 12 and 13 students studying A-level Physics and Psychology were invited to express an interest in taking part after a short 5-minute presentation at the start of their lessons. This presentation briefly outlined the aims of the research and what they might benefit from participating. Students who expressed an interest where then invited to a longer presentation (20 minutes) that outlined in more detail the rationale behind the study. Students were also made aware of the likely time commitments expected of them, given examples of interview scenarios (including exemplar questions), and informed of who their data will be shared with and for what purpose. At that point, students were asked to consider 'opting in' to the research. It was stressed that no student was under obligation to participate in this study and if they did decide to take part, they could withdraw at any time and refuse consent for any data collected to be used. Moreover, if they did withdraw, this would have no bearing on their future assessment. This information was also provided for students in a letter, which they were encouraged to take away to give themselves time to make an informed choice about participation with no pressure from the presence of the researcher. The information letter also included a written consent form, which students completed and returned, if they decided to opt-in and become a participant (see Appendix 10.1). The rationale behind this process of consent is explained in section 3.3.2.4.

3.3.2.3.3 Process for Consent to Participate

The students asked that participated in this study were aged between 16 and 18 years. In addition, this study used methods that did not ask students to share sensitive personal information and were likely to be reasonably familiar to the students through the usual monitoring of their understanding by their teachers. For example, students are routinely asked about their understanding of certain words or to explain physical phenomenon (e.g., the motion of thrown object). For these reasons, the participants in this study were considered competent youths (Oxford, 2018). Hence, only the explicit written consent of the participants, rather than their parents, was obtained.

3.3.2.3.4 Data Protection

Data was captured, transferred, and stored securely as per CUREC Approved Procedure 25 (Oxford, 2018). No reference to individuals was or will be made within any final report to uphold anonymity. All consent forms and individual paper records were kept in a locked filing cabinet. Anonymity data keys were stored separately from datasets. The DPhil researcher stored electronic anonymised data on password encrypted storage media. There was no perceived harm, risk, or possible injury anticipated from either the pilot or main study. The participants were made aware of this data protection procedure.

3.3.2.4 Ethical Considerations in Measuring the Sign Networks

This phase of the pilot and main study presented a low risk to both participants and researcher (Oxford, 2018). However, the student-researcher's dual role as a teacher at the target school and a DPhil researcher introduced specific tensions (BERA, 2014) which raised several ethical issues. How these issues were addressed is described below.

3.3.2.4.1 Unequal Power-Relations and Pressure to Participate

The DPhil researcher is a member of the target school community. As a science teacher, he was teaching or had taught many of the students who were potential participants at the time the research took place. Hence, in several cases, a teacher-student relationship existed between the researcher and the potential participants before the research began. This relationship was an advantage when recruiting participants because a level of trust to act in students' best interests already existed between the researcher and potential participants (see section 3.3.2.3). However, as part of the teacher-student relationship, students are generally expected to comply with a teacher's instructions and to engage with activities that the teacher considers beneficial to their

learning. There is also a chance that a student will want to act in way they think will please the teacher. Especially when a teacher is responsible for monitoring attainment grades, UCAS predicted grades and references, and attitude to learning scores, as is the case in the target school. Hence, the existing teacher-student relationship could lead to students feeling obliged or pressure to participate. To avoid this, several 'cooling off' stages were built into the study, which gave students the opportunity to withdraw from the study, without giving reasons to do so.

These 'cooling off' stages are described below:

1. Students were encouraged to take away information letter from the introductory session and return the consent form to the DPhil researcher at a later date. This gave students time to consider whether they wanted to participate without any pressure from the presence of the researcher. This also required students to actively choose to participate, rather than simply giving in a form at the end of a session as part of politely leaving a room.
2. After handing in a consent form, an initial online survey was sent to the participant. If a participant did not respond to this survey, they were sent a reminder email and asked if they had received the survey. If they still did not respond to this initial reminder email, it was assumed that they had decided to withdraw.
3. The data collection process took longer than expected (see section 3.3.3.2). At this point the DPhil researcher became very aware of the need to limit the 'bureaucratic burden' on the participants (BERA, 2018, article 21), which might risk impacting on their attainment. Hence, if a student stopped responding to online surveys, this was taken as an indication that they might want to withdraw from the study. At this point, the DPhil researcher had spoken with all the participants on several occasions about the broad aims of the research in terms of enhancing our understanding of learning science and hence, possibly improving science teaching. It had also been repeatedly stressed by the researcher that this research should not have an impact on their attainment by taking up too much of their time. Therefore, it was considered appropriate at this time to ask a participant whether they were happy to continue with the research. Most participants

stated that they were happy to proceed. However, if subsequently they did not respond to another survey reminder, again, it was assumed they had withdrawn from the study.

Table 3.4 shows the attrition rate for participants and the stage (in terms of word-association iteration) they left the study. Three of twenty-three participants signed a consent form but did not respond to an initial survey. Nine out of twenty participants who took the initial surveys did not complete the data collection process. Whilst this was disappointing, this does indicate that participants did not feel undue pressure to continue to participate, because of the existing teacher-student relationship. Hence, the cooling off stages build into the data collection process were successful in avoiding ethical issues because of unequal power-relationships between the researcher and the participants. Eleven participants did complete the data collection process (six physics and five psychologists).

Table 3.4 The data collection process for all participants in the main study. The shaded participants completed the data collection process. If a participant dropped out, the table shows the stage of the data collection process (in terms of word-association iterations) that this occurred. If no iteration has been completed, the participant did not complete the initial seed word survey. Three participants did not complete initial surveys. Nine participants did not complete the data collection process. Eleven participants did complete the data collection process (six physics and five psychologists).

Cohort	Participant	Iteration						Start of Data Collection	End of Data Collection	Dropped out
		0	1	2	3	4	5			
Physics	12P01	×	×	×	×			Feb 2019	Nov 2019	Yes
	12P02	×	×	×	×	×		Feb 2019	Oct 2019	No
	12P03	×	×	×	×			Feb 2019	Mar 2020	No
	12P04	×	×	×				Feb 2019	Oct 2019	No
	12P05	×	×	×	X			Feb 2019	Mar 2020	No
	12P06							Feb 2019	Feb 2019	Yes
	12P07							Feb 2019	Feb 2019	Yes
	12P08	×	×	×				Apr 2019	Jun 2020	No
	12P09	×	×					Apr 2019	May 2019	Yes
	12P10	×						Apr 2019	Jul 2019	Yes
	12P11	x						Apr 2019	Sept 2019	Yes
	12P12	×	×	×	×	×		Apr 2019	Nov 2019	No
	12P13							Apr 2019	Jul 2019	Yes
Psychology	12S01	×	×	×	×			May 2019	Jan 2020	No
	12S02	×	×	×	×	×		Jun 2019	Feb 2020	No
	12S03	×	×	×	×			May 2019	Mar 2020	No
	12S04	×						May 2019	Dec 2020	Yes
	12S05	×	×	×	×	×		May 2019	Oct 2019	No
	12S06	×	×	×				May 2019	Jul 2019	Yes
	12S07	×	×	×	×			May 2019	Nov 2019	Yes
	12S08	×	×					May 2019	Sept 2019	Yes
	12S09	×	×	×	×	×	×	Jun 2019	Feb 2019	No
	12S10	x						May 2019	Sept 2019	Yes

3.3.2.4.2 Misconceptions Over the Use of Data

In his role as a teacher at the target school, the DPhil researcher is required to assess students to measure their attainment and effort and report this data in various ways (working at grades, UCAS predicted grades, attitude to learning scores). It was made clear to all students who participated that the data collected as part of the research did not contribute to these forms of assessment.

3.3.2.4.3 Anonymity

As a member of the target school community and teacher of many of the participants, it was impossible for the participants to remain anonymous to the DPhil researcher. Moreover, using students' names is considered good practice for any teacher to establish positive relationships with their students (Rogers, 2015). This approach is consistent with Article 20 of the BERA Ethical Guidelines for Educational Research (BERA, 2018), which states that steps should be taken to make participants feel at ease in any data collection process. Hence, the lack of anonymity and the use of participants' names was considered unavoidable and possibly an advantage in the data collection process, especially if a student's ongoing participation in the study needed to be discussed (see unequal power-relations discussion). However, the participants of this study, including students and the school, were made anonymous for all subsequent data analysis and research reports. All the data collected was kept strictly confidential and available only to the DPhil researcher and his supervisors (who only saw anonymous data).

3.3.3 Snowball Sampling Procedure for Measuring Network Vertices (Signs) and Edges (Sign Connections)

3.3.3.1 Selecting the Initial Seed Words

3.3.3.1.1 Pilot Study Procedure

In line with the procedure used in the Morais et al. (2013) study, ten words were selected to act as initial 'seed' words to begin the process of capturing the word-associations that were used to create the sign networks in the pilot study. The selection procedure for these words was as follows.

First, two lists of the most frequent words in two mechanics chapters (*Motion* and *Newton's Laws*) of the A-level course textbook (AQA Physics, Breithaupt (2015)) used by the target school were made using the *Quanteda* text analysis package (Benoit et al., 2018). The most frequent words were selected as seed words because frequently used words are more likely to be hub-words in a semantic network (De Deyne & Storms, 2008a; Steyvers & Tenenbaum, 2005). Given that hub-words are likely to have a significant fraction of the total connections in a sign network, if the network does have a scale-free or similar structure, these words were good starting points to reach as many of the parts of an individual's sign network during subsequent iterations. Second, frequently occurring words that were clearly present due to their grammatical importance, and not their conceptual importance, were removed (e.g., *this*, *the*, *a*). Third, the ten most frequent words that appeared in both lists were selected as the seed words in the 0^{th} iteration for the pilot study. These words were: *velocity*, *force*, *acceleration*, *time*, *speed*, *constant*, *motion*, *distance*, *mass*, *resultant*.

1.1.1.1.1 Main Study Procedure

The procedure for selecting seed words in the main study broadly follow the procedure adopted in the pilot study (see above). First, the most frequent words in relevant sections of the textbook used by the target school for A-level study (AQA Physics, Breithaupt (2015)) were listed. Second, grammatical, rather than conceptual words were removed from this list and the remaining ten most frequent words became the initial seed words in the 0^{th} iteration. However, some modifications to the procedure were made. First, the chapters in the textbook were expanded to include *Forces in Equilibrium*, *Force and Momentum* and *Work, Energy, and Power*. These chapters were included because the participants in the pilot study seemed to use aspects of force, energy, and momentum interchangeably when explaining the motion of objects. This

tendency has also been noted by McCloskey (1983). Furthermore, Carey (2009) points out the weight and density are also used interchangeably by children. She goes on to argue that developing an appropriate sign structure around *weight* and *density* is the first stage in these concepts becoming differentiated. Hence, it was considered important to attempt to measure sign networks in this expanding mechanics domain of knowledge. Second, listing the most frequently occurring words in each chapter separately, and then identifying words that appeared on both lists was an overly complex procedure. Instead, for the main study, a single list of the most frequently occurring words was made and the initial seed words were selected from this list. The seed words were *velocity*, *distance*, *energy*, *acceleration*, *weight*, *force*, *speed*, *mass*, *time*, *momentum*. These words were presented in random order to the participant in the 0^{th} iteration, rather than in frequency order.

3.3.3.2 The Testing Procedure

3.3.3.2.1 Pilot Study Procedure

The testing procedure used in this pilot study closely followed the snowball sampling method procedure used by Morais et al. (2013). In the pilot study, this snowball sampling method procedure began with an initial testing session where five of the initial seed words were presented as cue words to students. The five initial seed words were selected randomly from the set of ten seed words and were the same for each participant. This was the first sampling iteration. The response words each participant gave in the first sampling iteration became the cue words for that participant in the second sampling iteration, as well as the five remaining initial seed words. All ten seed words were not presented to the participants in the first testing session to minimise the chance that these closely related words would influences later responses. Finally, words given as a response in the second iteration became the cue words in the third iteration.

In the Morais et al. (2013) study, stimulus words were presented to participants on a computer screen. A participant then had one minute to respond to a cue word with as many words that came to mind. In contrast, in the pilot study, stimulus words were presented to participants on paper and in a classroom with the DPhil researcher present. Furthermore, participants were explicitly told to consider their response words. Therefore, no limit was placed on the time given to participants to respond.

This change in procedure signals a subtle shift between the aim of this study and the Morais et al. (2013) study. The goal of the latter study was to measure individual semantic networks and,

hence, create a maps of individuals' semantic memory. It is usually assumed that the faster an item is recalled in response to a given stimulus, the closer this item is stored in memory to the stimulus (Braisby & Gellatly, 2012). Hence, by giving participants a time-limit, Morais et al. (2013) ensures only the response words that are closest to the stimulus words in an individual's semantic memory are included in the networks they measure. However, this study aims to measure the sign structure that form around certain scientific words that might be part of the process of bootstrapping of a scientific concept (Carey, 2009), which might not be fully encoded into an individual's semantic memory. In this case, the first response words that come to mind might not be the most relevant in this sign structure. For instance, word association norms list *pepper* as a highly probably first response to *salt* (Nelson et al., 1999). However, in terms of bootstrapping a scientific understanding of *salt,* associating *salt* with *sodium* or *compound* might be more relevant. Maskill and Cachapuz (1989) argue these types of response are more likely to be given when concepts are being learnt if participants are asked to consider their answer and are given time to do so. However, this procedure does not require participants to fully justify their response words by citing specific scientifically accepted links (e.g., operational definitions, logical deductions) to the stimulus word (Koponen & Pehkonen, 2009). Therefore, these considered word associations capture connections between words when individuals know that these words are important to understand each other, but perhaps before they can fully explain why this is.

Hence, this procedure is likely to measure placeholder sign structure, rather than a fully developed concept network. This discussion reflects the distinction made between sign network studies in the literature review (see section 2.6.2) On the continuum based on the conditions for drawing edges between signs defined in the literature review, considered word associations would sit between the very loose condition of free word associations (e.g., De Deyne & Storms, 2008a; Morais et al., 2013; Steyvers & Tenenbaum, 2005) and the stringent conditions applied for concept networks (Koponen & Pehkonen, 2009) (see Figure 3.5).

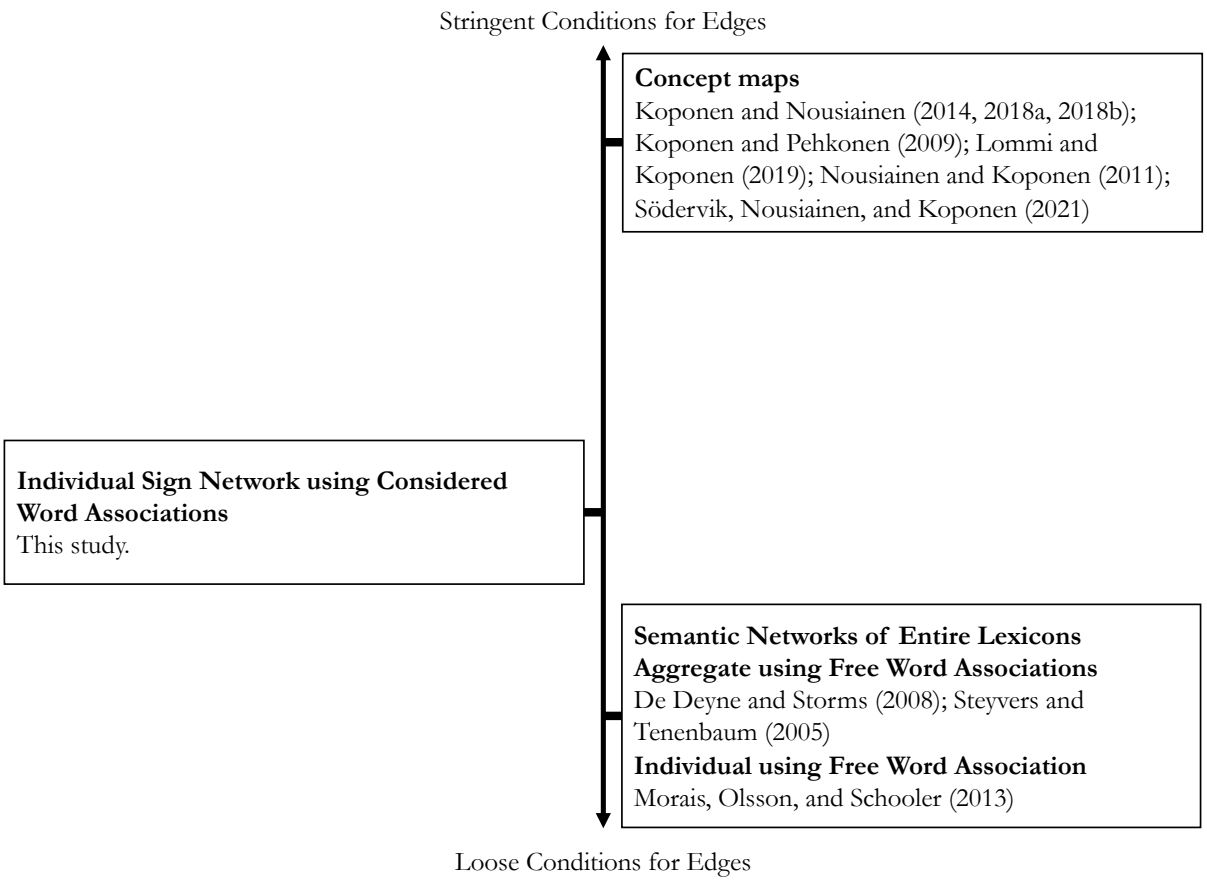


Figure 3.5 Continuum of conditions for drawing edges between signs. The most stringent conditions define concepts maps. The loosest conditions are free word associations. Placeholder sign structure is likely to be revealed by considered word associations.

Without a time-limit to give responses, it was possible the total data collection time could quickly become burdensome if the participants spent too longer considering their responses to each stimulus word. Hence, a limit of five response words for each stimulus word was set for two of the participants (one physics and one psychology student). However, any limit – time or number of words – has the potential to distort the network structure, particularly the in and out degree for vertices, by limiting the number of connections (edges) that can be made between a stimulus and response word. Hence, the other two participants were allowed to list as many responses as they thought necessary. However, these participants were told the aim of the task was not to list as many words as possible.

The response times for each stimulus word were recorded (Table 3.5) to allow comparisons between each method. These times indicate the data collection timeframes were reasonable for both the limited and unlimited participants, with all data collection times taking place within four one-hour study sessions. However, Figure 3.6 illustrates that physics and psychology participants often respond with more than five words to a stimulus word when they are given

the opportunity to do so, and typically give the maximum number of responses when they are limited to five responses. Hence, in the case of limited response, the out-degree is distorted (see Figure 3.6) and other network parameters (e.g., average shortest path length, L, clustering coefficient, C, and the diameter, D) are also likely to be inaccurately measured. Therefore, the pilot study suggests limiting participants to a certain number of response words is not preferable, given that the response times are reasonable.

Table 3.5 Total number of iterations and response time for each participant. Note novice participants were studying A-level Psychology without Physics and expert participants were studying A-level Physics.

Participant	Group	Limit to Number of Responses	Total Iterations	Total Number of Stimulus Words	Total Response Time (min)
1	Novice	5	3	64	72
2	Novice	Unlimited	3	83	204
3	Expert	5	3	59	150
4	Expert	Unlimited	3	132	235

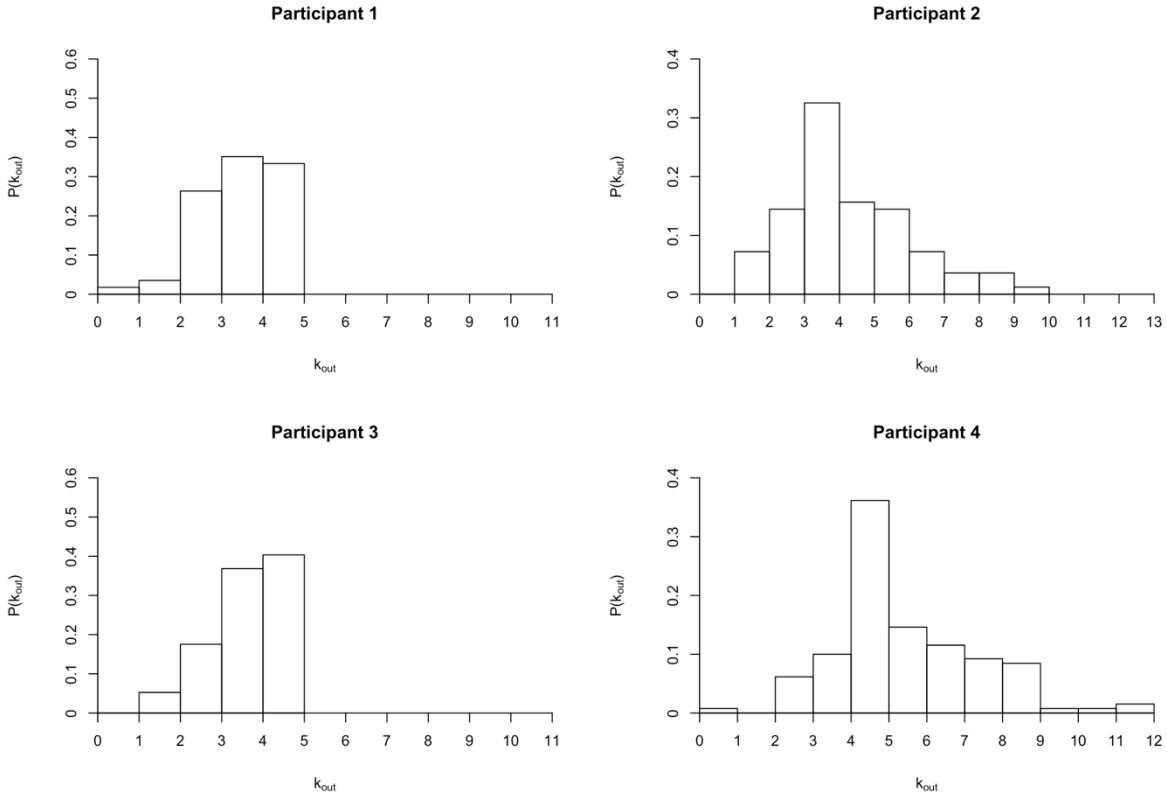


Figure 3.6 The out-degree, k_{out}, distributions for the four participants in the pilot study. Participants 1 (Physics) and 3 (Psychologist) are restricted to five responses for each stimulus word. Participants 2 (Physics) and 4 (Psychologist) were allowed to give an unlimited number of responses. Comparing the distribution demonstrates that participants often respond to stimulus words with more than five response words. Hence, limiting responses to five distorts the out-degree distributions. This is clear from the distributions for participants 1 and 3.

To enable data collection to take place within the allocated four study periods, the pilot study limited the number of iterations completed by the participants to three. The mean number of stimulus words presented to participants across the three iterations was 85. This is well below the 1000 words that would give the 66% coverage of the 1500 words[6] in a mechanics domain of knowledge, which the simulation work carried out by Morais et al. (2013) suggest is required for accurate parameter estimates. Although, the estimates of key network parameters in the pilot study (see Table 3.1), are consistent with values obtained by Morais et al. (2013) for their individual networks and also for aggregate networks (Steyvers & Tenenbaum, 2005) which indicates they may not been underestimated, as was the case for early iterations (low word coverage) in the simulation work. However, these simulations also demonstrate that parameter estimates from sample networks become more accurate as the number of iterations increases. Hence, the pilot study indicated that it would be beneficial to attempt more iterations of word association collection in the main study to obtain more accurate estimates of network parameters.

Furthermore, one of the aims of this research is to establish whether individual's sign networks have a scale-free structure, which includes highly connected hub-words. This requires a power-law model to be a good fit to network degree distributions and the degree exponent (α) to be within 2 and 3 (Barabási & Pósfai, 2016). A power-law model is a good fit to all the sign networks measured in the pilot study (see Table 3.2). However, there is a large range in the uncertainties of the estimated values of α, often below 2 and above 3. Hence, the analysis of the network in the pilot study does not establish with certainty whether these networks are scale-free. Increasing the number of iterations would increase the sampling of the tail of the degree distribution (Morais et al., 2013) and, hence, increase the range that the power-law model is fitted to the data. This could reduce the uncertainty in the estimates of α (Clauset et al., 2009).

Moreover, Broido and Clauset (2019) argue that claims about whether a network is scale-free are often made without stringent checks, which should include rigorous tests of whether other distributions (e.g. power-law with exponential cut-off, logarithmic, and exponential) can model the data. These tests have not been carried out in this pilot study but, for this reason, will be required in the main study. Clauset et al. (2009) describe a principle statistical framework of

[6] Note the number of words in a mechanics domain of knowledge is estimated by counting the number of unique word occurrences in the mechanics chapters of the textbook (Breithaupt, 2015) that accompanies the mechanics section of the A-level Physics curriculum (AQA, 2015b).

tests that can robustly determine whether a power-law model is best fit to a distribution or whether other models are equally as good or better. These tests are more reliable if a data set samples more of the tail of a distribution (in the case of networks, high vertex degrees), where the divergence for power-law behaviour is greatest. For instance, using the framework proposed by Clauset et al. (2009), Morais et al. (2013) find, for three of their networks, that a power-law with an exponential cut-off is a better fit to the degree distribution. A power-law model with an exponential cut-off implies that words with a high-degree (i.e. hub-words) would have a lower degree than would be expected in a pure power-law model (Morais et al., 2013). These words are found in the extreme tail of the degree distribution, which is more accurately sampled when a greater number of response words are presented as stimulus words. Hence, the pilot study suggests that more than three iterations of word association collection are needed to obtain estimates of α with reduced uncertainty and to robustly establish whether a power-law model is the best fit to the degree distribution. This would allow the main study to establish with greater confidence whether any measured sign network was scale-free.

3.3.3.2.2 Main Study Procedure

The testing procedure used in the main study closely followed the pilot study procedure. Initially, participants were presented with the ten seed words (one-by-one) and asked to list the words that helped them understand these seed words. No time or word limit was set for participants' responses. This was justified by the results of the pilot study that suggested that response for an appropriate number of iterations could be collected in a reasonable timeframe, without the need to set conditions that would distort important network metrics (e.g., vertex out-degree). However, participants were asked to consider their choices carefully and not to simply list as many words as they could. This instruction was also given to minimise the risk of chaining, which occurs when participants begin to respond with association to the previous response word and not the initial stimulus word (Maskill & Cachapuz, 1989).

The findings of the pilot study demonstrated that more than three iterations of word-association may be required to characterise the nature of the vertex degree distributions and to obtain more accurate estimates of network parameters. The pilot study also suggested these increased numbers of iterations of word-associations may be difficult to collect using the supervised paper-based method. For example, approximately 180 stimulus words would have to be included in a fourth interaction for participant 4. Hence, a supervised session of around three hours would have to be arranged for this data collection to take place. Alternatively, this session could be split (e.g., three one-hour sessions). However, in both cases, this would result in a

significant loss of study time for a participant if these sessions took place within the school day. This is unlikely to be ethically acceptable when the time a participant would have already committed to the study during school hours is considered. Also, these sessions would be difficult to timetable for the potential number of participants in the main study (21 at the start of the data collection process). Moreover, all the participants in the pilot study asked if they could take the paper survey away because they would prefer to complete it in their own time. They also indicated they would be willing to commit to the extra time required for a more iterations if they were able to do this.

Consequently, an online data-collection method, using the Jisc website (Jisc, 2022), was used in the main study to provide the flexibility asked for by the participants of the pilot study, by allowing them to complete the word-association task at times and in locations that are convenient for them (Lefever, Dal, & Matthíasdóttir, 2007), and to avoid the ethical and logistical issues of timetabling supervised sessions during the school day. It is also possible that the online method encouraged more students to become participants, and to continue to participate, because they could spread their commitment over a longer period (e.g., three twenty minute session in a week, rather than a one-hour long session). All these measures made collecting further iterations of word associations possible in the main study and, hence, provided a strong justification for the use of online data collection.

Online data collection also offered logistical advantages in data storage and processing. For instance, significant time did not have to be spend manually transferring data, collected via pen and paper, into an electronic format (Granello & Wheaton, 2004). Moreover, data could be stored electronically, and password protected.

However, online data collection methods are also open to abuses (Granello & Wheaton, 2004; Lefever et al., 2007; Topp & Pawloski, 2002). First, participants may not answer truthfully. Second, participants may use resources (e.g., textbooks) to help them think of their word-associations, rather than relying on their memory. Third, there is no guarantee that a word-association task has been completed by the intended participant. There is a chance a participant may have asked a friend to complete the task. Fourth, participants may not be able to ask for clarification if they do not understand the instruction in the task.

In the case of the first abuse, Lefever et al. (2007) argue this also true of any paper-based task. So, for both paper and online data collection, the participants need to be trusted to answer honestly. In the case of the second and third abuses, the high-level of control afforded by

supervising the pilot study participants for the paper-based collection ensured that these are not committed. However, again, these concerns could be eliminated if the participants were trusted to act appropriately. The participants in the main study opted-in to participate, were free to withdraw at any stage and have been informed that the results of this task will not be used as a measure of their academic attainment. Therefore, they had little incentive to abuse the online system in the manner identified by Granello and Wheaton (2004), Lefever et al. (2007) and Topp and Pawloski (2002). This assumption is supported by the word-association data, which shows little evidence of only producing responses that might be expected if participants were using resources to help them answer. Furthermore, all the participants completed several surveys over several months. Hence, it would have been challenging for a participant to ask a friend to take on this data collection responsibility for the entire process.

To provide clear guidance for the participants, the instructions for completion of the survey were set out at the start of phase of online data collection. However, the researcher was not present when these surveys were complete. Hence, the lack of an opportunity for participants to ask for clarification of instructions when taking a survey was a disadvantage for the online method. However, for each participant, the DPhil researcher needed to use data from an iteration of word-association to identify the stimulus for the next iteration. Hence, throughout the data collection process, there were regular opportunities for the researcher to check whether participants were following the instructions for completing the surveys and to intervene if he felt they were not. All participants generally followed the instructions and complete the surveys correctly, some did not fully adhere to the requirement not to use compound words or phrases. The procedures developed to deal with these incidences are described in section 3.3.4.2.

3.3.3.2.3 Participant Attrition

The pilot study demonstrated that the time commitment for participants to complete three iterations of word association collection was reasonable. However, the results of the pilot study also suggested that four iterations of data collection in the main study would be preferable. As has been argued in previous sections, this number of iterations could improve the accuracy of network parameter estimates and robustness of any conclusions regarding the best fitting model to the network degree distributions. For the participants who were allowed to give an unlimited number of responses to a stimulus word, the mean number of responses was 3 (equal to the out-degree, k_{out}). However, in the main study the mean number of responses was higher for most participants (mean $k_{out} = 4$). Hence, in most cases, by the 2nd or 3rd iteration, participants were being asked to respond to over 150 stimulus words. The pilot study suggested a participant

took approximately one minute to respond to each stimulus word. Using this time estimate for each stimulus word, it was clearly not feasible to ask participants to spend over two hours responding to surveys of over 120 words. Hence, in later iterations, which included large numbers of words, stimulus words were presented to participants in 25-word online surveys. This made the data collection time for each survey more manageable but spread data collection times over a longer period. Hence, data collection times were longer than predicted. This had two implications. First, data collection took place not just in a limited time after the A-level Physics student had just taught the mechanics topic, but throughout a significant period of their A-level studies (from April 2019 to March 2020, see also Table 3.4). Second, participant attrition was higher than expected (see Table 3.4), and nine participants withdrew at this stage following stage 3 of the cooling off protocol.

3.3.4 Student Sign Network Construction

3.3.4.1 Constructing and Finishing a Directed Sign Network

Directed networks for each participant were constructed using the cue-response word associations for each participant. In these networks, out-going edges were drawn between a stimulus word and each of the associated response words. Hence, an in-going edge for any word, or vertex, in the directed network indicates that that word has been given as a response to a particular stimulus word. Figure 3.7 shows a small part of a hypothetical sign network to illustrate the directed network construction technique.

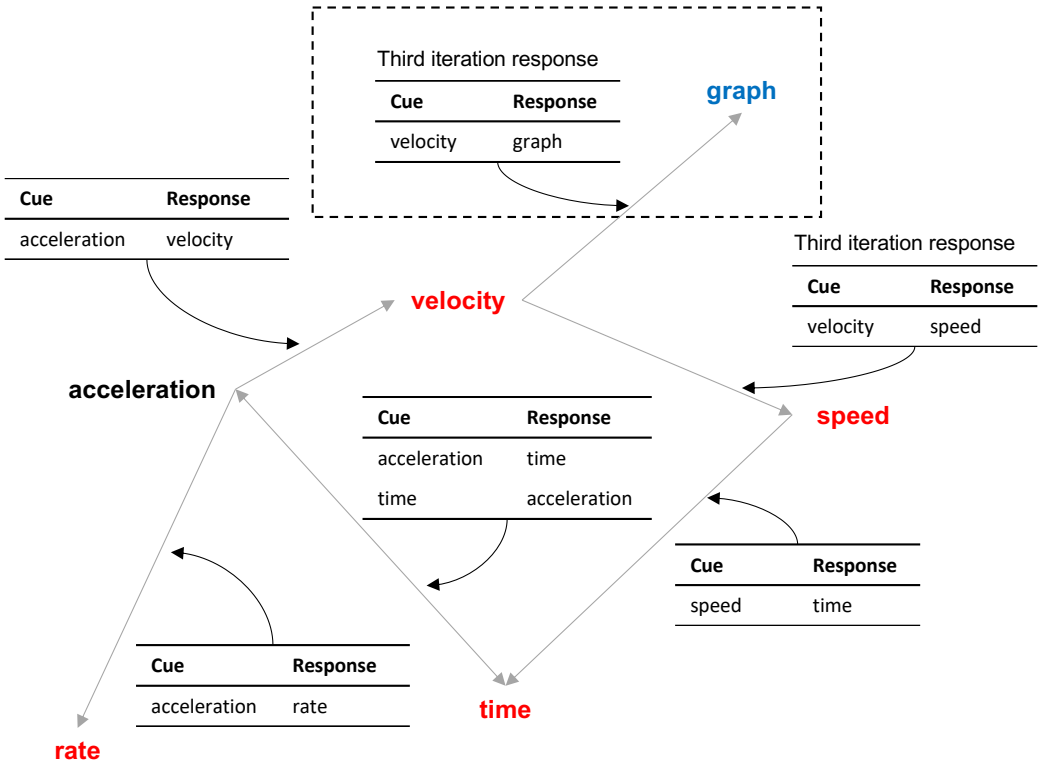


Figure 3.7 Construction of a directed semantic networks using cue and response words. First, second and third iteration cue words are shown in black, red, and blue respectively. Note the vertex in the dashed box would be removed by the network finishing technique because it is a third iteration response to a second iteration cue word with no link back into the semantic network.

Figure 3.7 also highlights a problem with the iterative snowball sampling techniques, which arises due to the limited time available to collect stimulus-response word associations. These time-constraints will always limit the number of possible iterations (to three in the case of the pilot study). Hence, the last, or n^{th}, iteration response words are not presented as stimulus words in a n+1 iteration and, therefore, response words for these stimulus words are not collected. This will distort any sign network because a large proportion of words, or sign

vertices, will not have out-going edges. For example, the out-degree distribution will show all the nth response words vertices with an out-degree of zero ($k_{out}=0$) solely because these words have not become stimulus words in the n+1 iteration. This may also distort many of the measures of network structure (e.g., shortest average path length, L, and diameter, D) as the nth iteration vertices are dead ends.

In the pilot study, a 'fix' was used to address this problem. This fix assumed by the n^{th} iteration (3^{rd} in this case); the stimulus-response word associations were no longer part of a sign network that was closely associated with the mechanics domain of knowledge. This assumption allows any network to be finished by only keeping n^{th} iteration response words that connect back to the network via an out-going edge (i.e., a word given as a response in the nth is already present in the network). Hence, only response words that have already been used as stimulus-words in previous iterations are kept in the n^{th} iteration. Again, this network construction technique is illustrated in Figure 3.7.

Unfortunately, whilst this technique does allow a sign network made to be finished, it does have an impact on the structure of the network. For example, removing any n^{th} iteration response words that have not been previous presented as stimulus words, reduces the number of out-going edges (and hence the out-degree, k_{out}) for n^{th} iteration stimulus words. This is illustrated by another hypothetical section of sign network shown in Figure 3.3 and the values of k_{in} (vertex in-degree) and k_{out} for the n^{th} and $(n-1)^{th}$ vertices in this network section shown in Table 3.6. Whilst this is not demonstrated by the k_{in} values in Table 3.6, it is also possible that the k_{in} values for the $(n-1)^{th}$ vertices will be lower than measured, as many n^{th} iteration response words will not been presented as stimulus words, and these words are likely to include n^{th} iteration stimulus words as their response words. Put more simply, there is often a stimulus-response word symmetry with edges between these words in both directions. For instance, *electron* is likely to be a response for *atom*, and *atom* is likely to be a response for *electron*. However, it is not possible to quantify the reduction on k_{in} because the n^{th} iteration response words have not been given as stimulus words and, hence, the responses to these words are not known.

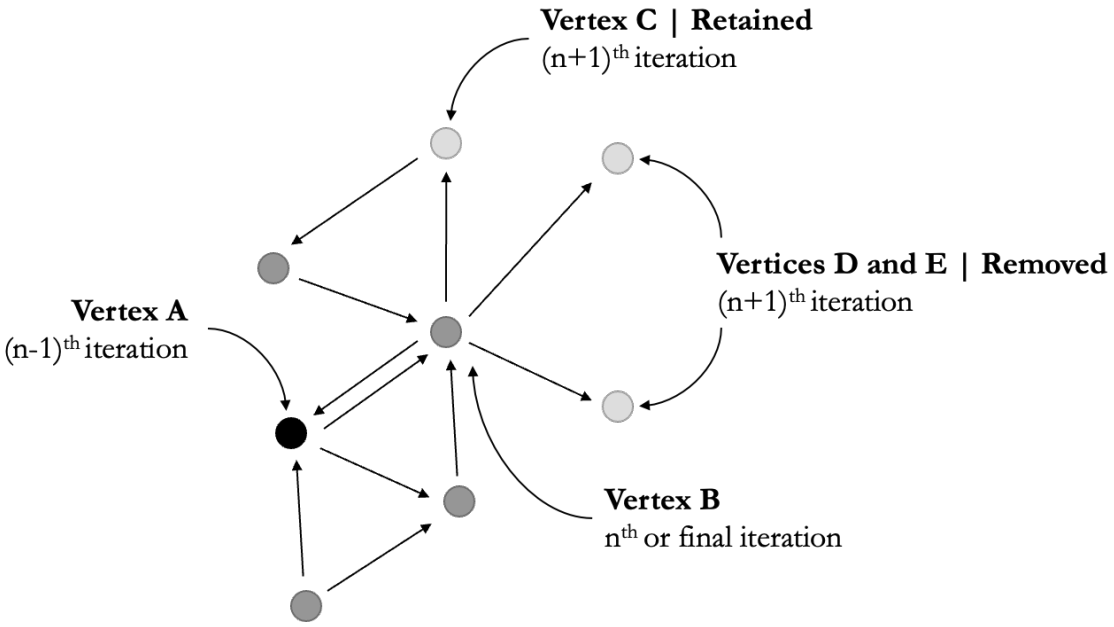


Figure 3.8 A hypothetical section of a sign network. The $(n+1)^{th}$ iteration vertices are shown in light grey. These vertices represent response words to stimulus words in the final n^{th} iteration (dark grey vertices) of word association collection. However, these words (light grey vertices) have no response words because they have not been presented as stimulus words in an $(n+1)^{th}$ iteration. Vertices D and E are removed when the network is constructed because they do not connect back into the network, while vertex C is retained. Table 3.6 The effect of this network finishing technique on the in and out degrees for vertex A (stimulus word in penultimate $(n-1)^{th}$ iteration) and vertex B (stimulus word in final n^{th} iteration).

Table 3.6 To finish the networks in this study a 'fix' is applied during their construction (see section 3.3.4.1). This table illustrates the effect of this fix on the in-degree and out-degree for vertex A (stimulus word in penultimate $(n-1)^{th}$ iteration) and vertex B (stimulus word in final n^{th} iteration) (see Figure 3.8).

Vertex	Before Fix		After Fix	
	In-degree (k_{in})	Out-degree (k_{out})	In-degree (k_{in})	Out-degree (k_{out})
A	2	2	2	2
B	3	3	3	1

3.3.4.2 Vertex Decisions

When selecting the stimulus words for the next iteration and constructing the networks, two issues arose that required decisions to be made about how a participant's response word should be represented as a sign, or vertex, within the network. First, a participant may respond to different stimulus words with the same word but in a different grammatical tense. For instance, participant 12P03 responded to the stimulus word *optic* with *bend* (12P03: *optic* | *bend*) and *refraction* with *bending* (12P03: *refraction* | *bending*). Similarly, a participant may give a response that is a plural of a word what has already been a response, or a different class of a related word (e.g., *measure* versus *measurement*). Hence, a decision whether to merge these similar responses into a single sign (single vertex) or keep them as separate signs (two vertices) was needed.

Second, some participants responded to stimulus words using compound words (e.g., 12P03: *electron* | *matter-wave*, 12P03: *threshold* | *work-function*) or phrases (12P04: *graph* | *from a table*), despite being asked not to in the survey instructions. Again, a decision was required about how to represent these compound words and phrases in the network. For instance, matter-wave could be represented as a single sign (one vertex – matter-wave) or as two signs (first vertex – matter, second vertex – wave). The procedures for making these vertex decisions consistently and the rationale behind them is outlined in the sections below.

3.3.4.2.1 Possible word merges

The vertex and edges represent the basic units of analysis in any network and, therefore, determine the level of analysis that can be achieved with the network (see section 2.6.1.1.1). Hence, when decisions were made about whether to merge vertices in sign networks, these were in essence decisions about the basic units of analysis in the network and what level of analysis is possible. However, these decisions were not necessarily driven by a desire to achieve the finest grain in the sign network. Rather, they were guided by the aims of this study, which was to compare the sign networks of individuals to analyse the role of signs-sign structures in conceptual development. Merging incidences of plural, tense and word class inflections was considered most consistent with these aims. The rationale for why each of these word groups were merged is given below.

3.3.4.2.1.1 Plurals

Plurals were most common form of possible words merges and were merged into a single vertex representing one sign (e.g., *electron* and *electrons* were both represented with a single vertex; *electron*). These merges reflect a decision not to distinguish between plural signs in this analysis of an individual's sign structure, because this distinction was not considered to be pivotal in any bootstrapping of a scientific concept and could give rise to a misleading sign structure. For example, when examining the sign network around the word *atom*, it would be more salient to note that two individuals have made connections to *electron*, rather than distinguishing between the fact that one individual connected *atom* with *electron* (singular), and the other with *electrons* (plural).

Moreover, distinguishing between plurals might also distort the network structure because words often form clusters within individual sign networks. For instance, the sign networks measured in the pilot study typically had clustering coefficients between 0.22 and 0.33. Again, using the example of *atom,* this suggests that if *atom* connects with *electron* and *proton*, then it is

likely that *electron* and *proton* will also be connected. As had been said above, plural variations in response to different stimulus word are very common. Hence, in this example, a participant's response could be *atom* | *electron* (singular) and *proton* | *electrons* (plural). If the plurals were not merged, *atom*, *proton*, and *electron*, would not form a cluster leading to a different, and arguably misleading network structure (see Figure 3.9). Furthermore, if *electron* and *electrons* are kept as separate signs, this creates two vertices with an in-degree of one, rather than a single vertex with and in-degree of two. Again, the latter case, where the sign *electron* is connected to the words *atom* and *proton* and has an in-degree of two seems to be intuitively more informative about which signs are considered by an individual in this example to help them understand *electron*. Hence, this is a more appropriate level of analysis to investigate the sign structure needed to bootstrap an understanding of these concepts. Steyvers and Tenenbaum (2005) adopt a similar approach when they construct their sign networks, arguing that ignoring difference between verbs and noun forms of words makes their analysis. They argue that whilst this decision is:

> *'[…] insensitive to a great deal of interesting linguistic structure, [it enables] us to see the forest for the trees—to pick up on general properties of meaning in language that might be missed in [a] more fine-grained [network] […]' (Steyvers & Tenenbaum, 2005, p. 54)*

Like Steyvers and Tenenbaum (2005), this study adopts a general approach that constructs networks with a more course grained structure, but are more likely to reveal important network structure in particular domains of knowledge. Hence, for the reasons listed above, plural inflections of words were merged to form a single sign vertex.

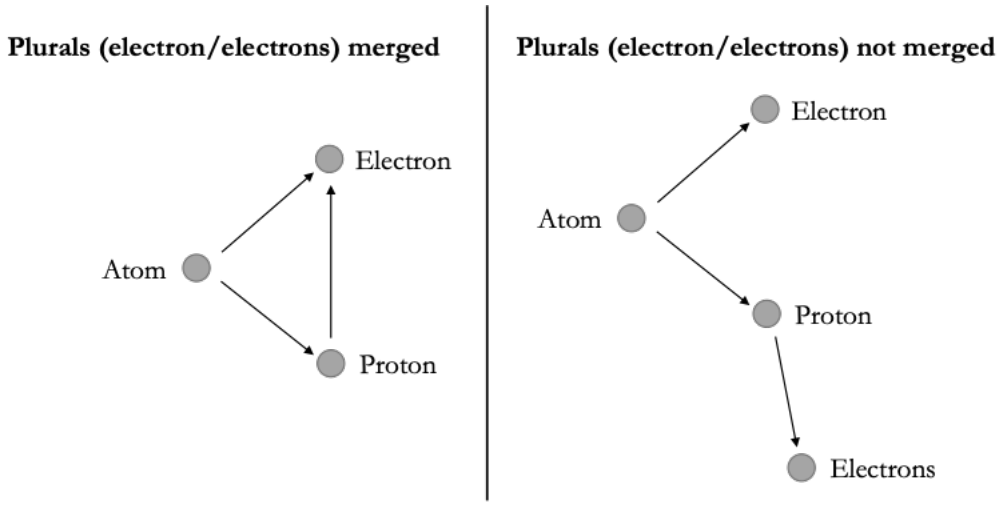


Figure 3.9 The effect of merging plural inflections of words on a network structure. Merging plurals forms a cluster whilst preserving the plural distinction does not.

3.3.4.2.1.2 Different tenses of the same word

Different tenses of the same word are also frequently given as responses to different stimuli words. As with plurals, these tense inflections of the same word were also merged. The rationale for this decision follows the same argument given above for plurals, where it was not considered appropriate to distinguish between singular or plural forms of a word. Similarly, for different tenses of the same word, the primary concern when measuring and comparing sign structures was not distinguishing between whether a process or action took place in the past, present, or future. Rather, recognising that the same connection exists between, for example, *light* | *reflecting* and *light* | *reflected* in two different networks, and preserving the clustering structure between *light* | *reflected* and *refraction* | *reflects* in an individual network, was considered a more informative level of analysis to investigate the possible bootstrapping of concepts. This is consistent with the approach of using a courser grain of analysis to make any sign network structure more visible.

3.3.4.2.1.3 Different word classes

Response words to different stimulus words that belong to different word classes but share the same word stem (e.g., *measure* and *measurement*) occur less frequently than plural and tense inflections. However, these incidences are not uncommon and occurred across all participants. Hence, a coherent and systematic procedure was also needed to decide whether to merge different word classes with the same word stem into a single sign vertex. Traditional definitions of word class (e.g. noun: '[…] a class of people, place or thing,' (OED, 2018); verb: 'an action, state, or occurrence, and forming the main part of the predicate of a sentence', (OED, 2018)) suggest different word classes refer to different ontological categories. This suggests responses such as *construct* and *construction*, *measure* and *measurement* should be treated as different and, hence, keep as separate sign vertices. However, Crystal (2003) highlights several examples which illustrate difficulties in applying traditional definitions of word classes to functional uses of language. For instance, in the case of nouns, there is some doubt whether abstract qualities and actions (e.g., *measurement*) can be described as 'things' and should, therefore, be belong to this word class. Hence, different word classes might not strictly indicate that different signs are required in a sign network. For example, *measure* is a verb and is defined as 'Ascertain the size, amount, or degree of (something) by using an instrument or device marked in standard units', (OED, 2018). In contrast, *measurement* is a noun and is defined as 'the action of measuring something', (OED, 2018). Therefore, using separate sign vertices for *measure* and *measurement* defines a level of analysis that distinguishes between 'ascertaining the size' and 'the action of

ascertaining the size'. As was case for plural and tense inflections of response words, this level of analysis could over-complicate a sign network, hiding and distorting the structure, rather than enabling a clearer and more informative analysis. Again, this is consistent with the approach used by Steyvers and Tenenbaum (2005).

The decision to merge verb and noun forms of response words is also justified by the process of nominalisation (Halliday, 2004). Halliday identifies this as a common rhetorical technique in scientific writing, where a complex process is given a name – or noun form – to linguistically package up the process into a single semiotic entity which is easier to use grammatically. For example, Isaac Newton employs this technique when sharing his ideas about how light travels in different mediums: 'the bending of light as it crosses a boundary between materials with different densities' becomes 'refraction' (Halliday, 2004). This nominalised form is then often used a verb in its own right (e.g., 'the ray refracts'). Hence, there is certain level of fluidity between noun and verb forms of scientific words, which justifies the decision to merge noun and verb forms of response words to form a single sign vertex (e.g., in the above example, *refract* and *refraction* would be merged).

Crystal (2003) also points out that there is also a certain amount of fluidity, or fuzziness, in the class of adjective. For example, an adjective is used to 'qualify a noun' (Crystal, 2000, p. 206) and 'express some feature or quality of a noun' (Crystal, 2000, p. 211). However, in a sentence, a noun can act as an adjective (e.g., 'The London station'). In this example, *London* is no longer strictly a noun or an adjective. It '*forms a mixed word class*' (Crystal, 2000, p. 211). This can also apply to verbs. For instance, in the sentence 'the reflected wave', *reflected* is functioning as an adjective. Again, this ambiguity in word class reinforces the decision to merge response words with the same word stem to form a single sign vertex.

During the construction of the participants' networks, several procedures were carried out to ensure that words were being merged according to the principles outlined above and consistently across different participants. These stages are shown in Figure 3.10. Initially, response words are checked to see if they were an exact match to any stimulus words using a R script and the text matching capabilities of the *Quanteda* R package for managing and analysing textual data (Benoit et al., 2018). Any response words that were not an exact match were then checked to see if they share a word stem with a stimulus word. The word stems were determined by the *Quanteda* R package and, again, a R script was used to identify matches. At this point a decision was made about whether to merge the words, guide by the rationale for word groups outlined above. If they could not be merged and it was not the last iteration, a

new stimulus word was added to the next iteration, or if it was the last iteration, the word was removed from the sign network (see network finishing section 3.3.4.1).

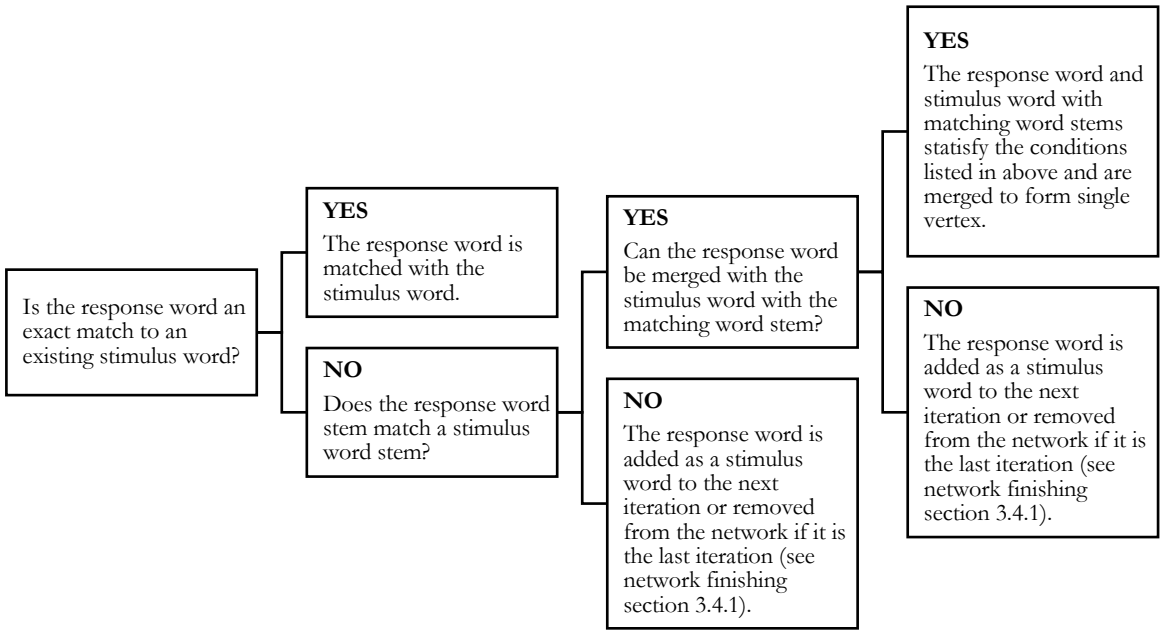


Figure 3.10 The procedure used in the study to identify word within a network that should be merged according to the principles outlined above. An R script using the Quanteda R package for managing and analysing textual data was used to identify matches between new response words and existing stimulus words and between stemmed words.

The merged word tables for all participants are shown in Appendix 10.2. The checking procedures did highlight some cases when exceptions needed to be made to the merging rules. These exceptions are also listed in Appendix 10.2 in Table 10.4, Table 10.5, and Table 10.6, with the reasons why they are exceptions.

3.3.4.2.2 Compound Words and Phrases

The instructions in the online survey asked participants to not use compound words or phrases. Despite this instruction, most of the participants responded to some stimulus words in ways that could be interpreted as using a compound-word or phrase (e.g., 12P03: *electron* | *matterwave* or 12S09: *parent* | *caregiver*). These incidences were defined as a combined response. When they occurred, combined responses raised the question of whether a participant viewed the compound word or phrase as one sign, and therefore it should be treated a one vertex, or whether the participant has simply ignored the instructions and the compound word or phrase should be split into two vertices, representing two signs. There were also incidences when compound words or phrases were presented as sequential responses to the same stimulus word (e.g., 12S03: *task* | *hard*, *task* | *work*). On these occasions, these separate responses could be

left as separate sign vertices or combined into a single sign vertex. Hence, a procedure was developed to decide systematically and coherently how to represent compound word or phrases as vertices in the network, when they were presented either as a combined or sequential response. This procedure and the rationale for it are described below.

The presentation (combined or sequential) of a participant's response was the primary factor in deciding how to represent a compound word or phrase. For instance, participant 12P03 gave a combined response of *matterwave* to the stimulus word *electron*. In this case, it is feasible to treat the combined response as a single sign vertex (i.e., *electron* | *matterwave*), or as two separate sign vertices, both connected to the stimulus word (i.e. *electron* | *matter* and *electron* | *wave*). However, the latter case requires altering the participant's response, which implies a greater amount of interpretation of what the participant intended from the researcher. Similarly, changing 12S03 sequential response of *task* | *hard* and *task* | *work* to *task* | *hard-work*, requires more interpretation of 12S03's intention than leaving the responses to *task* as *hard* and *work*.

To avoid these incidences of guessing what a participant really meant, a conservative approach to altering combined (word presented a without a space) and sequential compound word (word presented with a space or carriage return) or phrase responses was adopted: combined responses were represented by a single sign vertex, and sequential responses were represented by two or more vertices. This approach ensured participants' original response determined the vertex decisions in the first instance.

However, there were associations given by participants when respecting the combined or sequential presentation that would have led to clearly nonsensical connections in a sign network, compared to if a change was made. For instance, participant 12P03 gave a sequential response of *milky* then *way* to the stimulus word *galaxy*. In this case, *galaxy* | *milky-way* makes more sense as a connection in the sign network than *galaxy* | *milky* and *galaxy* | *way*. Hence, a protocol, show in Figure 3.11 was developed to identify when it was deemed necessary to override the principle of respecting the presentation of the participant response. The rationale for this protocol draw upon the arguments made by Murphy (2002) in a discussion about how conceptual combination are formed, which is in turn informed by extension analysis.

Extension analysis states that the extension of a word is the group of objects, or set, that it refers to (Murphy, 2002, p. 444). Hence, the extension of the word *dog* is the set of all dogs, and the extension of the word *yellow* is the set of all yellow objects. Hence, *yellow dog* can be understood as the intersection of the dog and yellow set. More simplistically, yellow dog is

formed by connecting the extensions of dog and yellow. So, if *yellow dog* is used as a response to a stimulus word *Labrador* it is reasonable to connect *Labrador* | *yellow* and *Labrador* | *dog* to avoid the use of the compound word.

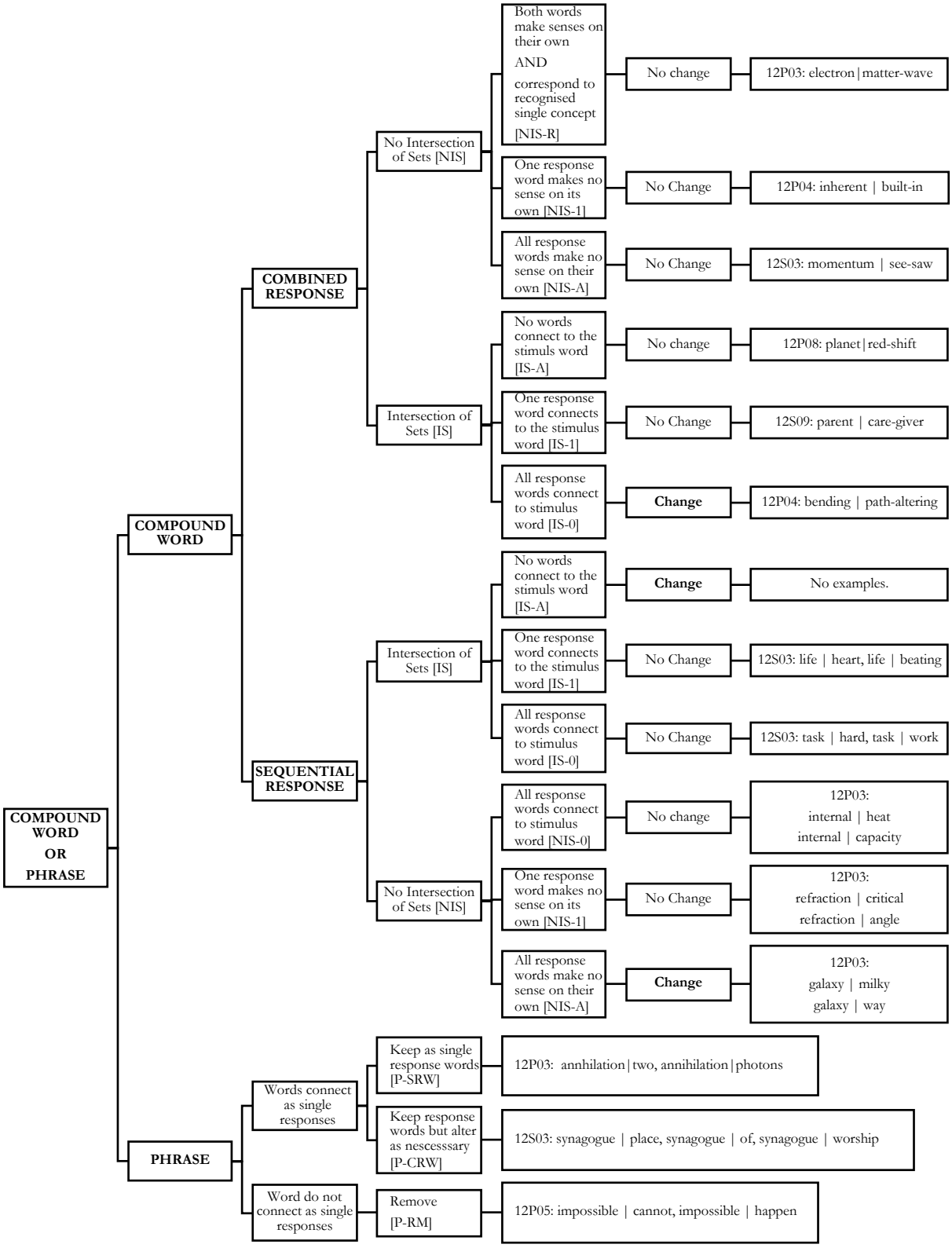


Figure 3.11 The procedure used to decide when it was necessary to override the principle of respecting the initial presentation of a participant's response word. Note the codes in the figure (e.g. NIS-A) are used to identify the reason for representing a compound word or phrase as a vertex or as vertices in a network listed in Table 10.7, Table 10.8, and Table 10.9.

As the above example illustrates, intersection of sets works well for predicating adjective-noun compound words or phrases, which can be tested by seeing if *yellow dog* can be rephrased as a sentence; *the dog is yellow* (Murphy, 2002, p. 445). There are examples of participant responses that are adjective-noun compound word or phrases. For example, 12P03 associates *race* with *downforce*. This response passes the sentence test (*the force acts down*) and is therefore an example of an intersection of sets. Hence, this combined response could be split into two sign vertices (*race* | *down* and *race* | *force*).

In contrast, Murphy (2002) points out that relative adjective and non-predicating adjectives compound words and phrases pose a problem for an extensional approach, and therefore the interpretation that these compound words or phrases represent the intersection of sets. For instance, *large dog* (relative adjective) implies a comparison between other objects in the set, in this case dogs, rather than an absolute property of a dog. Similarly, *nuclear physicist* (non-predicating adjective) cannot be interpreted as the intersection, of nuclear objects and physicists. Note that examples of non-predicating adjectives compound words or phrases fail the sentence test (i.e., the *physicist is nuclear* does not have the same meaning as *nuclear physicist*). Participant 12P03's sequentially presented response *milky way* to the stimulus word *galaxy* is an example of non-predicating adjective compound word that fails the sentence test (i.e., *the milky is way* is nonsensical). Hence, this compound word cannot be explained as simply the intersection, or connection, of *milky* and *way*. This provides the justification for changing this association in the sign network from connections between *galaxy* and two sign vertices *milky* and *way* (*galaxy* | *milky; galaxy* | *way*), to a single connection between *galaxy* and a single sign vertex *milky-way* (*galaxy* | *milky-way*).

Figure 3.11 and the above discussion outlines a second layer to the protocol for addressing participants' use of compound words and phrases, which checks whether the first principle of respecting the presentation of response generates nonsensical connections in a sign network and should therefore be overridden. In general, an association was queried if it was a combined response and could be defined as an intersection of sets (e.g., *downforce*), or the opposite, if it was sequential response and not an intersection of sets (e.g. *milky-way*).

A third and final layer of checking followed this intersection of sets check. This final layer identified associations that did not fit neatly into two general categories which have just been defined. For example, a sequential response could appear to be a compound word or phrase and not be an intersection of sets, but one or all parts of the sequential response could be a reasonable connection with the stimulus word (12P03: *refraction* | *critical*, *refraction* | *angle*). In

this case, changing the association to *refection* | *critical-angle* again involves a greater degree of interpretation of the participants' response, than leaving the association as *refraction* | *critical*, *refraction* | *angle*. Hence, the first principle in the protocol of being primarily guided by the presentation of a participants' response implies this association should be left as it was presented. Similarly, but following the reverse of this logic, combined presentation responses where only changed if they could be defined as an intersection of sets and both parts of the compound word or phrase could be said to connect to the stimulus word (12P04: *bending* | *path altering* becomes *bending* | *path*, *bending* | *altering*).

Finally, there were combined and sequential responses that could be clearly identified with recognised scientific or mathematical coefficients, phenomena, laws, or theories. When these responses were given sequentially, they were combined into a single sign vertex (e.g., 12P03: *statistics* | *standard*, *statistics* | *deviation* was changed to *statistics* | *standard deviation*).

Figure 3.11 also shows the procedures for dealing with phrases. If all the words used in the response phrase appear to reasonably connect with the stimulus word, then both response words became vertices (e.g., 12P03: *annihilation* | *two photons* became *annihilation* | *two*, *annihilation* | *photons*). Any response words did not sensibly connect were removed (e.g., 12S03: *synagogue* | *place of worship* became 12S03: *synagogue* | *place*, *synagogue* |*worship*). If none of the response words in the phrase sensibly connected with the stimulus word when they are presented sequentially, the phrase was removed (e.g., 12P05: *impossible* | *cannot*, *impossible* | *happen*). The changes were made to participants compound word and phrase responses are shown in Table 10.7, Table 10.8, and Table 10.9 in Appendix 10.3.

3.3.5 Summary

This study aims to measure and then compare sign networks for individuals in a mechanics domain of knowledge to investigate the role of sign structure in conceptual development. This aim is captured by research questions stated in section 2.7. This section has set out an adapted version of the Morais et al. (2013) snowball sampling method which was used measure individual sign networks and address the stated research questions.

A pilot study was carried out as a proof of concept of the adapted snowball sampling method. This pilot study demonstrated the efficacy of the adapted method by showing that it could sample individual sign networks in a way that was appropriate for participants in a secondary school setting. Moreover, the network parameters estimated using these sample networks were

consistent with the values obtained for networks in previous studies (e.g. De Deyne & Storms, 2008a; Morais et al., 2013; Steyvers & Tenenbaum, 2005) (see Table 3.1) and, therefore, did not appear to be significantly underestimated, which was identified as possible consequence of under sampling by the stimulation work carried out by Morais et al. (2013). Therefore, the pilot study demonstrated that the adapted snow-ball sampling method could address research question 1 and would provide networks that could be analysed to answer research question 2. Hence, this method was used in the main study. However, there are several important findings from the pilot study that informed the design of the main study. These are:

- The simulations carried out by Morais et al. (2013) suggested that accurate network parameters required word associations for approximately 66% of the words in a population to be sampled. For the mechanics domain of knowledge this is approximately 1000 words, if the number of words in the mechanics domain of knowledge is defined as the total number of unique words in the mechanics chapters of the textbook. The sign networks measured by the pilot study contain 64, 83, 59, and 132 words. Hence, more iterations of word association collection were used in the main study. This increased data collection time prompted a switch to online data collection.
- The pilot study demonstrated that the sign networks of physicists and psychologists had similar global structure but had different connections between words. The degree distributions for these networks could also be fitted with a power-law. This suggested the structure of the networks might be scale-free and contained highly connected hub-words, which account for many of the structural properties of the networks. These hub-words were different for physicists and psychologists. These results are broadly consistent with the bootstrapping theory of conceptual development proposed by Carey (2009), where the development of a concept begins with the formation of a sign structure around a key placeholder value, which then becomes the concept. The potential for empirically testing Carey's theoretical ideas using sign networks for physicists and psychologist justifies using these cohorts of participants in the main study.
- The construction of the sign networks during the pilot study highlighted an important limitation of the snowball sampling method. This limitation arose from the decision to remove response words given in the final iteration of any sampling if these words had not already been given as a stimulus word. This avoided a large number of words (vertices) being included in a final network with an unrealistic out-degree of zero and

was justified by assuming that these response words were out of the mechanics domain of knowledge. This assumption and, therefore, this method for finishing the network was adopted in the main study. However, the discussion in section 3.3.4.1 describes some possible effects on the degree distribution for a network. More widely, this discussion illustrated that while the snowballing sampling technique may be an effective way of exploring the structure of an individual's sign network, it will always sample, rather than completely measure the structure of that network. Hence, the network parameters for any network created using a snowballing sampling technique, including the degree-distributions, will always be estimates and should be treated as such (Newman, 2018).

- The analysis of pilot study participants' explanations of objects motions indicated that they may be using putative mechanical concepts like *force, energy,* and *momentum* in an undifferentiated way (these pilot findings are presented in the interview methodology chapter). Hence, the proposed mechanics domain of knowledge was expanded in the main study by including the textbook chapters *Forces in Equilibrium*, *Force and Momentum* and *Work, Energy, and Power*, which describe *energy* and *momentum* and how these concepts are related to *force.* The inclusion of these chapters changed rank order of the most frequent words in the textbook chapters, and hence, change the initial seed words used in the 0^{th} iteration of the main study. These became *velocity*, *distance*, *energy*, *acceleration*, *weight*, *force*, *speed*, *mass*, *time*, *momentum*.

Apart from the switch to online data collection and the changes to the initial seed words, the main study procedure followed the procedure outlined in the pilot study. However, the data collection and construction of the networks in the main study highlighted some issues that were not anticipated by the pilot study. For instance, the increased mean number of response words to each stimulus word for some participants, which greatly increased the data collection time for each iteration and most likely led to increased participant attrition (see section 3.3.3.2). Furthermore, the increased use of compound words, phrases, and different word inflections (e.g., plurals, tense, and class) were not often encountered in the pilot study. The procedures used to decide how these types of response should be represented in the sign-networks were described in section 3.3.4.2.

4 Results of the Sign (Word) Network Analysis

4.1 Introduction

The network construction methodology chapter has described how the sign networks for individual participants were measured and constructed, hence, creating the sign networks addresses research question 1.

> *What are the sign networks of individuals within a scientific domain of knowledge?*

This chapter addresses research question 2.

> *What are the similarities and differences between the sign networks for individuals with different levels of expertise in mechanics?*

This question is addressed in two parts. Section 4.2 establishes that physics and psychology students have different levels of expertise in mechanics but similar vocabulary knowledge. Section 4.3 analyses the similarities and differences in the networks for these two groups.

4.2 Physicists' and Psychologists' Understanding of Mechanics and Vocabulary

In this study, cohorts of students studying A-level physics and A-level psychology were chosen because, as previously argued, it was expected that these groups would have a different understanding of mechanics and might therefore have different sign networks. This hypothesis was tested using the Force Concept Inventory (FCI) (Hestenes et al., 1992). The application of this test and the results are described in section 4.2.1.

In contrast, it was assumed that these groups of students would have similar vocabularies because all participants were similar ages, were studying for A-level qualifications (hence they had all met the entry criteria of 5 GCSE grades at level 6 or above, and GCSE English Language at grade 5), and had attended the same or similar school settings prior to taking part in the study. However, it is possible that this assumption may not be correct. For instance, the entry grade requirements for A-level physics are slightly higher than A-level psychology (physics: grade 6 in Combined Science or Physics, and a grade 6 in Mathematics; psychology: grade 5 in Combined Science or any individual GCSE Science). Moreover, A-level psychology assessment includes more 'long-answer' essay-style questions (AQA, 2017), which require students to develop

writing skills. In contrast, A-level physics assessments have a greater focus on problem solving using calculations (AQA, 2022). These differences could lead to differences in how the general vocabularies of participants in different groups develop.

Any differences in general vocabulary might affect the sign networks. For instance, smaller vocabularies might limit the number of responses to each stimulus word in the word association data collection. This would affect the network in and out degree. Hence, establishing whether there are significant differences in the vocabularies of the two groups, which are not related to their expertise in mechanics, is important. This was investigated using tests of receptive (Nation & Beglar, 2007) and productive vocabulary (Laufer & Nation, 1999). The application and results of these tests are described in section 4.2.2.

4.2.1 The Force Concept Inventory

4.2.1.1 The Design and Structure of the Force Concept Inventory

The Force Concept Inventory (FCI) (Hestenes et al., 1992) is an established test to measure conceptual understanding of mechanics (Savinainen & Scott, 2002), which has been used in numerous studies (e.g., Bani-Salameh, 2016; Bruun & Brewe, 2013). The FCI test (1995 version) (Halloun, Hake, Mosca, & Hestenes, 1995) uses 30 multiple-choice questions to establish whether an individual has a knowledge and understanding of mechanics that is consistent with Newtonian mechanics. Hestenes et al. (1992) identify six 'conceptual dimensions' that make up a Newtonian understanding of mechanics. These dimensions are Kinematics, Newton's First Law, Newton's Second Law, Newton's Third Law, Superposition Principle and Kinds of Forces. Table 4.1 lists further details of the knowledge and understanding the FCI test in each of these dimensions and test items that are relevant to each dimension.

Table 4.1 The conceptual dimension in the Force Concept Inventory (FCI) (Hestenes et al., 1992).

Dimension			Test items
1	Kinematics	Velocity discriminated from position	20
		Acceleration discriminated from velocity	21
		Constant acceleration entails parabolic orbit and changing speed	23, 24, 25
		Vector addition of velocities	7
2	First law	With no force, velocity direction constant	4, 6, 10, 26
		With no force, speed constant	8, 27
		With cancelling forces, velocity direction constant	18, 28
		With cancelling forces, speed constant	18, 28
3	Second law	Impulsive force	6, 7
		Constant force implies constant acceleration	24, 25
4	Third law	For impulsive forces	2, 11
		For continuous forces	13, 14
5	Superposition principle	Vector sum	19
		Cancelling forces	9, 18, 28
6	Kinds of force	Solid contact: passive, impulsive, friction opposes motion	9, 12. 15, 29
		Fluid contact: air resistance, buoyant (air pressure)	22, 12
		Gravitation: acceleration independent of weight, parabolic trajectory	5, 9, 12, 17, 18, 22, 1, 3, 16, 23

In addition to six conceptual dimensions of Newtonian mechanics, Hestenes et al. (1992) identify a taxonomy of misconceptions that are associated with mechanics. This taxonomy is a comprehensive summary of alternative conceptions (Savinainen & Scott, 2002) that have been identified by previous empirical studies (e.g., Halloun & Hestenes, 1985a). Hestenes et al. (1992) use their taxonomy of misconceptions to form the alternative choices in each multiple-choice question. Hence, each test item in the FCI represent a 'forced choice' (Hestenes et al., 1992, p. 142) between a Newtonian understand of mechanics and, what Hestenes et al. (1992) call, a *common-sense* understanding of mechanics. Hestenes et al. (1992) stress that this common-sense view is based on reasonable hypotheses grounded in everyday experience and has been held by several important historic great thinkers (e.g., Aristotle and, at times, Galileo). Hence, the FCI is not a test of intelligence, but a probe of these different belief systems. Furthermore, Hestenes et al. (1992) view the development of a Newtonian understanding of mechanics as a process where individuals develop a greater understanding of each of the six conceptual dimensions, but also develop greater coherence between different dimensions. Hence, they

argue that the FCI test scores need to be interpreted as a whole score, rather than in each dimension. Moreover, Hestenes et al. (1992) suggest that a score of 60% (18 out of 30) indicates an entry level understanding of Newtonian mechanics, whereas scores lower than this suggests a common sense understanding. Later, Halloun et al. (1995) suggested that a score of 85% (25 out of 30) indicates a 'mastery' level of Newtonian understanding (Savinainen & Scott, 2002).

4.2.1.2 The Reliability, Validity, and Interpretation of the Force Conceptual Inventory

The FCI is based on an earlier instrument designed to assess students' knowledge and understanding of mechanics called the Mechanics Diagnostic Test (MDT) (Halloun & Hestenes, 1985b). The development of the MDT took place over three years, during which several versions of the test were written and tested on more than 1000 students in an introductory, college level physics course in the USA. Early versions of the MDT required written answers. These written responses were used to identify phrasing of answers that reflected the common misconceptions identified by Halloun and Hestenes (1985a). These phrasings became the alternative answers included in the multiple-choice questions, which were developed for the later version of the MDT to allow the test to be easily graded.

Halloun and Hestenes (1985b) used several methods to establish the face and content validity of the MDT. First, physics professors and graduate physics students examined the early versions of the test and their suggestions for improvement were incorporated into later versions. Second, eleven graduate students took the test and all these students agreed on the correct answer for each item. Third, twenty-two introductory physics students sat the test. They were then interviewed. These interviews confirmed that the students understood the questions and the alternative answers. Fourth, the responses of thirty-one of the highest attaining introductory level students (grade A) were analysed. This analysis did not find any common misunderstanding because of the formulation of the questions.

Halloun and Hestenes (1985b) established the reliability of the MDT by interviewing students who had taken the test. During these interviews, students 'repeated the answers they had given in the written test virtually without exception' (Halloun & Hestenes, 1985b, p. 1044) and were not easily dissuaded from their answers. Hence, Halloun and Hestenes (1985b) concluded that these responses represented stable beliefs, rather than flippant answers. Halloun and Hestenes (1985b) also carried out a statistical analysis of the test results, which used a Kuder-Richardson (KR) Test to compare the test score distributions for comparable groups that took the test at

different times. The KR reliability coefficients were 0.86 for pre-test use and 0.89 for post-test use (Halloun & Hestenes, 1985b, p. 1044), which indicate a highly reliable test.

The claims of Halloun and Hestenes (1985b) and Hestenes et al. (1992) about the validity and reliability of the FCI have been challenged. For instance, Dancy (2000) points out that the original version of the FCI only includes approximately half of the questions from the MDT. Hestenes et al. (1992) claim the new questions in the FCI improve the efficacy of the test to completely identify misconceptions. Moreover, further changes were made to the original FCI in 1995 to remove ambiguities in the questions and reduce the likelihood of false positives (Hake, 1998). However, the extensive measures used to establish the reliability and validity of the MDT were not repeated for the 1992 and 1995 versions of the FCI (Savinainen & Scott, 2002). Hence, Dancy (2000) argues this brings into question the validity and reliability of the FCI because the reliability and validity of the updated items in the FCI has not been explicitly tested. However, Savinainen and Scott (2002) argue that the extensive use of the FCI (e.g., Hake, 1998) in various institutions have now effectively established the reliability of this test.

Whilst the FCI is considered a reliable test, there remains a debate around the validity of the FCI. The claim that the FCI is a valid test is based on the view of numerous physics instructors who agree that it '*measures students' understanding of the force concept'* (Savinainen & Scott, 2002, p. 50). However, several researchers using the FCI have questioned whether the test *does* measure a conceptual understanding of force (Huffman & Heller, 1995) that is consistent with the view of this concept proposed by Hestenes et al. (1992). For instance, Huffman and Heller (1995) carried out two separate factor analyses of 145 high school and 750 university students (in the USA) FCI responses. They argued that if *force* is a single unity concept, then their analysis should only identify one single factor. However, if the force concept is based on six conceptual dimensions, as Hestenes et al. (1992) suggest, their factor analysis should identify six factors, associated with each dimension. In contrast with these two hypotheses, they identified two factors associated with Newton's Third Law and Kinds of Force in the high-school analysis and one factor associated with Kinds of Force in the university analysis. There were also several (>6) insignificant factors in both analyses. Hence, Huffman and Heller (1995) concluded that the FCI '*does not necessarily measure a single force concept or the six conceptual dimensions proposed by Hestenes et al. (1992)*' (p140). However, Eaton and Willoughby (2018) later claimed to have confirmed the six dimensions.

The questions raised about the validity of the FCI highlights, and specifically what the FCI *actually measures*, highlights an important limitation of the instrument. The multiple tests of

validity of the MDT (Halloun & Hestenes, 1985b) generally test whether a student's answer to a question is viewed as correct by a physicist, who is accepted as having an authoritative understanding of Newtonian mechanics. Hence, the tests of validity for the MDT, and therefore the FCI, establish that these instruments can measure whether the individual 'sees' the world in a way that is consistent with Newtonian mechanics. diSessa and Sherin (1998) argue *seeing* the world in a particular way is a criterion for having a particular concept. However, they argue that this ability is a result of being about to 'read out' information and then 'coordinate' this information. If the six dimensions proposed by Hestenes et al. (1992) are correct, this would suggest 'seeing' the world in a Newtonian way involves reading out information about kinematics, kind of forces, etc… and coordinate this information. The debate around factor analysis of Huffman and Heller (1995) and Eaton and Willoughby (2018) indicates that this is not certain. Moreover, it is possible that the six conceptual dimensions might themselves contain concepts (units of thought) (e.g., *velocity*, *speed*, *acceleration*). This lack of clarity reflects the lack of agreement on a definition or appropriate model of scientific concept, like *force* (e.g., Rusanen, 2014; Sherin, 2017a). Hence, given that there is considerable debate around what the content of scientific concepts might be and what content the FCI measures, is not clear then that the FCI is a valid measure of conceptual content.

However, this does not necessarily preclude the use of FCI as a useful test if this limitation is acknowledged. For instance, many classic experiments on conceptual understanding which are used to probe the content of the concepts (mental representation) begin by testing whether participants 'see' the world in a particular way (e.g., McCloskey and Glucksberg (1978) asks participants whether an object 'is a fish' to test typicality effects). The FCI is accepted as a valid test whether an individual 'sees' the world in a particular way (e.g., as consistent with a 'common sense' or Newtonian mechanics), even though it is not clear if the assumptions the test makes about the conceptual content that is used to arrive at that view are correct. Hence, the FCI is a valid instrument for determining whether individuals have a view of the world that is consistent with experts in mechanics. This can then be used as a starting point for probing the conceptual content that these individuals use to arrive at this view. This interpretation of the FCI is adopted in this thesis.

Finally, Mahajan (2005) and Steinberg and Sabella (1997) both question whether the FCI is a valid measure of an individual 'seeing' the world in a way that is consistent with Newtonian mechanics. Mahajan (2005) found that undergraduates students who scored very highly in the FCI, and hence were judged to have a Mastery Newtonian understanding of mechanics, could

not replicate this reasoning for more complex problems in one-to-one tutorials. Similarly, Steinberg and Sabella (1997) found that undergraduate students did not use the same reasoning to solve problems that required the same use of Newtonian mechanics but were presented in a different context. Hence, these studies raise questions about whether FCI adequately tests two performance characteristics of conceptual understanding identified by diSessa and Sherin (1998): the ability to coordinate information in a single situation (integration), and coordinating different features in different information to find the same information (invariance). However, Savinainen and Scott (2002) argue that it is, perhaps, unreasonable to expect any instrument to take into account all contexts and presentations and, hence, they stress that the FCI is a valid test. However, this argument does highlight that the thresholds used in the FCI to define Common Sense, Entry Newtonian and then Mastery Newtonian thinking can give a false impression that these are dichotomous views of the world. Therefore, the FCI scores are viewed as defining a spectrum of conceptual understanding in this analysis.

4.2.1.3 The Implementation of the Force Concept Inventory

The FCI is intended to be a pen and paper test taken under tightly controlled conditions (Madsen & McKagan, 2022). It was intended to use the FCI in accordance with this guidance after the interviews with participants had taken place. However, the last interview with a participant took place on the 20th March 2020 and the first UK national lockdown, due to the COVID-19 pandemic, begun on the 25th March 2020. All UK schools closed on this date. Hence, it was not possible to administer the FCI as a face-to-face pen and paper test in this study. To enable the FCI to still be used, it was converted to an online test administered through the JISC online survey website (Jisc, 2022).

There are some risks involved with the collection of online data. For instance, it cannot be guaranteed that the participants complete the test themselves or do not use aids, such as textbooks, to help them (Granello & Wheaton, 2004; Lefever et al., 2007; Topp & Pawloski, 2002). However, the participants had already extensively use online data collection when completing the word-association surveys and many of the arguments that were used to justify this method of data collection (see chapter 3) were still deemed to be valid. For instance, each participant was a willing volunteer and were unlikely to gain anything from not completing the test honestly.

The FCI scores for each participant are presented in Table 4.2. The scores in the physics cohort are higher than those in the psychology cohort. Whether this difference in the scores is

statistically significant is tested in Section 4.2.3. However, an initial inspection of the scores in the two groups suggests that the physicist have a more conceptually developed understanding of mechanics than the psychologists. Note that participants 12P04 and 12S09 did not complete the test. These participants were sent two reminder emails. However, as the testing window was during the COVID-19 pandemic, and therefore these participants could have been facing considerable stress in their personal lives, they were not pressed for a response.

Table 4.2 The raw scores for each participant in the Force Concept Inventory (Hestenes et al., 1992) (1995 version). Hestenes et al. (1992) group questions into six conceptual dimensions (kinematics, Newton's first law, Newton's second law, Newton's third law, superposition of vectors and kinds of forces) that test different aspects of a Newtonian understanding of mechanics. The raw scores for each participant in each conceptual dimension are given. Note some questions test more than one conceptual dimension and, hence, are included in the score for more than one dimension. Therefore, the total raw score is not the sum of the total scores in each dimension. Hestenes et al. (1992) recommend that the total raw score is used to assess a participant's understanding of mechanics. If a participant achieves a raw score of ≥ 18 (60%), they are considered to have a 'Entry Newtonian' understanding of mechanics and a raw score of ≥ 26 (85%) indicates a 'Mastery Newtonian' understanding. A score < 18 suggests a 'common sense' understanding.

		Conceptual Dimensions							
		1	2	3	4	5	6		
Cohort	Participant	Kinematics/7	Newton's First Law/9	Newton's Second Law/5	Newton's Third Law/4	Superposition Principle/5	Kinds of Forces/13	Total Raw Score/30	Level
Physics	12P02	7	8	5	4	4	10	27	Mastery Newtonian
	12P03	7	8	3	4	4	13	28	Mastery Newtonian
	12P04	-	-	-	-	-	-	-	-
	12P05	3	4	3	2	4	8	16	Common Sense
	12P08	5	9	4	1	5	11	24	Entry Newtonian
	12P12	5	8	5	4	4	9	26	Mastery Newtonian
Psychology	12S01	2	3	1	0	1	3	7	Common Sense
	12S02	1	0	0	1	0	4	6	Common Sense
	12S03	0	1	0	2	0	1	4	Common Sense
	12S05	0	2	1	0	1	2	5	Common Sense
	12S09	-	-	-	-	-	-	-	-

4.2.2 Vocabulary Tests

Two tests were used to assess students' vocabulary. Each test measured a different aspect of word knowledge, which are likely to be learnt at different times (Bloom, 2000), thereby giving a more complete picture of an individual's vocabulary. The receptive vocabulary test (Nation & Beglar, 2007) measures an individual's ability to recognise the meaning of a word when presented with the word in a given context. Given the relatively loose conditions for drawing edges between signs (i.e., considered word associations, see section 3.3.3.2), this test is likely to identify the impact of any difference in participants' general vocabulary on the sign networks. The productive vocabulary test measures an individual's ability to use a word confidently when producing their own language (Laufer & Nation, 1999). Hence, this test could identify any possible impact differences in participants' general vocabulary could have on their interview performance. The method and administration of each of these tests with the results is presented in sections 4.2.2.1 and 4.2.2.2.

4.2.2.1 Receptive Vocabulary Test

The receptive vocabulary test developed by Nation and Beglar (2007) estimates the size of an individual's vocabulary by testing their understanding of words taken from word families defined by the frequency that they occur in a corpus of English texts. Each family contains 1000 words, and the families range from the 1st to the 20th most frequently used words. The test assesses an individual's understanding of 5 words sampled at random from each family (100 words in total) using a multiple-choice question for each word (see Figure 4.1 below).

Azalea: This **azalea** is very pretty.

a. Small tree with many flowers growing in groups
b. Light material made from natural threads
c. Long piece of material worn by women in India
d. Sea shell shaped like a fan

Figure 4.1 Example question from the receptive vocabulary test (Nation, 2012).

Each question contains distractors, which Nation and Beglar (2007) claim makes the assessment more rigorous than previous receptive vocabulary tests, and ensures an individual has to have a '*moderately developed idea of the meaning of the word*' (p11) to answer the question correctly. The estimate of an individual's vocabulary size is calculated by multiplying their final score in the test by 200.

The validity and reliability of the test has been confirmed by several pilot studies described in Nation (2012). However, the results of these pilot studies are not described by the author.

The scores for the test for each participant are shown in Table 4.3. Unlike the FCI scores, there appears to be little difference between the scores in the physics and psychology cohort. The scores have been converted to a vocabulary size for each participant, and the highest word family level that this corresponds to is also given.

Table 4.3 The raw scores for each participants from the receptive vocabulary size test (Nation & Beglar, 2007). The vocabulary size estimate for each participant is calculated by multiplying a raw score by 200 (Nation, 2012). The highest level indicates the highest word frequency family a participant has receptive knowledge of (e.g., 16 indicates a participant has knowledge of the 16th word frequency family). Note – indicates that a participant did not return an online test.

Cohort	Participant	Raw Score	Vocabulary Size Estimate	Highest Level
Physics	12P02	84	16800	16
	12P03	71	14200	14
	12P04	-	-	-
	12P05	81	16200	16
	12P08	77	15400	15
	12P12	86	17200	17
Psychology	12S01	80	16000	16
	12S02	-	-	-
	12S03	75	15000	16
	12S05	94	18800	18
	12S09	83	16600	16

4.2.2.2 Productive Vocabulary Test

The productive vocabulary test designed by Laufer and Nation (1999) assesses the ability of an individual to use a word in a controlled context. The context is controlled because a participant's understanding of each word is assessed by whether they can 'fill-in' the word when given a sentence as context (see Figure 4.2 below), rather than whether a word is used correctly when freely when speaking or writing.

The 2000 word level.

1. I'm glad we had this opp_____ to talk.
2. There are a doz_____ eggs in the basket.

The 10,000 word level

1. The new vic_____ was appointed by the bishop.
2. If your lips are sore, try lip sal_____, not medicine.

Figure 4.2 Example questions from the productive vocabulary test (Laufer & Nation, 1999).

Like the receptive word test, this test also uses word families that are defined by the frequency of word occurrence in a corpus of text (Nation, 1983, 1990). However, in this test there are five word families: 2000 most frequently used words, 3000 most frequently used words, 5000 most frequently used words, University Level, and 10,000 most frequently used words. Hence, the test measures whether an individual can correctly use progressively less frequently occurring words. Eighteen words are selected at random from each of these families to form the test items in each work family. A participant is considered to have productive work knowledge of a family if they score 15 out of 18.

The reliability and validity of the tests was investigated by comparing the scores of the productive vocabulary test with measures of language proficiency (Laufer & Nation, 1999). The results of this comparison showed the scores of the productive vocabulary test consistently correlated with language proficiency scores. Hence, Laufer and Nation (1999) conclude that the productive vocabulary test is a reliable and valid measure of an individual productive vocabulary ability. In terms of this study, the correlations of these scores with measures of language proficiency is reassuring because these scores were used to judge the impact any productive vocabulary differences might have on participants' interview performance.

The total scores for the productive vocabulary tests are shown in Table 4.4, as well as the scores for each word family. Table 4.4 also shows whether they are working at a particular level using this criterion and the highest level they have achieved. Like the receptive vocabulary scores, there appears to be little difference between the scores in the two cohorts. The similarity in the receptive

and productive vocabulary test scores in the physics and psychology cohorts suggests these groups have similar vocabulary knowledge.

*Table 4.4 The raw scores for each participant from the productive vocabulary test (Laufer & Nation, 1999) in each word frequency family. The vocabulary size estimate for each participant is calculated by multiplying a raw score by 200 (Nation, 2012). The highest level indicates the highest word frequency family a participant has productive knowledge of (e.g., University indicates a participant has knowledge of the university word frequency family). A participant is considered to have knowledge of a family if they score 15 in that word family section. There are 18 questions in each word family section. A * indicates that a participant did not score highly enough to have productive knowledge of a word family and – indicates that a participant did not return an online test.*

Cohort	Participant	2000	2000-3000	3000-5000	University	5000-10000	Total Raw Score	Highest Word Family Level Attained
Physics	12P02	17	18	17	16	18	86	5000-10000
	12P03	18	18	16	14*	12*	78	3000-5000
	12P04	-	-	-	-	-	-	-
	12P05	17	17	17	17	14*	82	University
	12P08	17	17	18	15	13*	80	University
	12P12	17	18	16	18	17	86	5000-10000
Psychology	12S01	18	13	15	14*	11*	71	3000-5000
	12S02	-	-	-	-	-	-	-
	12S03	17	17	18	16	12*	80	5000-10000
	12S05	17	16	16	16	12*	77	5000-10000
	12S09	-	-	-	-	-	-	-

4.2.3 Differences in the Conceptual Understanding of Mechanics and Vocabulary Knowledge in Physics and Psychology Cohorts

The previous sections have listed the individual scores for the participants in the physics and psychology cohorts for the FCI and vocabulary tests. These scores are informative. However, to address research question 2, it is important to establish whether there are systematic differences in the conceptual understanding of mechanics and vocabulary knowledge in the two cohorts.

The descriptive statistics for each of the tests are shown in Table 4.5. The difference between the mean scores for the physics and psychology FCI scores is larger than the standard deviations in each of these cases (physics: M=24.20, SD=4.83; psychology: M=5.50, SD=1.29). There are also differences in the mean scores for the receptive and productive vocabulary tests (receptive test – physics: M=79.80, SD=5.97; psychology: M=83.25, SD=8.50, productive test – physics: M=82.40, SD=3.58; psychology: M=76.00, SD=4.58). However, these mean scores are within the range one standard deviation. This suggests the differences in mean scores for the receptive and productive vocabulary test could be explained by statistical fluctuations, but this is unlikely to be the case for the FCI scores.

Table 4.5 Descriptive statistics for the FCI and vocabulary test scores with the KS statistics and probabilities.

	Force Concept Inventory					Receptive Vocabulary Size					Productive Vocabulary				
Cohort	n	M	SD	KS	p	n	M	SD	KS	p	n	M	SD	KS	p
Physics	5	24.20	4.82	1.00	0.02*	5	79.80	5.97	0.25	0.99	5	82.40	3.58	0.67	0.38
Psychology	4	5.50	1.29			4	83.25	8.50			3	76.00	4.58		

A Kolmogorov-Smirnov (KS) test was used to determine whether these difference in the mean scores were significant by testing whether the null hypothesis that the scores in each group can be drawn from the same distributions (Crawley, 2013). A KS test was selected because it is a non-parametric test that makes minimal assumptions about the data (e.g., the scores are normally distributed), which are difficult to verify with small sample sizes (i.e., physics: n=5, psychology: n=4, for the FCI test) (Crawley, 2013). Moreover, other non-parametric tests are not suitable for small sample sizes (Field, 2013).

The null hypothesis can be rejected for the FCI scores, but not for the scores of the receptive and productive vocabulary test. This suggests that the difference between the FCI mean scores in the

physics and psychology cohorts is significant, but the difference between the mean scores for the vocabulary tests is not significant. Therefore, the physicists have a better conceptual understanding of mechanics, but the vocabulary knowledge (receptive and productive) in the two cohorts is similar.

In addition to the KS test, a Fisher exact test was used to determine if there were any statistically significant differences between the two cohorts. The Fisher exact test establishes whether there are significant differences in the frequency of occurrences in different categories (Crawley, 2013). For instance, the Fischer exact test can establish if there are any significant differences in the number of participants with a Newtonian understanding of mechanics (FCI core > 18) between the physics and psychology cohorts. To determine whether any differences are significant, the Fisher exact test compares expected and observed frequencies of occurrence in each cohort. The expected values are computed using probability calculations and are, therefore, exact. Hence, the Fisher exact test is reliable for small expected frequencies (e.g., ≤ 5), which is the case in physics and psychology groups because of the small sample sizes.

The criteria used to define categories for the FCI, receptive vocabulary, and productive vocabulary tests, the frequency of occurrence in these categories, the p-values from the Fischer exact tests are show in Table 4.6. The only significant difference ($p<0.05$) is between the number of participants with a Newtonian understanding physics in the physics cohort compared with the psychology cohort. This confirms the results of the KS tests that there are systematic differences in the conceptual understanding of mechanics between physics and psychology cohorts. The FCI scores suggest the physics have a more developed conceptual understanding of mechanics than the psychologists. In contrast, there are no systematic difference between the groups for measures of vocabulary knowledge. Hence, any differences in the sign networks for the two cohorts is likely to be due to their understanding of mechanics, rather than differences in their general vocabulary.

Table 4.6 The Fisher exact test for the FCI, receptive vocabulary size test, and productive vocabulary size. For the FCI, the number of students achieving a Newtonian understanding of mechanics (score>18) are listed. The number of participants in each group that achieve the criteria for word knowledge (receptive and productive) in each work frequency family is given for the receptive and productive vocabulary tests. The total number of participants that took each test is as follows: FCI, $n_{physics}=5$, $n_{pyschology}=4$; receptive vocabulary test, $n_{physics}=5$, $n_{pyschology}=4$; productive vocabulary test, $n_{physics}=5$, $n_{pyschology}=3$.

Test	Contingency	Physics	Psychology	p
Force Concept Inventory	Newtonian	4	0	0.048*
Receptive Vocabulary Size	14th Word Frequency Family	5	4	1.00
	15th Word Frequency Family	4	4	1.00
	16th Word Frequency Family	2	4	0.17
	17th Word Frequency Family	1	1	1.00
	18th Word Frequency Family	0	1	0.44
Productive Vocabulary	2000	5	3	1.00
	2000 - 3000	5	2	0.38
	3000 - 5000	5	3	1.00
	University	4	2	1.00
	5000 - 10000	2	0	0.46

4.3 Sign Network Analysis

4.3.1 Introduction to the Sign Network Analysis

The networks for participants 12P02 (physics) and 12S01 (psychology) are shown in Figure 4.3 and Figure 4.4. It is clear from these figures that visual inspection cannot identify the similarities and differences between the networks. Hence, to address the second part of research question 2, this analysis follows the procedure of using defined mathematical measures to reduce the structure of each network to a few key parameters, which are easy to compare (see section 2.6.1.2 and Newman (2018)).

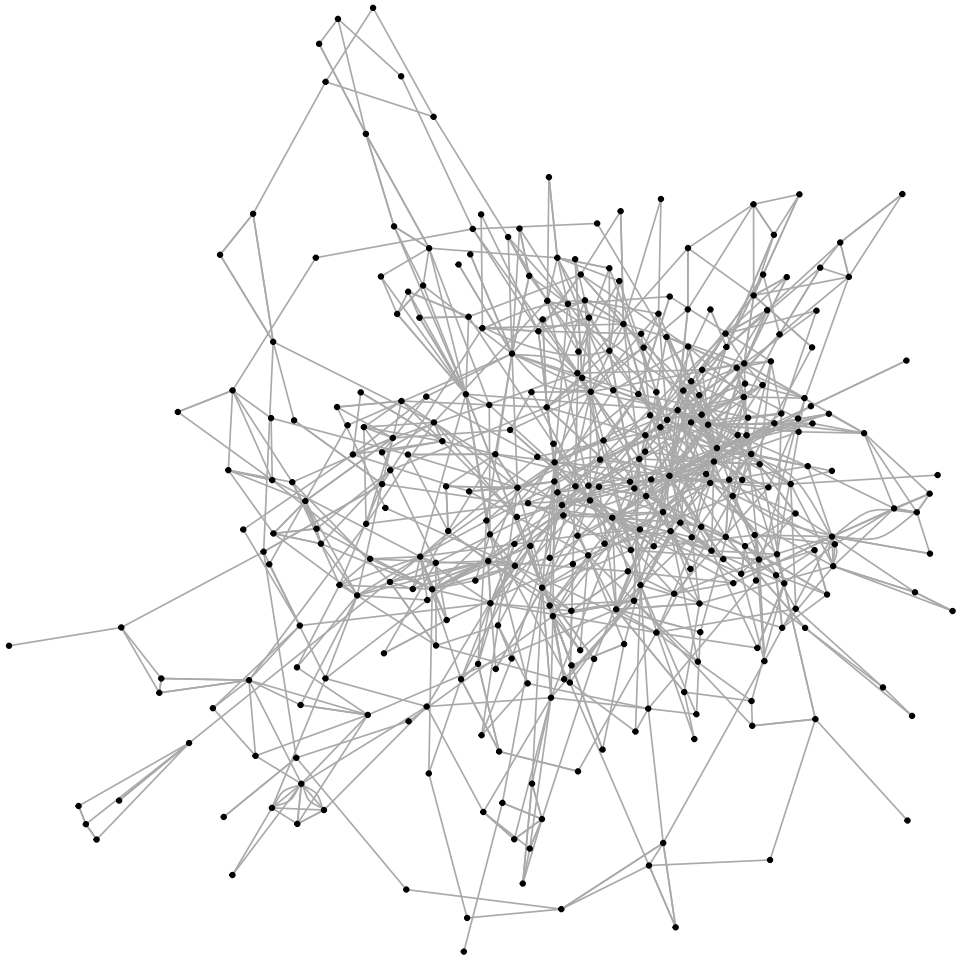


Figure 4.3 The sign network for 12P02 (physics). Note, the number of vertices (n=331) and edges (m=1104) in this network makes it difficult to display this network clearly. Section 4.3 describes how the sign networks are analysed, given the difficulties with visual analysis.



Figure 4.4 The sign network for 12S01 (psychology). Note the number of vertices (n=250) and edges (m=790) in this network makes it difficult to display this network clearly. Section 4.3 describes how the sign networks are analysed, given the difficulties with visual analysis.

Initially, this analysis uses widely used measures of network structures (Barabási & Pósfai, 2016; Caldarelli, 2007; Newman, 2018). These are: sparsity, average shortest path length, mean degree and clustering coefficient. These measures are widely used in other studies (e.g., De Deyne & Storms, 2008a; Morais et al., 2013; Steyvers & Tenenbaum, 2005) that have created and analysed sign networks. Hence, the analysis of the individual sign networks in this study (presented in section 4.3.2) can easily be compared with the finding of these studies.

Moreover, using these measures allows this analysis to refer to models that have been proposed to explain these measures of network structure (Newman, 2018). These models provide a more detailed picture of network structures and often incorporate theories of network development (Barabási & Albert, 1999; Barabási & Pósfai, 2016; Caldarelli, 2007; Newman, 2018) . Hence, they facilitate a greater understanding of a network. This analysis evaluates whether the model of sign

network growth developed by Steyvers and Tenenbaum (2005) is applicable to the sign networks for the physicists and psychologists.

The insight provided the Steyvers and Tenenbaum (2005) model informs a key aspect of this analysis because this model specifies how the placeholder sign network needed by TOOC might grow. This model specifically predicts a scale-free structure, and hub words, as 'signatures' of the type of network growth. Hence, as well as using key measures to look at similarities and differences, this analysis also aims to identify whether the sign networks in the physics and psychology groups have a scale-free structure and whether they contain hub works. Although, whether the networks contain hub-words is arguably the more important insight in terms of the placeholder sign structure, rather than whether the structure is scale-free, because the presence of hub-words would suggest the sign structure is 'growing' around or being coordinated by these key signs. Detecting hub-words and scale-free structure involves analysing the degree distributions of the networks. This analysis is described in section 4.3.3.

4.3.2 Basic Network Properties

Table 4.7 shows the measures for each participant's directed network which are used to define the basic properties of a network. These measures were obtained using the igraph package (Csardi & Nepusz, 2006) in R (R Development Core Team, 2013). The descriptive statistics for each of these measures are given in Table 4.8. These mean values for the average degree, diameter, mean shortest path length, and the clustering coefficient are similar in physics and psychology cohorts.

Table 4.7 The basic network parameters for each sign network. Note, n is the network order (number of vertices), m is the network size (number of edges in a network, <k> is the mean vertex degree, ρ is the density of the network ($\rho = m/m_{max}$), D is the network diameter, L is the average distance or average shortest path length, and C is the clustering coefficient. L_r and C_r are the average shortest path length and clustering coefficient for a random network of the same order and size.

Cohort	Participant	n	M	m_{max}	$\boldsymbol{\rho}$	<k>	D	L	L_r	C	C_r
Physics	12P02	331	1104	109230	0.01	3.34	13	5.74	3.75	0.18	0.02
	12P03	348	2160	120756	0.02	6.21	10	3.91	7.80	0.18	0.00
	12P04	187	813	34782	0.02	4.35	10	4.03	4.14	0.20	0.01
	12P05	278	1135	77006	0.01	4.08	11	4.98	6.01	0.21	0.00
	12P08	344	2850	117992	0.02	8.29	7	3.27	8.95	0.22	0.00
	12P12	397	1238	157212	0.01	3.12	16	6.04	7.02	0.12	0.01
Psychology	12S01	250	790	62250	0.01	3.16	14	5.77	6.01	0.25	0.01
	12S02	270	941	72630	0.01	3.49	18	6.96	4.86	0.25	0.01
	12S03	490	2059	239610	0.01	4.20	12	5.15	5.15	0.18	0.00
	12S05	324	1206	104652	0.01	3.72	13	5.45	3.00	0.20	0.00
	12S09	308	797	94556	0.01	2.59	19	8.01	7.20	0.22	0.01

Table 4.8 The descriptive statistics for selected network parameters. Note, n is the network order (number of vertices), m is the network size (number of edges in a network, <k> is the mean vertex degree, D is the network diameter, L is the average distance or average shortest path length, and C is the clustering coefficient. M and SD indicate the mean and standard deviations for each of the parameter. KS and p indicated the KS statistic and p-values for the KS tests carried out to test the null hypothesis that the values of these parameters in the physics and psychology groups are drawn from the same distributions.

Cohort	n				m				<k>				D				L				C			
	M	SD	KS	p	M	SD	KS	p	M	SD	KS	p	M	SD	KS	p	M	SD	KS	p	M	SD	KS	p
All	321	80	-	-	1372	680	-	-	4.230	1.646	-	-	13	4	-	-	5.391	1.371	-	-	0.200	0.037	-	-
Physics	314	73			1550	784			4.896	1.988			11	3			4.661	1.101			0.185	0.037		
Psychology	328	95	0.467	0.474	1159	531	0.433	0.591	3.431	0.605	0.500	0.357	15	3	0.667	0.177	6.268	1.193	0.667	0.108	0.219	0.029	0.433	0.685

4.3.2.1 Mean Degree

The mean degree for the physics network is higher than the mean degree for the psychology networks (physics: <k>=4.896, psychology: <k>=3.401). This would suggest that on average the physicists made more connections between signs than the psychologists.

4.3.2.2 Density and sparseness

The density of each of the networks is given in Table 4.7. The values of $\rho \sim 0$ in each case. Hence, the sign networks for all the participants are sparse (Newman, 2018). The sign networks measured by Steyvers and Tenenbaum (2005), De Deyne and Storms (2008a) and (Morais et al., 2013) are also sparse.

4.3.2.3 Diameter and average shortest path length

Low average shortest path lengths (between ~ 3 - 7) are a characteristic of the sign networks measured and analysed by Steyvers and Tenenbaum (2005), De Deyne and Storms (2008a) and Morais et al. (2013). The average shortest path lengths for each of the networks in both the physics (range: 3.27 – 6.04, $L_{mean} = 4.661$, $L_{sd} = 1.101$) and psychology (range: 5.15 – 8.01, $L_{mean} = 6.268$, $L_{sd} = 1.193$) cohorts are consistent with these short path lengths. Furthermore, the average path lengths of each participant's network are consistent with the average path lengths for random networks with the same number of vertices and edges (L_r). This suggest that a random network model can account for the average shortest path length measured for the sign network in both the physics and psychology cohorts.

The diameter of a network is defined as the longest shortest path in the network. As with the average shortest path lengths, the diameter of the networks in the physics (range: 7-16, $D_{mean} = 11$,$D_{sd} = 3$) and psychology (range: 12-19, $D_{mean} = 15$, $D_{sd} = 3$) cohorts) are consistent with the diameters of the larger sign networks for whole vocabularies measured and analysed by Steyvers and Tenenbaum (2005) and De Deyne and Storms (2008a).

4.3.2.4 Clustering Coefficient

The clustering coefficient is a measure of the probability any two randomly chosen words in the network will both be connected to the same neighbouring word (Newman, 2018). The calculation for the clustering coefficient is different for directed and undirected networks (Caldarelli, 2007). However, Morais et al. (2013) collapse their directed sign networks to form undirected networks

to calculate the clustering coefficients for each of their participant's networks. To allow for comparison, the same procedure is carried out in this analysis.

As with the average shortest path length and diameter, the values of the clustering coefficient for the network in the physics and psychology groups are consistent with those measured by Morais et al. (2013) for individual networks and Steyvers and Tenenbaum (2005) for aggregate networks. Moreover, these values are higher than would be expected for a randomly connected network with the same number of vertices and links. Hence, the high values of clustering coefficients for the sign networks is the first 'signature' feature that suggests the connections in these networks may not be made randomly. Again, the clustering coefficients are very similar in the physics and psychology groups (range: 0.12-0.22, $C_{mean} = 0.19$,$C_{sd} = 0.04$) and psychology (range: 0.18-0.25, $C_{mean} = 0.22$, $C = 0.03$) cohorts).

4.3.2.5 Summary of the Analysis of the Basic Network Properties

The aim of this study is to identify similarities and difference between the sign networks for physics and psychologists. The discussion above, which outlines an initial inspection of the measures that are typically used to define basic properties of networks, suggest that these cohorts share many basic properties (i.e., low average shortest path length, diameters, and clustering coefficients). This initial observation can be formalised in a hypothesis that the values of <k>, L, D, and C, measured in each cohort are drawn from the same distribution. This hypothesis was tested using a Kolmogorov-Smirnov (KS) test because of the small sample sizes (see section 4.2.3 for further rationale for using the KS test).

The results of the KS test are presented in Table 4.8. The p-values for each of the measures indicate that the null hypothesis cannot be rejected and, therefore, the values of all these measures are drawn from the same distribution. Therefore, the basic properties of the networks in the physics and psychologist networks appear to be the same, despite the difference in their conceptual understanding of mechanics, which was highlighted by the FCI tests. This is (perhaps) not surprising because these basic measures essentially characterise the macro or global structure of a sign network. It is likely that this macro structure is similar for all vocabularies. This would explain the similarities between the values of L, D, and C, measured here and those measured by other studies of sign networks (De Deyne & Storms, 2008a; Morais et al., 2013; Steyvers & Tenenbaum, 2005).

4.3.2.6 Using Aggregate Networks to Examine the Microstructure of the Sign Networks

It is possible (or likely) that any difference between the physics and psychology networks exist at a more micro-scale, despite the networks having a similar macrostructure. For instance, in the connections (and therefore the structure) around individual words. In terms of the representation of the sign networks in Figure 4.3 and Figure 4.4, this would imply that at the level of 'dots' for vertices and 'lines' for edges, the dot-and-line structure for the physics and psychology networks is similar (macrostructure). But different signs (or words) could be represented by the individual dots (microstructure). However, the complexity of the individual networks, shown in Figure 4.3 and Figure 4.4, illustrates the difficulties in making any visual comparison. Therefore, aggregate physics and psychology networks were created to find common connections between signs within each group and, hence, facilitate an initial visual comparison of the microstructure of the networks.

These aggregate networks were created by collapsing the directed sign networks into undirected networks. In these collapsed undirected networks, any bidirectional edges between signs were merged into a single edge. The individual undirected networks were then used to create combined vertex and edge lists for the physics and psychology groups. The igraph package (Csardi & Nepusz, 2006) was then used to create the aggregate physics and psychology networks from these vertex and edge lists. Weighted edges were used to indicate how many participants made connections between signs. For instance, a weight of one was allocated to an edge if the edge was present in only one of the individual networks in the groups, and a weight of two was allocated if the individual network of two participants contained the edge, and so on.

The physics and psychology aggregate sign networks are shown below in Figure 4.5 and Figure 4.6. The edges with a weight of 5 or more are shown in the physics network, and 4 or more in the psychology network. These weights ensured that only edges that occur in over 80% (5 out of 6 participants in the physics group; 4 out of 5 participants in the psychology group) are shown. The 80% is an arbitrary number. However, it defines a criterion that sets the most stringent condition for the present of common edges in the aggregate network, without requiring edges to be present in all networks. Hence, the aggregate placeholder sign networks show the most frequently used signs, and the connections between them, in the physics and psychology groups.

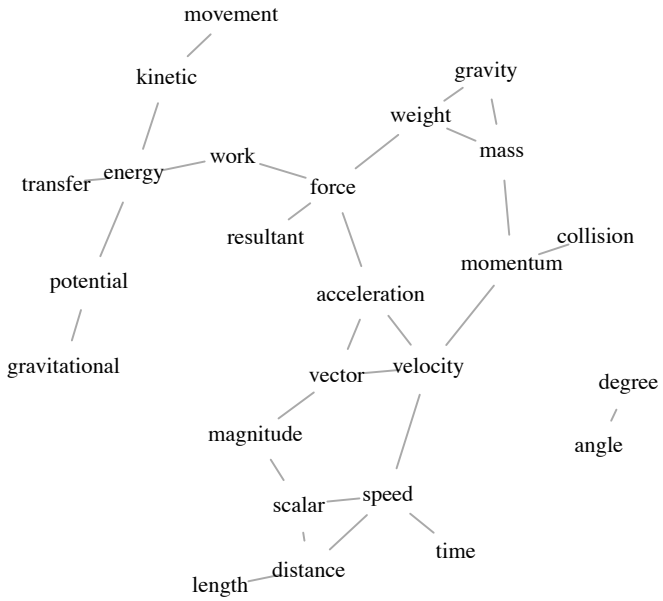


Figure 4.5 An aggregate sign network for the physics cohort. The edges in the network are weighted according to the number of participants that made the connection between two signs. This network shows the connections between signs made by 5 out of the 6 physics participants.

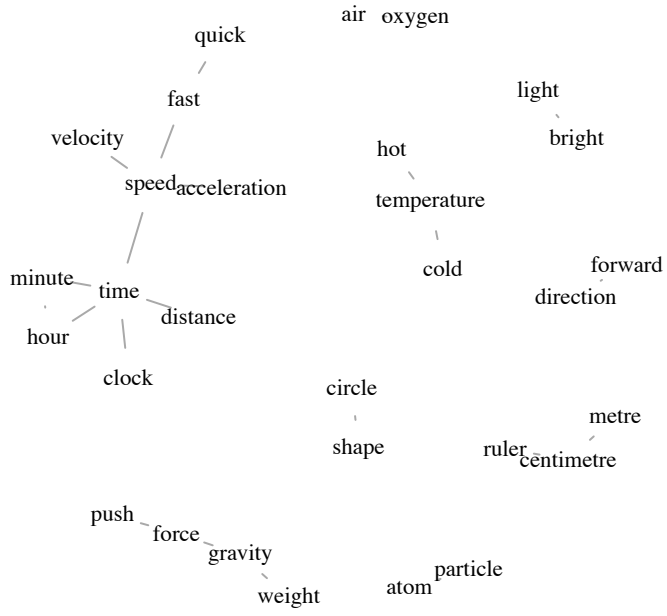


Figure 4.6 An aggregate sign network for the psychology cohort. The edges in the network are weighted according to the number of participants that made the connection between two signs. This network shows the connections between signs made by 5 out of the 6 psychology participants.

A visual inspection of the networks shows that the aggregate psychology network consists of several components that are unconnected. In contrast, the aggregate physics network contains of a single large, connected component and a small sub-component (*angle* | *degree*). This result would

appear to contradict the finding of that the physicist and psychologist individual sign networks have a similar macrostructure. However, the aggregate networks show shared edges between signs. Therefore, these apparently contradictory findings suggest that the physicists' sign network contain more shared microstructure (common vertices and edges) than the psychologists' networks.

There are differences between vertices and edges that are frequently present in the physicist and psychology sign networks. For instance, physicists connect *force* with *resultant* and *acceleration*, whilst psychologists do not. However, there are some similar structures. For example, both physicists and psychologists frequently connect *speed* to *time* and *velocity*. There is also some indication that the physicists' aggregate network contains signs, *force*, *energy,* and *speed*, that are associated with key physical concepts and have the highest degree in the network. These signs could be hub-words. The next section investigates whether the individual sign networks for physicists and psychologists contain hub-word by analysing the degree distributions. This section also establishes whether the sign network have a scale-free structure.

4.3.3 Network Vertex Degree Distributions

The vertex degree probability distribution, or degree-distribution, for any network is a key indicator of how a network is structured and, hence, how the connections in that network are achieved. As has been discussed in section 4.3.1, if the vertices in the network are randomly connected, the degree-distribution follow a Poisson distribution. However, if the network has a scale-free structure, with hubs, the degree-distribution will follow a power-law. A scale-free structure is a specific feature predicted by the model of sign network growth developed by Steyvers and Tenenbaum (2005). Hence, when looking at similarities and differences in the sign networks, this analysis aims to establish whether the vertex in and out-degree distributions follow a Poisson distribution or a power-law.

However, Broido and Clauset (2019) argue that a statement that a network has a scale-free structure if its distribution is a power-law is too simplistic. They stress that the precise criteria used in previous studies to define a scale-free structure vary. Stronger claims require a fraction of nodes to follow a power-law and a power-law exponent, $2 < \alpha < 3$ (Dorogovtsev & Mendes, 2002; Newman, 2005) or that the vertices evolve through preferential attachment (Barabási & Albert, 1999). Weaker claims only require that a power-law with $\alpha > 1$ (Barabási & Pósfai, 2016), a power-law only in upper tail (Willinger, Alderson, & Doyle, 2009), or a power-law with an exponential cut-off (Pastor-Satorras & Vespignani, 2002), or only that a power-law is a more plausible fit that exponential or log-normal distribution (Barabási, Jeong, & Albert, 2000). Broido

and Clauset (2019) argue this lack of agreement creates confusion as researchers applying different criteria are treating slightly different notions of 'scale-free structure' as the same. To avoid this issue, they suggest a 'severe test' for scale-free structure that incorporates an agreed set of quantitative criteria, which represent the different strength and types of evidence used to establish claims that a network has a scale-free structure. This analysis uses the strict test developed by Broido and Clauset (2019). The procedure and the evidence this strict test uses to define criteria for a scale-free structure is described below.

Broadly, the evidence used to establish that a network has a scale-free structure usually comes in two types. First, that a power-law model is relatively a better fit to the data than other models and, second, that a power-law itself is a good fit to the data (Broido & Clauset, 2019). Establishing these two types of evidence is the approach this analysis takes. Sections 4.3.3.3 and 4.3.3.4 determine whether a power-law model is a good fit to the in-degree, k_{in}, and out-degree, k_{out}, vertex distributions respectively. These power-law models are compared with other relevant distributions (Poisson, power-law with exponential cut-off, log-normal, exponential) in sections 4.3.3.3 and 4.3.3.4.

Broido and Clauset (2019) point out that these two types of evidence are not equally as strong. The comparison between models is only indirect evidence because it does not require a power-law model to be a good fit to the data. Hence, this evidence relies on the assumption that a scale-free structure is more plausible and other non-scale-free structures (Broido & Clauset, 2019). In contrast, the second type of evidence is stronger because it requires a power-law model to be a good fit to the data and, hence, is direct evidence for a scale-free structure. Hence, evidence that a power-law model is a good fit to the data is graded higher than evidence that a power-law is a comparably better fit than alternative distributions.

Broido and Clauset (2019) also identify a grading of the strength of evidence used to justify a scale-free structure when a power-law model is a good fit to a distribution. This grading has already been mentioned. For instance, requiring $2 < \alpha < 3$ is stronger evidence than just requiring $\alpha > 1$. Broido and Clauset (2019) use this grading of evidence to create a taxonomy of scale-free definitions. This taxonomy is show in Table 4.9 and is used in this analysis to define any claims about scale-free structures.

Table 4.9 The taxonomy of scale-free network definitions identified by Broido and Clauset (2019) which they use to define a 'strict test' for whether networks have a scale-free structure.

Strict test criteria	
In-direct	**Super-weak** For at least 50% of networks, no alternative model to the distribution is favoured.
Direct	**Weakest** For at least 50% of networks, a power-Law distribution model cannot be rejected (P>0.1) **Weak** The power-law region contains at least 50 vertices. **Strong** For at least 50% of the networks, $2 < \alpha < 3$ **Strongest** Requirements of strong for at least 90% of networks and of super-weak for 95%.

The severe test developed by Broido and Clauset (2019) also requires the use of robust procedures to fit a power-law model. This is an important aspect of the test because other methodologies for fitting, such as least squares fitting, can give inaccurate estimates of power-law parameters (Clauset et al., 2009). Hence, as part of their severe test they use the rigorous statistical method to determine power-law developed by Clauset et al. (2009). This procedure also contains a methodology for comparing power-law fits with other distributions. The implementation of the power-law fitting, and comparison of model is described in section 4.3.3.1.

4.3.3.1 Procedure for Fitting and Comparing the Vertex Degree Distributions

The Clauset et al. (2009) method for fitting discrete long-tailed distributions was used to test whether the k_{in} and k_{out} distributions can be modelled with a power-law. This method tests whether the values of k_{in} and k_{out} are drawn from a probability distribution:

$$P(k) \propto k^{-\alpha}$$

where α is known as the exponent or scaling parameter. This model was fitted to each of the k_{in} and k_{out} distributions using the poweRlaw package (Gillespie, 2015), which implements the Clauset et al. (2009) recommended fitting procedure in R.

This procedure estimates the power-law exponent α using a maximum likelihood estimator, which for a discrete power-law is:

$$\alpha \cong 1 + n\left[\sum_{i=1}^{n} ln\frac{k_i}{k_{min} - \frac{1}{2}}\right]^{-1}$$

Where n is the number of vertices, k_i is a vertex with degree, i, and k_{min} is the minimum vertex degree the power-law model is fitted to. Hence, k_{min} defines the beginning of the tail of the distribution.

It is clear from the maximum likelihood estimator equation that an accurate estimate of α requires an accurate estimate of the value of k_{min}. Clauset et al. (2009) recommend using a Kolmogorov-Smirnov (KS) method to obtain this estimate. This method calculates a KS statistic, which is the maximum distance between the cumulative distribution functions (CDF) for the data ($S(k)$) and the fitted model ($P(k)$):

$$D = \max_{k \geq k_{min}} |S(k) - P(k)|$$

The estimate of k_{min} is the value of k that minimises the KS statistic, D.

A goodness-of-fit test is used to determine whether the hypothesised power-law model is plausible. This goodness-of-fit test uses a standard approach that calculates the 'distance' between the distribution of the empirical data and the hypothesised model, and then compares this distance measurement with measurements for comparable synthetic data sets. The p-value used to quantify the plausibility of the hypothesis is fraction of the synthetic distances that are larger than the empirical distance. Statistical fluctuations only can account for the distance between the model and the data if p is close to 1. This is not the case if p is small ($p \cong 0$). Hence, the model is a plausible fit to the data if p is above a threshold value. This argument can framed as a hypothesis test (Gillespie, 2015):

H_0: data is generated from a power-law distribution.

H_1: data is not generated from a power-law distribution.

Clauset et al. (2009) suggest the power-law hypothesis (H_0) can be rejected if $p<0.1$.

The poweRlaw programme implements a bootstrapping procedure to carry out a goodness-of-fit test. The KS statistic is used to calculate the distance between a model and data (empirical and synthetic). The bootstrapping procedure was used to obtain the 95% confidence limits for each of the values of k_{min} and α.

In addition, the procedure developed by Clauset et al. (2009) provides a robust methodology for comparing the power-law fits to other models (assuming the power-law model cannot be rejected), including a power-law with an exponential cut-off, a Poisson distribution, an exponential distribution, and a log-normal distribution. The form of each of these models are described in Table 4.10.

Table 4.10 Definitions of distributions fitted to the in and out degree distributions (Clauset et al., 2009).

Distribution	
Poisson	$\frac{\mu^x}{x!}$
Power-law with exponential cut-off	$x^{-\alpha}e^{-\lambda x}$
Exponential	$e^{-\lambda x}$
Log-Normal	$\frac{1}{x}exp\left[-\frac{(lnx-\mu)^2}{2\sigma^2}\right]$

To compare each of these distributions with a power-law model, Clauset et al. (2009) uses a likelihood ratio test. This computes the likelihood of the data under two competing distributions and computes a test statistic that is the ratio of the logarithm of these of these likelihoods. If this log likelihood ratio, $\mathcal{R} > 0$ the power-law model is a better fit to the data and, conversely, the alternative distribution is a better fit to the data if $\mathcal{R} < 0$. If $\mathcal{R} = 0$, the test is inclusive, and the data cannot distinguish between either model. Therefore, a p-value is also computed using a method proposed by Vuong (1989) to quantify if any deviation in $\mathcal{R}$ from zero are statistically significant. If $p \leq 0.1$, Broido and Clauset (2019) suggest the null hypothesis that $\mathcal{R} = 0$ can be rejected, and the models can be evaluated using the sign of $\mathcal{R}$. Again, this test procedure is implemented in R using the poweRlaw package for the Poisson, exponential, and log-normal distributions. However, the power-law with exponential cut-off is a nested version of the power-law model. Hence, a modified likelihood ratio test is used (Clauset et al., 2009). This test is not

included in the poweRlaw package. Hence, it is implemented by developing original code using functions written in R by Cosma Shalizi (Clauset, 2022).

4.3.3.2 Summary of Procedure for Degree Distribution Analysis

Broido and Clauset (2019) claim they have developed a strict test to establish whether a network has a scale-free structure. A scale-free structure and cluster are two features of sign networks that the model developed by Steyvers and Tenenbaum (2005) model of sign network growth predict. Therefore, using the strict test in this analysis is a robust way of establishing whether one of features predicted by the Steyvers and Tenenbaum (2005) model is present in the participants' sign networks. However, Voitalov et al. (2019) have questioned the definition of scale-free adopted by Broido and Clauset (2019) and some of their statistical methods. These criticisms raise concerns over the validity of the strict test to establish whether a network has a scale-free structure.

However, the possibility that placeholder sign structure grows around a hub-word is the central insight provided by the Steyvers and Tenenbaum (2005) model in terms of the Quinian bootstrapping process, rather than whether this structure is scale-free. Hence, the pivotal question in the analysis of the analysis of the degree distributions is whether a Poisson distribution can be rejected in favour of heavy tailed distributions (i.e., a power-law, log-normal, exponential, power-law with exponential cut-off). The strict test outlines robust statistical procedure to answer this question. Hence, despite the concerns over whether it can establish whether a network is scale-free, the strict test is adopted as the procedure for analysing the degree distributions.

Moreover, the strict test uses the same statistical procedure that Morais et al. (2013) employed in their analysis to demonstrate that a power-law with exponential cut-off was a better fit to both their sign networks and the networks created by Steyvers and Tenenbaum (2005) than a pure power-law. This finding is significant as it suggest that the Steyvers and Tenenbaum (2005) may have to be modified if it is to completely explain the formation of sign networks. Hence, using the strict test in this analysis enables the findings or the Morais et al. (2013) study and this analysis to be easily compared. This is important because if this study confirms the findings of the Morais et al. (2013) study (i.e., the sign networks might not have a scale-free structure), this will suggest the Steyvers and Tenenbaum (2005) model will have to be modified to explain the formation of sign networks. Hence, whether the networks meet the strict test's criteria for a scale-free structure is still considered as part of this analysis. However, this discussion takes place with the caveat about the validity of the strict test in mind.

4.3.3.3 K_{in} Distributions

4.3.3.3.1 Initial Power-Law Fits

The power-law model fits to the cumulative k_{in} distributions for each participant are shown in Figure 4.7, and the fit parameters are summarised in Table 4.11. The null hypothesis that a power-law model is a good fit to the data cannot be rejected for most of the distributions ($p \geq 0.1$ for all these distributions). However, a power-law model can be rejected for the k_{in} distributions of 12P02 and 12P03. Therefore, the criteria that a power-law model cannot be rejected for 50% of the networks in the strict test for a scale-free structure is satisfied in the physics and the psychology groups.

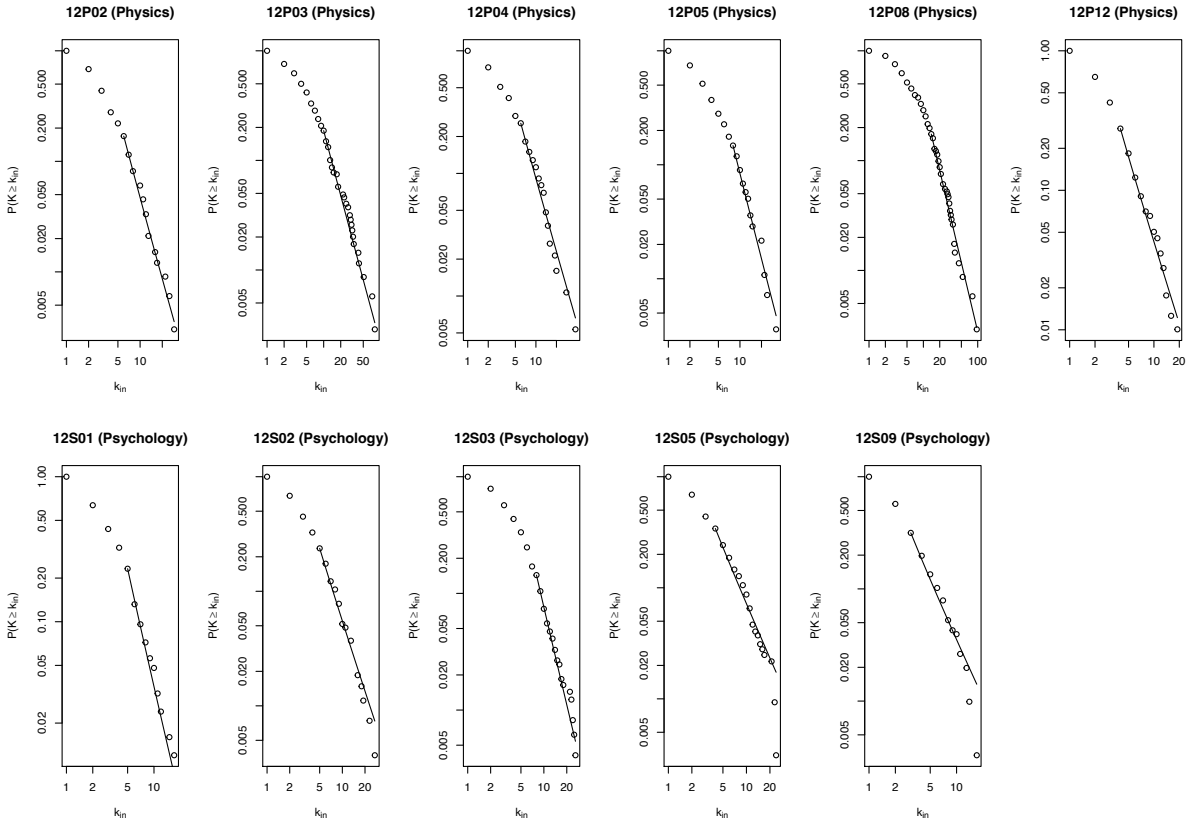


Figure 4.7 The cumulative vertex in-degree (k_{in}) distributions for the sign networks of each participant. The distributions are plotted with log-log axis and the solid line represent a best-fitting power-law model. Note the power-law models are fitted the tail of the distributions above a cut-off value, k_{min}.

*Table 4.11 The parameters and Kolmogorov-Smirnov statistics (KS) for a power-law model fit to the vertex in-degree distributions for each participant. Note * indicates that the hypothesis that a power-law model is good fit to the data can be rejected. Note, α is the scaling parameter (or exponent) for each power-law fit, k_{min} is the cut-off value for k in each fit that defines the tail of the distribution, and n_{tail} is the number of vertices tail of the distribution above k_{min}.*

Cohort	Participant	α	95% Confidence Limits		k_{min}	95% Confidence Limits		n_{tail}	KS	P
			α_{lower}	α_{upper}		$k_{min\ lower}$	$k_{min\ upper}$			
Physics	12P02	3.36	2.24	4.73	6	2	9	56	0.08	0.05*
	12P03	2.90	2.20	3.80	10	4	16	65	0.06	0.48
	12P04	2.91	2.12	4.21	6	2	10	48	0.08	0.14
	12P05	3.40	2.10	4.99	8	2	12	41	0.06	0.46
	12P08	3.08	2.51	4.01	14	8	20	60	0.07	0.22
	12P12	2.89	2.49	3.54	4	3	7	110	0.05	0.27
Psychology	12S01	3.52	2.80	5.29	5	4	8	58	0.05	0.46
	12S02	2.96	2.27	4.00	5	2	8	64	0.07	0.13
	12S03	3.69	2.90	5.07	8	5	12	70	0.04	0.75
	12S05	2.57	2.18	3.11	4	2	8	110	0.07	0.05*
	12S09	2.66	2.30	3.23	3	2	5	96	0.05	0.22

The power-law models, which have been fitted to the k_{in} distributions, are characterised by the power-law exponent, or scaling factor, α, and the region of the distribution the power-law model is applied to, which is determined by the minimum degree k_{min}. The descriptive statistics for these values and the number of vertices in the tail, n_{tail}, of each distribution are given in Table 4.12. As with the basic structural measures of each network (described in section 4.2.3), a Kolmogorov-Smirnov test was used to determine whether there whether any statistically significant differences between the distribution of these values in the physics and the psychology cohorts. The results are also shown in Table 4.12. The p-values (α, p=0.955 and k_{min}, p=0.503) for the α, k_{min} tests suggests that the null hypothesis that the values of α and k_{min} are drawn from the same distribution cannot be rejected. Hence, the differences in the power-law models fitted to the k_{in} distributions for physics and psychologists can be explained through statistical fluctuations. This suggests the power-law model that fits the physicists' and psychologists' k_{in} distributions is essentially the same. This again suggest the networks for both groups have a similar structure, which could be scale-free as 50% of the distributions in both groups follow a power-law.

Table 4.12 The descriptive statistics for the power-law fits to the vertex in-degree (k_{in}) distributions. Note, α is the scaling parameter (or exponent), k_{min} is the cut-off value for k in each fit, and n_{tail} is the number of vertices tail of the distribution above k_{min}. M and SD indicate the mean and standard deviations for each of the parameter. KS and p indicated the KS statistic and p-values for the KS tests carried out to test the null hypothesis that the values of these parameters in the physics and psychology groups are drawn from the same distributions.

Cohort	Alpha				K_{min}				N_{tail}			
	M	SD	KS	p	M	SD	KS	p	M	SD	KS	p
All	3.08	0.36	-	-	6.64	3.20	-	-	70.73	23.87	-	-
Physics	3.09	0.24	0.400	0.688	8.00	3.58	0.633	0.224	63.33	24.41	0.500	0.503
Psychology	3.08	0.50			5.00	1.87			79.60	22.33		

The strict test specifies several other criteria that need to be satisfied to strengthen the claim that a type of network has as scale-free structure. The mean scaling factor for all is networks is 3.08, for physics is 3.09, and for psychology is 3.08. These values fall to 2.99, 3.04, and 2.92 respectively if the networks for 12P02 and 12S05, where a power-law model can be rejected, are excluded. These mean values for α are either below, or just above the upper limit of 3 in the range of values of $2 < \alpha < 3$ which defines a scale-free structure. However, in the physics, the values of α range from 2.89 to 3.36 (excluding 12S05, where the power-law model can be rejected, with three networks with < 3 (n=5), and in psychology from 2.57 to 3.69 (excluding 12S05, where the power-law model can be rejected), and two networks with < 3 (n=4). Hence, in both groups all the values of α do not violate the lower bound of $2 < \alpha$ and the criteria that at least 50% of networks have $2 < \alpha < 3$ is satisfied in both groups. However, the uncertainty in the values of α is large in most cases and does not exclude the possibility that $\alpha > 3$ is all networks. Hence, the evidence is weak that the $2 < \alpha < 3$ criteria for at least 50% of networks has been satisfied.

4.3.3.3.2 Comparison Between a Power-Law and Poisson Model

The previous section establishes that a power-law model cannot be ruled out as an appropriate model of the k_{in} distributions all but two cases (12P02, 12S05). This is consistent with the distributions in both the physics and psychology cohort having a scale-free structure. Moreover, there is some evidence that the criteria in the strict test of scale-free structure for at least 50% of networks to have $2 < \alpha < 3$ is satisfied. Furthermore, most of the distributions contain at least 50 vertices, which adds weight to the claim that the networks could have a scale-free structure. However, the strict test requires that other alternative models are found to be worse fits to the distributions than a power-law. This section test whether a Poisson distribution is a better fit to the data (other distributions are compared in the next section). This comparison is at the centre

of this analysis because it established whether the k_{in} network structures can arise from simply making connections randomly or whether a different underlying non-random mechanism (e.g., preferential attachment) is needed to form the networks.

The results of the comparison between the power-law and Poisson distribution models of the k_{in} distribution are presented in Table 4.13 and are shown in Figure 4.8. The p-values for all the distribution are $p<0.1$ and often $p<0.05$. Hence, the null hypothesis that $\mathcal{R} = 0$ as a result of statistical fluctuations in this random variable can be rejected (Clauset et al., 2009) and the test can distinguish between a power-law or a Poisson distribution is a better fit in all cases. The log likelihood ratio statistic $\mathcal{R} > 0$ for all the distributions, which indicates that the power-law model is a better fit to all the k_{in} distributions. This suggests that a purely random model cannot explain the k_{in} network structure and a scale-free structure might be more appropriate. However, to strengthen this claim, alternative models for the distributions need to be considered.

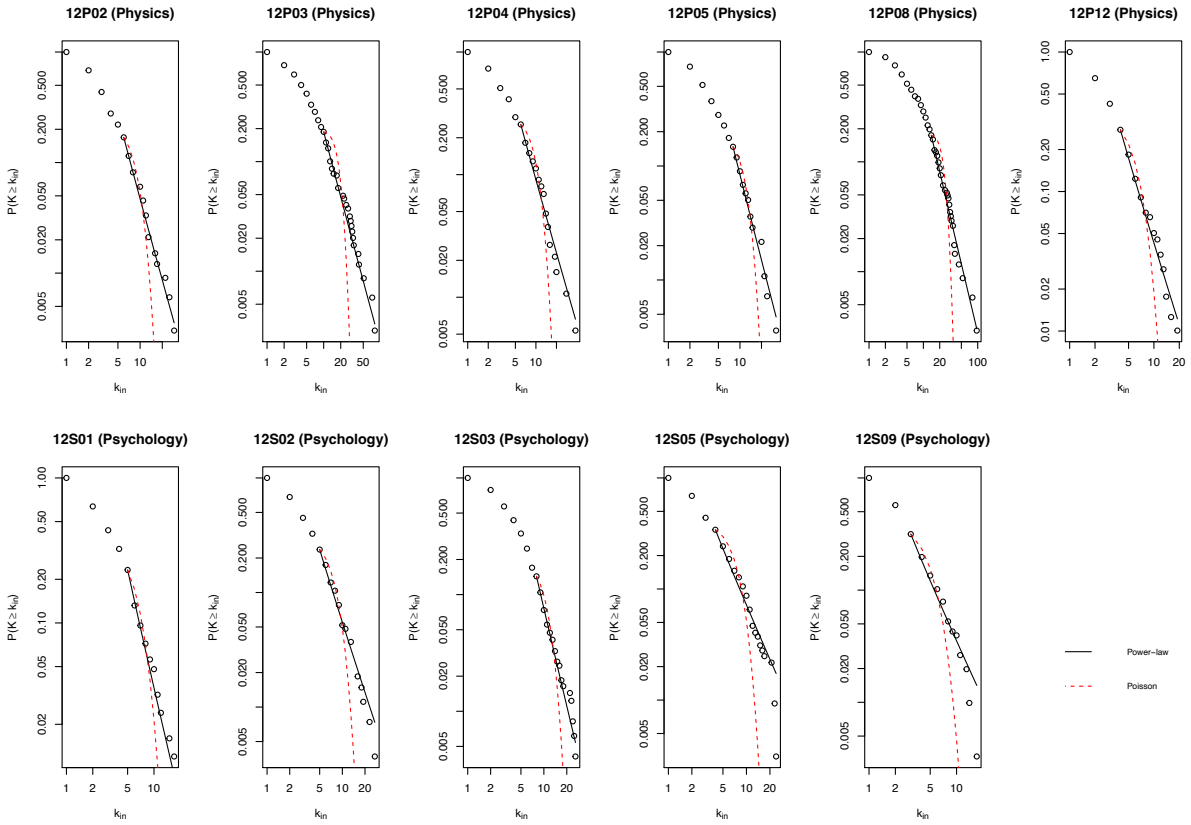


Figure 4.8 The cumulative vertex in-degree (k_{in}) distributions for the sign networks of each participant. The distributions are plotted with log-log axis and the solid line represent a best-fitting power-law model and the broken line a Poisson distribution. Note the models are fitted the tail of the distributions above a cut-off value, k_{min}.

*Table 4.13 The test statistic for the comparison between a power-law model and a Poisson distribution fit to the vertex in-degree (k_{in}) distributions for the sign networks. k_{min} is the cut-off value for k in each fit that defines the tail of the distribution. α is the scaling parameter (or exponent) in the power-law model and λ is the rate parameter in the Poisson distribution. R is the log likelihood ratio and P is the associated p-value. * indicated that the null hypothesis that R=0 can be rejected ($P \leq 0$).*

Cohort	Participant	k_{min}	α	λ	R	P
Physics	12P02	6	3.36	8.50	2.03	0.04*
	12P03	10	2.90	18.79	2.76	0.01*
	12P04	6	2.91	10.06	1.71	0.09*
	12P05	8	3.40	11.53	1.66	0.10*
	12P08	14	3.07	24.27	2.07	0.04*
	12P12	4	2.89	6.18	3.39	0.00*
Psychology	12S01	5	3.52	6.28	2.47	0.01*
	12S02	5	2.96	7.79	1.93	0.05*
	12S03	8	3.69	10.80	3.25	0.00*
	12S05	4	2.57	7.41	3.40	0.00*
	12S09	3	2.66	4.79	2.50	0.01*

4.3.3.3.3 Comparison with Other Distributions

This section compares a power-law model fit to the k_{in} distributions with other relevant distributions, which are often overlooked as plausible models and, therefore, viable alternatives to a power-law (Broido & Clauset, 2019). There is some evidence that this is the case in previous studies of sign networks. For instance, Morais et al. (2013) reanalysed the sign networks created by Steyvers and Tenenbaum (2005) and found that a power-law with an exponential cut-off is a better fit to the degree distributions than a power-law. This analysis follows this lead and compares a power-law model with a power-law with exponential cut-off. But also compares a power-law with an exponential and log-normal distribution, as suggested by Broido and Clauset (2019).

The test statistics for the comparisons between a power-law model and the alternative models are shown in Figure 4.9 and Table 4.14. The log likelihood ratio statistics $\mathcal{R} < 0$ in the comparison between a power-law and a power-law with an exponential cut-off for all distributions. However, the p-values ($p<0.1$) suggests the null-hypothesis of $\mathcal{R} = 0$ can only be ruled out in one case for the physics distributions (12P12) and for two cases for the psychology distributions (12S02 and 12S09)[7]. Hence, the data can only distinguish between the two models of the distributions for

[7] Note that the p-value for 12S05 was also <0.1 (p = 0.000). However, a power-law model can be ruled out as plausible model for this distribution. Hence, a valid comparison between a power-law model and a power-law with an exponential cut-off cannot be made in this case.

these networks. Similarly, $\mathcal{R} < 0$ in all the comparisons between a power-law and a log-normal distribution. But again, $\mathcal{R}$ is only significantly different from zero in one case: 12S05. However, this is not a valid comparison because a power-law model can be rejected as an appropriate fit to this distribution. Similarly, in the comparison between a power-law model and exponential distribution, the p-values only significant for 12P03. In this case, $\mathcal{R} > 0$, indicating that a power-law model is a better fit. For most of the comparisons between a power-law model and the three alternative models, the data is not able to distinguish between these models.

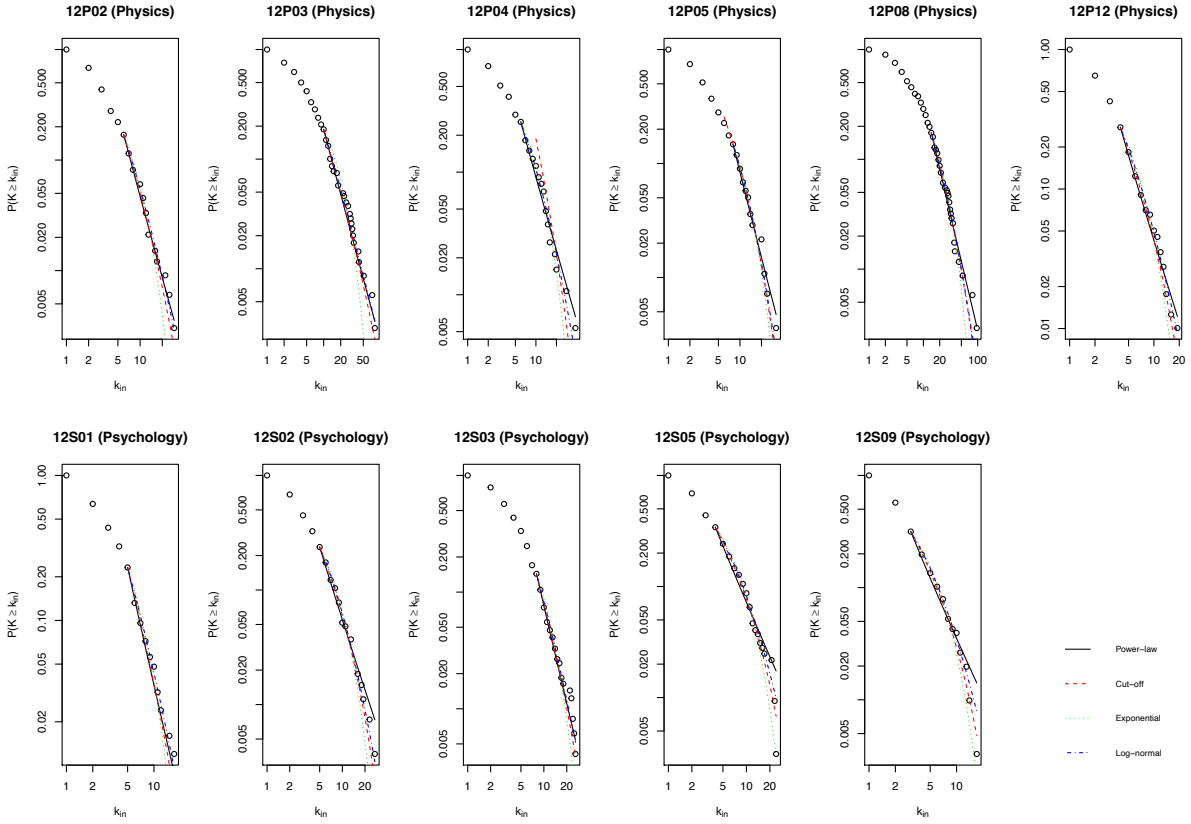


Figure 4.9 The cumulative vertex in-degree (k_{in}) distributions for the sign networks of each participant. The distributions are plotted with log-log axis and the solid line represent a best-fitting power-law model and the broken lines represent alternative models (see key). Note the models are fitted the tail of the distributions above a cut-off value, k_{min}.

*Table 4.14 The test statistic for the comparison between a power-law model and alternative model fits to the vertex in-degree (k_{in}) distributions for the sign networks. k_{min} is the cut-off value for k in each fit that defines the tail of the distribution. α is the scaling parameter (or exponent) in the power-law model. In the power-law with exponential cut-off model, γ is the rate parameter. In the exponential distribution, λ is rate parameter. In the log-normal distribution, μ is the expected value and σ is the standard deviation. For each comparison, R is the log likelihood ratios and P is the associated p-value. * indicated that the null hypothesis that R=0 can be rejected ($P \leq 0$).*

Cohort	Participant	K_{min}	α	Power-Law vs Power-Law with Cut-Off				Power-Law vs Exponential			Power-Law vs Log-Normal			
				α	γ	R	P	λ	R	P	μ	σ	R	P
Physics	12P02	6	3.36	2.62	0.05	-0.36	0.40	0.27	0.71	0.48	-0.86	1.20	-0.47	0.64
	12P03	10	2.90	2.64	0.01	-0.19	0.54	0.11	1.71	0.09*	-11.29	2.77	-0.21	0.84
	12P04	6	2.91	1.80	0.07	-0.97	0.16	0.20	0.10	0.92	0.85	0.96	-0.74	0.46
	12P05	8	3.40	1.48	0.12	-0.91	0.18	0.21	-0.38	0.71	1.50	0.71	-0.87	0.38
	12P08	14	3.07	2.28	0.02	-0.63	0.26	0.09	0.54	0.59	1.21	1.04	-0.67	0.50
	12P12	4	2.89	1.73	0.12	-2.36	0.03*	0.31	0.10	0.92	0.39	0.98	-1.34	0.18
Psychology	12S01	5	3.52	2.62	0.09	-0.40	0.37	0.36	0.64	0.52	-1.45	1.22	-0.46	0.64
	12S02	5	2.96	1.49	0.12	-1.69	0.07*	0.26	-0.29	0.77	1.06	0.82	-1.13	0.26
	12S03	8	3.69	2.74	0.06	-0.43	0.35	0.24	0.73	0.47	-0.76	1.14	-0.53	0.60
	12S05	4	2.57	1.30	0.11	-4.19	0.00*	0.24	-0.49	0.62	0.94	0.94	-1.69	0.09
	12S09	3	2.66	1.24	0.18	-3.58	0.01*	0.36	-0.66	0.51	0.66	0.88	-1.61	0.11

4.3.3.4 K_{out} Distributions

The procedure for fitting the k_{out} distributions is identical to the procedure used to fit k_{in} distributions, which aimed to establish whether the criteria specified in the strict test for scale-free structure were satisfied for the physics and psychology sign networks. As with the k_{in} distributions, this section describes how the evidence used to assess whether these criteria had been met was collected.

4.3.3.4.1 Initial Power-Law Fits

The power-law model fits to the cumulative k_{out} distributions for each participant are shown in Figure 4.10, and the fit parameters are summarised in Table 4.15. The null hypothesis that a power-law model is a good fit to the distribution cannot be ruled out for ten of the eleven distributions. The distribution for 12P08 is the only distribution where a power-law model can be rejected. This result is similar to the results for the k_{in} distributions. Hence, the criteria that a power-law model cannot be rejected for a least 50% of the networks is satisfied, for both physics and the psychology groups. However, unlike the k_{in} distributions the values for alpha for all the distributions fall outside the range $2 < \alpha < 3$ that specifies a scale-free distribution, even when the uncertainties, which are large, are taken in account. Hence, this criterion is not met for the k_{out} distributions. Moreover, the mean values of alpha for both the physics and the psychology groups are larger than the mean values for the kin distributions. But like the k_{in} distributions, the KS test suggest that any difference between the mean values of alpha and k_{min} arise from statistical fluctuations and, hence, these values are drawn from the same distributions. Therefore, there are no significant differences in the power-law model for the two groups.

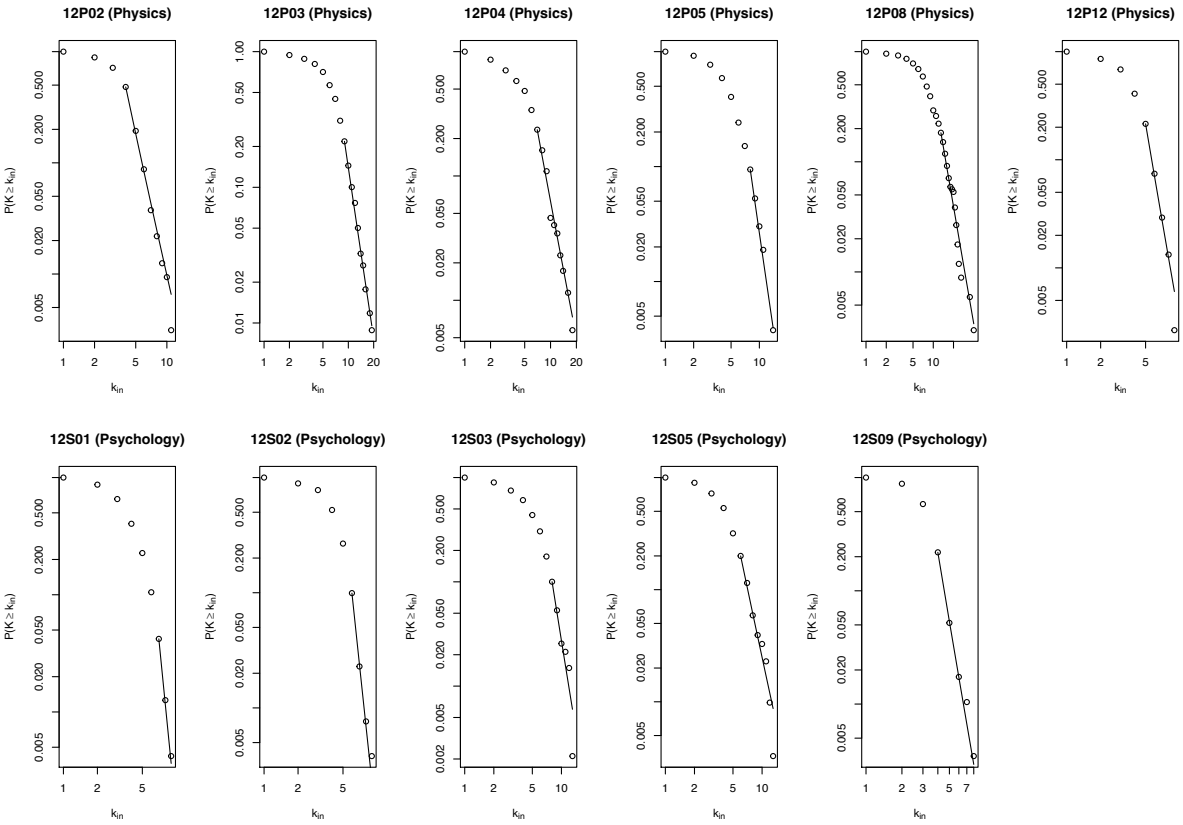


Figure 4.10 The cumulative vertex out-degree (k_{out}) distributions for the sign networks of each participant. The distributions are plotted with log-log axis and the solid line represent a best-fitting power-law model. Note the power-law models are fitted the tail of the distributions above a cut-off value, k_{min}.

*Table 4.15 The parameters and Kolmogorov-Smirnov statistics (KS) for a power-law model fit to the vertex out-degree distributions for each participant. Note * indicates that the hypothesis that a power-law model is good fit to the data can be rejected. Note, α is the scaling parameter (or exponent) for each power-law fit, k_{min} is the cut-off value for k in each fit that defines the tail of the distribution, and n_{tail} is the number of vertices tail of the distribution above k_{min}.*

Cohort	Participant	α	95% Confidence Limits		k_{min}	95% Confidence Limits		n_{tail}	KS	P
			α_{lower}	α_{upper}		$k_{min\ lower}$	$k_{min\ upper}$			
Physics	12P02	4.96	4.27	6.48	4	4	6	154	0.02	0.44
	12P03	5.04	4.01	7.00	9	7	12	74	0.05	0.17
	12P04	4.52	3.37	7.75	7	5	10	41	0.08	0.12
	12P05	6.42	3.88	11.23	8	5	9	25	0.06	0.26
	12P08	4.47	3.13	6.16	13	7	17	62	0.08	0.05
	12P12	6.70	5.41	10.27	5	5	7	81	0.02	0.72
Psychology	12S01	10.25	4.19	12.31	7	4	7	10	0.03	0.80
	12S02	9.76	5.12	13.24	6	4	6	26	0.01	0.91
	12S03	6.55	4.78	10.32	8	6	10	47	0.06	0.11
	12S05	4.84	3.56	7.64	6	4	8	61	0.04	0.45
	12S09	6.78	4.82	9.55	4	3	5	63	0.02	0.71

Table 4.16 The descriptive statistics for the power-law fits to the vertex out-degree (k_{out}) distributions. Note, α is the scaling parameter (or exponent), k_{min} is the cut-off value for k in each fit, and n_{tail} is the number of vertices tail of the distribution above k_{min}. M and SD indicate the mean and standard deviations for each of the parameter. KS and p indicated the KS statistic and p-values for the KS tests carried out to test the null hypothesis that the values of these parameters in the physics and psychology groups are drawn from the same distributions.

Cohort	Alpha				K_{min}				N_{tail}			
	M	SD	KS	p	M	SD	KS	p	M	SD	KS	p
All	6.39	2.00	-	-	7.00	2.57	-	-	58.55	38.57	-	-
Physics	5.35	0.97	0.633	0.177	7.67	3.20	0.333	0.922	72.83	44.89	0.500	0.357
Psychology	7.64	2.30			6.20	1.48			41.40	22.94		

4.3.3.4.2 Comparison with Power-Law Distribution

The comparison with between a power-law and Poisson model is show in Figure 4.11 and fit statistics are given in Table 4.17. The p-values for all distributions are >0.1. Hence, the null hypothesis that $\mathcal{R} = 0$ cannot be rejected and, hence, the data cannot distinguish between a power-law model and a Poisson distribution in all cases, in contrast to the k_{in} distributions were a power-law was a better fit in all instances. Moreover, whilst the $\mathcal{R}$ are not significant, they are not uniformly > 0. Hence, there are instances when a Poisson distribution could be a better fit than a power-law to the k_{out} distributions (see 12P05, 12S03, and 12S01)[8].

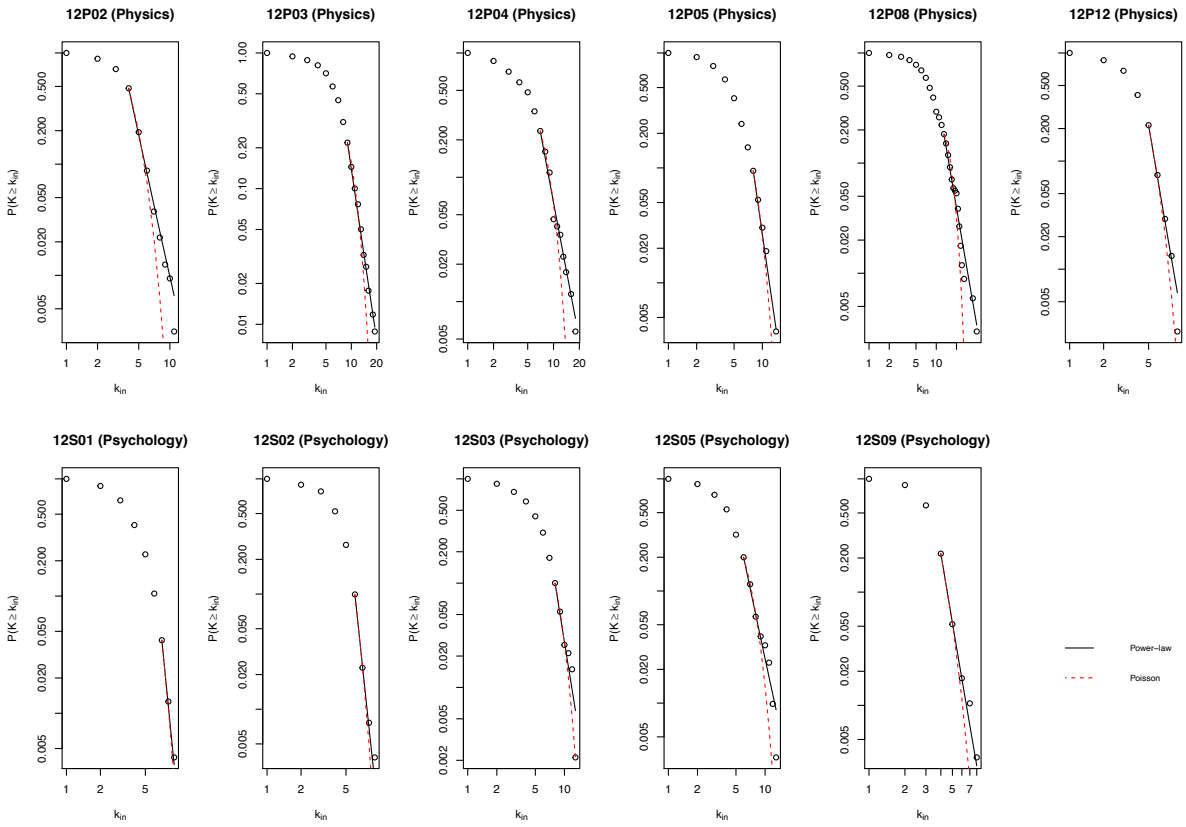


Figure 4.11 The cumulative vertex out-degree (k_{out}) distributions for the sign networks of each participant. The distributions are plotted with log-log axis and the solid line represent a best-fitting power-law model and the broken line a Poisson distribution. Note the models are fitted the tail of the distributions above a cut-off value, k_{min}.

[8] Note that for 12P12, $\mathcal{R} = -0.25$. However, a power-law model can be ruled out as plausible model for this distribution. Hence, a valid comparison between a power-law model and a Poisson distribution cannot be made in this case.

*Table 4.17 The test statistic for the comparison between a power-law model and a Poisson distribution fit to the vertex out-degree (k_{out}) distributions for the sign networks. k_{min} is the cut-off value for k in each fit that defines the tail of the distribution. α is the scaling parameter (or exponent) in the power-law model and λ is the rate parameter in the Poisson distribution. R is the log likelihood ratio and P is the associated p-value. * indicated that the null hypothesis that R=0 can be rejected ($P \leq 0$).*

Cohort	Participant	k_{min}	α	λ	R	P
Physics	12P02	4	4.96	2.64	1.03	0.30
	12P03	9	5.04	9.05	0.72	0.47
	12P04	7	4.52	7.16	0.65	0.52
	12P05	8	6.42	5.92	-0.54	0.59
	12P08	13	4.47	15.92	1.24	0.22
	12P12	5	6.70	2.45	-0.25	0.80
Psychology	12S01	7	10.25	2.48	-0.60	0.55
	12S02	6	9.76	1.95	0.45	0.66
	12S03	8	6.55	5.83	-0.40	0.69
	12S05	6	4.84	5.29	0.58	0.56
	12S09	4	6.78	1.54	1.06	0.29

4.3.3.4.3 Comparison with other Distributions

As with the k_{in} distribution and in accordance with the criteria of the strict test, the power-law model fits with k_{out} distributions were compared other appropriate models. In almost all comparisons, $\mathcal{R} < 0$. This indicates that power-law with exponential cut-off, exponential and log-normal might all be better fits to the distributions than the power-law model. However, the p-values for all the comparisons indicate that the null hypothesis that $\mathcal{R} = 0$ cannot be ruled out. Hence, the data cannot distinguish between any of these models.

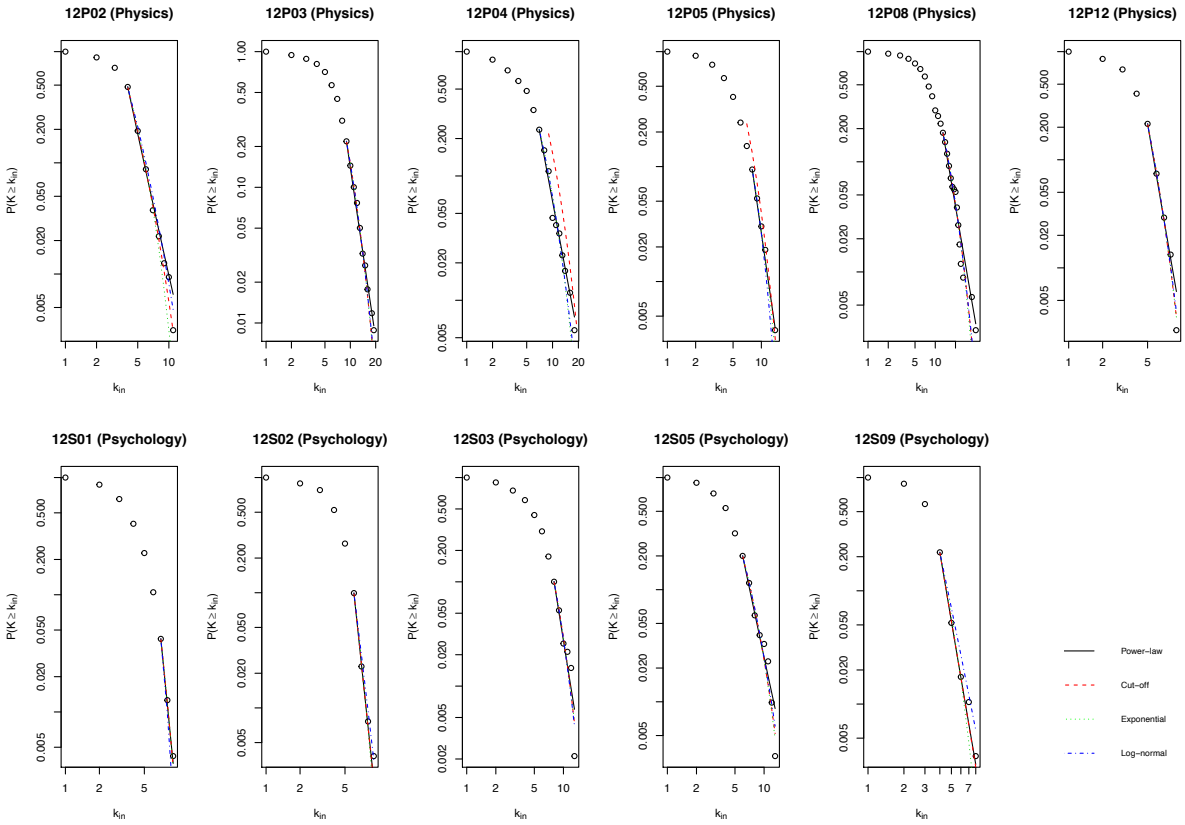


Figure 4.12 The cumulative vertex in-degree (k_{out}) distributions for the sign networks of each participant. The distributions are plotted with log-log axis and the solid line represent a best-fitting power-law model and the broken lines represent alternative models (see key). Note the models are fitted the tail of the distributions above a cut-off value, k_{min}.

*Table 4.18 The test statistic for the comparison between a power-law model and alternative model fits to the vertex out-degree (k_{out}) distributions for the sign networks. k_{min} is the cut-off value for k in each fit that defines the tail of the distribution. α is the scaling parameter (or exponent) in the power-law model. In the power-law with exponential cut-off model, γ is the rate parameter. In the exponential distribution, λ is rate parameter. In the log-normal distribution, μ is the expected value and σ is the standard deviation. For each comparison, R is the log likelihood ratios and P is the associated p-value. * indicated that the null hypothesis that R=0 can be rejected ($P \leq 0$).*

Cohort	Participant	K_{min}	α	Power-Law vs Power-Law with Cut-Off				Power-Law vs Exponential			Power-Law vs Log-Normal			
				α	γ	R	P	λ	R	P	μ	σ	R	P
Physics	12P02	4	4.96	2.71	0.37	-0.88	0.18	0.84	-0.05	0.96	0.15	0.64	-0.81	0.42
	12P03	9	5.04	0.16	0.36	-1.35	0.10	0.37	-1.11	0.27	1.72	0.46	-1.03	0.30
	12P04	7	4.52	0.72	0.34	-0.77	0.21	0.41	-0.74	0.46	1.51	0.49	-0.83	0.41
	12P05	8	6.42	-1.00	0.70	-0.60	0.27	0.61	-1.08	0.28	1.90	0.28	-0.69	0.49
	12P08	13	4.47	1.14	0.16	-1.05	0.15	0.21	-0.59	0.55	2.07	0.52	-0.84	0.40
	12P12	5	6.70	0.10	1.01	-0.70	0.24	1.03	-0.86	0.39	0.98	0.39	-0.79	0.43
Psychology	12S01	7	10.25	-1.00	1.38	-0.12	0.63	1.25	-0.70	0.49	1.79	0.17	-0.42	0.68
	12S02	6	9.76	8.38	0.19	0.00	0.94	1.36	0.26	0.80	-7.08	1.03	-0.07	0.94
	12S03	8	6.55	-1.00	0.71	-0.74	0.22	0.62	-1.15	0.25	1.67	0.35	-0.85	0.39
	12S05	6	4.84	0.90	0.43	-0.92	0.17	0.54	-0.86	0.39	1.16	0.51	-0.93	0.35
	12S09	4	6.78	6.40	0.07	0.00	0.93	1.29	0.73	0.47	-26.63	2.24	-0.53	0.59

4.3.3.5 Summary of the Analysis of the Vertex Degree Distributions

The aim of this section of the analysis was to look at the similarities and differences in the structures of the sign network of the physicists and psychologists by examining the vertex degree distributions. More specifically, this analysis aimed to establish whether the networks have a scale-free structure with hub-words, which would be 'signature' of a network growing through a Steyvers and Tenenbaum (2005) preferential attachment model that incorporates ideas from TOOC[9]. The strict test developed by Broido and Clauset (2019) was used to establish whether the networks have a scale-free structure. The collated results for the strict test are show in Table 4.19. The criteria for each network in the k_{in} and k_{out} distributions are shown in Table 10.10 and Table 10.11 in Appendix 10.4.

Table 4.19 The criteria for the strict test for a scale-free structure in a network developed by Broido and Clauset (2019). Whether the criteria have been met for the sign vertex in and out degree networks in each cohort is listed.

Strict test criteria		Cohort	
		Physics (n=6)	Psychology (n=5)
In-direct	**Super-weak** For at least 50% of networks, no alternative model to the distribution is favoured.	Inconclusive	Inconclusive
Direct	**Weakest** For at least 50% of networks, a power-Law distribution model cannot be rejected (P>0.1)	Met	Met
	Weak The power-law region contains at least 50 vertices.	Partially Met	Met
	Strong For at least 50% of the networks, $2 < \alpha < 3$	Met	Met
	Strongest Requirements of strong for at least 90% of networks and of super-weak for 95%.	Not Met	Not Met

[9] As section 2.6.2.3.1 explains, whether a sign network contains hub-words is the more salient finding in terms of network growth, rather than whether structure of the network is scale-free. However, a scale-free structure is a specific prediction of the Steyvers and Tenenbaum (2005) model. Hence, this is still considered in this analysis, despite the debate over whether the strict test is a valid means of testing whether the structure of the network is scale-free (Voitalov et al., 2019). The strict test is used because it closely follows the procedure used by Morais et al. (2013) to analyse the sign network in their study. Hence, these results can be easily compared with their findings. Moreover, the procedure in the strict test compares different models of the degree distribution, and importantly provide a means establishing whether a heavy-tailed distribution or a Poisson distribution is better fit. Hence, establishing whether a network contains hub-words (see section 4.3.3.2).

The strict test splits evidence for scale-free into direct and indirect evidence. In the direct category, most of the criteria for the strict test are satisfied or mostly satisfied (there are two physics k_{in} networks that do not satisfy the $n_{tail}>50$ criteria, but are close to this threshold, $n_{tail\ 12P04}=41$, $n_{tail\ 12P05}=48$) for the k_{in} distributions in both groups. This suggest that a power-law fit is a good fit to the k_{in} and $2 < \alpha < 3$. However, several caveats should be kept in mind. First, the uncertainties in the values of α are large, and often extend beyond the $2 < \alpha < 3$ range. Second, the sign networks are samples, rather than complete networks. This sampling could distort the measured values of α. Third, the snowball sampling technique requires the sign networks to be 'finished' using a technique described in section 3.3.4.1. Again, this technique could distort the shape of the degree distributions and affect the measured values of α.

There are no significant differences in α between cohorts. Hence, the evidence from the power-law fits suggests a similar scale-free structure with hub words for the k_{in} networks in both groups. However, the number of networks in each group is small. Therefore, the lack of any systematic differences in the power-law model to the distributions might be due to the inability of the KS test to detect these differences with small samples, rather than the lack of actual differences.

The indirect evidence that the k_{in} networks have a scale-free structure is less conclusive because the data is not able to distinguish between alternative distributions to a power-law model for most networks. Therefore, according to the strict test, the claim that k_{in} networks have a scale-free structure cannot be strong because the super-weak criteria is not met. In addition to the caveats raised about the accurate estimation of α, this raises some doubts about the robustness of the claim that the networks are scale-free.

However, the evidence that the k_{in} contain highly connected hub-words is more conclusive for two reasons. First, a power-law model is a better fit the k_{in} distributions than a Poisson distribution in all networks. This suggests a purely random model without hub-words cannot explain the structure of the networks. Second, there is evidence of systematic better fits to power-law with exponential cut-off and log normal (i.e., all values of $R<0$), even though the p-values for these fits are > 0.1. This is consistent with the analysis of sign networks carried out by Morais et al. (2013), which also found that a power-law with an exponential cut-off is significantly better fit in some networks. However, the power-law with an exponential cut-off and the log-normal distribution are both what Barabási and Pósfai (2016) define as cross-over distributions . This is because these distributions sit between distributions that are exponentially bounded, like the Poisson distribution, which do not contain hubs, and a power-law, which is the best-known example of a

heavy tailed distribution, which do contain hubs. Therefore, these cross over distributions contain hubs but the number of connections they can form is limited in some way. Hence, it is likely that the k_{in} networks contain hub-words because a power-law, power-law with an exponential cut-off, and a log-normal distribution can all fit the k_{in} distributions, but a Poisson distribution cannot.

The direct and the indirect evidence for a scale-free structure in the k_{out} distribution is less strong in both cases. In the direct category, a power-law model for the data cannot be rejected, but the tails of the distribution typically do not contain 50 vertices and the values of alpha are not $2 < \alpha < 3$. Hence, the direct evidence claim is 'weakest' at best. Moreover, the data is not sufficient to distinguish between any of the distributions, including the Poisson distribution. Hence, even a random structure for the k_{out} distribution cannot be ruled out. Hence, this analysis cannot claim with any certainty that k_{out} distribution have a scale-free structure and contain hub-words.

Following this discussion, the main findings from this analysis are:

- The k_{in} network may have structures that contain hub-words. This is a possible signature of a sign network growing through preferential attachment to these hub-words.
- The k_{in} networks for the physicists and psychologists are strikingly similar. As with the basic network parameter, this might suggest that the macro network structure is similar for both groups and the differences exist at a microscale, in the differences between words and the connections between individual words.

These two findings suggest that hub-words might play an important role in the growth of the sign network and in explaining any difference between the networks. This is looked at in the next part of this analysis, described in the next section (4.3.3.5).

4.3.4 Hub Words in the k_{in} Distributions

The analysis of the k_{in} distributions established that the sign networks could contain hub-words. This section begins to look at the microstructure of the networks by looking at whether the physics and psychologists share the same hub-words. In this analysis, hub-words are defined as the most highly connected signs in a network. Hence, they are the signs that have the highest degree. Typically, hubs have at least 1% of the edges in a network compared with 0.1% for most vertices (Newman, 2018). However, in the sign networks in this study, the mostly highly connected words (ranked 1 to 5) typically have ~ 1.5% of the connections. Hence, initially 1.5% of connections is used as a threshold to define the mostly highly connected hub-words in a network.

Table 4.20 show mostly commonly occurring hub-words in the physics and psychology group above this 1.5% threshold. These hub-words are *force*, *energy*, *mass*, *measurement*, *particle*, *time*, and *movement*. *Force*, *energy*, and *mass* are the top three ranking hub-words, which clearly correspond to key concepts in the mechanics topic. A Fisher exact test was used to determine whether if there were any statistically significant differences in the occurrence hub-words between the physics and the physics groups. There, were no significant differences at 0.05%. However, both *force* and *mass* were significant at 0.1%, with these words occurring more frequently in the physics group.

Growth is a key aspect of the Steyvers and Tenenbaum (2005) model and of Carey's theory of the origin of concepts (Carey, 2009). Hence, a lower threshold for hub-words of 1.0% and 0.5% of connections were applied to the networks. These lower thresholds attempted to capture any words in either group that might be developing into hub-words. The hub-words for both the lower thresholds are also shown in Table 4.20. Again, a Fischer exact test was carried out to determine if there were any significant difference in the occurrence of words in the physics and psychology cohorts. For the 1.0% threshold, the highest-ranking hub-words are: *force*, *energy*, *time*, *particle*, *mass*. There is no longer a significant difference between the occurrence of *force* in the two groups. This could suggest that *force* is a developing hub-word in the psychology group. However, the difference between occurrence of *mass* remains and has become more significant ($p<0.05\%$) because *mass* is now included as a hub-word for one more physics student. This suggest that *mass* might be a developing hub-word for the physicists but not the psychologists. For the lowest 0.5% threshold, the top five ranking hub-words are: *time*, *force*, *distance*, *energy*, *gravity*, *movement*, *particle*, *speed*, *direction*, and *mass*. Again, the only significant difference in the occurrence of hub-words is with *mass* and now also with *particle*. At the time of completing the word-association surveys, used to complete the sign networks, the physicists were completing a particle physics module. Hence, *particle* could also be a developing hub-word.

The key finding from this analysis of the k_{in} network hub-word are summarised below:

- The 1.5% threshold defines the tightest condition for hub-words. Under this condition, *force*, *energy*, and *mass* are the top three ranking hub-words. These words correspond to key concept in mechanics. Similarly, hub-words that are less highly ranking and appear as developing hubs under the lower thresholds (e.g., particle, time, measurement, movement) also correspond to key concepts.
- There are significant differences in the occurrence of hub-words between the physics and psychology cohorts. However, *mass* is the only hub-word that remains a significant

difference under all threshold for hub-words. Other differences in the occurrence of hub-words are only significant under difference thresholds. For instance, *force* is only significant under the highest 1.5% threshold. This could suggest *force* is a growing hub-word.

- Despite some significant differences in the occurrence of hub-words in the physics and psychology groups, there is more commonality in the occurrence with hubs than might be expected. *Force, energy, time,* and *particle* are commonly occurring hub-words in both the physics and psychology networks. This is despite the significant difference in the conceptual understanding of mechanics in the two groups demonstrated by the FCI tests. This result could be consistent with Carey's theory of the origin of concepts, which predicts that sign network structure develops before a fully coherent conceptual understanding.

Table 4.20 Top five ranking hub-words in the vertex in-degree sign networks (k_{in}) for the 1.5%, 1.0%, and 0.5% of connections thresholds. There are 6 participants in the physics cohort and 5 participants in the psychology cohort.

Rank	1.5%					1.0%					0.5%				
	Word	Physics	Psychology	Total	P	Word	Physics	Psychology	Total	p	Word	Physics	Psychology	Total	p
1	force	6	2	8	0.061	force	6	3	9	0.182	time	6	5	11	1.000
2	energy	4	2	6	0.567	energy	5	3	8	0.545	force	6	4	10	0.455
3	mass	4	0	4	0.061	time	4	3	7	1.000	distance	5	4	9	1.000
											energy	6	3	9	0.182
											gravity	6	3	9	0.182
											movement	5	4	9	1.000
4	measurement	1	2	3	0.545	particle	4	2	6	0.567	particle	6	2	8	0.061
	particle	2	1	3	1.000										
	time	1	2	3	0.545						speed	4	4	8	1.000
	movement	1	2	3	0.545										
5	change	2	0	2	0.455	mass	5	0	5	0.015	direction	3	4	7	0.545
	direction	2	0	2	0.455										
	distance	1	1	2	1.000						mass	6	1	7	0.015
	motion	2	0	2	0.455										
	small	0	2	2	0.182						work	4	3	7	1.000
	light	1	1	2	1.000										

4.3.5 Summary of Sign Network Analysis

The analysis and results in presented in this chapter address research question 2.

> *What are the similarities and differences between the sign networks for individuals with different levels of expertise in mechanics?*

The findings from this results chapter, and how they relate to this research question, are summarised as follows:

- The FCI scores demonstrate that the conceptual understanding of the participants in the physics cohort is consistent with a Newtonian understanding of mechanics. The participants in the psychology group have a common sense understanding of mechanics. In contrast, the vocabulary test scores indicate that there are no significant differences in the vocabulary knowledge between the two cohorts. Hence, any difference in the sign network for the two cohorts are likely to be due to difference in their level of expertise in mechanics.
- Key measures of the network features were used to provide a basic characterisation of the structure of the sign networks. There are no significant differences in the short average path lengths, clustering coefficients, mean degree, or diameter of the networks in the physics and psychology cohorts. Hence, the macrostructure of the network appears to be similar.
- The in and out degree distributions were analysed to determine whether the structure of the in and out sign networks could be scale-free. The results of this analysis are inclusive. Many of the criteria of the strict-test for a scale-free structure (Broido & Clauset, 2019) are satisfied, but alternative heavy tailed models for the distributions, which would indicate different network structures, cannot be ruled out by the data. However, for in-degree networks, purely random structure can be ruled out and all the distributions that fit the data are all heavy tailed, which strongly suggests the networks contain highly connected hub-words. Again, there are no significant differences between the networks in the physics and psychology cohorts, which again indicates that the macrostructure of the sign networks is similar.
- The analysis of the occurrence of hub-words in the networks is the first stage in examining the microstructure of the networks (i.e., what the individual 'dots' or vertices (signs) are the network and how they are connected to other vertices (signs), rather than just the overall structure represented by the 'dots' and 'lines' edges). There are several hubs words

that appear frequently in both the physics and psychology networks (*force*, *energy*, *time*, *particle*). *Mass* is the only hub-work that is nearly ubiquitous in the physics network but absent as hub-word in the psychology networks. The aggregate physics and psychology network also reveal some of the differences in the microstructure of the sign networks for these two different groups. First, the physics network appears to be more connected because the physicists share more connections than the psychologist. Some of these shared connection highlight salient differences. For instance, most physicists connect *force* with *acceleration,* and *mass* with *weight* and *gravity*. However, like the shared hub-word, the aggregate networks suggest the physicists and psychologist share some microstructure. For example, some of the connection between the signs in what seems to be motion cluster (e.g., *speed* with *time* and *velocity*).

5 Methodology to Investigate the Function of Hub-Words

5.1 Introduction

The analysis of the sign networks (described in chapter 4) demonstrated the presence of highly connected hub-words in the sign networks, which are associated with what could be considered key concepts in mechanics. Moreover, the Force Concept Inventory (FCI) scores suggests that these hub-words are present in networks before individuals develop a fully coherent conceptual understanding of mechanics. This is consistent with Carey's Theory of the Origin of Concepts (Carey, 2009) and the model of sign network development, developed by Steyvers and Tenenbaum (2005), that suggests that conceptual development is driven by the growth of sign structure around key words, which initially act as placeholder values rather than fully developed concepts. Within Carey's theory, the low FCI scores of the psychologists, which indicate a non-Newtonian conceptual understanding of mechanics, could be explained if the hub-words in their networks are still associated with placeholder values. Whereas these placeholder values have developed into fully formed concepts in the Physicist networks. This could, therefore, explain the high FCI scores, which indicate a Newtonian conceptual understanding of mechanics.

As discussed in section 4.2.1, the FCI is widely accepted as a valid and reliable measure of whether an individual solves problems in a way that is consistent with expert physicists. However, as a multiple-choice test, it does not directly capture the thought processes individuals use to solve problems. Hence, the inferences that can be drawn from the FCI about an individual's conceptual understanding are based on a priori assumptions made in construction of the instrument about the nature of the conceptual resources that are needed to arrive at the correct solutions to the multiple-choice problems. There is disagreement about whether the FCI's assumptions about the nature of these concepts are correct (e.g., Huffman and Heller (1995); Eaton and Willoughby (2018)).

Therefore, the FCI is limited in its ability to investigate the nature of an individual's conceptual understanding and, hence, is not the ideal instrument to explore the suggestion that the hub-words in the networks are associated with fully formed concepts in the physicist's network but placeholder values in the psychologist's networks. Instead, this suggestion, and the more general

question concerned with how individuals use hub-words to reason, is addressed in the second phase of this research, which answers research question 3.

> *Do individuals use hub-words when they reason about a scientific demonstration or experiment? If so, how?*

Using a demonstration or experiment is a well-established method of investigating how participants reason within a several scientific domain of knowledge (e.g. Black & Lucas, 1993; Driver et al., 1985; Driver et al., 1994; Levrini & DiSessa, 2008; Seah, 2015) and also specifically within mechanics (e.g., diSessa & Sherin, 1998; Ioannides & Vosniadou, 2002; Parnafes, 2007). Often participants are asked to describe and/or explain what they have seen in the demonstration, in either an oral or written format. Hence, a demonstration is a stimulus for participants to produce examples of language use. These examples can then be used as proxies for participants' thinking and can be used to identify patterns of thought that reveal possible connections between concepts. These systems of connection define a concept in an intuitive view of scientific concepts. Moreover, analysis of language can reveal when concepts in intuitive theories are incommensurate (Carey, 2009).

The methodological approach used in these studies is broadly the approach used in this section of research to investigate the function of the hub-words identified in the sign networks. Several demonstrations were used to prompt participants to describe and explain what they have seen and, therefore, capture examples of participants' use of language. These examples of language use were then analysed to identify possible patterns or connection in the participants' use of hub-words. These connections were used to build an understanding of how hub-words are used to construct descriptions and explanations of dynamic physical phenomena. This analysis was used to identify any systematic differences in the use of hub-words between the physics and psychology cohorts. Moreover, this analysis was used to explore whether hub-words seem to be associated with a recognisable mechanics concept, or a placeholder value.

The selection of the demonstrations and associated questions are discussed in section 5.1.3. In previous studies, participants' use of language has been captured through writing or by interview. In this study, interviews are used, rather than written tasks. The justification for using interviews, the interview structure, and discussion of validity and reliability are described in section 5.1.2. The analysis of the interview data is described in section 5.2. Again, the proposed methodology for this phase of the research was tested in a second phase of the pilot-study. The results of this pilot

study have informed the design of the main study and are presented here when they are relevant. The results of the analysis are presented in Chapter 6.

5.1.1 Participants

To answer research question 3, the participants used in the first phase of the study (investigation of sign networks) needed to be used again in this second phase (interviews). The description of the participant sample and the reasons for selecting these participants was discussed in section 3.3.2. Note the ethical issues regarding recruitment and student participation in this study have also been discussed in section 3.3.2.4. All six of the physics participants and all five of the psychology participants took part in the interview phase.

5.1.2 Procedure

To address research question 3, participants were shown demonstrations of different instances of moving objects during a semi-structure interview. During the interview, they were asked to describe and then explain these instances of motion. Each interview was audio recorded and then transcribed to enable analysis of the participant's use of hub-words (see section 5.1.4).

Several studies have asked participants to write explanations of natural phenomena (e.g., the germination of seeds) to obtain examples of their use of scientific language (e.g. Seah, Clarke, and Hart (2015)). However, Wellington and Osborne (2001) point out that many students are unfamiliar with the style of scientific writing and can therefore find writing about their ideas particularly difficult. Hence, participants' writing may not accurately reflect their use of spoken language to reason and, therefore, their conceptual understanding. This claim is consistent with the findings of (Taylor, 2012). In this study, 15-16-year-old students were asked to write written explanations of the motion of a film-canister rocket. Subsequent interviews suggested some students had a reasonable understanding of key terminology (e.g., resultant force), but they chose not to use this terminology in their written explanations.

In contrast with written tasks, interviews involve an oral exchange of ideas that allow students to verbalise their description and explanations. In an interview, as with any written task, language is being used as a proxy for thought. Some caution is needed when equating language with thought because individuals may begin to use words before they have a fully developed understanding of any concept they represent (e.g., see Bloom (2000)). However, this lag between the use of a word in spoken language and a fully developed conceptual understanding is an expected feature of

TOOC. Hence, any unexpected use of language in an interview could reveal that a participant is still developing their conceptual understanding. This may not the case with written language, when formulating these thoughts in writing adds additional cognitive strain, which is not anticipated to be part of the conceptual development process (Wellington & Osborne, 2001).

Moreover, Ginsburg (1997) and diSessa (2007) argue that interviewing can often provide opportunities to enhance data collection by exposing more of the participants' thinking by reframing problems, reflecting their thinking back to them and acting in opportunistic ways to explore previously unconsidered situations. Again, this view is supported by Taylor (2012), who found that some questioning was often required for some students in their study to move beyond simple one word or phrase answers and add detail to their explanations and descriptions. Furthermore, as has already been noted, further questioning in interviews can probe participants' understanding of key terminology. This option is important in this research given that it has been suggested that some participants may still associate hub-word with placeholder values, rather than fully formed concepts. Hence, questioning around these hub-words could establish the extent of a participant's conceptual understanding of a concept associated with these terms.

This freedom to move in a direction determined by circumstance to freely explore participants' use of hub-words is consistent with a less structured approach to interviewing (Cohen et al., 2018, p. 509). However, to address research question 3, a comparison between the use of participants' hub-words was required. Hence, some structure to the interviews was needed to enable this comparison to take place by capturing examples of participants' use of language when they were reasoning about similar physical phenomenon (Cohen et al., 2018, p. 509). This structure was achieved by using the same demonstrations in each interview and using the same questions to begin any discussion and similar question prompts to sustain and develop any discussion. The development of the demonstrations, interview schedule, and starting and prompting questions are described in the materials section (5.1.3).

This section has argued that semi-structured interviews can capture the instances of language use when participants describe and explain instances of motion that will be required to answer research question 3. However, diSessa (2007) highlights three criticisms that have been made of interviews concerning their ability to collect valid and reliable data. These criticisms are addressed in the next three sections.

5.1.2.1 Ecological Validity

Interviews provide ecologically valid data if they are able to reveal participants' thinking that applies in more usual situations (diSessa, 2007). Therefore, any discussion about whether an interview is likely to gather ecologically valid data needs to consider how any task, completed during an interview, differs from what a participant might be expected to do in more usual circumstances. In the pilot and main study interviews, each participant was asked to describe and explain what happens during the several demonstrations. The interviewer would ask questions to attempt to make each participant's thinking more explicit. Hence, in a sense, the interviewee and interviewer were working together to construct a description and explanation of what happened during each demonstration.

diSessa (2007) argues that this '*inquiry and mutual inquiry [to generate an explanation] are natural activities in which people frequently and completely engage'* (p541). Mercer (2013) goes further arguing that using our 'social brains' to work together to solve problems collaboratively is a distinct feature of human cognition. Hence, during the interviews, the participants are not being asked to perform a task that is completely unfamiliar to them or that they do not attempt outside interview environment. Moreover, Mortimer and Scott (2003) identify similar patterns of interaction, where a teacher initiates a response with a question and then provides feedback that encourages a student to develop their thinking, in science lessons in the UK. Therefore, the interviews are likely to generate explanations similar to those in a science classroom or in joint discussion.

Furthermore, the problems presented during the task (e.g., explaining the motion of the cup-cake cases) are unlikely to be entirely unfamiliar to the participant as they will have met similar explanations of falling objects during their GCSE studies (AQA, 2015a, 2015b). Therefore, it is reasonable to assume that these interviews will generate ecologically valid data, which are at least applicable to how students might reason about the scientific problems they encounter as part of a school curriculum. Table 5.1 list excerpts from pilot study interview support this assumption. These excerpts demonstrate how participants and interviewers respond and ask questions to develop an atmosphere of mutual enquiry that generates descriptions and explanations that are consistent with 'expected' answers for what is taught in the classroom (i.e., an explanation of terminal velocity). Moreover, non-physics answers can be identified with other misconceptions (e.g., heavier objects fall faster). This is evidence that semi-structured interviews built around asking participants to describe and explain what they see in a demonstration can generate ecologically valid data.

Table 5.1 Excerpts from participant interview demonstrating 'mutual enquiry' by interviewee and interviewer. Different levels of scaffolding are shown by each excerpt.

Mutual enquiry leading to an 'accepted' explanation of terminal velocity. *Minimal Scaffolding*	**12P12:** That [the cup-cake case] has, umm well, accelerated to some, some maximum, whatever that is, and then it's going to hit the floor and decelerate again very quickly to be zero. **RT:** OK. Let's just put a number on that. This is interesting because what you said there is you said it accelerates until it reaches. I think we said some maximum. **12P12:** Yeah, terminal velocity. **RT:** Right. OK. Can you explain why it reaches this terminal velocity? **12P12:** So as, kind of discussed earlier, air resistance is proportional to velocity. And in this case, the driving force mg is constant. So, eventually, as the cake tin accelerates air resistance will increase until it is equal to mg, at which point there will be no resultant force, so no further acceleration will take place.
Mutual enquiry leading to an 'accepted' description of the motion of cup-cake case when it lands on a desk. *Extensive Scaffolding.*	**12P08:** [The cup-cake case] is stopped by the, umm, the upward force from the table, the resistance force from the table, some of its energy is expelled as sound. Yeah, and it just becomes stationary. **RT:** Yeah, we've said it becomes stationary. So, what we're saying is, it stops. **12P08:** Yeah. **RT:** When it stops it, it's become stationary. And does it speed change then? **12P08:** Yes. **RT:** Okay. So, what happens to its speed? **12P08:** So, it becomes zero. **RT:** So, does it accelerate when it hits the desk? **12P08:** Yes, because a force from the table is enacted on it. It's going to be very fast acceleration.

However, the extracts from Table 5.1 also illustrate how mutual enquiry can potentially led participants in a particular direction or enable them to co-construct a description or explanation they would not be able to independently arrive at, because the interviewer's questions scaffold their thinking (see Howe (2013) and references within). Therefore, the potential scaffolding of a response does raise concerns about the ecological validity of the interview data. Hence, the extent to which follow up question scaffold students' thinking needs to be carefully considered. This issue is addressed in section 5.1.3.2, which outlines how follow up questions are used and phrased to avoid overly scaffolding participants' responses.

Finally, diSessa (2007) also argues that for the condition of 'mutual enquiry' to be satisfied, participants must be comfortable with expressing their ideas freely, without concern for the legitimacy of what they are saying or coercion from the interviewer. Creating these conditions during any interview is potentially more problematic because of the relationship between the researcher and interviewer, and the participants. Hence, this issue may limit the efficacy of interviews to collect valid data. This problem is discussed in the next section.

5.1.2.2 Coercive/Seductive practice

For the condition of 'mutual enquiry' to be satisfied, participants must acknowledge that the interview is designed to help freely explore their thinking (diSessa, 2007). Moreover, they need to be willing to share this thinking with the interviewer. Unfortunately, the researcher's position as a teacher within the target school may create a situation where a participant might not feel able to freely express their ideas because of the relationship between students and teachers. Particularly, because of the nature and role of questioning between teachers and students in a school context. Questioning in a school context is often not used to facilitate a genuine exploration and sharing of ideas (Wellington & Osborne, 2001), in contrast to what was observed by Mortimer and Scott (2003). Instead, teachers typically ask questions (with a known answer) to check their students' understanding and to make judgements about their attainment (Alexander, 1997).

The issues that arise because of the nature of questioning within a school environment can usefully framed and developed using the three conceptions of an interview identified by Kitwood (1977) as cited in Cohen et al. (2018). The first conception is as a means of '*pure information transfer*' (Cohen et al., 2018, p. 507). This conception aligns with the condition of mutual enquiry, where an interviewee and interviewer freely share their thinking to develop descriptions and explanation of the demonstration, which is required to collect valid data. However, Kitwood (1977) also argues an interview can also be conceived as a transaction in which the interviewer and interviewee will behave in ways influenced by motivational factors. In these interviews, the interviewer is also a teacher who has responsibility, or is seen as part of the system, for deciding high-stakes attainment and performance indicators (such as UCAS predicated grades and effort grades). Hence, there is a motivational factor for students to want to appear to be knowledgeable and correct. Put more simply, the participants may want to 'look good' in front of the teacher who is responsible for assessing their attainment. This is especially true for the physicists, who were also taught by the researcher. Hence, these students may be reluctant to volunteer answers unless they are sure they are correct. This would negate the conditions required for mutual enquiry.

A closely related issue can be understood in terms of the third conception of an interview as an encounter that shares many aspects of everyday life. In their everyday experience with a school context, students experience questioning as an exercise where the teacher knows the answer to their question and is checking students' understanding. Students recognise this and play the game of giving the 'correct' answer the teacher wants to hear. Hence, the participant could conceptualise the questioning in an interview as the 'assessment game' that is played in lessons, rather than as a

genuinely exploratory, knowledge seeking process. Again, this would invalidate the conditions required for mutual enquiry.

Thus, the usual teacher/student relationship has the potential to invalidate the condition of mutual enquiry required to gather valid data. Therefore, several measures were taken to facilitate a shift in the relationship from student/teacher to participant/interviewer-researcher to enable participants to share their ideas more freely. These measures are outlined below. They were informed and developed using pilot study results.

5.1.2.2.1 Explaining the Purpose of the Interviews to the Participants

During the pilot study, the 'mutual enquiry' aim of the interview was briefly explained to each participant before they were shown any demonstrations. Specifically, each participant was told that there 'are no wrong answers, it is about what you think' and reminded that this interview would not contribute to any form of school assessment. This approach was largely a success. There are several instances in participants 1, 3 and 4's interviews when they change their mind about their explanations. This would seem to indicate that they are happy to explore ideas freely without fear of giving only correct answers. However, this was not the case with participant 2. This participant regularly asks, 'Is that right?', during the interview, which suggested an anxiety and a need to be correct remained.

In light of this, changes were made in the main study. First, the explanation of mutual enquiry was kept. However, a more explicit explanation of the purpose of questioning was given at the start of the interview. Specifically, participants were told that their answers may be reflected back to them. It was stressed that was an opportunity for them to hear their responses and see if they wanted to add anything. It was not an indication that they might be wrong. Second, the teacher/student assessment game of a teacher asking a question and expecting a specific answer was referred to. It was emphasised that this was not the aim of the interview. Rather, the goal of the interview was to freely explore ideas and that there was not an answer the interviewer wanted to arrive at. Third, during an interview, if a participant was struggling to answer a question, the interviewer intervened to remind them that it was perfectly acceptable to say, 'I don't know.'

5.1.2.2.2 Soft Measures to Indicate a Change of Relationship

Several 'soft' measures were taken to subtly delineate between a teacher/student relationship and participant/interviewer relationship. These included allowing participants to schedule their own interview time and the interviewer/teacher making the participants hot drinks and providing

biscuits for the interviews. These measures were used to develop a more relaxed atmosphere which contrasted with the usual classroom environment.

These soft measures and providing clear and frank explanation of the aims of the interviews were intended to develop mutual trust to facilitate free exchange of ideas, leading to mutual enquiry. There were several indications that this condition of mutual enquiry had been achieved in the interviews. For instance, several participants were willing to change their mind (see first extract below), to say 'I don't know' (see second extract) or ask clarifying questions (see third extract).

> *Again, I see if it's frictionless, if there's no air resistance, no friction, then that force isn't going to decrease. So it's the same, there's going to be the same resultant for the entire time, which I guess would mean it would keep going. So that can't make sense (12P04).*
>
> *Well, there's probably more air pressure when it's heavier. I don't know. I'm not really that sure (12S09).*
>
> *No, the acceleration? Or do you mean the deceleration force? (12P05).*

5.1.2.3 Questionable invariance of knowledge

Invariance of knowledge refers to whether an interview reveals thinking that a participant will apply in a wide range or circumstances rather than just the specific example of the interview. In this case, this would mean that each participant's hub-words would be used in a similar way in a variety of mechanics explanations rather than just the explanations of the motion of the cup-cake cases. As this pilot study only used the cup-cake case task, it was impossible to test this invariance issue. This flaw in the design of the interviews became apparent in the analysis of the pilot study interview data. As a result, further demonstrations were developed in the main study to test this invariance of knowledge by giving participants the opportunity to apply similar reasoning in a variety of scenarios. The development of these demonstrations is described in the next section (section 5.1.3.1)

5.1.3 Materials

5.1.3.1 Demonstrations

This section describes the selection and design of the demonstrations participants asked to describe and explain during their interviews. The use of demonstrations to capture participants' use of language when they reason about an object's motion was trialled during the pilot study. The initial design and rationale for using a particular type of demonstration is described in the next section. The development of additional demonstrations for the main study is then outlined.

5.1.3.1.1 Initial Design of a Demonstration in the Pilot Study

During an interview, each participant was shown a demonstration of three different types of cup-cake case falling from 1m to a desk. The different types of cup-cake case are as follows:

- Cup-cake case with no alterations.
- Cup-cake case with four paperclips added.
- Screwed up cup-cake case with no added paperclips.

In each demonstration, each cup-cake case initially accelerated towards the ground and then reached a maximum speed or terminal velocity. This terminal velocity depends on the mass of the cup-cake case, which affects its weight, and the surface area, which affects the amount of air resistance or drag acting on the cup-cake case at any given speed. Hence, each cup-cake cases reached different terminal velocities and, therefore, reached the desk in different times.

Each participant was initially shown the unaltered cup-cake case falling and asked to describe and then explain what they had just witnessed. Then, they were shown the unaltered cup-cake case and the cup-cake case with paperclips falling together, and asked if they noticed any differences and, if so, how could they explain those differences. This step was then repeated with all three cup-cake cases falling.

The cup-cake demonstration was chosen for several reasons. First, the equipment (cup-cake cases, one metre rule with 1cm division marked, paperclips) used in the demonstration was reasonably familiar to the participants. This minimised the risk of the participants not understanding how the equipment worked (e.g. thinking that an air track blows a slider along horizontally) or focusing on novel aspects of the equipment. This familiarity would increase the ecological validity of the interviews.

Second, there were several variables, which vary in their abstraction, that participants with different expertise might identify as salient to describing or explaining what they have witnessed. For instance, they might simply note that one cup-cake case has been screwed up. Or more abstractly, they might note that the screwed-up cup-cake case has less surface area and, hence, less air resistance at the same speed. diSessa (1993) identifies an increasing level of abstraction as an important feature of developing conceptual knowledge.

Third, the motion of the cup-cake cases can be explained with what might be considered both novice and expert explanations, even if they identify the same variable that could affect the motion. For instance, a participant might explain the cup-cake case with the paperclips falling faster by referring to the extra weight. This would reflect a common misconception that heavier objects fall faster. However, the expert and correct explanation would require a participant to recognise that the cup-cake case with the paperclips accelerates for longer and, therefore, has a higher maximum speed because the air resistance or drag must be larger to balance the extra weight. In contrast, novices might find it very difficult to explain physical phenomena that are completely unfamiliar. For instance, it would be unreasonable to expect a psychology student, with only every day and GCSE knowledge, to describe what happens when an electron and a positron annihilate. This would limit the opportunity for novices to articulate their thinking and, hence, limit the efficacy of the interview to capture novice use of language.

The cupcake case demonstration was designed to provide participants with options for selecting how they choose to describe and explain the motion of the cup-cake case. Hence, it was hoped that the demonstration would act as stimulus for capturing examples of participants' language use that revealed how participants connect hub-words to other ideas. These connections are likely to define a scientific concept, as a has previously been discussed (see section 2.3.3). Hence, this approach was adopted because it could reveal the nature of any scientific concept associated with a hub-word. The results of the pilot study indicated this goal could be achieved with this demonstration. For instance, in the first extract below, the participant describes the cup-cake case as accelerating as it falls and identifies the weight as the cause of this acceleration. In the second extract, a different participant describes the cup-cake case as falling slowly due to the large surface area and then refers to the cup-cake case falling faster because the weight is 'forcing' the air particles away (third extract). Both participants see the weight as a cause of the cup-cake case's motion. However, they associate weight with causing very different processes (i.e., acceleration versus forcing particles away leading to faster movement).

The cup-cake has mass. Therefore, it is attracted by gravity. So it has weight. The weight then causes acceleration because F=ma (Participant 1).

Urrr, it has a large surface area means that the air particles are more likely to hit it – ummm – resulting in it falling slower (Participant 2).

[the cup-cake case falls faster] because they're being forced away from the bottom a lot faster than if there were no paperclips because of the added weight (Participant 2)).

However, these extracts also demonstrate that both participants were drawn to the weight of the cup-cake case as a cause of the motion, perhaps, because of the obvious addition of the four paperclips. It was not possible to investigate whether weight, and other causes of motion identified by participants, were used in different contexts because the pilot interviews only used one demonstration. Hence, other demonstrations were developed in the main study to test participants' invariance of knowledge.

5.1.3.1.2 Development of Demonstration in Main Study

Several other demonstrations were used in the main study to test whether participants applied similar reasoning in different mechanical scenarios and to ensure a range of mechanical knowledge tested. The broad criteria used in the pilot study were used to design the new demonstrations. These criteria were:

1. Familiar equipment.
2. The variables used to describe and explain the motion of objects varied from concrete to abstract.
3. Familiar context or at least accessible to novice.

In addition to the falling cup-cake case used in the pilot study, the two additional demonstrations designed using these criteria are described below.

5.1.3.1.3 Film Canister Rocket

This demonstration involves using a 35mm film canister as a rocket. The canister is filled with a small amount of water and an effervescent vitamin tablet. The lid of the canister is then replaced. The tablet reacts with the water and produces a gas. The pressure from build-up of gas in the canister eventually 'pops' the lid. At this point, the gas escapes from the canister and the canister acts as a rocket, rising into the air, reaching a maximum point, and then descending. This is a standard demonstration to GCSE students at the target school when they study mechanics and is

also discussed as part of the A-level programme. Hence, the demonstration and the equipment used in the demonstration were familiar to the psychology and physics students. Moreover, this demonstration had been previously used by the researcher in the target school to investigate students' explanations of motion (Taylor, 2012; Taylor & Hillier, 2018). This study demonstrated that students focused on different concrete to abstract aspects of the demonstration to describe explain the motion of the film canister (e.g., the 'fizz from the gas' making the canister 'go up' versus a 'resultant force' making the canister 'accelerate'). Furthermore, in each interview, the participants were shown the canister launching 'canister body' first and subsequently 'lid' first. This gave the participant the opportunity to identify a range of variables, again from concrete to abstract (e.g., size of the lid versus mass of the lid) that might explain any differences in the motion.

5.1.3.1.4 Dynamics Trolley Rolling Down an Inclined Plane and then along a Flat Surface.

During this demonstration, a dynamics trolley was rolled down an inclined plane and then onto a flat surface. The trolley slowly came to a halt on the flat surface. The distance the trolley travelled on the flat surface was varied by adding masses to the trolley or changing the surface of the flat section (e.g., wood versus carpet). As with the other demonstrations, dynamics trolleys are regularly shown to students during their GCSE and A-level studies to illustrate laws of mechanics (e.g., Newton's 2nd Law). Furthermore, they are required to carry out practical work with this equipment as part of their GCSE (AQA, 2015a, 2015b). Moreover, there are a range of explanations that could be given for why the trolley moves further on the flat surface that draw upon knowledge that varies in its abstraction (e.g., 'the carpet is rougher, so it slows down quicker' versus 'the resultant force on the trolley is greater because the frictional force is greater, so the acceleration of the trolley is greater.').

To check whether all the demonstrations covered a range of mechanics' knowledge and whether this knowledge is drawn upon in different contexts, an 'ideal' description and explanation of the motion of the objects in each were written in collaboration with other physics teachers in the target school and using the course textbook (Breithaupt, 2015) (see Table 5.2 and Appendix 10.6). The conceptual dimensions and sub-dimensions defined in the FCI were used as a framework to identify the nature of the knowledge being used in each explanation and to ensure coverage across demonstrations (also see Appendix 10.7).

Table 5.2 The explanation of a film canister rocket launch agreed by Physics teacher and the FCI inventory cross-references in this explanation.

Agreed Explanation of Film Canister Launch	FCI Cross-Reference
When the film canister 'pops', gas escapes in a downward direction and exerts an equal and opposite force on the canister.	Newton's Third Law Impulsive forces
This force is greater than the weight of the canister.	Superposition Principle Vector sum and cancelling forces.
Therefore, there is a resultant force upwards which causes the canister to accelerate upwards. Hence, the canister is given a large velocity in this direction.	Newton's Second Law. Constant force implies constant acceleration.

This analysis demonstrated that the 'ideal' description and explanations of the motion of the falling cup-cake case, film canister rocket, and dynamics trolley in the three demonstrations draw upon a wide range of mechanical knowledge. Moreover, this knowledge is applied in a different context in each demonstration. Hence, the invariance of participants' mechanical knowledge can be tested.

5.1.3.2 Interview Schedule and Questioning Prompts

Semi-structured interviews were used to balance the ability to freely explore each participant's use of hub-words, through their descriptions and explanations of what they see in the demonstrations, with the need to compare the use of hub-word between participants. This balance was achieved by presenting the demonstration to each participant in the same order and using a scripted introduction to each phase. Moreover, participants were always asked to describe what had happen immediately after they had seen a demonstration and were then asked to explain what they had just described. However, within this structure, questions were asked to encourage participants to develop, clarify or sustain aspects of their description and explanations.

This free questioning of participants allowed for a more thorough exploration of participants' use of hub-words and how they connect these words with other ideas. It was also consistent with the notion of mutual enquiry defined by diSessa (2007). However, care was required when using questioning not to provide different levels of scaffolding for different participants and, therefore, making any comparison between participants' use of language invalid. Hence, some structure was used to guide the further questioning of participants to elicit more detail in their responses. This structure was developed using the pilot study responses and the researcher's classroom experience. It broadly involved echoing participants' initial response back to them and progressively asking for more detail, as set out in Table 5.3.

Table 5.3 Types of questioning prompts (shown in bold) used during interviews and examples of dialogue associated with these prompts.

Questioning Type		Interview Extract
1	Participants are generally asked for more detail	**RT:** Okay, so what we've got here, is one metre rule and I've got markings on it. And what we've got here is a muffin case or cupcake case. I always get the words mixed up. So we put it there. And what we do, is we drop it. **12S05:** It's going to float down **RT:** So, you've just described what happens. It floats down. **Can you unpick that a bit more for me?** **12S05:** So, because it's, kind of, a very light thing. It's going to, kind of, there's not much air resistance, I guess. Because there's, umm. So obviously heavier things fall more quickly because there's more weight acting down. But with, it's so light that it's going to go at a slower rate than something that is heavier.
2	Participants are asked about a specific aspect of the demonstration.	**RT:** Right. So, can you just describe what happened there? **12P04:** OK. So, when you put the tablet into the water, the tablet started to dissolve and the gas was produced, and when you then put the cap on the end, you seal all the air within the capsule or whatever you want to call it, the film case, in the film case. So as the tablet produced more and more gas, the air pressure within the capsule thing increased until the, the pressure force pushing out on the container was greater than the force of the cap holding it in, at which point all the gas escapes. The pressure was released, and the force was great enough to push up into the air. **RT:** OK, right. So, we've got the idea of the gas escaping, then the force was enough to push it into the air. I want to come back. I just want to note that down. So, the gas escaped. Force enough into air. OK, well, I want to do is I want to pick a little bit about what happens before and after this pop. So, this is the moment of the pop. And this is before. **Can you just tell me about the way it's moving before the pop?** **12P04:** So, before the pop, it's stationary.
3	Participants are asked about specific terminology they might have already used in their description or explanation.	**12P02**: So, it's got this sort of quite large force acting in all directions. And then because the cap isn't held on very securely, it's forced away. So, you've still got this force acting on all points along here, but you don't have a force acting on the cap. So, the resultant force of all of these is going to be an upwards motion. **RT**: OK, so there's a resultant force going up. **12P02**: Yeah. **RT**: Yeah. **What do what do you mean by resultant force?** **12P02:** So, the resulting force is the sum of all the individual forces on the object. So, they replace all the forces just with one single force and have the same motion.
4	Participants are asked about specific terms (acceleration, speed, velocity), which they have not already mentioned.	**RT:** So, I'm absolutely clear, how do you know it slows down? **12S03**: Because it becomes stationary. **RT:** Ok, umm, **does it accelerate when it hits the desk?** **12S03:** No **RT:** No. Why not? **12S03:** Because it's no longer moving, and something can only accelerate if it's moving.

Note that participants are only asked about specific terms at the end of the question prompt progression outlined in Table 5.3. However, all participants were asked about their understanding of acceleration at the end of the interview if their understanding of this term had not already been discussed. This question was in response to the findings of Taylor (2012) that suggested that student often define acceleration with 'getting faster' rather than the rate of change of velocity.

The full interview schedule is given in Appendix 10.8.

Participants were free to take part in interviews during any of their free lessons. A lesson in the target school is fifty minutes. The pilot study demonstrated that the cup-cake demonstration discussion lasted around twenty minutes. Therefore, it was assumed that each demonstration in the main study would last a similar length. Hence, a maximum of two demonstrations could be fitted within a lesson. Therefore, it was intended to use two interviews. The first interview included the film canister rocket and cup-cake case demonstration. The second interview contained the dynamics trolley demonstration, which required more setting up than the film canister rocket and cup-cake case demonstrations, and hence logistically required a separate interview.

The first interviews took place between the 4th March 2020 and 16th March 2020. The second interviews were scheduled for week beginning 23rd March 2020. However, on the 20th March 2020 the target school was closed as part of the UK government's response to COVID-19. Hence, the second interview, which included the dynamics trolley demonstration, was not completed for any participants.

The loss of the second interviews for participants was unfortunate. However, the first two demonstrations still gave participants at least two opportunities to use similar aspects of their knowledge of mechanics in different contexts (see Appendix 10.6 and Appendix 10.7). Furthermore, the falling cup-cake and film canister rocket demonstrations both encouraged participants to use a wide range of their knowledge of mechanics (again, see Appendix 10.6 and Appendix 10.7). Weighed against these two arguments, it was not considered ethically appropriate to approach participants to ask if they would be willing to participate remotely. Especially given the uncertainty at the time around how the grades of participants' A-level qualifications would be awarded, with the cancellation of formal examinations, and the probable involvement of the researcher in any alternative procedure, which could have caused participants' stress.

5.1.4 Interview Transcription

The interviews were transcribed to enable the analysis of participants' use of hub-words. The pilot interviews were transcribed by the researcher. However, this was time consuming and by the time the main study interviews had been completed, voice recognition software for transcription was readily available and considered reliable (Lumivero, 2023). Hence, NVivo transcription was used to generate the initial interview transcripts. These were then manually checked by the researcher. Using NVivo transcription also allowed the interview audio to be aligned with the text transcription. Hence, the audio could be played when using NVivo to code text transcription.

This reduced the chance that subtle nuances in the data indicated by, for instance, voice inflections would be missed (Cohen et al., 2018).

5.2 Data Analysis

The text of the interview transcripts was analysed by combining two analytical techniques. First, a content analysis was used to identify instances when a participant had used each of the hub-words (*force*, *energy*, *mass*, *time*, *particle*). This first phase of the analysis addressed the 'do individuals use hub-words' part of the research question. Second, a lexicogrammatical (LG) analysis using systemic functional linguistics (SFL) was then used to categorise how participants use these hub-words in their explanations and descriptions. Hence, this analysis addressed the 'how do individuals use hub-words' part of the research question. The LG analysis enabled a further phase of content analysis that systematically compared how hub-words are used by participants in the physics and psychology cohort. A rationale for using this mixed-method analysis is given in section 5.2.1 below.

5.2.1 Rationale for Content and Lexicogrammatical Analysis

Seah et al. (2011) highlight the range of analytical techniques used by studies interested in students' and teachers' use of scientific language. For instance, several studies that have examined the content, structure and patterns of interaction in classroom talk have used thematic analysis (Lemke, 1990), discourse analysis (Chin, 2006; Mortimer & Scott, 2003), critical discourse analysis (e.g. Hanrahan, 2006) and conversation analysis (Tan, 2009). Seah et al. (2011) note that there is less variety in the techniques used to analyse students' written language, which they attribute to the nature of writing as a monologue, rather than a dialogic interaction between two or several participants. Hence, with analysis of written language, there is no requirement to understand any pattern of interaction between participants. This research is not concerned with the patterns of interaction between the interviewer and the participant, but rather the extracts of monologue when the participant explains or describes what is happening in the demonstration. Hence, the types of analysis used for written techniques are more appropriate in this case.

Seah et al. (2011) argue that the techniques used by studies to examine written scientific language can be placed on a continuum from macro to micro-level analysis. For instance, at a macro-level, content analysis has been used by Grimberg and Hand (2009) to identify types of reasoning (e.g. analogy, deduction, examples) used by students. At a micro-level, Christie and Derewianka (2008) used LG analysis to chart the development of students' writing from early childhood to late

adolescence in a range of curriculum areas. Here, LG refers to both 'the vocabulary (both technical and non-technical) and the grammatical resources' (Seah et al., 2011, p. 854) individual use to create and communicate meaning.

Using LG analysis, Christie and Derewianka (2008) identified particular genres of scientific writing that students were expected to use at different ages and the vocabulary and grammar that characterised these genres. However, several studies have used both macro and micro-level of analyses to provide complimentary perspectives, which add richness and depth to findings. For instance, Keys (1999) used content analysis to identify different types of information (e.g. information, procedures, inferences) presented in students' report writing. Keys (1999) then used LG analyse to identify similarities and differences between students' use and scientific use of language.

A similar methodology, was used in studies by Seah, Clarke, and Hart (2014) and Seah (2015). Seah et al. (2014) used content analysis to identify three aspects of an explanation of thermal expansion (macro, micro and causal references) that 12-13 year-old students were expected to learn. LG analysis was then used to identify differences between students' use of language and the expected scientific use of language in each aspect. Similarly, Seah (2015) used content analysis to identify different types of explanations (e.g. explanation that used purely observable causes versus explanations that used inferred causes) used by younger students (age 10) to explain why seeds covered with oil and water did not germinate. Then, LG analysis was used to examine patterns in how language was used in each type of explanation.

Answering research questions 3 requires instances when participants are using hub-words to be identified. Then these instances then need to be systematically categorised to allow for a robust comparison between their use by participants in the physics and psychology cohorts. The studies described above demonstrate the efficacy of using a mixed-level content and LG analysis to achieve this goal. Hence, this analytical approach was used in this study. The next section describes the LG analysis of hub-word use in more detail.

5.2.2 Lexicogrammatical Analysis of Hub-Word Use

The LG analysis was carried out using Systemic Functional Linguistics (SFL), which has been used in several studies that were discussed in the previous section (e.g., Seah, 2015; Seah et al., 2014). SFL, as the name suggests, specifies the systems of options that individuals have within language which allow language to perform certain functions (Halliday & Matthiessen, 2004). In SFL there are three functions known as meta-functions (Halliday & Matthiessen, 2004). The ideational meta-function allows individuals to build and maintain a theory of existence – to describe and account for what we see and experience. Language is also a mechanism that allow speakers to enact personal and social relationships. This function of language is known as the interpersonal meta-function. Finally, the textual meta-function manages the interaction between the ideational and interpersonal meta-functions.

In characterising a language as a semiotic tool that allows individuals to perform certain functions (Halliday & Martin, 1993) by following syntactic rules, SFL is theoretically underpinned by the same principles as the Representational Theory of Mind (RTM). In this theory, reasoning is identified as a computational process on sentence-like structures in a mental 'language of thought' (A. Clark, 2014), which is sometimes called mentalese (Pinker, 2009). In mentalese the basic units of thought are concepts. In contrast, words (or signs as Carey (2009) called them) are the basic units that are put together grammatically to generate descriptions and explanations in spoken and written language. However, words or signs in language are assumed to be equivalent to concepts in mentalese (Carey, 2009; A. Clark, 2014). Moreover, word and signs are associated with a corresponding concept (Carey, 2009). Hence, this shared theoretical background and association between word and concepts makes SFL suitable for use in this research because it specifies the systems of grammatical and lexical choices individuals have to perform certain functions in language, which reflect their reasoning in mentalese with concepts. As Carey (2009) demonstrates, these lexical choices can reveal the nature of concepts in intuitive theories and when concept in different theories are incommensurate.

When studying scientific language, the ideational meta-function is of particular interest because it is this meta-function that enables individuals to build their description and explanations of the physical world. Indeed, Halliday (2004) use SFL and the ideational meta-function to characterise the social language of science. Moreover, SFL and the ideational meta-function is frequently used by in the studies listed above, which examine similarities and differences between students' use and the scientific use of language. A brief outline of SFL in terms of the ideational meta-function is given in section 5.2.2.1 and the how SFL is used LG analysis is described in section 5.2.2.2.

5.2.2.1 A Brief Outline of Systemic Functional Linguistic Analysis

In SFL, the ideational meta-function is split into two components: the experiential and logical functions (Halliday & Matthiessen, 2004). The experiential and logical functions explore how individuals use language to describe and explain their experiences by assuming that these experiences can be represented as a flow of events, which can then be accounted for. For instance, a student might identify the cup-cake case falling as a flow of events. This flow of events is '*chunked into a series of figures'* (Halliday & Matthiessen, 2004, p. 170), with each figure defining a discrete change or 'quantum of change' (Halliday & Matthiessen, 2004, p. 170). For instance, in the cup-cake case demonstration, the case initially being held in a stationary position might be seen as one figure and then the case falling through the air when it is dropped as another subsequent figure. Hence, linking these two events creates a flow of events that describes the motion of the cup-cake case. Furthermore, a participant might explain these figures in terms of gravity pulling the cup-cake case down or the weight of the cup-cake case minus any drag force leading to a resultant force that accelerates the case. The experiential and logical functions provide a framework to systematically analyse similarities and differences in the events participants use to describe the motion of the film-canister rocket and how they explain these events.

The experiential function operates at a clause level and uses a transitive grammatical model to defines individual figures. In the transitive model, a figure (or 'quantum of change') is realised through a process unfolding through time and the participants that are involved in that process. Processes are always associated with verbal groups and participants typically with nominal groups. The process and the participants form the 'experiential centre' (Halliday & Matthiessen, 2004, p. 176). However, circumstantial elements can augment this experiential centre in several ways. For instance, a circumstantial element can add, for instance, temporal, spatial, and causal information. By adding this information, circumstantial elements either expand or project upon the experiential centre. In an expansion, a secondary clause adds information to the experiential centre through elaboration, extension, or enhancement. In contrast, a projection ties the experiential centre to speech (locution) or thought (idea). These circumstantial functions are summarised in Table 5.4.

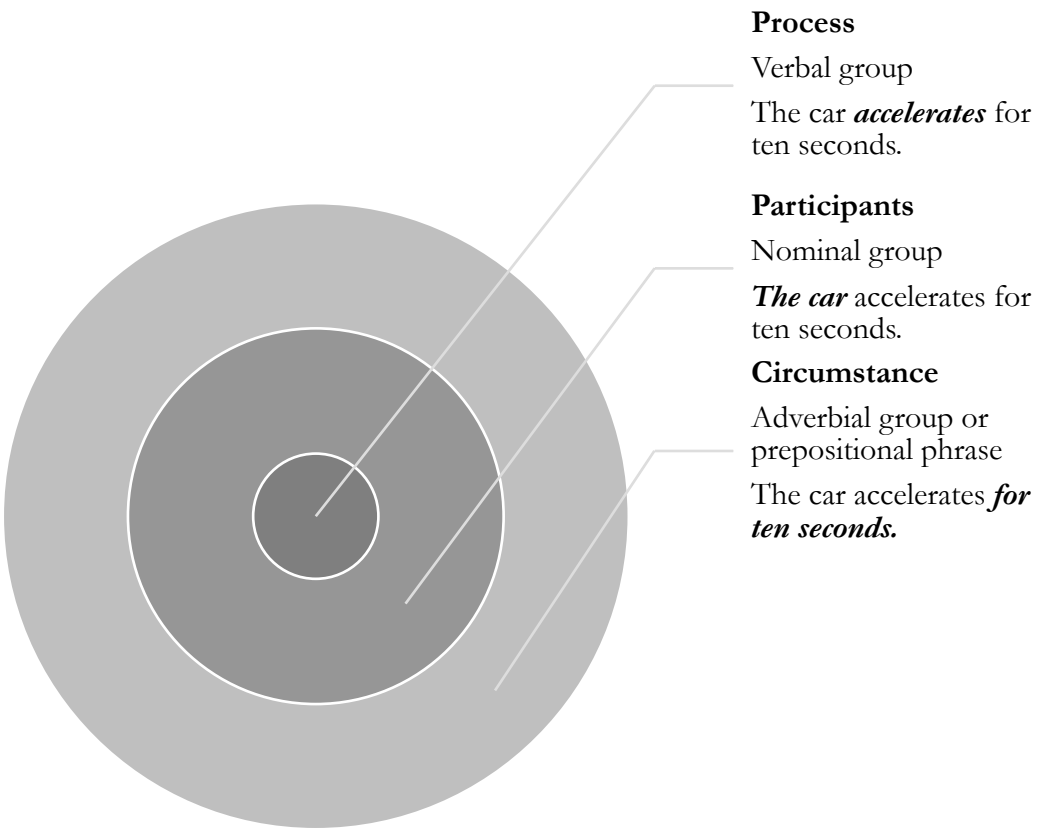


Figure 5.1 Central and peripheral elements in the experiential structure of the clause (adapted from figure 5.4, Halliday and Matthiessen (2004, p. 176)).

Table 5.4 The transitive grammatical model for characterising a figure within a clause adapted from Halliday and Matthiessen (2004, p. 291). The ergative grammatical model equivalent is also shown.

	Typical Preposition	**Ergative Model**		**Transitive Model**						
				Material	**Behavioural**	**Mental**	**Verbal**	**Relational: Attributive**	**Relational: Identifying**	**Existential**
Process		1. Process								
Participants		2. Medium		Actor [mid.]; Actor or goal [eff.]	Behaver	Senser	Sayer [mid.]; Target [eff.]	Carrier	Token	Existent
	By	3. Agent		Initiator or Actor [eff.]	-	Phenomenon ['please']	Sayer [eff.]	Attributor	Assigner	-
	To, for	4. Beneficiary		Recipient; Client	-	-	Receiver	(Beneficiary)		-
	At, on, etc…	5. Range		Scope	Behaviour	Phenomenon ['like']	Verbiage	Attribute	Value	-
Circumstances	For, over, across, etc…	Enhancing	6. Extent	Duration, frequency (temporal), distance (spatial)				How long, how far, how often?		
	At, in, on, from, etc…		7. Location	Time (temporal), place (spatial)				When? Where?		
	With, by, like		8. Manner	Means, quality, comparison, degree				How? With what? In what why? Like what? To what extent?		
	Through, for, etc…		9. Cause	Reason, purpose, behalf				Why? What for? Who for?		
	In case of, etc…		10. Contingency	Condition, concession, default				Under what conditions?		
	With, besides, etc…	Extending	11. Accompaniment	Commutation, addition.				Who/what with? Who/what else?		
	As, into, etc…	Elaborating	12. Role	Guise, product				What as? What into?		
	About, etc…	Projection	13. Matter					What about?		
	According to, to, etc…		14. Angle	Source, viewpoint				Who says? Who thinks?		

The type of process determines the type of experience described by the clause. For instance, '*She ate an apple*' defines an experience that takes place in the 'outer' world. In contrast, '*she likes apples*' defines an 'inner' experience because these experiences these take place in our heads. Halliday and Matthiessen (2004) call the latter type of clause a material clause and the former type of clause a mental clause. A third type of clause, relational clauses, link experiences. For example, '*the apple is on that tree'* connects the apple with its location. Finally, Halliday and Matthiessen (2004) identify behavioural, verbal, and existential clauses. Halliday and Matthiessen (2004) argue all types of clause are needed for individuals to build a coherent theory of their experience. However, this research is primarily concerned with material clauses because they describe figures or changes that take place in the physical world, and relational clauses, which help to characterise and identify figures.

Table 5.4 lists the participants associated with each type of process in the transitive model. However, this table also lists the equivalent participants and circumstantial elements in an ergative grammatical model. The ergative model is important in this analysis because an ergative grammatical pattern in more appropriate than the transitive model for understanding how modern English scientific registers of text construct meaning (Halliday, 2004; Halliday & Martin, 1993). To understand the difference between an ergative and transitive grammatical patterning, it is useful to compare an ergative and transitive analysis of a possible explanation of film canister rocket's motion given by a 15 year old student, taken from (Taylor, 2012).

The canister pops and climbs into the air. Then gravity pulls the canister down.

In the first clause, the canister is the actor but in the second clause, gravity is the actor, and the canister is the goal. Note that the canister does not have the same role across both clauses in the transitive model. Halliday and Matthiessen (2004) argue that this feature of the transitive system has led to many registers of modern language, including Scientific English, moving away from a transitive grammatical model, and adopting an ergative model. This change reflects the ergative model's ability to consistently identify a participant that is central to unfolding of events described by a series of processes, which allows the cause-and-effect aspect of a process to be emphasised. For instance, in the above example, the canister can be consistently identified as a medium – the participant through which the process is actualized (Halliday & Matthiessen, 2004, p. 284). Hence, the ergative model is better suited to describe a flow of events happening to one object that are caused by many different external agents. Therefore, an ergative model was used to analyse the clauses in the written explanations. The ergative model is summarised in Table 5.4 and the equivalent transitive interpretation is also shown.

At a clause level, the experiential function characterises each figure, which describes the 'quantum of change' that defines an event. However, to fully describe our experience of the world, figures are also are strung together or expanded to form clause complexes to describe a flow of events and give account of this flow (Halliday & Matthiessen, 2004). The logical function in SFL characterises how these clause complexes are formed. Figure 5.2 summarise the main elements of the logical function.

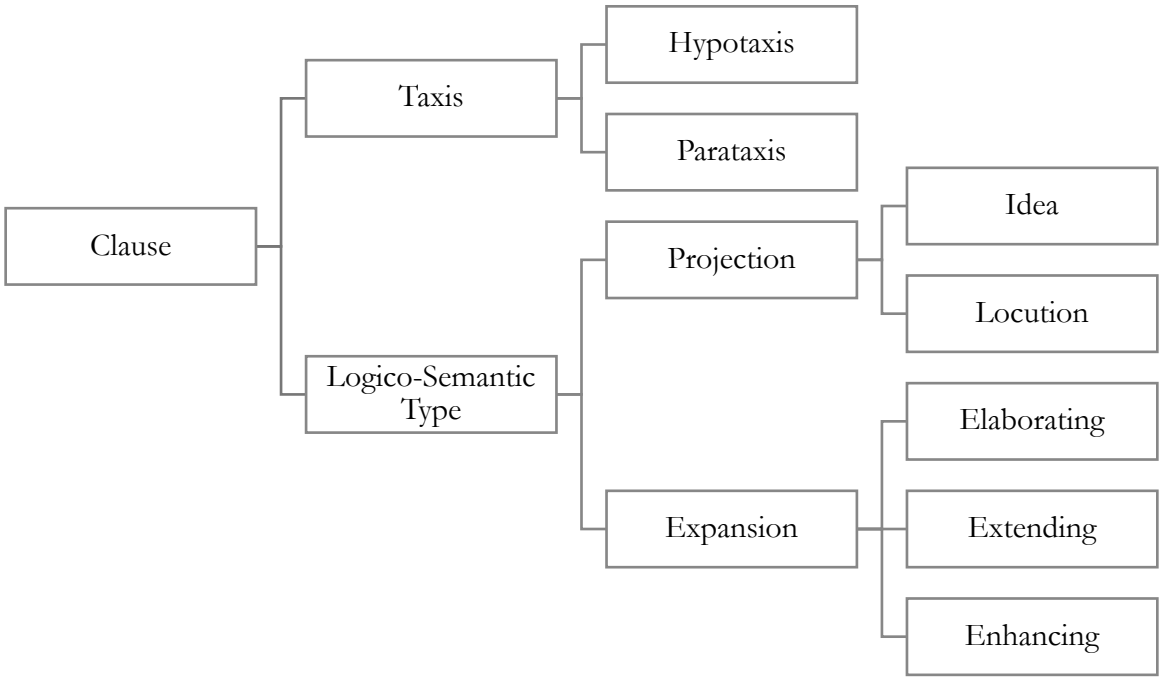


Figure 5.2 The system of clause complexing taken from Halliday and Matthiessen (2004, p. 373)

The taxis, or interdependency, defines the status of the clauses in relation to each other. The relationship between clauses is paratactic when the clauses have equal status, and hypotactic when one clause acts a main or dominant clause. Hence, the relationship between clauses is unequal. Halliday and Matthiessen (2004) refer to any pair or chain of clauses as a clause nexus.

The logico-semantic types define how each clause relates to each other. Halliday and Matthiessen (2004) identify a wide range of logico-semantic relationships, which can be grouped into a small number of general types based on two fundamental relationships: expansion and projection, which can then be sub-divided respectively into idea and locution, and elaborating, extending, and enhancing. The function of these relationships has already been identified when describing the function of circumstantial elements in the experiential function (see Table 5.4). This is because a clause nexus can perform the same function a circumstantial element in a single clause (Halliday & Matthiessen, 2004). For example, in the explanation*, the canister pops because of the pressure build up*,

the pressure build up is a causal circumstantial element that explains why the canister pops. However, the same explanation is given by the clause nexus, *the pressure builds up and makes the canister pop*. The nature of any causal relationship between clauses is of particular interest in this analysis, which focuses on how participants explain a flow of events in a scientific demonstration. Hence, Table 5.5 summarises the enhancement logico-semantic types, which include causal types, in more detail. The next section describes how the experiential and logical functions are used to categorise incidences of participants' use of hub-words.

Table 5.5 The enhancement logico-sematic types taken from Halliday and Matthiessen (2004, p. 291)

Enhancement	**Category**	**Meaning**	**Paratactic**	**Hypotactic**		
				Finite	**Non-finite: conjunction**	**Non-finite: preposition**
Temporal	Same time	A meanwhile B	(and) meanwhile; (when)	[extent] as; while	While	In (the course of/process of)
				[point] when, as soon as, the moment	When	On
				[spread] whenever, every time	-	-
	Different time: later	A subsequently B	(and) then, and + afterwards	After, since	Since	After
	Different time: earlier	A previously B	And/but + before that/first	Before, until/till	Until	Before
Spatial	Same place	C there is D	And there	[extent] as far as		-
				[point] where	-	-
				[spread] wherever, everywhere	-	-
Manner	Means	N is via/by means of M	And + in that way; (and) thus	-	-	By (means of)
	Comparison	N is like M	And + similarly; (and) so, thus	As, as if, like, the way	Like	

Causal Conditional	**Cause: reason**	**Because P, so Q**	**[cause^effect] (and) so; and + therefore**			
			[effect^cause] for; (because)	Because, as since, in case, seeing that, considering		With, through, by at, as a result, because of, in case of
	Cause: purpose	Because intention Q, so action P.	-	In order that, so that	-	(in order/so as) to, for (the sake of), with the aim of, for fear of
	Cause: result			So that	-	To
	Condition: positive	If P then Q.	(and) then; and + in that case	If, provided that, as long as	If	In the event of
	Condition: negative	If not P then Q	Or else; (or) otherwise	Unless	Unless	But for, without
	Condition: concessive	If P then contrary to expectation Q	[conscession^consequence] but; (and) yet, still; but + nevertheless [consequence^concession] (though)	Even if, even though, although	Even if, even though, although	Despite, in spite of, without

5.2.2.2 Using SFL to Characterise Hub-Word Use in Participant's Reasoning

The ergative model, at a clause level, and taxis and logico-semantic relationships between clauses, define the experiential and logical functions of language that individuals use to describe the flow of events and the reasons for these events. Hence, if the motion of each object in a demonstration is viewed as a flow of events, SFL provides a comprehensive and powerful analytical tool for characterizing and comparing how participants use hub-words to reason about these demonstrations. Figure 5.3 summarises how SFL was used in the analysis to define a coding decision tree, which was applied to categorise any instance of a participant using a hub-word. The coding decision tree was formed from Table 5.4 and Table 5.5, which also provide examples of language use that can guide any coding decision. Table 5.6 gives examples of coding decisions for the frequently used categories.

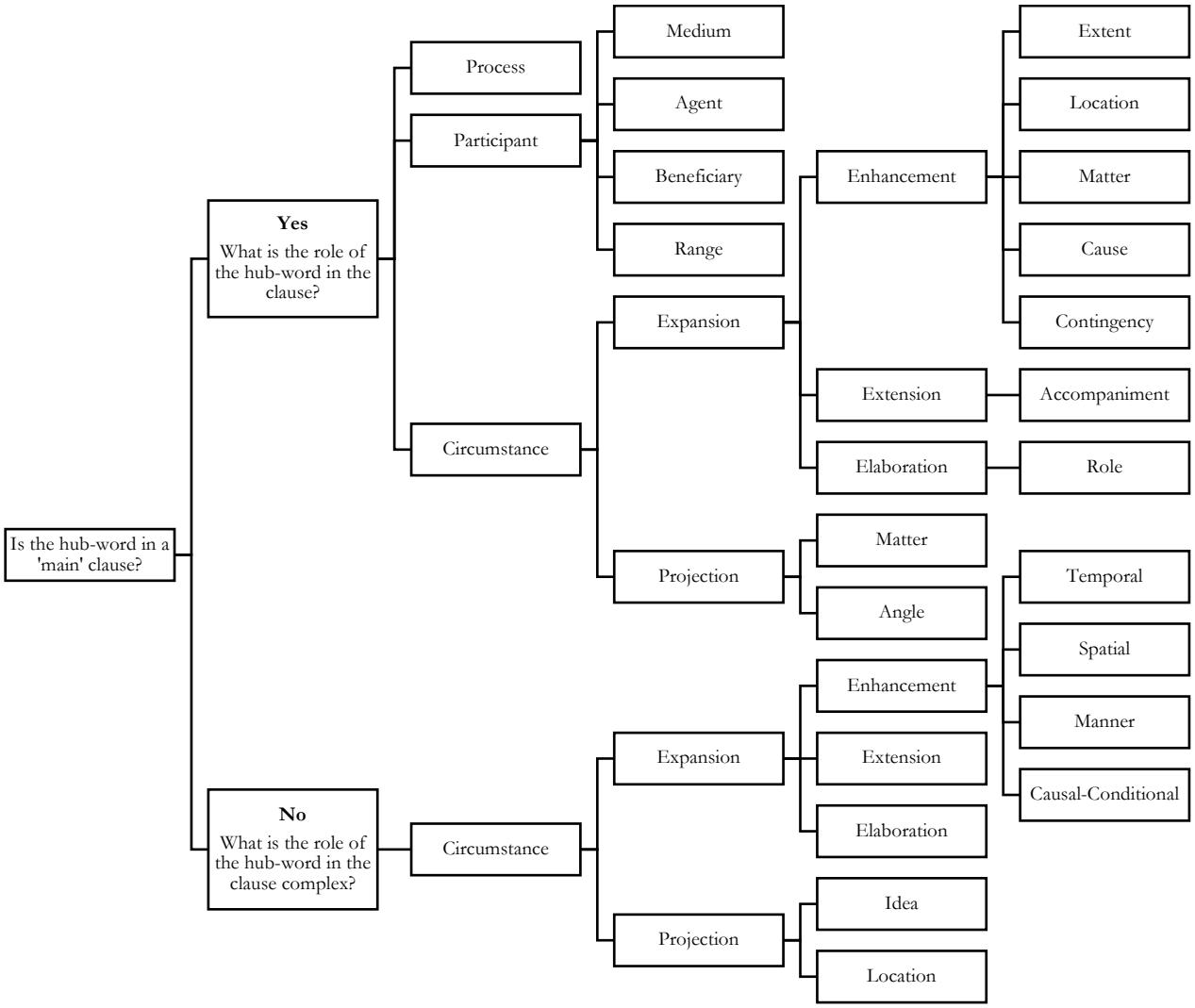


Figure 5.3 Coding decision tree for categorising function of hub-words.

Table 5.6 Examples of frequently used codes for categorising the function of hub words. The main clause is shown in bold, and the hub word is underlined.

Code	Interview Extract	Participant
Process Main Clause	**The pressure is <u>forcing</u> it upwards**	12P05
Medium Participant Main Clause	**The <u>force</u> going up wouldn't necessarily be from the pressure**	12P03
Range Participant Main Clause	**It is fo<u>rce</u> acting per kilogram.**	12P04
Agent Participant Main Clause	**The <u>force</u> from particles is pushing it upwards.**	12P08
Cause Circumstance Main Clause	**It is immediately being deaccelerated as soon as that reaction <u>force</u> is no longer there.**	12P03
Cause Circumstance Clause Complex	It's experiencing a <u>force</u> of mg, which will begin to **accelerate it.** At that point where it kind of runs out of the gas pressure, its kind of, comes back down because the gravity is stronger than the <u>force</u> pulling it up. So therefore, **it comes down.**	12P08 12S05

Figure 5.3 make clear that the first decision when coding an instance of the use of a hub-word is whether the hub-word occurs in what is termed the 'main clause' in any clause nexus. Taxis is central to the definition of the main clause in this analysis. The system of taxis relates to whether clauses in a clause nexus have equal status, as has been explained in section 5.2.2.1. The two clause complex interviews excerpts in Table 5.6 (last row) illustrate how in any explanation of the motion of an object, the clause that describes the motion of the object is likely to exist in a hypotactic relationship to other clauses in the nexus. This is because, in sense, the other clauses only exist to explain this motion. Hence, in this analysis, the clause that describes the motion of an object in a demonstration is defined as the main clause.

This definition of a main clause is a strong interpretation of the relationship between clauses in a clause nexus, considering that cause-and-effect paratactic logico-semantic relationships are allowed in SFL (see Table 5.5). However, defining the clause that describes the motion of an object of as main clause is also a useful technique to prevent the analysis of participants' explanation from becoming too fragmented, and therefore less informative. For instance, in the excerpt from Table 5.6, *'It's experiencing a force of mg, which will begin to accelerate it.'*, the hub-word *force* functions as a

participant (range) in the first clause if the first clause is considered in isolation. However, this does not capture why the participant considered it important to state that the object experiences a force of mg. In this case, it is because this force causes the acceleration of the object. Hence, the more informative function of the hub-word (cause rather than range) has been emphasised by defining the motion of the object as the main clause.

5.3 Summary

This chapter has described and justified the methodology used in this study to investigate whether participants use hub-words when they explain and object's motion, and if they do, how these hub-word are used (research question 3). The results of this this analysis are presented in the next chapter.

6 Results of the Interview Analysis to Investigate the Function of Hub-Words

This chapter presents the result of the content and lexiogrammatical (LG) analysis of the interview transcripts. The initial content analysis identified incidences of any use of hub-words in participants' interview transcripts. This analysis addresses the 'do individuals use hub-words' part of the research questions and is presented in section 6.1. Each incidence of hub-word use was then coded using the protocol described in section 5.2.2. This LG analysis answered the 'how are the hub-words used?' part of the research question 3 and is presented in section 6.2. A second phase of content analysis systematically and quantitatively analyses the similarities and differences in participants' use of hub-words in the physics and psychology cohorts. The NVivo (QSR International Pty Ltd., 2020) and Quanteda (Benoit et al., 2018) programmes were used to facilitate this analysis.

6.1 Frequency of Hub-Word Use in Interviews

The frequency counts for each hub-word were collated from the word frequency counts for each participants calculated using the Quanteda package (Benoit et al., 2018) in R (R Development Core Team, 2013). The frequency counts and normalised frequency counts for the hub-words *force*, *mass*, *energy*, *particle,* and *time* are shown with the interviewer's speech included in Table 10.15[10] and with the interviewer's speech excluded in Table 6.1. The normalised frequency counts express the likely number of occurrences of the hub-word per 500 words for each participant. Hence, these normalised frequencies allow for comparison between different participants and hub-words taking in account the different lengths of the interviews. Note that these normalised frequencies are expressed per 500 words, rather than per 1000 words, as the interviews are only a few thousand words long. Hence, using per 500 words avoids inflating the normalised word frequencies (Biber, Reppen, & Conrad, 1998). *Force* is the mostly frequently occurring hub-word for all participants when the interview's dialogue is included and excluded. *Time* is also frequently used. However, *energy* and *particle* occur infrequently, apart from 12P04 who uses *energy* 43 times during their interview. *Mass* is often used by physicists but is rarely used by psychologists (12S02 and 12S05 do not refer to *mass* at all).

[10] This table is presented in Appendix 10.9 as the inclusion of the interview's speech does not affect the finding of this analysis.

The descriptive statistics for the normalised hub-word frequencies for the physicist and psychology cohorts are shown in Table 6.2. A Wilcoxon rank sum test was used to establish if there were any significant differences in the normalised word frequencies between the physics and psychology cohorts (Lijffijt et al., 2014). Consistent with the frequency counts for individual participants, *force* is the most frequently used hub-word in both the physics and psychology cohorts (M_{phys}=28.23, SD_{phys}= 4.669; M_{phys}=9.294, SD_{phys}=4.507). However, *force* is used more frequently in the physicist cohort compared to the psychology cohort and this difference is significant (W=30, p=0.004). There is also a significant difference between the normalised frequency counts for *mass* between the two cohorts (W=30, p=0.008). In the physics cohort, the normalised frequency count for *mass* is noticeably higher than the normalised frequency counts for other hub-words. Hence, mass is the second most frequently used hub-word in the physics cohort. However, in the psychology cohort, the mean normalised frequency count for mass is comparable with the mean normalised frequency counts for the *energy*, *time*, and *particle*. There are no significant differences between the mean normalised frequency counts in the physics and psychology cohorts for *energy*, *time*, and *particle*. These results are invariant when the interviewer's speech is included or excluded.

Table 6.1 The frequency counts for hub-words in each participant's interview. T is the total number of words in the interview, C is the raw frequency count, C/T is the fractional frequency count (raw count/ total number of words in an interview), N is the normalised work frequency count (hub-word occurrence per 500 words), and R is the hub-word rank in the list of most frequently occurring words in an interview. Note the normalised word frequency count is given per 500 words and not per 1000 words as most interview are ~ 500 to 1500 words (Biber et al., 1998). These descriptive statistics exclude the interview's speech.

		T	Force				Mass				Energy				Time				Particle			
Cohort	Participant		C	$\frac{C}{T}$	N	R	C	$\frac{C}{T}$	N	R	C	$\frac{C}{T}$	N	R	C	$\frac{C}{T}$	N	R	C	$\frac{C}{T}$	N	R
Physics	12P02	841	61	0.073	36.266	1	7	0.008	4.162	24	0	0.000	0.000	-	5	0.006	2.973	37	0	0.000	0.000	-
	12P03	1288	70	0.054	27.174	1	10	0.008	3.882	29	4	0.003	1.553	74	12	0.009	4.658	22	0	0.000	0.000	-
	12P04	1521	94	0.062	30.901	1	10	0.007	3.287	38	43	0.028	14.135	2	7	0.005	2.301	51	4	0.003	1.315	85
	12P05	1158	54	0.047	23.316	3	16	0.014	6.908	16	0	0.000	0.000	-	2	0.002	0.864	89	3	0.003	1.295	73
	12P08	983	52	0.053	26.450	1	12	0.012	6.104	19	1	0.001	0.509	137	2	0.002	1.017	86	10	0.010	5.086	21
	12P12	891	45	0.051	25.253	1	14	0.016	7.856	10	0	0.000	0.000	-	4	0.004	2.245	50	1	0.001	0.561	137
Psychology	12S01	1125	17	0.015	7.556	9	0	0.000	0.000	-	0	0.000	0.000	-	0	0.000	0.000	-	0	0.000	0.000	-
	12S02	551	10	0.018	9.074	11	2	0.004	1.815	53	0	0.000	0.000	-	3	0.005	2.722	41	1	0.002	0.907	81
	12S03	670	21	0.031	15.672	6	4	0.006	2.985	32	0	0.000	0.000	-	2	0.003	1.493	66	2	0.003	1.493	66
	12S05	1249	27	0.022	10.809	6	0	0.000	0.000	-	1	0.001	0.400	152	9	0.007	3.603	32	1	0.001	0.400	152
	12S09	744	5	0.007	3.360	27	1	0.001	0.672	108	0	0.000	0.000	-	3	0.004	2.016	47	0	0.000	0.000	-

*Table 6.2 Descriptive statistics for the normalised word frequency counts for the hub-words in the participants interviews. A Wilcoxon Rank Sum test was used to establish whether any difference between the normalised frequency counts in each group for each hub-word were significant. * indicates p<0.05, which suggests that the null hypothesis that the normalised word frequency counts in the physicist and psychology cohorts for each hub-word are drawn from the same distribution can be rejected with 95% confidence.*

		Force				Mass				Energy				Time				Particle			
Interviewer Dialogue	Cohort	Mean	SD	W	p	Mean	SD	W	p	Mean	SD	W	p	Mean	SD	W	p	Mean	SD	W	p
Included	Physics	19.841	3.497	30	0.004*	3.107	0.935	30	0.008*	1.871	3.866	21	0.245	2.368	0.679	14	0.931	0.745	1.090	18	0.640
	Psychology	8.209	3.500			0.887	0.846			0.040	0.089			2.185	1.185			0.345	0.507		
Removed	Physics	28.23	4.669	30	0.004*	5.367	1.849	30	0.008*	2.699	5.635	21	0.245	2.343	1.394	17	0.792	1.376	1.909	18	0.640
	Psychology	9.294	4.507			1.094	1.291			0.080	0.179			1.967	1.355			0.560	0.641		

The word frequency counts for the twenty most frequently used words are show in Table 6.3 for the physics cohort and Table 6.4 for the psychology cohorts. *Force* is the most frequently used word in the physics interviews (rank 1). *Force* is also typically the highest ranked word when the word frequencies are calculated individually for physics interviews (see Table 6.1). In the physics interviews, *velocity* (rank 4), *acceleration* (rank 6), *resistance* (rank 7), *weight* (rank 9), *pressure* (rank 18), and *mass* (rank 19) are also frequently occurring words, which can be associated with recognisable physical concepts, rather than words that sustain dialogue (e.g., yes) or other grammatical functions (e.g., because). In the psychology interviews, *pressure* (rank 8) and *force* (rank 9) are the highest-ranking words that are associated with physical concepts. *Gravity* (rank 11) and *speed* (rank 13) are also included in the twenty most frequently occurring words.

Table 6.5 shows the count of the number of times a word occurs in the twenty highest-ranking words in individual interviews in the physics and psychology cohorts. *Force* is the most commonly occurring word in both the physics and psychology cohorts. A Fischer exact test was used to determine if there were statistically significant differences in the occurrence of top twenty highest-ranking words between the two cohorts. There are significant differences between the occurrence of *acceleration* ($p=0.015$), *weight* ($p=0.015$), *velocity* ($p=0.002$), *gravity* ($p=0.015$), and *speed* ($p=0.015$) at a 95% confidence level. *Acceleration* ($count_{phys}=6$, $count_{psy}=1$), *weight* ($count_{phys}=6$, $count_{psy}=1$) and *velocity* ($count_{phys}=6$, $count_{psy}=0$) occur more frequently as the twenty highest-ranked words in the physics cohort. Whereas *gravity* ($count_{phys}=0$, $count_{psy}=4$) and *speed* ($count_{phys}=0$, $count_{psy}=4$) occur more frequently as high-ranking words in the psychology cohort.

Table 6.3 Word frequencies for physics participant interviews. C is the raw frequency count, C/T is the fractional frequency count (raw count/total number of words in an interview), N is the normalised work frequency count (hub-word occurrence per 500 words), and R is the word rank in the list of most frequently occurring words in an interview. Note the top 20 ranked words are shown. Words in bold are recognisably associated with physical concepts in mechanics. The interview's speech is excluded.

Word	C	$\frac{C}{T}$	N	R
Force	376	0.053	26.599	1
Go	228	0.032	16.129	2
Because	224	0.032	15.846	3
Velocity	195	0.028	13.795	4
Yeah	191	0.027	13.512	5
Acceleration	145	0.021	10.257	6
Resistance	119	0.017	8.418	7
Greater	118	0.017	8.347	8
Weight	116	0.016	8.206	9
Well	106	0.015	7.499	10
Air	100	0.014	7.074	11
Act	99	0.014	7.003	12
One	82	0.012	5.801	13
Yes	76	0.011	5.376	14
Just	74	0.010	5.235	15
Downward	72	0.010	5.093	16
Pressure	71	0.010	5.023	17
Mass	69	0.010	4.881	18
Lid	68	0.010	4.810	19
Upward	67	0.009	4.740	20

Table 6.4 Word frequencies for psychology participant interviews. C is the raw frequency count, C/T is the fractional frequency count (raw count/total number of words in an interview), N is the normalised work frequency count (hub-word occurrence per 500 words), and R is the word rank in the list of most frequently occurring words in an interview. Note the top 20 ranked words are shown. Words in bold are recognisably associated with physical concepts in mechanics. The interview's speech is excluded.

Word	C	C/T	N	R
Yeah	1007	0.061	30.359	1
Like	603	0.036	18.179	2
Go	470	0.028	14.169	3
Because	450	0.027	13.566	4
Just	294	0.018	8.863	5
Ok	274	0.017	8.260	6
Pressure	269	0.016	8.110	7
Force	260	0.016	7.838	8
Kind	247	0.015	7.446	9
Move	246	0.015	7.416	10
Gravity	230	0.014	6.934	11
Got	223	0.013	6.723	12
Speed	218	0.013	6.572	13
Say	217	0.013	6.542	14
One	215	0.013	6.482	15
Know	187	0.011	5.638	16
Air	178	0.011	5.366	17
Think	169	0.010	5.095	18
Umm	157	0.009	4.733	19
Okay	153	0.009	4.613	20

Table 6.5 Frequency count of the highest ranked most frequency occurring words in the participant interviews grouped by cohort. The interviews speech was excluded in the word frequency counts. The p values were obtained using a Fischer exact test and indicate whether any difference between the frequency of occurrence of a word in the top 20 most frequent words is significant (indicates p<0.05 and the difference is significant). Word associated with a significant difference are shown in bold. Note words that are associated with dialogue, rather than a description or explanation of a demonstration have been excluded (just, yeah, like, well, kind, know, umm, got, think, bit, get, guess, obvious, ok, yes, say, really).*

	Cohort			
Word	Physics	Psychology	Total	p
Force	6	4	10	0.455
Go	6	3	9	0.182
Air	4	4	8	1.000
Acceleration	6	1	7	0.015*
One	4	3	7	1.000
Pressure	3	4	7	0.545
Resistance	5	2	7	0.242
Weight	6	1	7	0.015*
Move	2	4	6	0.242
Velocity	6	0	6	0.002*
Lid	3	2	5	1.000
Act	3	1	4	0.545
Gravity	0	4	4	0.015*
Greater	4	0	4	0.061
Speed	0	4	4	0.015*
Come	0	3	3	0.061
Faster	0	3	3	0.061
Mass	3	0	3	0.182
Point	2	1	3	1.000

Upward	3	0	3	0.182
Back	0	2	2	0.182
Bigger	0	2	2	0.182
Canister	1	1	2	1.000
Downward	2	0	2	0.455
Equal	2	0	2	0.455
Increase	2	0	2	0.455
Less	2	0	2	0.455
Mg	2	0	2	0.455
Push	1	1	2	1.000
Resultant	2	0	2	0.455
Start	0	2	2	0.182

6.2 The Function of Hub-Word in Descriptions and Explanations

This section presents the results of the LG analysis that characterises the function of *force* and *mass* in interviews. *Force* and *mass* were selected for this analysis because they are the most frequently used hub-words. This includes the analysis of participants' descriptions and explanation of the film canister rocket (demonstration 1) and the falling cup-cake case (demonstration 2). NVivo was used to facilitate this analysis through coding and then content analysis of these codes (looking for systematic differences).

6.2.1 Force

Table 6.6 shows the coding frequency counts for the LG analysis of participants' use of *force* in their interviews. The normalised frequency counts are the fraction of the total codes for force for each participant. In addition to the coding described in section 5.2.2, codes for fragmentation, correction, and querying have been added in response to these unanticipated uses of *force*. These codes refer to when a participant has used *force*:

1. in a fragmented series of words, which do not form a complete clause (e.g., 12P12; '*The resistance force is, well, equal to* [participant stops]'),
2. when they have corrected themselves in an explanation by changing their use of force within a clause complex, making the original explanation redundant (e.g., 12P05: '*And then the resistance forces are only going to get bigger because, actually. OK, yeah, no…*'),
3. when they have asked a question (12P04: '*No, the acceleration, or do you mean the deceleration force?*').

The descriptive statistics for the coding frequency counts for *force* are shown in Table 6.7 and the mean normalised frequency counts are plotted in Figure 6.1. *Force* is most frequently used in both the physics and psychology cohorts as a circumstantial cause in a clause complex, where the main clause describing the motion of the object in the demonstration. For example;

> *So, the instant it hits, the reaction force would be greater than the weight force in order to have a resultant force* ***to decelerate the object to a stationary position.*** *(12P04)*
>
> *Gravity is a force acting on it that causes* ***it to come back down*** *(12S03).*

In the above example, the main clause is in bold. There are no statistically significant differences in LG function of *force* for the physics and psychology participants.

Table 6.6 Coding frequency counts for instances of participants' use of the hub-word force. Normalised counts are fraction of the total occurrence for each participant.

	Raw Counts											Normalised Counts										
	Physics (n=6)						Psychology (n=5)					Physics (n=6)						Psychology (n=5)				
Codes	P02	P03	P04	P05	P08	P12	S01	S02	S03	S05	S09	P02	P03	P04	P05	P08	P12	S01	S02	S03	S05	S09
Process																						
Main Clause	1	0	1	3	0	0	0	1	0	6	0	0.026	0.000	0.020	0.094	0.000	0.000	0.000	0.143	0.000	0.286	0.000
Medium																						
Participant																						
Main Clause	4	8	8	1	3	0	1	1	2	1	0	0.105	0.200	0.163	0.031	0.097	0.000	0.083	0.143	0.133	0.048	0.000
Range																						
Participant																						
Main Clause	3	5	1	4	2	2	0	0	1	4	0	0.079	0.125	0.020	0.125	0.065	0.087	0.000	0.000	0.067	0.190	0.000
Agent																						
Participant																						
Main Clause	4	4	0	1	2	1	1	0	1	1	0	0.105	0.100	0.000	0.031	0.065	0.043	0.083	0.000	0.067	0.048	0.000
Cause																						
Circumstance																						
Main Clause	2	1	0	0	2	1	0	0	0	0	0	0.053	0.025	0.000	0.000	0.065	0.043	0.000	0.000	0.000	0.000	0.000
Cause																						
Circumstance																						
Clause Complex	24	22	33	18	21	18	10	5	8	5	3	0.632	0.550	0.673	0.563	0.677	0.783	0.833	0.714	0.533	0.238	0.750
Correction	0	0	1	2	0	0	0	0	0	0	0	0.000	0.000	0.020	0.063	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Fragment	0	0	4	3	1	1	0	0	3	4	1	0.000	0.000	0.082	0.094	0.032	0.043	0.000	0.000	0.200	0.190	0.250
Query	0	0	1	0	0	0	0	0	0	0	0	0.000	0.000	0.020	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Total	38	40	49	32	31	23	12	7	15	21	4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

*Table 6.7 Descriptive statistics for the coding of the hub-word force in the interviews. A Wilcoxon Rank Sum test was used to establish whether any difference between the normalised coding frequency counts in each group for each coding were significant. * indicates p<0.05, which suggests that the null hypothesis that there is no difference between the normalised coded frequency counts in the physicist and psychology cohorts for each code can be rejected with 95% confidence.*

	Physics (n=6)			Psychology (n=5)			Significance	
Codes	Total Raw Count	Normalised Mean	Normalised SD	Total Raw Count	Normalised Mean	Normalised SD	W	p
Process								
Main Clause	5	0.023	0.036	7	0.086	0.128	13.5	0.842
Medium								
Participant								
Main Clause	24	0.099	0.076	5	0.081	0.060	17.5	0.714
Range								
Participant								
Main Clause	17	0.083	0.040	5	0.051	0.083	22.0	0.230
Agent								
Participant								
Main Clause	12	0.057	0.041	3	0.040	0.038	18.0	0.645
Cause								
Circumstance								
Main Clause	6	0.031	0.027	0	0.000	0.000	25.0	0.045
Cause								
Circumstance								
Clause Complex	136	0.646	0.086	31	0.614	0.237	14.0	0.931
Correction	3	0.014	0.025	0	0.000	0.000	20.0	0.223
Fragment	9	0.042	0.040	8	0.128	0.119	10.0	0.400
Query	1	0.003	0.008	0	0.000	0.000	17.5	0.465
Total	213			59				

As discussed in 5.2.2, circumstantial information can be given within the clause or as part of a clause nexus complex (Halliday & Matthiessen, 2004). In this analysis, *force* acts as a circumstantial element both within the clause and as part of a clause complex. In both cases, cause is the circumstantial function. Similarly, *force* occasionally also acts as an agent within the ergative structure of the clause. In this instance, *force* is again functioning in a causal way. These incidences of *force* acting in a causal role were grouped as a single coding of 'cause'.

The coding frequency counts (raw and normalised) for each participant with the grouped coding 'cause' are shown in Table 6.8. The descriptive statistics for the normalised coding frequencies with this new grouped coding of cause are shown in Table 6.9 and the mean normalised coding frequencies are plotted in Figure 6.1. In both the physics and psychologist cohorts, cause is the most frequently used function for force (normalised $mean_{phys} = 0.735$, normalised $mean_{psy} = 0.653$). Again, there are no statistically significant differences between the frequencies in the LG function of force in the physics and psychology cohorts.

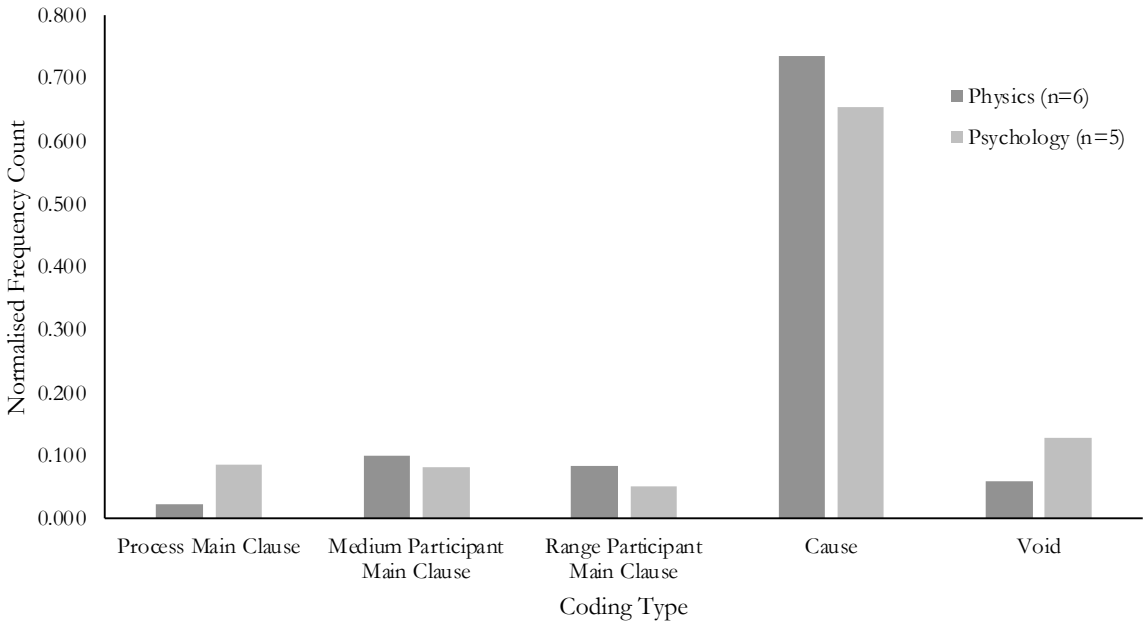


Figure 6.1 Mean normalised coding frequency counts for instances of participants use of the hub-word force with cause codes grouped. Normalised counts are fraction of the total coding for force for each participant.

Table 6.8 Coding Frequency counts for instances of participants use of the hub-word force with the cause and void codes grouped. Normalised counts are fraction of the total occurrence for each participant.

	Raw Counts											Normalised Counts										
	Physics (n=6)						Psychology (n=5)					Physics (n=6)						Psychology (n=5)				
Codes	P02	P03	P04	P05	P08	P12	S01	S02	S03	S05	S09	P02	P03	P04	P05	P08	P12	S01	S02	S03	S05	S09
Process																						
Main Clause	1	0	1	3	0	0	0	1	0	6	0	0.026	0.000	0.020	0.094	0.000	0.000	0.000	0.143	0.000	0.286	0.000
Medium																						
Participant																						
Main Clause	4	8	8	1	3	0	1	1	2	1	0	0.105	0.200	0.163	0.031	0.097	0.000	0.083	0.143	0.133	0.048	0.000
Range																						
Participant																						
Main Clause	3	5	1	4	2	2	0	0	1	4	0	0.079	0.125	0.020	0.125	0.065	0.087	0.000	0.000	0.067	0.190	0.000
Cause	30	27	33	19	25	20	11	5	9	6	3	0.789	0.675	0.673	0.594	0.806	0.870	0.917	0.714	0.600	0.286	0.750
Void	0	0	6	5	1	1	0	0	3	4	1	0.000	0.000	0.122	0.156	0.032	0.043	0.000	0.000	0.200	0.190	0.250
Total	38	40	49	32	31	23	12	7	15	21	4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

*Table 6.9 Descriptive statistics for the coding of the hub-word force in the interviews with the cause and void coding grouped. A Wilcoxon Rank Sum test was used to establish whether any difference between the normalised coding frequency counts in each group for each coding were significant. * indicates p<0.05, which suggests that the null hypothesis that there is no difference between the normalised coded frequency counts in the physicist and psychology cohorts for each code can be rejected with 95% confidence.*

	Physics (n=6)			Psychology (n=5)			Significance	
Codes	Total Raw Count	Normalised Mean	Normalised SD	Total Raw Count	Normalised Mean	Normalised SD	W	p
Process								
Main Clause	5	0.023	0.036	7	0.086	0.128	13.5	0.842
Medium								
Participant								
Main Clause	24	0.099	0.076	5	0.081	0.060	17.5	0.714
Range								
Participant								
Main Clause	17	0.083	0.040	5	0.051	0.083	22.0	0.230
Cause	154	0.735	0.103	34	0.653	0.235	17.0	0.792
Void	13	0.059	0.065	8	0.128	0.119	10.0	0.400
Total	213			59				

The LG analysis of participants' use of *force* suggests *force* predominantly functions as a cause in both the physicists' and psychologists' explanations. However, the FCI scores indicate the physicists and psychologists have different understandings of mechanics. The physicists' scores demonstrate that they answer mechanics problems in a way that is consistent with expert physicists. Hence, they are considered to have a Newtonian understanding of mechanics. In contrast, the psychologists derive solutions to problems that are consistent with an everyday understanding of mechanics. The fact that the physicists and psychologists both see *force* as a cause but arrive at different solutions to the same mechanics problems begs the question of whether the physicists and psychologists agree about what *force* causes. This question was investigated using another LG analysis. This analysis focused on nature of the process in the ergative model associated with the main cause. Some examples of different processes with *force* as a cause are shown in Table 6.10 below.

Table 6.10 Examples of process codes when force is identified as a cause of the process. The main clause is in bold and the process is underlined.

Participant	Cohort	Interview Extract	Coding
12P02	Physics	And the force is still acting. So now **it begins to <u>accelerate downwards</u>.**	Acceleration downwards.
12P08	Physics	[…] **its <u>downwards acceleration is going to decrease</u> way faster** and a lot more because of the air resistance being greater and bigger and counteracting that weight force faster.	Acceleration downwards decreases
12P04	Physics	**It <u>reaches a terminal velocity,</u>** when there's a balance or an equilibrium between the weight component and the air resistance frictional component. So, if they're equal, then there's zero resultant force.	Velocity terminal
12S09	Psychology	Because there's more force, I guess, pushing it down because it's heavier. And so there has to be more force going upward, then **it just <u>moves faster.</u>**	Motion faster
12P08	Physics	**It <u>fell</u>** due to its weight force from gravity.	Fall
12S01	Psychology	The force like going up, it's going to be a bit bigger. So, it causes **it to like, <u>pop up.</u>**	Colloquial Pop-up

These examples illustrate how the process in the ergative model was used to create initial 'loose' codes for what a participant identifies force as causing. These codes were applied to each instance of *force* with a 'cause' code in the first phase of this LG analysis. These loose codes were then grouped as shown in Table 6.11.

Table 6.11 The process group codes for when force is identified as a cause and the loose coding that form each group.

Grouped Process Code	Loose Process Codes
Acceleration	Accelerate, accelerate change, accelerate constant, accelerate decrease, accelerate downward decrease, accelerate downward increase, accelerate negative, no acceleration, accelerate positive, accelerate upwards, decelerate, decelerate constant.
Velocity	Velocity, velocity constant, velocity decrease, velocity increase, velocity opposing, velocity terminal
Momentum	Momentum, momentum change, momentum same
Equilibrium	Equilibrium
Stationary	Stationary, stationary becomes
Force	Force exerted, force into air, force overcome, force gravity more powerful
Other Physical Process	Gas escapes
Motion (moves)	Motion, motion faster, motion quicker, motion slower, motion same, motion upwards, motion parabolic, motion how far up.
Go	Go higher, go up, go straight up, go up higher, go upwards longer, go quickest, go keep, go through the desk.
Fall	Fall, fall slower, fall not
Pull	Pull down quicker
Push	Push, push down, push up
Stop	Stop, stop from sinking
Colloquial	Blowing up, comes down, flies off, pop up, propel up, reach a great height, shoots back up, speedy bit, travel further.

The raw and the normalised code frequency counts for the processes each participant associate with *force* (as a cause) are shown in Table 6.12. The descriptive statistics for the normalised codes frequency counts are shown in Table 6.13 and the mean normalised frequency counts for the processes are plotted in Figure 6.2. In the physicist cohort, acceleration (raw counts = 85, normalised mean = 0.552, normalised SD = 0.130) and velocity (raw counts = 26, normalised mean = 0.155, normalised SD = 0.065) are most frequency occurring process associated with *force* as a cause. Moreover, there is a statistically significant different in the normalised frequency counts for these processes between the physics and psychology cohort ($W_{acceleration} = 30$, $p_{acceleration} = 0.007$; $W_{velocity} = 30$, $p_{velocity} = 0.005$), with only one reference to force causing acceleration in the psychology cohort and no reference to velocity.

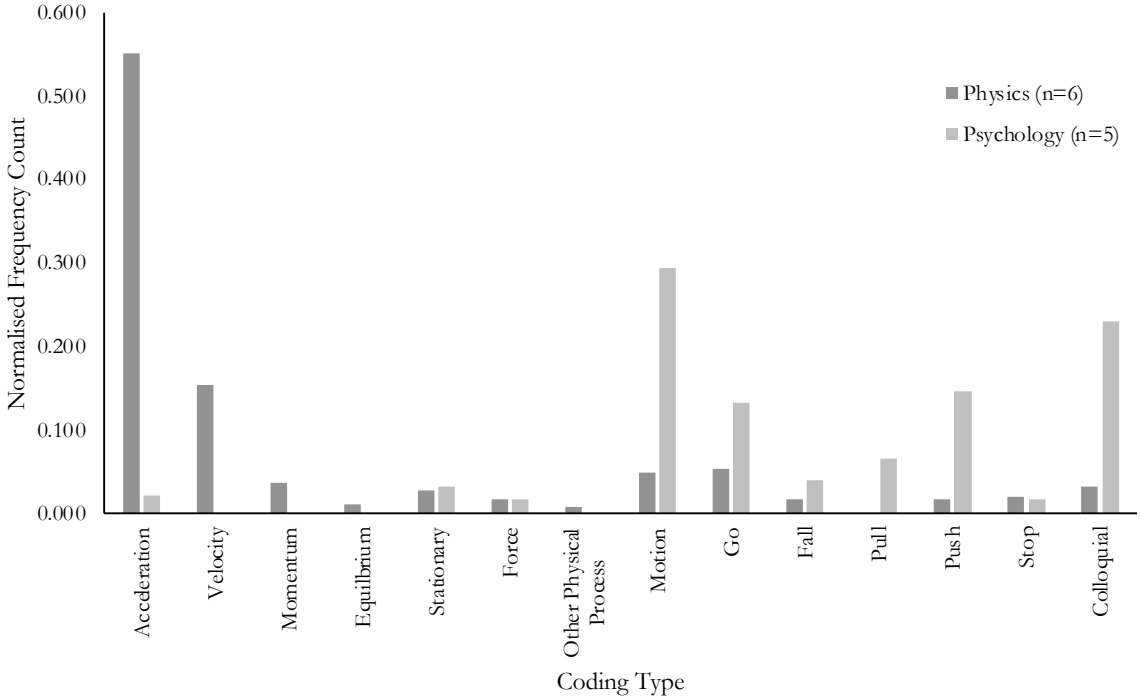


Figure 6.2 Normalised coding frequency counts for the process associated with each participants use the hub-word force as a cause. Normalised counts are fraction of the total codes for processes for each participant.

Table 6.12 Coding frequency counts for the processes associated with each participants' use of the hub-word force as a cause. Normalised counts are fraction of the total occurrence for each participant.

	Raw Counts											Normalised Counts										
	Physics (n=6)						Psychology (n=5)					Physics (n=6)						Psychology (n=5)				
Codes	P02	P03	P04	P05	P08	P12	S01	S02	S03	S05	S09	P02	P03	P04	P05	P08	P12	S01	S02	S03	S05	S09
Acceleration	13	16	17	13	11	15	0	0	1	0	0	0.433	0.571	0.486	0.650	0.423	0.750	0.000	0.000	0.111	0.000	0.000
Velocity	6	6	6	2	5	1	0	0	0	0	0	0.200	0.214	0.171	0.100	0.192	0.050	0.000	0.000	0.000	0.000	0.000
Momentum	3	0	0	0	2	1	0	0	0	0	0	0.100	0.000	0.000	0.000	0.077	0.050	0.000	0.000	0.000	0.000	0.000
Equilibrium	2	0	0	0	0	0	0	0	0	0	0	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Stationary	0	2	2	0	1	0	0	0	0	1	0	0.000	0.071	0.057	0.000	0.038	0.000	0.000	0.000	0.000	0.167	0.000
Force	0	2	1	0	0	0	1	0	0	0	0	0.000	0.071	0.029	0.000	0.000	0.000	0.083	0.000	0.000	0.000	0.000
Other Physical Process	0	0	0	0	0	1	0	0	0	0	0	0.000	0.000	0.000	0.000	0.000	0.050	0.000	0.000	0.000	0.000	0.000
Motion	5	0	0	1	2	0	1	0	5	1	2	0.167	0.000	0.000	0.050	0.077	0.000	0.083	0.000	0.556	0.167	0.667
Go	0	1	3	4	0	0	4	0	0	2	0	0.000	0.036	0.086	0.200	0.000	0.000	0.333	0.000	0.000	0.333	0.000
Fall	0	0	1	0	2	0	0	1	0	0	0	0.000	0.000	0.029	0.000	0.077	0.000	0.000	0.200	0.000	0.000	0.000
Pull	0	0	0	0	0	0	0	0	3	0	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.333	0.000	0.000
Push	0	0	1	0	2	0	4	2	0	0	0	0.000	0.000	0.029	0.000	0.077	0.000	0.333	0.400	0.000	0.000	0.000
Stop	1	0	0	0	1	1	1	0	0	0	0	0.033	0.000	0.000	0.000	0.038	0.050	0.083	0.000	0.000	0.000	0.000
Colloquial	0	1	4	0	0	1	1	2	0	2	1	0.000	0.036	0.114	0.000	0.000	0.050	0.083	0.400	0.000	0.333	0.333
Total	30	28	35	20	26	20	12	5	9	6	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

*Table 6.13 Descriptive statistics for the coding frequency counts for the process associated with each participants' use of the hub-word force as a cause A Wilcoxon Rank Sum test was used to establish whether any difference between the normalised coding frequency counts in each group for each coding were significant. * indicates p<0.05, which suggests that the null hypothesis that there is no difference between the normalised coded frequency counts in the physicist and psychology cohorts for each code can be rejected with 95% confidence.*

	Physics (n=6)			Psychology (n=5)			Significance	
Codes	Total Raw Count	Normalised Mean	Normalised SD	Total Raw Count	Normalised Mean	Normalised SD	W	p
Acceleration	85	0.552	0.130	1	0.022	0.050	30.0	0.007*
Velocity	26	0.155	0.065	0	0.000	0.000	30.0	0.005*
Momentum	6	0.038	0.044	0	0.000	0.000	22.5	0.104
Equilibrium	2	0.011	0.027	0	0.000	0.000	17.5	0.465
Stationary	5	0.028	0.032	1	0.033	0.075	18.0	0.597
Force	3	0.017	0.029	1	0.017	0.037	16.0	0.908
Other Physical Process	1	0.008	0.020	0	0.000	0.000	17.5	0.465
Motion	8	0.049	0.066	9	0.294	0.298	6.0	0.111
Go	8	0.054	0.079	6	0.133	0.183	13.5	0.842
Fall	3	0.018	0.031	1	0.040	0.089	16.0	0.908
Pull	0	0.000	0.000	3	0.067	0.149	12.0	0.361
Push	3	0.018	0.031	6	0.147	0.202	12.0	0.597
Stop	3	0.020	0.023	1	0.017	0.037	18.0	0.597
Colloquial	6	0.033	0.045	6	0.230	0.177	5.5	0.092
Total	159			35				

6.2.2 Mass

Table 6.14 shows the code frequency counts (raw and normalised) for the LG analysis of participants' use of *mass* in their interviews. The descriptive statistics for the normalised frequency counts are shown in Table 6.15 and the mean normalised frequency counts for the physics and psychology cohorts are plotted in Figure 6.3. *Mass* is most frequently used as a casual circumstantial element in both the physics (normalised mean = 0.841, normalised SD = 0.117) and the psychology (normalised mean = 0.350, normalised SD = 0.487) cohorts. There are no statistically significant differences between the normalised frequency counts for all types of code in the physics and psychology cohorts.

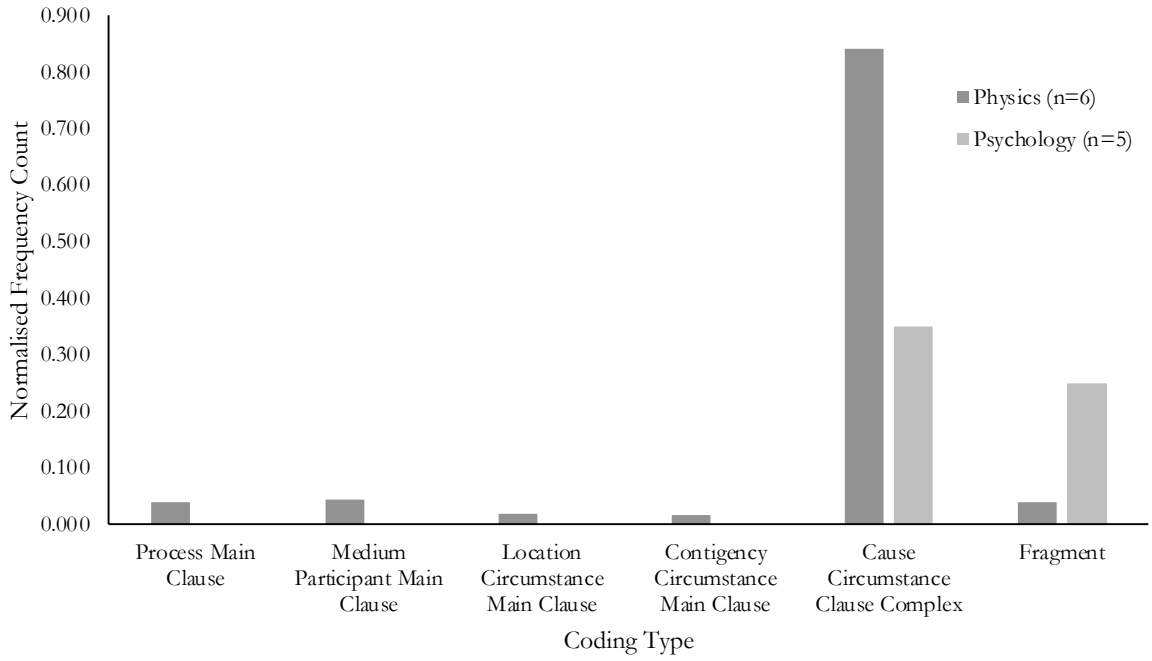


Figure 6.3 Mean normalised coding frequency counts for instances of participants use of the hub-word mass. Normalised counts are fraction of the total codes for mass for each participant.

As in the LG analysis for *force*, there are several codes that identify cause as the function of *mass* in participants' explanations (contingency circumstantial main clause, cause circumstantial clause complex). Again, these codes were grouped to form a single 'cause' code. The code frequency counts for *mass* with the grouped code for cause are shown in Table 6.16. Cause remains the most frequently used function of *mass* with the grouped code for cause in both the physics (cause: normalised mean = 0.858, normalised SD = 0.098) and psychology cohorts (cause: normalised mean = 0.350, normalised SD = 0.487) (see Table 6.17 and Figure 5.1). There remain no statistically significant differences between the normalised frequency counts for LG function of mass in the physics and psychology cohorts.

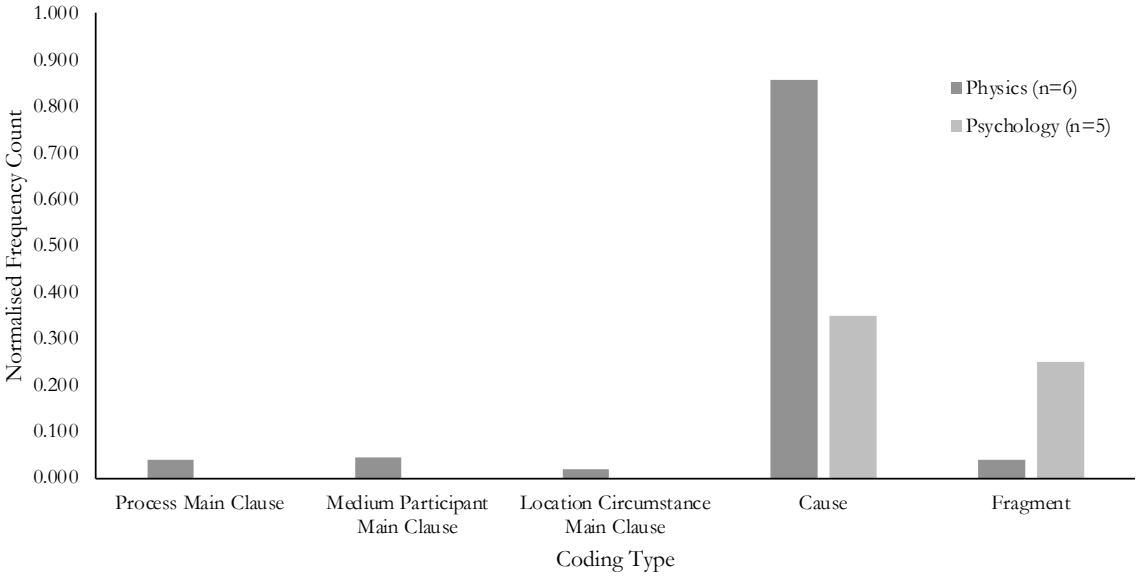


Figure 6.4 Mean normalised coding frequency counts for instances of participants' use of the hub-word mass. Normalised counts are fraction of the total coding for mass for each participant.

Table 6.14 Coding Frequency counts for instances of participants' use of the hub-word mass. Normalised counts are fraction of the total occurrence for each participant.

	Raw Counts											Normalised Counts										
	Physics (n=6)						Psychology (n=5)					Physics (n=6)						Psychology (n=5)				
Codes	P02	P03	P04	P05	P08	P12	S01	S02	S03	S05	S09	P02	P03	P04	P05	P08	P12	S01	S02	S03	S05	S09
Medium																						
Participant																						
Main Clause	0	1	0	0	0	1	0	0	0	0	0	0.000	0.143	0.000	0.000	0.000	0.100	0.000	0.000	0.000	0.000	0.000
Range																						
Participant																						
Main Clause	1	0	0	0	0	1	0	0	0	0	0	0.167	0.000	0.000	0.000	0.000	0.100	0.000	0.000	0.000	0.000	0.000
Location																						
Circumstance																						
Main Clause	0	0	0	0	1	0	0	0	0	0	0	0.000	0.000	0.000	0.000	0.111	0.000	0.000	0.000	0.000	0.000	0.000
Contingency																						
Circumstance																						
Main Clause	0	0	0	0	0	1	0	0	0	0	0	0.000	0.000	0.000	0.000	0.000	0.100	0.000	0.000	0.000	0.000	0.000
Cause																						
Circumstance																						
Clause Complex	5	5	7	10	8	7	0	2	3	0	0	0.833	0.714	1.000	0.909	0.889	0.700	0.000	1.000	0.750	0.000	0.000
Fragment	0	1	0	1	0	0	0	0	1	0	1	0.000	0.143	0.000	0.091	0.000	0.000	0.000	0.000	0.250	0.000	1.000
Total	6	7	7	11	9	10	0	2	4	0	1	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000	1.000	0.000	1.000

*Table 6.15 Descriptive statistics for the coding frequency counts for instances of participants' use of the hub-word mass. A Wilcoxon Rank Sum test was used to establish whether any difference between the normalised coding frequency counts in each group for each coding were significant. * indicates p<0.05, which suggests that the null hypothesis that there is no difference between the normalised coded frequency counts in the physicist and psychology cohorts for each code can be rejected with 95% confidence.*

	Physics (n=6)			Psychology (n=5)			Significance	
Codes	Total Raw Count	Normalised Mean	Normalised SD	Total Raw Count	Normalised Mean	Normalised SD	W	p
Medium								
Participant								
Main Clause	2	0.040	0.064	0	0.000	0.000	20.0	0.223
Range								
Participant								
Main Clause	2	0.044	0.072	0	0.000	0.000	20.0	0.223
Location								
Circumstance								
Main Clause	1	0.019	0.045	0	0.000	0.000	17.5	0.465
Contingency								
Circumstance								
Main Clause	1	0.017	0.041	0	0.000	0.000	17.5	0.465
Cause								
Circumstance								
Clause Complex	42	0.841	0.117	5	0.350	0.487	22.5	0.196
Fragment	2	0.039	0.063	2	0.250	0.433	12.0	0.597
Total	50			7				

Table 6.16 Coding Frequency counts for instances of participants' use of the hub-word mass with the cause codes grouped. Normalised counts are fraction of the total occurrence for each participant.

	Raw Counts											Normalised Counts										
	Physics (n=6)						Psychology (n=5)					Physics (n=6)						Psychology (n=5)				
Codes	P02	P03	P04	P05	P08	P12	S01	S02	S03	S05	S09	P02	P03	P04	P05	P08	P12	S01	S02	S03	S05	S09
Medium																						
Participant																						
Main Clause	0	1	0	0	0	1	0	0	0	0	0	0.000	0.143	0.000	0.000	0.000	0.100	0.000	0.000	0.000	0.000	0.000
Range																						
Participant																						
Main Clause	1	0	0	0	0	1	0	0	0	0	0	0.167	0.000	0.000	0.000	0.000	0.100	0.000	0.000	0.000	0.000	0.000
Location																						
Circumstance																						
Main Clause	0	0	0	0	1	0	0	0	0	0	0	0.000	0.000	0.000	0.000	0.111	0.000	0.000	0.000	0.000	0.000	0.000
Cause	5	5	7	10	8	8	0	2	3	0	0	0.833	0.714	1.000	0.909	0.889	0.800	0.000	1.000	0.750	0.000	0.000
Fragment	0	1	0	1	0	0	0	0	1	0	1	0.000	0.143	0.000	0.091	0.000	0.000	0.000	0.000	0.250	0.000	1.000
Total	6	7	7	11	9	10	0	2	4	0	1	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000	1.000	0.000	1.000

*Table 6.17 Descriptive statistics for the coding frequency counts for instances of participants' use of the hub-word mass with the cause codes grouped. A Wilcoxon Rank Sum test was used to establish whether any difference between the normalised coding frequency counts in each group for each coding were significant. * indicates $p<0.05$, which suggests that the null hypothesis that there is no difference between the normalised coded frequency counts in the physicist and psychology cohorts for each code can be rejected with 95% confidence.*

	Physics (n=6)			Psychology (n=5)			Significance	
Codes	Total Raw Count	Normalised Mean	Normalised SD	Total Raw Count	Normalised Mean	Normalised SD	W	p
Medium								
Participant								
Main Clause	2	0.040	0.064	0	0.000	0.000	20.0	0.223
Range								
Participant								
Main Clause	2	0.044	0.072	0	0.000	0.000	20.0	0.223
Location								
Circumstance								
Main Clause	1	0.019	0.045	0	0.000	0.000	17.5	0.465
Cause	43	0.858	0.098	5	0.350	0.487	23.5	0.140
Fragment	2	0.039	0.063	2	0.250	0.433	12.0	0.597
Total	50			7				

As was the case with *force*, both physicists and psychologists both frequently associate *mass* with a causal function, despite having significantly different FCI scores. Hence, again, a second phase of LG analysis was carried out to investigate what process participants identified *mass* causing. The procedure for this analysis followed the procedure carried out for *force*, which was described in section 6.2.1. Table 6.18 shows the initial loose coding and how these codes were grouped.

Table 6.18 The process group codes for when mass is identified as a cause and the loose coding that form each group.

Grouped Process Code	Loose Process Codes
Acceleration	Accelerate, accelerate constant, accelerate downwards, accelerate negative
Velocity	Velocity, velocity decrease, velocity higher downwards, velocity greater, velocity terminal
Momentum	Momentum transfer
Weight	Weight
Gravity	Gravity
Other physical process	Gas pushing up
Motion	Motion quicker, motion slower
Height	Height greater.
Go	Go up higher, go up how far.
Fall	Fall faster, fall vertically.
Colloquial	Landed first.

Table 6.19 shows the raw and the normalised coding frequency counts for the processes each participant associates with *mass* (as a cause). The descriptive statistics for the normalised coding frequency counts are shown in Table 6.20 and the mean normalised frequency counts for the processes are plotted in Figure 6.5. As with *force*, acceleration (raw counts = 18, normalised mean = 0.416, normalised SD = 0.182) and velocity (raw counts = 14, normalised mean = 0.307, normalised SD = 0.0163) are most frequently occurring processes physicists associate with being caused by *mass*. And again, there is a statistically significant difference in the normalised frequency counts for these processes between the physics and psychology cohorts ($W_{acceleration} = 30$, $p_{acceleration} = 0.005$; $W_{velocity} = 30$, $p_{velocity} = 0.005$), with no references to *mass* causing acceleration or velocity in the psychology cohort.

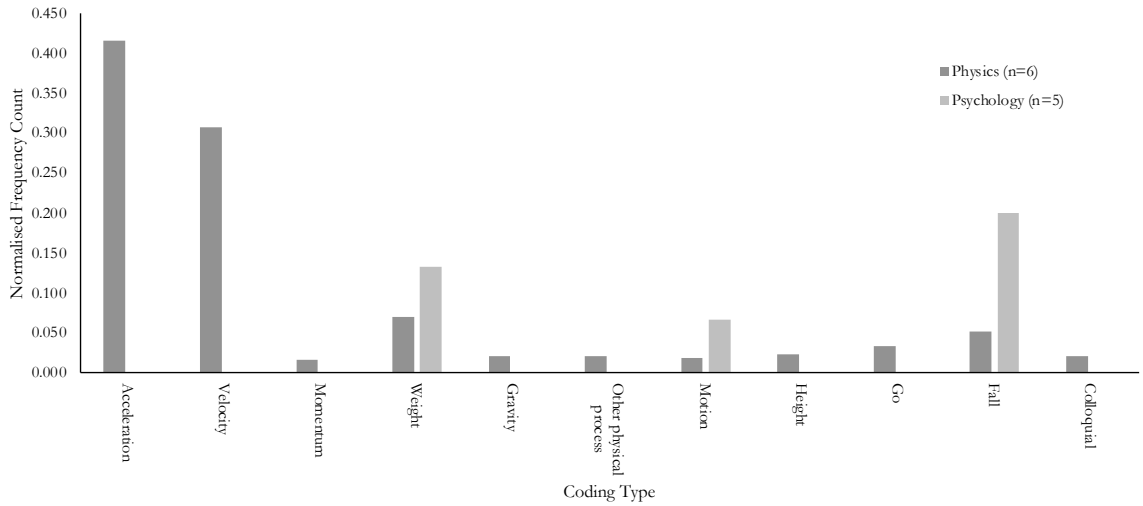


Figure 6.5 Normalised code frequency counts for the processes associated with each participant's use of the hub-word mass acting as a cause. Normalised counts are fraction of the total coding for process for each participant.

Table 6.19 Code Frequency counts for the processes associated with each participant's use of the hub-word mass acting as a cause. Normalised counts are fraction of the total occurrence for each participant.

	Raw Counts											Normalised Counts										
	Physics (n=6)						Psychology (n=5)					Physics (n=6)						Psychology (n=5)				
Codes	P02	P03	P04	P05	P08	P12	S01	S02	S03	S05	S09	P02	P03	P04	P05	P08	P12	S01	S02	S03	S05	S09
Acceleration	3	3	2	5	1	4	0	0	0	0	0	0.600	0.500	0.286	0.500	0.111	0.500	0.000	0.000	0.000	0.000	0.000
Velocity	1	2	3	2	5	1	0	0	0	0	0	0.200	0.333	0.429	0.200	0.556	0.125	0.000	0.000	0.000	0.000	0.000
Momentum	0	0	0	1	0	0	0	0	0	0	0	0.000	0.000	0.000	0.100	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Weight	0	1	1	0	1	0	0	0	2	0	0	0.000	0.167	0.143	0.000	0.111	0.000	0.000	0.000	0.667	0.000	0.000
Gravity	0	0	0	0	0	1	0	0	0	0	0	0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.000
Other physical process	0	0	0	0	0	1	0	0	0	0	0	0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.000
Motion	0	0	0	0	1	0	0	0	1	0	0	0.000	0.000	0.000	0.000	0.111	0.000	0.000	0.000	0.333	0.000	0.000
Height	0	0	1	0	0	0	0	0	0	0	0	0.000	0.000	0.143	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Go	0	0	0	2	0	0	0	0	0	0	0	0.000	0.000	0.000	0.200	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Fall	1	0	0	0	1	0	0	2	0	0	0	0.200	0.000	0.000	0.000	0.111	0.000	0.000	1.000	0.000	0.000	0.000
Colloquial	0	0	0	0	0	1	0	0	0	0	0	0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.000
Total	5	6	7	10	9	8	0	2	3	0	0	1	1	1	1	1	1	0	1	1	0	0

*Table 6.20 Descriptive statistics for the code frequency counts for the processes associated with each participant's use of the hub-word mass as a cause. A Wilcoxon Rank Sum test was used to establish whether any difference between the normalised coding frequency counts in each group for each coding were significant. * indicates p<0.05, which suggests that the null hypothesis that there is no difference between the normalised coded frequency counts in the physicist and psychology cohorts for each code can be rejected with 95% confidence.*

	Physics (n=6)			Psychology (n=5)			Significance	
Codes	Total Raw Count	Normalised Mean	Normalised SD	Total Raw Count	Normalised Mean	Normalised SD	W	p
Acceleration	18	0.416	0.182	0	0.000	0.000	30.0	0.005*
Velocity	14	0.307	0.163	0	0.000	0.000	30.0	0.005*
Momentum	1	0.017	0.041	0	0.000	0.000	17.5	0.465
Weight	3	0.070	0.079	2	0.133	0.298	18.0	0.597
Gravity	1	0.021	0.051	0	0.000	0.000	17.5	0.465
Other physical process	1	0.021	0.051	0	0.000	0.000	17.5	0.465
Motion	1	0.019	0.045	1	0.067	0.149	14.0	0.892
Height	1	0.024	0.058	0	0.000	0.000	17.5	0.465
Go	2	0.033	0.082	0	0.000	0.000	17.5	0.465
Fall	2	0.052	0.085	2	0.200	0.447	16.0	0.908
Colloquial	1	0.021	0.051	0	0.000	0.000	17.5	0.465
Total	45			5				

When *mass* was identified as functioning as a cause, it was noticed that *mass* often co-occurred in in this causal role with *force* and *energy*. For instance, in the first extract below, 12P02 identifies *mass* as causing greater weight (which they previous identified as a force). 12P02 then identifies this greater weight as being responsible for the cup-cake case taking longer to reach terminal velocity.

> *Umm, well it's got more mass. Urr. It's. Yeah, so its weight is greater. So, it takes longer for it to reach its terminal velocity. (12P05)*

> *Because when we have the smaller mass here and this force is the same because the pressure is the same, a smaller mass means that therefore acceleration would be greater. So, the lid will have accelerated greater as it left the floor. (12P12)*

Similarly, in the second extract, 12P12 explains that the smaller *mass* causes a greater acceleration for the lid but only because the *force* on the lid is the same. In both cases, the explanations use *mass* and *force* to explain the motion of the objects. The frequency of this interaction between *mass* and other hub-words, when *mass* functions as a cause, was investigated by all revisiting *mass* cause codes and coding for whether each incidence involved other hub-words or just *mass*. The types of interaction were also coded. For instance, the extracts above where coded as an interaction with force (weight) and force (Newton's 2nd Law) respectively. The extracts below show interactions with *energy* and *gravity*. Note that *gravity* was not treated as a *force* as the participants in this case had not explicitly referred to *gravity* as a force.

> *It's, it's doing work within the gravitational field, so it's, it's increasing its potential within the field as it moves away from the centre of the Earth, basically to the surface of the Earth. So, because of conservation of energy, energy is only going to be transferred. So that's why the kinetic energy goes to the potential energy of which kinetic energy is simplified and half mv squared. So, there's a component of velocity, that of which mass is constant. So, the velocity has to decrease if the kinetic energy is decreasing. (12P04)*

> *The one with the paper clips falls faster because it's now got a greater mass, which means there's more gravity acting on it. (12S02)*

Table 6.21 shows the raw and the normalised frequency counts for the types of interaction between *mass* and other hub-words, when *mass* functions as a cause, for the physics and the psychology cohorts. In both cohorts, *mass* is typically used in association with other hub-words (also see Figure 6.6). In the physics cohort, *mass* usually functions as a cause through an association with

weight (raw counts = 17, normalised counts = 0.395) or via Newton's 2nd Law (raw counts = 9, normalised counts = 0.209).

Table 6.21 The raw and normalised coding frequency counts for the mass-force interaction when mass is used as a cause. The normalised counts are the fraction of the total mass cause coding in each cohort.

		Physics (n=6)		Psychology (n=5)	
Type of interaction		Raw Total Count	Normalised Counts	Raw Total Count	Normalised Counts
Force	Weight	17	0.395	3	0.600
	Newton's 2nd Law	9	0.209	0	0.000
	Air resistance	3	0.070	0	0.000
	Acceleration due to gravity	3	0.070	0	0.000
	Newton's 1st Law	1	0.023	0	0.000
	Impulse	1	0.023	0	0.000
	Momentum	2	0.047	0	0.000
	Influences other forces	1	0.023	0	0.000
	Force has greater effect	2	0.047	0	0.000
Other Interaction	Energy	3	0.070	0	0.000
	Gravity	0	0.000	2	0.400
No interaction		1	0.023	0	0.000
Total		43		5	

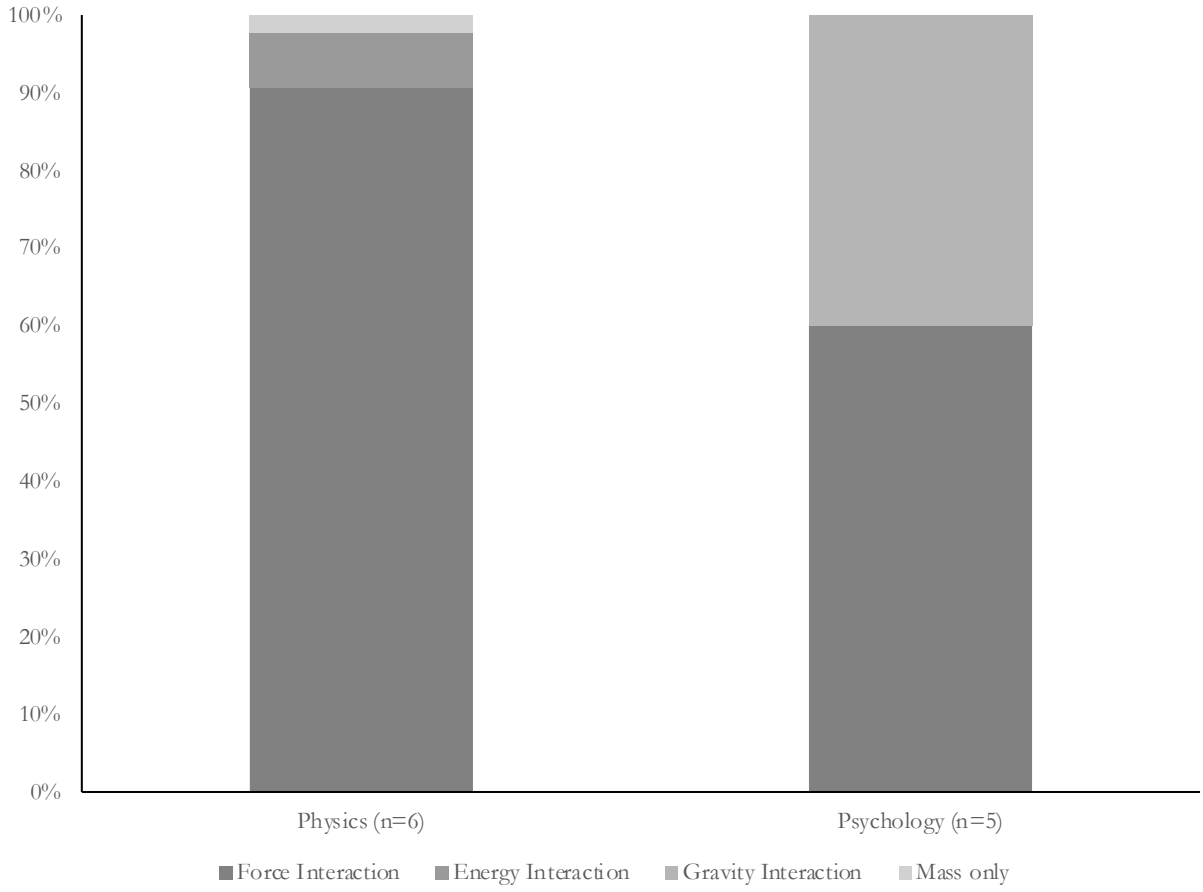


Figure 6.6 The percent mass codes for each cohort that interact with either force, energy, or gravity, or have no interaction (mass only).

6.3 Holistic Analysis of Explanations

The content and LG analysis of participants' interviews has enabled a quantitative analysis of the similarities and systematic differences in how hub-words are used by physicists and psychologists to describe and explain the motion of the objects in the demonstrations. However, the use of coding in this type of analysis can risk the loss of a more holistic picture of the data by fragmenting the dialogue into small segments (Cohen et al., 2018). This criticism could be applied to this analysis because codes are generated by categorising instances of hub-word use within a limited context (i.e., a maximum of a clause nexus). This section aims to address this criticism by presenting themes within the interview data that are noticeable when a more holistic 'whole interview' approach is taken to analysing participants' description and explanations of the motion of the objects in the demonstrations. These themes were developed from research notes taken by the researcher when checking the interview transcriptions and subsequently re-reading the interviews when coding. Each theme is supported by vignettes from the interviews.

6.3.1 Comparing Participants' Interview Responses with the FCI scores and Expected Explanations

The participants' FCI scores suggest that the physicists solve problems in a way that is consistent with having a Newtonian understanding of mechanics. In contrast, the psychologists apply reasoning that is incompatible with an understanding of mechanics that is based on Newton's Laws. The interview data are consistent with this interpretation of the FCI scores. For instance, in the excerpt below, 12S05 explains that a film-canister rocket is stationary on the desk because gravity is causing it to stay and there are no other forces acting, rather than in terms of the absence of a resultant force because the weight of the canister and the normal reaction force from the desk are equal and opposite.

> **12S05:** It's not moving.
>
> **RT:** It's stationary. It's not moving. So, can you explain why it is stationary?
>
> **12S05:** Because no forces are acting on it.

Moreover, psychologists often explain the motion of the film canister rocket in terms of a force or pressure overcoming gravity, until gravity 'takes over' again causes the rocket to fall to the ground. These explanations also invoke the idea that the force or pressure remains after the film canister rocket 'pops' and 'runs out' as the rocket moves upwards. Gravity takes over when either the upward force or pressure has 'run out'.

> **12S01:** The pressure hasn't made it pop yet because there's nothing, like a force, pushing it or moving it. Gravity might stay the same, like the force going up, it's going to be bigger. So, it causes it to pop up.
>
> **12S01:** In the initial pop, it's going to have the most force exerted on it and going to go the quickest and eventually, gravity is going to want to pull it back down when the force runs out.

This notion of one force 'winning' over another force is also present in psychologists' explanations of why a scrunched-up cup-cake case will fall faster than an unaltered cup-cake case.

> **12S01:** So maybe the area [of the cupcake case] should be smaller. So, the force going up is like decreased. So, then gravity is obviously more powerful, like plays a bigger part.

Physics participants can begin their explanations of an object's motion with statements that are more consistent with the everyday explanations the psychologist give. For instance, in this extract 12P03 initially explains the motion of the film canister rocket when it pops in terms of a pressure build-up and the canister 'shooting up'.

> **12P03:** And since the canister is contained, the gas cannot escape. So, you get a pressure build up in there. A then when the pressure reaches a certain amount, it blows the lid, pushes down, you get a reaction force, and it shoots up.

However, they then refine and expand their explanations using knowledge that is consistent with Newton's laws. In all cases, their explanations closely resemble the 'ideal' explanations written by physics teachers referring to the course textbooks (see the two examples in Table 5.2 and Appendix 10.6).

Table 6.22 Two examples of the explanation given by physicists that are consistent with the 'ideal' explanations written by physics teachers referring to the course textbooks.

Explanation	Interview Extract
Explanation of the motion of the film canister in the air after it has 'popped'.	**12P05:** The force only going up only acts in one instance, which is right at the start where it separates. And then after that, the only force acting vertically, that well, if, if air resistance is negligible, the only force acting vertically is the weight, so therefore you'd have a negative acceleration.
Explanation of why the falling cupcake case reaches a terminal velocity.	**12P08:** So as, when it starts move, it's going to start with a certain velocity and when the, so its weight is moving it downwards with, well, if it didn't have an air resistance force, it would be g, 9.81, that it's acceleration downward is. It's going to accelerate, while the weight force is less than the air resistance force. But gradually, as it gets faster, the air resistance force is going to get greater. And I don't know if it did in that case, but it might reach terminal velocity, which is when it's, the weight force downwards is equal to the air resistance force upwards. And so, it just stops accelerating and stays at that velocity until it reaches the surface.

6.3.2 Participants' Descriptions of Motion

Physics participants are confident in their description of the motion of the film canister rocket and the cup-cake case. Moreover, they describe the motion using the accepted physical definitions of velocity and acceleration. In contrast, several psychology participants are hesitant or unsure about the motion of the film-canister rocket or the cup-cake case. Moreover, when questioned, they are often not able to correctly identify when an object is accelerating, and they are unsure about the definitions of acceleration, velocity, and speed. Table 6.23 presents the interview vignettes that illustrate these observations.

Table 6.23 Vignettes from the psychologist interview that demonstrate the issue they have with using the concepts of speed, velocity, and acceleration to describe motion.

Motion Concept	Context	Misconception	Interview Vignette
Speed	12S05 is asked to justify what moving 'quicker' means when they refer to a cupcake case with paperclips attached to it being dropped and compared with a cupcake case with no paperclips attached.	Speed is described in terms of faster, rather than in terms of distance travelled in a unit time.	**12S05:** Yeah. So obviously that fell quicker and it's going to weigh more than the other one. **RT:** When you say something falls quicker, how do you know it falls quicker? **12S05:** It landed before the other one did. **RT:** Yeah. So, it is really what I'm getting at is what? What speed? **12S05:** Oh, quicker than the other one. **RT:** Yeah. When you say something is quicker. **12S05:** Yeah, fast. **RT:** What do we mean by fast? **12S05:** So, it moved. Um. Faster than the one.
Velocity	12S01 is asked what the velocity of object measures.	Velocity is distance.	**RT:** What is velocity? **12S01:** Isn't that, I think, it's like the distance.
Acceleration	12S01 is asked whether the cupcake case will accelerate when it hits the desks and stops.	Objects do not accelerate when they 'stop' (i.e., their speed decreases).	**RT:** What happens to the way it moves when it hits the desk? **12S01:** Um, it just is stationary, and, like, it doesn't move anymore. It just stops. **RT:** Right. It stops. So, what happens to speed when he stops? **12S01:** Goes to zero. Like it just stops. **RT:** Okay. Yeah. Does it accelerate then when it stops when it stops and hits the desk? **12S01:** No. **RT:** No. Why not? **12S01:** Um, because I feel like acceleration, you have to be moving to accelerate. And like obviously, when you're stopping, you can accelerate or decelerate because you're just there.

Acceleration	12S09 is describing the motion of the film canister rocket after it has 'popped' and is moving into the air.	Objects cannot not accelerate if they are moving up or down. They can only accelerate in a horizontal direction.	**RT:** OK. Right? So, when you say it goes faster, yeah. Does it speed change? **12S09:** What from when it's stationary? **RT:** Yeah. **12S09:** Yeah, it does. Because obviously it's not moving to start with. **RT:** Yeah. And then it gets faster. **12S09:** Yeah. **RT:** Okay. What about here? Is it accelerating from this point to this point? **12S09:** Umm, no. umm, yeah. No, I don't, because it's gone up into the air, but it's not really going, it's not like going in a direction. It's not like going left or right. But I guess, like, nothing's like being making it move, nothing is pushing it or anything. **RT:** OK, so nothing's pushing it what... **12S09:** it's pushing it up into the air, but it's not pushing it in, like, left, or right? **RT:** Right. Okay. Yeah. So, we can only accelerate if we're going left or right. **12S09:** That's how I view it. **RT:** Okay. **12S09:** I don't view going up in the air as accelerating. **RT:** Yeah, so go on then, tell me what you think accelerating is. If I said to you, this thing is accelerating, what would that mean? **12S09:** Something has to move in a direction that, that's not up or down.

There is also some evidence that psychologists' inability to describe motion correctly causes incoherence in their explanations. For example, in the vignette below, 12S09 associates pressure with causing the film canister rocket to move. They also suggest this pressure causes the rocket to move '*faster as it goes up'*. From this explanation, they imply that the rocket must have reached its maximum speed at its maximum height. However, they then recognise that this cannot be true. This inability to describe motion correctly could make it difficult for these students to correctly connect the forces acting on an object with its motion because their descriptions of the motion do not seem to be coherent.

RT: *OK, so it goes faster as it goes up.*

12S09*: Yeah.*

RT: *Why does it go faster as it goes up?*

12S09: *Because the pressure is causing it to move.*

RT: *OK. Yeah. So, the pressure again, the pressure is causing, it to move and it goes faster as it goes up.*

12S09: *Yeah*

RT: *Okay. How is it moving at this point here at the top?*

12S09: *Well, it doesn't really stay there for a long amount of time. It, just kind of, reached its peak speed or whatever and then comes back down.*

RT: *So, it reaches a peak speed up here?*

12S09: *Yeah.*

RT: *Okay. And then once, so it is going fastest at the top?*

12S09: *No, it's not, because that's, like it gets to the fastest bit and then it comes back down. It's not moving fastest as when it's going round the bit there if that makes sense.*

6.4 Summary

This chapter has presented the results of the analysis of whether participants use hub-words when they explain how an object moves, and how they use these words. Therefore, these results address research question 3. These results are summarised below.

- Physicists and psychologists frequently use the hub-word *force* in their explanations. However, *force* is used more often by the physicists and this difference is statistically significant. Moreover, *force* is highest ranking word used by physicists. *Force* also ranks highly amongst the word used by psychologists. However, *pressure* is ranked highest above *force*. The hub-word *mass* is also used frequently by the physicists but not by the psychologists. Again, this difference is significant. There are also significant differences in the participants' use of other words. Physicists frequently use *acceleration, weight,* and *velocity* in their explanations. However, *speed* and *gravity* are used more frequently by psychologists.
- The hub-word *force* acts as a cause of a process in physicists' and psychologists' explanations. However, the processes physicists associated with being caused by *force* are typically *acceleration* or changes in *velocity*. In contrast, psychologists identify *force* as causing more generally processes associated with motion (e.g., falling).
- The hub-word *mass* also acts as a cause in physicists' explanations. However, *mass* typically acts as a cause in association with *force*. Again, physicists associate mass and force as causing processes that involve *acceleration* or changes in *velocity*. When psychologists use *mass*, they do identify this hub-word as a cause, but again as a cause of processes that describe more general types of motion.
- Physicists' explanations are typically consistent with a Newtonian understanding of mechanics. Whereas psychologists' explanations are consistent with an everyday understanding. Furthermore, psychologists seem to confuse key concepts that physicists use to give differentiated descriptions of motion (e.g., acceleration, velocity).

7 Discussion

Learning science requires conceptual change (see Amin and Levrini (2017a) and section 2.1). Therefore, improving pedagogy requires knowledge of the process of conceptual change and, perhaps more importantly, the mechanism that drives conceptual change. Susan Carey's Theory of the Origin of Concepts (TOOC) (Carey, 2009) has been adopted in this thesis as the theoretical framework that provides the most complete account of conceptual change, both in terms of how concepts develop and the mechanism that drives this development (Amin, 2017). Carey frames the TOOC in terms of three theses that address the basic questions any theory of conceptual development should answer. These are:

1. The innate conceptual repertoire cannot and does not solely consist of sensory/perceptual representations. It contains core concepts, such as *object*, *number*, *agent*, and *cause*.
2. Humans can build conceptual representations that are more powerful than antecedent representation (e.g., core concepts) or intuitive theories that are formed using new concepts that are incommensurate with antecedent concepts.
3. Quinian bootstrapping is the mechanism that drives conceptual development. It is a learning process that does not rely on hypothesis testing.

The Quinian bootstrapping process as a driver of conceptual change is the central focus of this thesis and of this discussion. The bootstrapping process begins when an individual connects a mental representation with a newly coined or learned external symbol or sign, usually a word. Initially, the mental representation associated with the new sign is far from a fully developed concept. The sign is initially a 'placeholder', whose meaning is solely defined by its set of connections or interrelations between other signs (Carey, 2009, 2011). This set of interrelations in the placeholder sign structure guide modelling activities (e.g., analogy monitoring and construction, thought experiments and inductive inference) that allow an individual to develop a full understanding of the concept the placeholder sign represents.

Carey uses the example of how young children learn to represent the natural number series to explain how Quinian bootstrapping drives conceptual development. This explanation draws upon a detailed empirical knowledge of children's core cognition of number (CS1) and the final integer concept of number (CS2). However, the steps that a child needs to take to move from CS1 to CS2 are identified by Carey using knowledge of the initial placeholder or sign structure that guides

the learning of CS2. Moreover, Carey uses knowledge of these steps to marshal empirical evidence that supports her explanation of development of CS2 from CS1.

Carey also argues that Quinian bootstrapping is the mechanism that drives the development of concepts in initiative theories. However, Carey's evidence for this is based on historical argument and analysis of the efficacy of teaching sequences (e.g., Smith et al. (1985), Smith (2007) and see section 2.5.3.1.2). She does not provide the direct evidence for this in terms of the placeholder sign structure that needs to develop, as was the case in her detailed explanation of the development of the concept of a natural number series. Hence, this thesis aimed to examine the placeholder sign structure, which develops when 17-18 year-old students develop their intuitive theory of mechanics and use this knowledge to develop a more detailed and evidenced picture of the Quinian bootstrapping mechanism, which drives the development of the concepts associated with this intuitive theory. The research questions were designed to capture this aim, and were:

1. What are the sign networks of individuals within a scientific domain of knowledge?

2. What are the similarities and differences between the networks for individuals who have different levels of scientific understanding (novices and experts)?

    a. Are hub-words responsible for the majority of connections in the networks?

    b. If so, are the hub-words the same for different individuals?

3. Do these individuals use hub-words from their networks when they reason about a scientific experiment or demonstration? If so, how?

The following discussion brings together the findings from answering these research questions as a series of claims. These claims highlight important features of the placeholder sign structure but also integrate these features with findings about participants' conceptual understanding. In doing so, this discussion attempts to develop Carey's ideas by providing a more detailed account of how placeholder sign structure guides conceptual development in intuitive theories. Moreover, by examining the relationship between placeholder structure and conceptual understanding, this discussion highlights aspects of these findings that provide empirical support for the TOOC.

7.1 Claim 1: The Physicists and Psychologists have a Different Conceptual Understanding of Mechanics

Two methods were used to investigate the participants' conceptual understanding of mechanics. These were the force conceptual index (FCI) (Hestenes et al., 1992) and interviews, where participants were asked to explain what they had seen during a demonstration of physical phenomena (i.e., the motion of a film canister rocket). Both methods demonstrated that the physicists and psychologists had a different conceptual understanding of mechanics.

The force conceptual inventory (FCI) is widely accepted as a reliable and valid measure of whether an individual solves problems in a way that is consistent with expert physicists, with a Newtonian understanding of mechanics, or whether they solve problems using an everyday understanding of mechanics (Savinainen & Viiri, 2008). Hence, an FCI score is an indication of whether an individual has an everyday (CS1) or Newtonian (CS2) conceptual system (Hestenes et al., 1992). The FCI scores of the physicists indicated they had a Newtonian conceptual system (CS2), whilst the psychologists scores were consistent with having an everyday conceptual system (CS1).

The FCI is designed to be easily administered to a large number of participants (Hestenes et al., 1992). To achieve this aim, it uses a multiple-choice question format. However, this multiple-choice format introduces an important limitation of the FCI because this type of question does not directly capture the reasoning that a participant uses to arrive at their chosen solution for each item. Instead, for each item, a participant is forced to choose a solution to a problem that that is consistent with the participant having either an everyday or Newtonian conceptual system. Hence to design multiple-choice questions in the FCI, the authors had to make assumptions about the nature of CS1 and CS2 (Hestenes et al., 1992). These assumptions were based on research into the conceptual systems in mechanics that are held by novice and experts. However, the FCI is still based on a priori assumptions about the nature of the conceptual systems individuals possess. Therefore, any inferences about the nature of the conceptual system of a participant taking the FCI is based on these assumptions. The validity of these assumptions has been questioned (e.g., Huffman & Heller, 1995). Hence, interviews in which participants were asked to explain demonstrations of physical phenomena using their understanding of mechanics were used to explore participants' conceptual understanding more freely.

The interview dialogue was analysed using content and lexiogrammatical analysis (LG) using systemic functional linguistics (SFL) (see section 5.2). SFL provides a robust method for analysing how language is used to perform certain functions. SFL characterises language as a tool that uses

semiotic and syntactic rules to perform certain functions, such as describing and explaining experiences (ideational meta-function). In the Representational Theory of Mind (A. Clark, 2014), describing and explaining is also achieved through computational processes on concepts that follow syntactic rules. Hence, the language spoken by students in interview is taken to be a proxy for their thinking about the demonstrations, then SFL can be used to systematically analyse how the hub-words, and any concepts associated with these words, function in their explanations.

The findings of this LG analysis of the interview transcripts confirm the FCI findings. However, the LG analysis also adds detail to how the psychologists' and physicists' conceptual understanding differs. The discussion of these differences is presented in section 7.4.1. However, the claim that physicists and psychologists have different conceptual understanding is highlighted at this point to emphasise that when any similarities and differences in the sign networks of these two groups are discussed, these are differences between two groups with different conceptual understanding. Moreover, moving from CS1 to CS2 defines a transition between what could be seen as a novice to an expert conceptual understanding of mechanics. Hence, whilst the two groups are independent, comparing the psychology and physics groups can be seen as a proxy for a trajectory in conceptual development in mechanics. This is broadly consistent with the academic status of the two groups, with the physics group containing students studying A-level Physics, whilst the psychology group contains students who ended their formal study of Physics with the completion of their GCSEs.

7.2 Claim 2: The Networks have Similar Macrostructure but there are Differences in the Microstructure

The initial analysis of the sign network employed widely used basic network parameters to characterise the structure of the networks and identify similarities and differences between them (Barabási & Pósfai, 2016; Newman, 2018). These parameters were the network order, density, mean degree, diameter, average shortest path length, and clustering coefficient. There were no significant differences between in the basic network parameters for the physicist and psychologist sign networks, despite the differences in their conceptual understanding. Furthermore, these parameters are consistent with the same parameters that describe the structure of networks that have been created using looser requirements for identifying and connecting signs (i.e., free word associations), and for entire lexicons for both individuals (Morais et al., 2013) and groups of speakers (i.e., native English or Dutch speakers) (De Deyne & Storms, 2008a; Steyvers & Tenenbaum, 2005). The similarity in the basic structure of the sign networks made for a specific domain of knowledge and a whole vocabulary, which would include many domains of knowledge, suggests that the insights into conceptual development, which are outlined in this discussion by analysing sign structure in mechanics, could also be applicable more widely as general principles of conceptual learning. This is encouraging because the TOOC is a general explanation of how concepts develop, rather than a theory that applies specifically to mechanics.

The basic network parameters indicate that the physicist and psychologist sign networks share several key structural properties. This is what is meant by the first part of claim 2 that states that both groups share macrostructure. However, the second part of claim 2 states that the microstructure of physicists' and psychologists' sign networks does have differences. This refers to the differences in the individual vertices or signs in the network, and the connections between them. The key differences in this microstructure are shown by the aggregate placeholder sign networks (see Figure 7.1 and Figure 7.2, which are reproduced below for ease of reading).

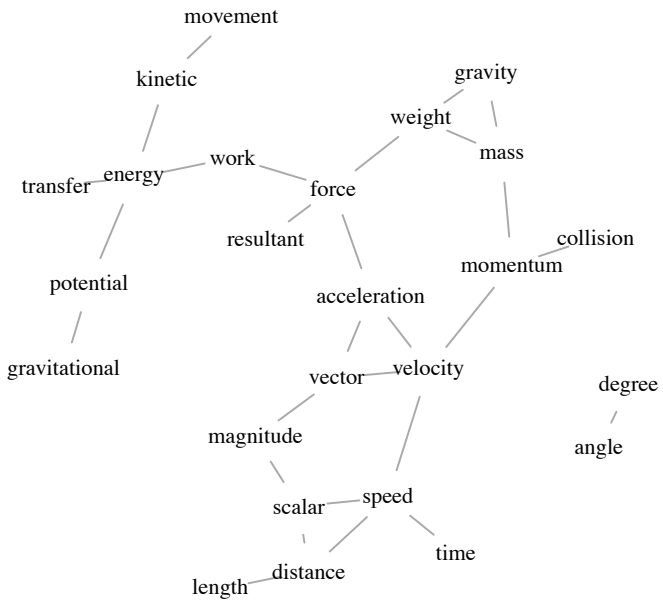


Figure 7.1 An aggregate sign network for the physics cohort. The edges in the network are weighted according to the number of participants that made the connection between two signs. This network shows the connections between signs made by 5 out of the 6 physics participants.

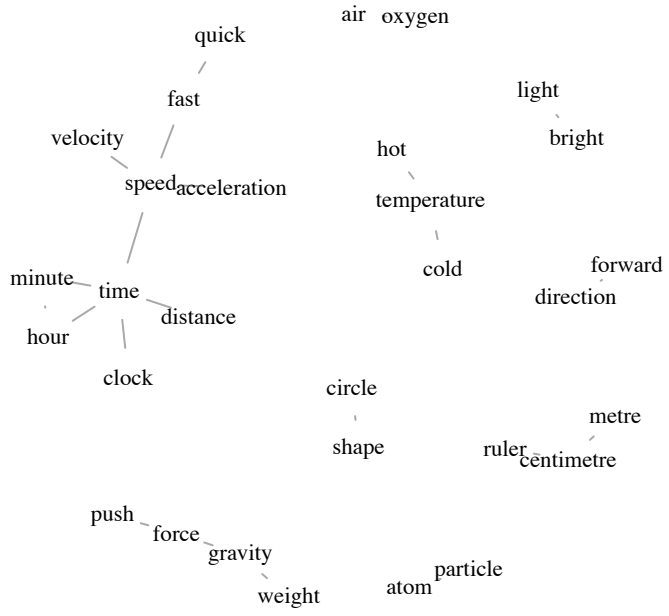


Figure 7.2 An aggregate sign network for the psychology cohort. The edges in the network are weighted according to the number of participants that made the connection between two signs. This network shows the connections between signs made by 5 out of the 6 psychology participants.

The aggregate psychology network consists of several components that are unconnected. In contrast, the aggregate physics network contains of a single large, connected component and small sub-component (*angle* | *degree*). This result would appear to contradict the finding that the physicists' and psychologists' individual sign networks have a similar macrostructure. However, the aggregate networks show shared edges between signs (shared by 5 out of the 6 physics participants and 4 out of the 5 psychology participants). Therefore, these apparently contradictory findings suggest that the physicists' sign network contain more shared microstructure (common vertices and edges) than the psychologists' networks, which contains greater variation in connections between signs. For instance, most physicists connect *force* with *resultant* and *acceleration*, whilst most psychologists do not. This would suggest that some common connections between signs need to be made to develop a Newtonian conceptual understanding (CS2) of mechanics. This is consistent with the Quinian bootstrapping process in TOOC.

Despite the differences in the microstructure in the aggregate sign networks, there are also some similar structures. For example, both physicists and psychologists frequently connect *speed* to *time* and *velocity*. This could suggest that placeholder sign structure precedes conceptual understanding, which is an important prediction made by TOOC. This possibility is discussed in greater detail in section 7.4.1.

The following sections discuss the implications of the shared features of the macrostructure and the difference in the microstructure in terms of how the Quinian bootstrapping process could drive conceptual development.

7.2.1 The Networks are Approximately the Same Size

The networks in both groups have approximately the same number of signs and edges, despite the groups having different conceptual understanding (claim 1), although there is some variation within both groups. It might be expected that the physicists would have larger networks, both in terms of size and order, if they have a more developed conceptual understanding of mechanics. However, both groups have similar general vocabularies (see section 4.2.2). Hence, the similarity in the size and order of the networks for physicists and psychologists might suggest that the snowball sampling technique is moving beyond the bounds of the mechanics domain of knowledge for the psychologists and into general vocabulary. This is consistent with some of the signs and connection between in the aggregate psychologist network (e.g., *circle* | *shape*) (see Figure 7.2). This interpretation implies that the growth of a specifically mechanics sign structure takes place within the context of general vocabulary sign structure.

7.2.2 The Networks are Sparse and have a Low Average Degree

The physicists' and psychologists' sign networks are sparse, which indicates most signs only connect to a small fraction of the total number of available connections (Newman, 2018). This structural property is also reflected by the average degree of the networks, which ranges between approximately 3 to 8. This finding suggests that the sign structure around most words is quite low. Hence, according to the TOOC and the Quinian bootstrapping process, the sign structure needed to develop lexical concepts (roughly associated with one sign) is minimal. This is illustrated by the aggregate sign networks. Whilst it should be noted that in aggregate networks each sign is likely to contain less connections than in an individual network, it is salient that concepts such as *weight* seem to be defined by connections to just *mass* and *gravity*, and kinetic with connections to *energy* and *movement*.

7.2.3 The Networks have Short Average Path Lengths

Despite the low number of connections per sign, all the sign networks have short average path lengths of typically six. Again, this is a very typical property of both sign and other networks (Morais et al., 2013; Newman, 2018; Steyvers & Tenenbaum, 2005). These short path lengths imply that any sign within the network can be reached from any other sign by on average a path length of 6 (i.e., 6 edges or 'jumps' between signs). This may imply whilst a minimal immediate sign structure is required to develop conceptual understanding (i.e., typically approximately three connections), it is important that signs are not isolated within the network and many signs can connect, and therefore influence each other, within a short number of edges. Again, the aggregate networks seem to illustrate this point. For instance, *energy* and *force* are connected via *work*.

Clustering is the fourth structural property that is implied by the basic network parameters. This is an important indicator for models of network growth. Therefore, clustering is discussed in greater depth in the next section.

7.2.4 The Significance of High Clustering Coefficients

Clustering in networks, measured using clustering coefficients, is a 'signature' structural property that can indicate that connections in a network are not being made randomly (Barabási & Pósfai, 2016; Newman, 2018). In the sign networks for all the physicists and psychologists, the clustering coefficients are higher than would be expected for a random network of the same size. Therefore, the clustering in the sign networks suggests that the connections between signs are not random,

and hence there is an underlying mechanism that determines how connections are made between signs (Newman, 2018; Steyvers & Tenenbaum, 2005).

The clustering coefficients in the sign networks are consistent with the clustering coefficients measured by Steyvers and Tenenbaum (2005) for their networks and, hence, could be explained by the model Steyvers and Tenenbaum (2005) develop that can account for their results. This model defines how a sign network grows in terms of three principles that govern how a new vertex entering a network connect to existing vertices (see section 3.3.4). These principles are:

1. When a new sign enters a network, it can form edges with several existing signs. However, after the first edge is formed between the new sign and an initial sign, any remaining edges must connect to signs that are within the neighbourhood of the initial sign.
2. The probability of a new sign forming an edge to an existing sign is proportional to the vertex degree.
3. A utility variable that is not necessary to produce any of the structural features of the network. However, it modulates the probability that a sign will be the target of any new edges being made in the network. This utility variable allows the model to be used to explore interactions between the structure of the network and other important properties of the signs within the network, such as usage frequency and age of acquisition (i.e., how early a word enters the network).

Figure 7.3 (initially introduced in section 2.6.2.3) shows how principles 1 and 2 operate to grow a sign network.

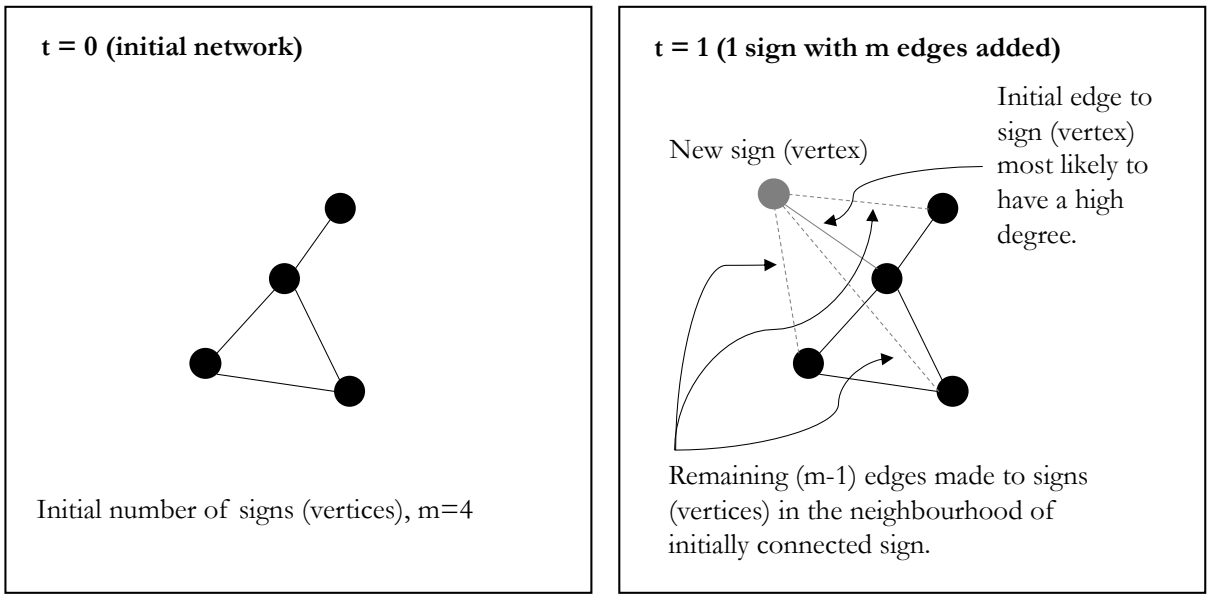


Figure 7.3 The preferential attachment model of sign network growth developed by Steyvers and Tenenbaum (2005).

In this discussion, the Steyvers and Tenenbaum (2005) model is used an analytical tool to understand how placeholder structure develops within a sign network. For instance, the first principle could explain the high clustering coefficients that are a feature of both the physicists' and psychologists' sign networks. This principle suggests that any new sign, which is initially associated with a placeholder value only rather than a fully formed concept, enters a network by connecting not only connecting to one existing sign, but also several of its neighbours. Hence, a new sign enters a sign network by forming a cluster. Hence, a clustering would appear to be an important feature of the sign placeholder structure that guides conceptual development.

Moreover, Carey suggests that the initial content of any placeholder value associated with a sign is only determined by it connections with the placeholder structure (Carey, 2009). Hence, the formation of a cluster when a sign first enters a network suggests that the initial content of a placeholder is determined by a small number of connections to closely related signs (and their associated concepts), rather than a single connection or many connections. Furthermore, the new connections that made to the existing signs in the cluster are likely to also modify the concepts that are associated with these signs (Steyvers & Tenenbaum, 2005). Therefore, this clustering within the sign networks is consistent with Carey's argument that the development of concepts within an intuitive theory involves simultaneous and coordinated changes in several closely related concepts. Again, this feature of placeholder structure and its role in conceptual development is illustrated by the clusters present in the merged physics networks. For example, it appears that *mass*, *weight*, and *gravity* form a cluster. The possible significance of this cluster is discussed in more depth in section 7.5. Similarly, there is an *acceleration*, *vector*, and *velocity* cluster and a *speed*, *scalar*, and *distance* cluster. It seems plausible that the entry of the sign *vector* into the *velocity/acceleration* cluster or the sign *scaler* into the *speed/distance* cluster could led to simultaneous changes in the concepts associated with each of these signs.

7.2.5 Comparison of Sign Networks with Concept Networks

In section 2.6.2 of the literature review a distinction was made between what could be called a sign and a concept network. This distinction was based on the criteria for drawing edges between vertices in the networks. Concepts maps require a high degree of declarative knowledge (e.g., knowledge of experimental or modelling procedures (Koponen & Pehkonen, 2009)) for two vertices to be connect. Hence, these networks are likely to map an individual conceptual understanding. In contrast, the free word associations used to construct the networks measured by De Deyne and Storms (2008a); Morais et al. (2013); Steyvers and Tenenbaum (2005) do not require any level of declarative knowledge beyond simply associating one word with another. This

study captured the connections between signs by asking participants to carefully consider words that they viewed helped them understand a stimulus word (e.g., *force*). However, participants did not have to give reasons why they listed any response words. Hence, this technique that sits in between the minimal declarative knowledge required for word association and the high level of declarative knowledge required for concept maps.

The process of Quinian bootstrapping, which drives conceptual development in TOOC, states that placeholder sign structure precedes the development of conceptual understanding, and therefore conceptual structure. Hence, in using relative loose requirements for drawing edges between signs, the methodological approach to capturing connection between signs in this study was designed to measure the placeholder sign structure that guides the development of conceptual structure. Therefore, it is encouraging that the low average degree (~3) and the clustering coefficients (~0.2) found in the sign networks are consistent with the values of these parameters measured by Koponen and Nousiainen (2018b) for individuals' concept networks in electricity and magnetism. The similarity in the structure of in both the placeholder sign network and concept networks is consistent with proposal that the structure of the sign network guides the development of any conceptual structure, which forms intuitive theories. Moreover, the commonality of the structure in the networks across different domains of knowledge (mechanics, electricity, and magnetism) suggests Quinian bootstrapping could be a general mechanism that drives conceptual development, as Carey states in TOOC (Carey, 2009). This idea would seem to be supported by some of the similarities between the Steyvers and Tenenbaum (2005) model and the model developed by Koponen and Nousiainen (2018b) to explain growth of electricity and magnetism concept networks. Both models use the principle that a vertex entering a network is likely to connect to neighbours of the original vertex it attaches to, to form clusters.

7.2.6 Summary

This claim stresses that the signs network of physics and psychologist sign networks are of a comparable size, order and have a complex but similar macrostructure. This claim is markedly different from the picture of complex, highly connected expert networks versus unconnected novice network with 'island' of simple structure suggested by early studies of sign networks (Cachapuz & Maskill, 1987, 1989). Moreover, the existence of antecedent sign structure is a subtle difference to how the placeholder structure of the count sequence (i.e., *one*, *two*, *three*…) guides the development of the concept of the natural number sequence. In the case of natural number, the count sequence (*one*, *two*, *three*) is presumably the first network of connections between the signs *one*, *two*, *three* etc… that infants learn. However, these findings suggest this is not the case for some

intuitive theories (e.g., mechanics). This claim implies that the growth of new placeholder structure that guides the development of new intuitive theories takes place within an already complex network of association between signs. This is equivalent to stating that any new sign, or word, that is learnt must find its place in relation to any existing vocabulary.

The sparseness and low average degree suggest that any new sign entering these complex sign network only to a low number (typically less than 3) of existing signs. Hence, the placeholder structure is economical in the sense that only a few immediate connections are required to guide the development of a new concept. However, the short path lengths in the networks suggest easy connections to less immediate sign structure is also important. Furthermore, any new sign entering the network is likely to form a cluster with its immediate neighbours. Hence, the initial content of a placeholder value associated with a new sign is typically determined by its connections to more than one immediate neighbour, with presumably any concepts associated with these signs also being modified by being connected to a new sign. Therefore, the macro structure of the sign network suggests that small but connected cluster of signs, with any associated concepts developing simultaneously, characterise the placeholder structure that guides the development of intuitive theories. This is consistent with the similarities and differences identified in the aggregate placeholder sign networks.

7.2.7 Implications for Teaching

Claim 2 emphasises that any new sign entering a sign network forms a cluster with existing signs, which are themselves are located within a complex network of signs that define an individual's vocabulary. Moreover, it could be that only a small number of connections between signs guide conceptual development. In terms of teaching, this implies that any new sign or word that is introduced to students is likely to initially derive meaning from their understanding of its immediate neighbours. Moreover, the introduction of this sign is likely to alter students' conceptual understanding of these neighbouring signs. For instance, introducing the term *vector* may change students' understanding of *acceleration* and *velocity*. However, their initial understanding of *vector* will also be influenced by their understanding of *acceleration* and *velocity*. There are two implications for teaching that follow from these observations.

First, these observations stress the importance of teachers having some initial understanding of students' understanding of the sign being introduced to them. It also highlights how any initial understanding can interfere with the development of any concept associated with a sign. For example, Jasien (2010, 2011) and Song and Carheden (2014) find that concepts that are associated

with signs or words that are used in everyday dialogue, rather than just scientific registers, can be harder for students to learn. Moreover, these observations also highlight the possible influence of what is sometimes referred to as Tier 2, rather than Tier 3 vocabulary (Wellington & Osborne, 2001). Tier 2 vocabulary is defined as vocabulary that is not part of a scientific register of terms but is associated with a more adult vocabulary or less frequent use (e.g., increase rather than bigger). It is likely that a sign entering a network will connect with Tier 2 and Tier 3 vocabulary (e.g., *acceleration* with *rate*). Hence, it is important that students have a sound understanding of this vocabulary. Unfortunately, Cassels and Johnstone (1985) have demonstrated that this might not be the case. Therefore, these findings suggest that general vocabulary development could also play a role in conceptual development.

Second, the simultaneous and coordinated changes of concepts implied by the clusters in the sign networks suggest that word (or signs) that are likely to connected need to be carefully considered when introducing students to a new sign and its associated concept. For instance, when teaching *acceleration*, students' understanding of *velocity*, *vector* and, perhaps, *rate* also needs to be simultaneous developed.

Both of the above implications are consistent with the language focused pedagogical approach used by experienced secondary science teachers in the UK and Brazil describe in a study by Mortimer and Scott (2003). In this approach, teachers use different patterns of language use, defined by different types of student/teacher interaction, to first explore, then work on and then consolidate students' understanding of scientific terms and any associated terminology.

7.3 Claim 3: The Sign Networks Contain Hub-Words

Like clustering coefficients, hub-words are a 'signature' property of networks that suggests connections are not being made randomly (Barabási & Pósfai, 2016; Newman, 2018). Moreover, if a network contains hub-words, they explain may of the structural properties of the network. For instance, the short paths length that are observed in the both the physicists' and psychologists' networks (Barabási & Pósfai, 2016). Hence, an important aspect of the network analysis was to determine that the networks had hub-words. The presence of hub-words in sign networks was determined by analysing the degree distributions. The results of this analysis are different for the in and out degree distribution. Therefore, the findings for these distributions are discussed separately below.

7.3.1 In-Degree Distributions

The analysis of the in-degree distribution demonstrates that a power-law distribution is a significantly better fit than a Poisson distribution. This result is significant because the degree distribution would follow a Poisson distribution if the connections between signs in the network were being made randomly. Hence, the rejection of the Poisson distribution fit to the in-degree is a 'signature' result that suggests there is an underlying mechanism that is responsible for determining how incoming connection are made to the signs in the network. Throughout this thesis, the model developed by Steyvers and Tenenbaum (2005) has been presented as a tool that can be used to understand and experiment with the mechanism that determine how connected are made in a sign network. This model specifically predicts that the degree distribution follows a power-law and, hence, has a scale-free structure. However, the analysis of networks suggests that other distributions (e.g., log-normal, exponential, power-law with exponential cut-off) cannot be ruled out as appropriate fits to the in-degree distributions. This weakens any claim that the sign networks have a scale-free structure. However, all these distributions are heavy tailed. Hence, the claim that the sign networks have hub-words in the in-direction is robust. This is consistent with the findings of Morais et al. (2013) who also find that the in-degree distributions for the sign networks they create follow heavy-tailed distributions and Poison distributions can be ruled out. However, often their analysis cannot distinguish between different types of heavy-tailed distributions (e.g., power-law versus exponential), but does often suggest that a power-law with an exponential cut-off is a better fit than a pure power-law.

Despite the fact that it is not certain in-degree distribution follows a power-law, the Steyvers and Tenenbaum (2005) model remains an important analytical tool in this discussion because it

indicates how the signature properties of high clustering coefficient and the presence of hub-words in the sign networks can be explained (see section 7.2.4). The first principle of the model explains the clustering the networks. The second principle explains the presence of hub-words. This principle assumes that any new sign entering a network is initially more likely to connect to a sign with a high degree. Hence, this principle is based on preferential attachment. Therefore, these findings suggest sign networks of the physicists and psychologists are essentially growing around these highly connected hub-words.

However, the fact that the Steyvers and Tenenbaum (2005) model predicts a power-law degree distribution, and other heavy-tailed distributions cannot be ruled out as appropriate fits to the data, is a caveat that needs to be kept in mind. This caveat indicates that the Steyvers and Tenenbaum (2005) model is not a complete description of how connections in the sign networks are formed. For instance, in the case of a power-law with exponential cut-off, the cut-off could be explained if there is a factor that limits the number of connections that can be made to hub-words (Morais et al., 2013).

7.3.2 Out-Degree Distributions

In contrast with the in-degree distributions, a power-law is not a statistically better fit to the out-degree distributions than a Poisson distribution. This finding needs to be viewed with some degree of caution because the process that is used to 'finish' the sign networks after the final iteration of snowball sampling lowers the out-degree of any vertices in the final iteration. This will possibly reduce the number of vertices with high degrees. Hence, the tail of the out-degree distribution may be distorted. This could explain why the analysis of the out-degree distribution cannot distinguish between a power-law (heavy-tailed) and a Poisson distribution.

However, bearing caveat in mind, the fact that the Poisson distribution cannot be ruled out as an appropriate fit to the out-degree distributions is unexpected. Moreover, this finding has not been reported by previous studies of sign networks (Morais et al., 2013; Steyvers & Tenenbaum, 2005), which have not analysed the in-degree distribution for their networks. Hence, this unexpected and previously unreported finding requires explanation. Superficially, this finding suggests that connections between signs are being made randomly in the out direction. This would appear to contradict the central claim in this thesis, and in the TOOC, that connection between signs is systematic and guides conceptual development. However, careful consideration of how the direction of edges is defined in the sign network might explain this result.

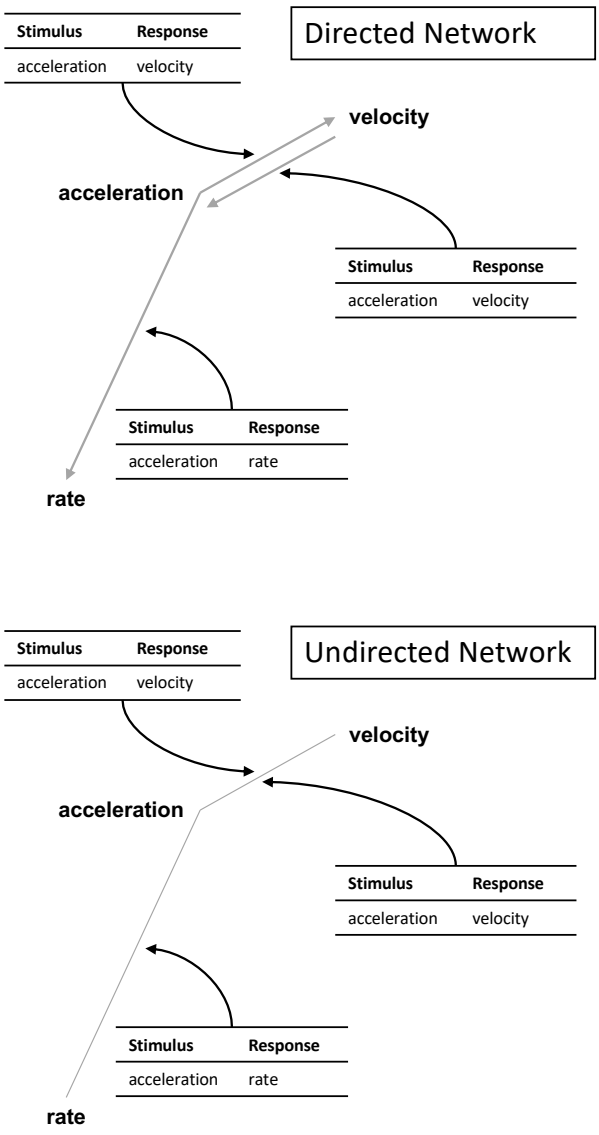


Figure 7.4 The difference between a directed and undirected semantic network.

Figure 7.4 illustrates how the directionality in the sign networks is defined from the word-associations for each stimulus word. A sign's in-degree is determined by the number of times it is given as a response to a stimulus word. In other words, the in-degree measures how frequently a sign has been named as helping a participant to understand a stimulus word. For example, 12P02 gives *velocity* as a response to the stimulus word of signs: *acceleration*, *momentum*, *motion*, *kinetic*, and

decelerate. Hence, the in-degree for *velocity* is five. In contrast, the out-degree measures how many signs were given as response when a sign was a stimulus word. For instance, 12P02 responded to *velocity* with *direction, motion, speed,* and *magnitude*. Hence, the out-degree for *velocity* is four.

The example of *velocity* from 12P02's sign network illustrates that the in-degree measure the number of times *velocity* has been cited by 12P02 as being important in understanding signs. In contrast, the out-degree can be seen as a measure of the number of signs 12P02 thinks they need to understand *velocity* (i.e., they need to list four words to understand velocity). It follows from this interpretation of the in and out degrees for a sign that the in-degree would seem to be associated with any placeholder structure that would guide conceptual development. The out-degree, and therefore the out-degree distribution, would represent the number of signs a participant typically needs to list to understand other signs. Hence, the fact that a Poisson distribution cannot be ruled out as a good fit to the sign network out-degree distributions could suggest that there is an average number of signs individuals typically need to list to develop the conceptual understanding of any given sign. However, these out-going connections are more likely to connect (i.e., preferentially attach) to signs with high degrees. Hence, the in-degree distribution follows heavy-tailed distributions and contains hub-words, whilst the out-degree distribution follows a Poisson distribution, and the signs can have an average out-degree.

This interpretation strengthens the importance of hub-words because it illustrates how these words disproportionally attract connections within the network. Hence, they play a significant role in organising the growth of the network and are, therefore, potentially important with the placeholder structure that Carey suggest guides conceptual development.

7.3.3 Summary

Claim 2 stressed that a new sign will enter an already complex network by through a small number of connections. However, these connections are within a small neighbourhood of signs. Hence, sign structure develops in clusters. Therefore, the structure that determines the initial content of any placeholder associated with a new sign is determined by connections within this cluster.

This claim adds to this picture by demonstrating that any new sign entering a network is more likely to initially connect to a high-connected hub-word. Hence, a sign network develops in clusters, but these clusters 'grow' around hub-words. Moreover, these hub-words explain some of the important structural properties of the networks, such as short path lengths. Hence, these hub-words are essential in terms of the movement of 'information' through the network. Again,

this is consistent with the findings from concepts network studies, which show that concept networks contain some vertices that have high degrees and that some vertices are pivotal in connecting different section of the networks (Koponen & Nousiainen, 2014, 2018b). As with the clustering in the network, these similarities in the features of the sign and concept network are consistent with the proposal that placeholder sign structure guides the development of conceptual structure.

7.3.4 Implications for Teaching

In terms of pedagogy, this claim highlights that there are certain sign or words that students will build their conceptual understanding of a topic around. Again, this is consistent with the pedagogical approach described by Mortimer and Scott (2003) that identifies that experienced science teachers will plan a series of lessons around developing an understanding of a key scientific term (e.g., rusting or particle). Moreover, these teaching sequences will often be structured around developing an understanding of key terminology associated with this key word or sign. Hence, the science teachers are in a sense following the structure of the sign networks by teaching lessons that develop an understanding of clusters of words that are drawn together through a key hub-word. An obvious question that arises from this analysis – what are the hub-words in a particular topics or domains of knowledge that could be used to organise and structure the delivery of lessons? This question is addressed by claim 4, which is described in the next section.

7.4 Claim 4: The Sign Networks Have Shared Hub-Words

The physics and psychology sign networks were analysed to determine whether the networks of these two groups shared hub-words. This analysed showed that several hub-words appear frequently in the both the physics and psychology networks. These hub-words were *force*, *energy*, *time*, and *particle*. There are two salient observations about these shared hub-words that are important to the argument that is made in this section regarding the significance of these shared hub-words.

First, hub-words are important structural features of sign networks (see section 2.6.1.3 and 7.3). Their presence explains many of the other structural properties of the networks. For instance, the short paths length between signs. Furthermore, as the previous section has made clear, the connections between the signs in the network are not made randomly. In contrast, it is more likely that a new sign entering the network will connect to these hub-words. Hence, the sign networks for both the physicists and psychologists grow around the shared hub-words of force, energy, time, and particle. These shared hub-words are therefore important shared placeholder structures, which are responsible for many features of the sign placeholder structure and, perhaps more importantly, are coordinating and organising the development of further placeholder structure (e.g., clusters) in individuals as they develop intuitive theories in a domain of knowledge.

Second, these hub-words are associated with recognisably important concepts in mechanics and other domains of physical knowledge, which are crucial to understanding physical theories. This is consistent with the findings of Koponen and Nousiainen (2014, 2018b) that identify key concepts in their electricity and magnetism concept networks that have betweenness high centrality and are, therefore, crucial to the flow of information in the networks. The fact that these hub-words are associated with fundamental and holistic concepts could explain why the signs associated with these concepts become hub-words. In this scenario, connections preferentially attach to these words because of their importance as concepts. However, this explanation is not consistent with the TOOC, which explicitly states that placeholder structure develops before conceptual understanding. Hence, if TOOC and Quinian bootstrapping is correct, hub-words should exist as placeholders before they are associated with a fully formed concept. The next section presents evidence that this is the case for the hub-word *force*, which seems to exist solely as a placeholder before it become associated with a concept.

7.4.1 Placeholder Structure Precedes Full Conceptual Understanding

The evidence that *force* initially exists as a sign associated with a placeholder rather than a fully formed concepts comes from the results of the FCI and the findings of the interviews. Despite sharing *force* as a hub-word, the FCI scores demonstrate that the physicists solve problems in a way that is consistent with having Newtonian conceptual system (CS2), whilst the psychologists have a common-sense conceptual system (CS1). However, both the physicists and psychologists share *force* as a hub-word. Hence, the FCI scores imply that the physicists and psychologists associate different concepts with the sign *force*.

The analysis of the interview transcripts demonstrates the similarities and differences between the concepts the physicists and psychologist associate with the sign *force*. For both cohorts, the content analysis demonstrate that *force* is a commonly occurring word in interview dialogue. For physicists, force is the most commonly occurring word that can be associated with a recognisable physical concept. This suggests any concept associated with *force* is seen as central in any explanation of the motion of the objects in the demonstration by both the physicists and psychologists. However, there is a significant difference between the normalised frequently of the use of *force* for the two cohorts, with the physicist using this hub-word more frequently. Moreover, *pressure* is most frequently occurring word in the psychologists' interview transcripts. This suggest that any concept associated with *pressure* plays an equally, or perhaps more important, role in psychologists' explanation of motion. This would be consistent with psychologists not having a fully developed theory of Newtonian mechanics.

The LG of the interview transcripts provides greater detail about how any concept associated with *force* is used by the participants to build explanations of what they see in the demonstrations. This analysis shows that both groups recognise *force* as cause of an object's motion (e.g., the upward movement of the film canister or the cupcake case falling to the ground). However, the descriptions of the motion (or process in SFL terminology) caused by *force* is different for the physicist and psychologists. Physicists typically identify *force* as causing acceleration, affecting velocity or momentum. These are recognisable physical concepts that give specific description of motion, rather than just general descriptions (i.e., the rocket accelerates towards the ground versus the rocket falls or move towards the ground). In this sense, these concepts are differentiating the concept of motion. In contrast, the psychologists associate *force* with causing with what could be called undifferentiated description of motion (e.g., it moves up, it falls, it stops). The findings from the LG analysis are supported by the content analysis of the interview transcript that show that *speed* is one of the most commonly occurring words in the psychologists' explanations (see

Table 6.4 and Table 6.5). In contrast, *acceleration*, and *velocity* are some of the most frequently occurring words in the physicists' explanations (see Table 6.3 and Table 6.5).

In addition to the differences in how the two cohorts describe motion, the physicists seem to recognise the important of differentiating between a force and a resultant force and associating resultant with acceleration. The psychologists do use the idea of an 'overall' force but, again, this is associated with undifferentiated motion or effects. Finally, psychologists do sometimes associate *force* with a cause that can 'run out'.

The difference between the concepts the psychologists and physicists associated with the sign *force* can be summarised in terms of what Carey calls translators gloss (see section 2.5.2.2)

- CS1: a force causes an object to move. It can 'run' out.
- CS2: a resultant force causes acceleration.

Clearly the translators gloss is different for physicists and psychologists. Hence, the concepts associated with *force* by physicists and psychologists satisfy the criteria defined by Carey to conclude that they are incommensurate. These findings are broadly consistent with previous studies of the differences between novice and expert conceptual systems in mechanics (Brookes & Etkina, 2009; Ioannides & Vosniadou, 2002; McCloskey, 1983). Moreover, the descriptions of CS1 and CS2 defined by the interview analysis align with the descriptions of CS1 and CS2 defined by the FCI. Hence, the interview findings agree with the FCI scores.

The findings of the interviews are consistent with Carey's account of how Quinian bootstrapping acts as a mechanism for conceptual changes and with the dual factor view of concepts, which is adopted by Carey. In this view and in terms of Quinian bootstrapping, any nascent concept is initially defined solely in terms of its placeholder structure. Hence, in this case, any representational content associated with the sign *force* is initially defined in terms of being a highly connected hub-word. However, Carey also argues that at the start of the bootstrapping process the content of these placeholder structure is only narrow and is therefore defined purely in terms of its conceptual role, which is identified with the most causally or inferentially central part of this role. Both the physicists and psychologists primarily recognise cause as the role of any concept associated with *force*. Hence, these findings suggest that the bootstrapping process begins with *force* being recognised as a centrally important sign, around which new structure develops. Moreover, the central conceptual role of *force* as a cause is recognised from the outset of the bootstrapping process.

Despite both groups sharing narrow content, the interview analysis suggest that the physicists have different broad content in terms of what they view force causes and how different types of force can be distinguished. This broad content recognises that resultant forces cause acceleration or changes in velocity, rather than force causing undifferentiated motion. This finding is significant because it suggests physicists appear to associate a concept with the hub-word *force* that is consistent with a fully developed concept of force within a Newtonian conceptual system (CS2). In contrast, the representational content the psychologist associate with the hub-word *force* is consistent with the narrow content that Carey argues characterises an initial placeholder value. Despite this both the physicists and psychologists share important sign structure in terms of *force* as a hub-word that is crucial for the development of conceptual understanding. Hence, the presence of important placeholder structure appears to precede the development of full conceptual understanding and hub-word (or sign) *force*, which is at the centre of this placeholder structure, is initially associated with a placeholder value.

This analysis also suggests that the move from CS1 to CS2 involves differentiating an undifferentiated concept of motion by recognising that motion can be described in terms of acceleration, velocity and connecting these concepts with force as a cause. In accordance with the prediction of the TOOC, this would initially involve connecting any sign structure associated with force with any sign structure associated with motion. The aggregate psychologist sign network suggests this has not yet happened for psychologists. However, the physicists and psychologists do share some sign structure around *acceleration*, *velocity*, and *speed*. Moreover, the interview analysis suggests this structure precedes a full understanding of these concepts (see Table 6.23). Again, this is evidence that sign structure precedes a full conceptual understanding.

Interestingly, psychologists' inability to accurately describe motion could create scenarios that makes it difficult for them to be convinced by any putative connection between force and motion, which are associated with a Newtonian conceptual understanding of mechanics. For instance, Taylor and Hillier (2018) identified instances in classroom dialogue when 16 year old students refused to accept that a resultant force could cause acceleration because they did not recognise that an object was accelerating (e.g., when a plasticine ball hits a desk and stops). Again, the findings of Taylor and Hillier (2018) and the claims made in this discussion are evidence for the need for simultaneous and coordinated changes in concepts when learning intuitive theories.

7.4.2 Summary

Claim 3 stressed the role of hub-words as signs that organise the growth of cluster of signs in the networks and explain the connectivity of these clusters in terms of the short path length between them. Hence, they are important features of the placeholder structure that the TOOC assumes is responsible for guiding conceptual development. This claim identifies hub-words that frequently occur in the sign networks of both physicists and psychologists. These hub-words are *force*, *energy*, *time*, and *particle*. They are recognisably associated with what researchers typically assume are fundamental concepts in both mechanics and several other physical domains of knowledge. Therefore, it is possible that *force*, *energy*, *time*, and *particle* become highly connected because of their importance as concepts. However, TOOC proposes Quinian bootstrapping as the mechanism that drives conceptual change. Quinian bootstrapping predicts that placeholder structure will precede conceptual development. This section presents an argument that supports this view, which is based on the evidence from sign networks, FCI and interviews that suggests that sign structures do appear to develop before conceptual understanding.

Carey's proposal that Quinian bootstrapping drives conceptual change places language at the centre of conceptual development, as presumably connections between signs are learnt from experiencing the use of language within a domain of knowledge. This is consistent with Carey's example of the development of the concept of integer number, which begins with the learning of the count sequence (e.g., *one*, *two*, *three*, etc…). This placeholder sign structure then guides the learning mechanisms needed to develop the necessary conceptual understanding. Hence, Quinian bootstrapping, through language learning, provides the mechanism that organises the learning process that reorders the system of conceptual elements into a 'better' conceptual system (see Rusanen (2014) and section 2.4.3). Hence, language learning is possibly one of the missing mechanisms that Rusanen (2014) suggests is required to explain conceptual change. Evidence for this view has been presented by studies that have focused on how effective science teaching focuses on language development (Brookes & Etkina, 2015; Mortimer & Scott, 2003) and the bootstrapping process (Smith, 2007). However, this claim provides direct evidence to support this claim by identifying placeholder structure that is shared by physicists and psychologists despite differences in their conceptual understanding (claim 1).

It could be significant that *force* is recognised as a cause by both the physicists and the psychologists. Cause is identified by Carey (2009) as a core concept and, therefore, part of the innate conceptual repertoire. After learning the initial count sequence that forms the sign placeholder structure (i.e., one, two, three, etc…), the second stage in the development of this concept is mapping the core

concept of number onto the signs for *one*, *two*, and *three*. Hence, the development of the concept of natural number and the concept of force are associated with mapping important placeholder structure onto core cognition. This is also consistent with the proposal by Laurence and Margolis (2002) (which is endorsed by Carey (2009)) that humans have an innate ability to recognise categories of concepts (i.e., objects). This capacity is realised in terms of placeholder structure, which are initially 'empty' but provide 'slots' that can be filled with appropriate content. This analysis seems to show that *force* is initially associated as a cause placeholder (narrow content) and is then filled with the appropriate broad content. Both stages in the development (i.e., the recognition of narrow content and the development of associated broad content) seem to be guided by placeholder sign structure, as proposed by Carey's Quinian bootstrapping process.

7.4.3 Implications for teaching

There has been an increasing awareness of the importance of language in science teaching (Amin, 2017; Brookes & Etkina, 2015). This claim strongly reinforces this view as it provides evidence that supports the view that language learning drives conceptual understanding through Quinian bootstrapping. More specifically, this study suggests students form a placeholder sign structure before they develop a full conceptual understanding of a domain of knowledge. Put more simply, students initially recognise that words or signs connected with each other before associating fully formed concepts with these signs. This process is illustrated in Figure 7.5 using acceleration as a hypothetical example, loosely based on the connection in the aggregate sign networks.

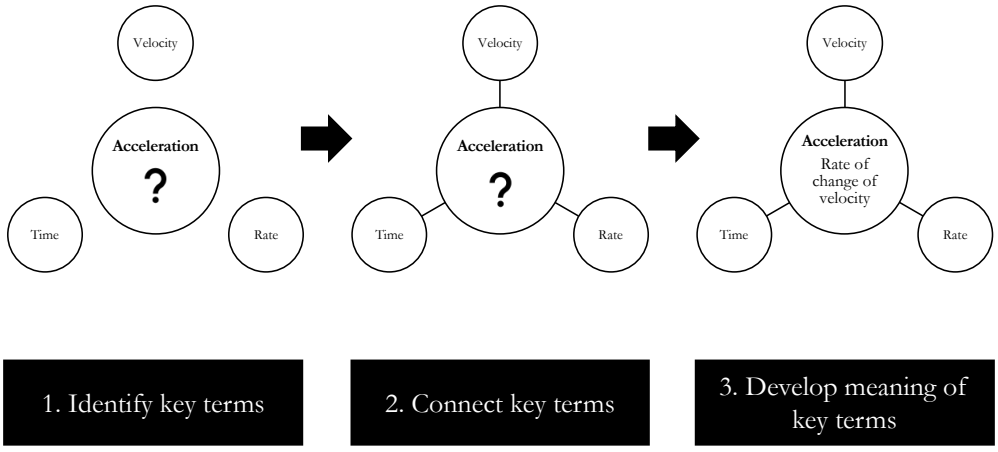


Figure 7.5 Sequence of developing conceptual understanding associated with signs suggested by TOOC (Carey, 2009) and the claims made in this discussion based on the results of this study.

There are several observations that can be made from the discussion of claim 4 that have implications for science teaching. These are outlined in what could be a learning sequence below.

1. Students recognise that force is an important cause of motion.
2. Students begin to form clusters of associated signs. These clusters grow around a hub-word. However, they do not have a full conceptual understanding of the concepts associated with these signs, which are placeholder values at this point.
3. The sign placeholder structure enables students to organise various learning processes that enable them to develop a new intuitive theory, which includes signs associated with fully formed concepts.

It seems from this analysis that despite having studied mechanics to GCSE level and obtained higher than average grades, the psychologists are stuck at stage 2. This suggests more teaching time needs to be dedicated to stage 3, where the activities that engage learning processes that develop conceptual understanding would take place. This study has not focused on what these learning processes might be, and therefore what activities might encourage these. However, the pedagogical approach identified by Mortimer and Scott (2003) could provide a framework to developing teaching sequence that encourage teachers to dedicate section of lesson to activities that develop the conceptual understanding associated with connected signs.

Furthermore, claim 4 and this sequence also highlights the importance of developing the concepts that form the clusters around each hub-word. In this case, the concepts that allow students to differentiate their descriptions of motion. Logically, if these concepts need to be connected with how forces cause motion, they need to be understood before Newton's Laws are introduced. Unfortunately, this is not the case in many GCSE standard texts (e.g., Breithaupt & Calder, 2016), which are also often form the basis of scheme of work provided by exam boards for teachers. In these texts, Newton's laws are often discussed *before* motion is introduced. In contrast, in the researcher's home school a new teaching sequence has been introduced for Key Stage 3 students (age 11-14), which is informed by the findings discussed here, and therefore uses several lessons to develop students' conceptual understanding of motion. The Mortimer and Scott (2003) approach has also been used to encourage appropriate modelling time during lesson. Initial feedback from teaching staff suggest it has been successful in improving students' understanding of mechanics. It is intended to formally assess this teaching approach.

This claims also stresses that sign structure, and therefore further learning, grows around hub-word. This raises the question of why certain word become hub-words and whether they can be identified prior to instruction. This question is tentatively addressed in the next section.

7.4.4 How do Words Become Hubs in a Sign Network?

7.4.4.1 Word Frequencies

The evidence presented in the section has reinforced the proposal made in TOOC that language learning plays a crucial role in concept development because it is through language that sign structure is initially learnt. Therefore, it might be possible to identify the sign structure that is eventually learnt by students in the spoken (e.g., classroom discourse) and written resources (e.g., textbooks) that are used in instruction. Again, the Steyvers and Tenenbaum (2005) model provides a useful tool for identifying what any analysis of these resource might focus on. In this model, a new sign entering a network is most likely to initially connect with a sign with high degree (second principle, see section 7.2.4). However, the new sign also has additional connections it can make. These connections are made to the signs in the neighbourhood of the sign that was initially connected to (first principle). Moreover, how these connections are distributed amongst the signs in the neighbourhood can be modulated according to a utility variable (third principle). The utility variable can be set according to word frequency occurrence in an existing corpus (e.g., a textbook). In this scenario, the most frequently occurring words in any pedagogical lexical resource (e.g., discourse, textbooks) are likely to become hub-words. Furthermore, is it well established that the frequency of distribution of words in natural language follows a power-law (Piantadosi, 2014). Hence, it is possible that this heavy-tailed frequency distribution of words in natural language somehow influences the heavy-tailed distribution of connections in the sign networks.

A speculative initial investigation was carried out to investigate the most frequently occurring words in the in the chapter of the textbook (Breithaupt, 2015) that is used to support the teaching of A-level Physics in the target school. Moreover, the robust statistical procedure used in section 3.3.4.1 was employed to determine whether the frequency distribution of words can be modelled with a power-law (Clauset et al., 2009). The word occurrence frequencies were measured using the Quanteda text analysis package for R (Benoit et al., 2018).

Figure 7.6 shows the word frequency distributions and the power-law fits for the mechanics chapters and all the chapters in the textbook. Table 7.2 show the fit parameter and the fit statistics, which indicate that a power-law model of the distribution cannot be rejected ($p<0.01$). Table 7.1

shows the most frequently occurring words in all the chapters of the textbook and the mechanics chapters. Note that often the most frequently occurring words in the textbook (mechanics chapters and all chapters) correspond to hub-word in students' sign network (*force*, *energy*, *time*, *mass*), when words that have syntactic roles are ignored (i.e., *the*, *a*, *for*) (see Table 7.1). Moreover, the word frequency distribution follows a power-law. Hence, as the Steyvers and Tenenbaum (2005) model suggests, the degree of signs in a sign network could be related to the frequency of word occurrence in natural language within a specific domain of knowledge. This would be consistent with conceptual development occurring through Quinian bootstrapping, where the first stage in this process is connection signs to each other without these signs being associated with fully developed concepts. Put more simply, students are perhaps noticing signs (words) that are frequently being used and how the signs they are often occur with, as a first step in learning the concept in a new intuitive theory, before they fully understand these concepts.

Table 7.1 Word frequencies with the A-level Physics course textbook (Breithaupt, 2015) for all chapter and the mechanics chapters. Word that are hub-words in the sign networks are show in bold.

All Chapters			Mechanics Chapters		
Word	Frequency	Rank	Word	Frequency	Rank
The	5552	1	The	2281	1
A	2147	2	Of	959	2
Of	2144	3	A	842	3
Is	1559	4	Is	660	4
To	1167	5	To	501	5
In	1019	6	**Force**	429	6
And	1002	7	And	407	7
It	686	8	In	351	8
Force	497	9	It	335	9
As	496	10	Object	310	10
At	464	11	On	283	11
Energy	434	12	At	224	12
On	430	13	An	195	13
An	423	14	As	179	14
For	389	15	**Energy**	169	15
From	387	16	By	165	16
That	381	17	Velocity	163	17
Object	371	18	For	152	18
Or	361	19	Be	150	19
By	361	19	**Time**	149	20
Be	351	21	Speed	143	21
Light	335	22	If	143	21
Are	321	23	From	135	23
When	294	24	That	135	23
This	283	25	Acceleration	131	25
If	273	26	Distance	120	26
Each	271	27	**Mass**	118	27
With	264	28	Direction	113	28
Wave	263	29	This	109	29
Can	252	30	Move	108	30

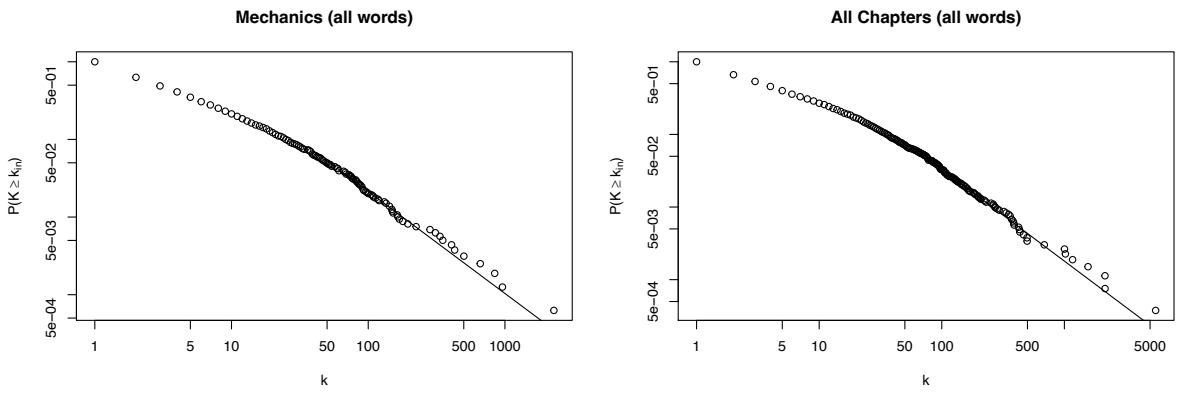


Figure 7.6 The cumulative word frequency distributions for mechanics chapters and all chapters of the physics A-level course text book (Breithaupt, 2015). The distributions are plotted with log-log axis and the solid line represent a best-fitting power-law model.

*Table 7.2 The parameters and Kolmogorov-Smirnov statistics (KS) for a power-law model fit to the word frequency distributions for the textbook chapters. Note * indicates that the hypothesis that a power-law model is good fit to the data can be rejected. Note, α is the scaling parameter (or exponent) for each power-law fit, n_{min} is the cut-off value for n in each fit that defines the tail of the distribution.*

Chapter	α	95% Confidence Limits		n_{min}	95% Confidence Limits		KS	P
		α_{lower}	α_{upper}		$n_{min\ lower}$	$n_{min\ upper}$		
Mechanics	2.312	1.580	2.572	60	1	65	0.050	0.336
All	2.264	1.927	2.548	67	17	101	0.030	0.875

7.4.4.2 Textbook Sign Networks

The ideas that students are noticing frequently occurring word and the words they often occur with can be investigate by examining the sign structure in the course textbook (Breithaupt, 2015). This aim can be achieved if some assumptions are made about how to connect words in the text. Following Christianson, Sizemore Blevins, and Bassett (2020), undirected sign networks can be constructed using text by drawing edges between words if they are in the same sentence. This is a very simple way of drawing edges between signs and other studies have used more sophisticated techniques based on syntactic structure to identify edges (e.g., Koponen & Nousiainen, 2019). However, the Christianson et al. (2020) technique is simple to implement for an initial analysis of the course textbook.

The degree distributions for the course textbook sign networks are shown in Figure 7.7 and the fit statistics in Table 7.3. The hypothesis that a power-law is a plausible model for the distribution cannot be ruled out in particles and quantum mechanics, materials, and electric circuits networks, but can be ruled out in the mechanics and waves networks, and in the network for entire textbook. The strict test has not been applied to these distributions yet, so whether alternative distributions are a better fit to the data has not been tested. However, the failure of a power-law to adequately

fit three of the distributions suggest that the structure of the textbook sign networks might not be scale-free. However, as with the sign networks, these distributions appear to be heavy tailed. Hence, they contain hub-words. Moreover, there is some overlap (*force*, *energy*, *mass*) with the hub-words in the students sign networks. Hence, it would be interesting to explore the similarities and differences between the textbook and student sign networks.

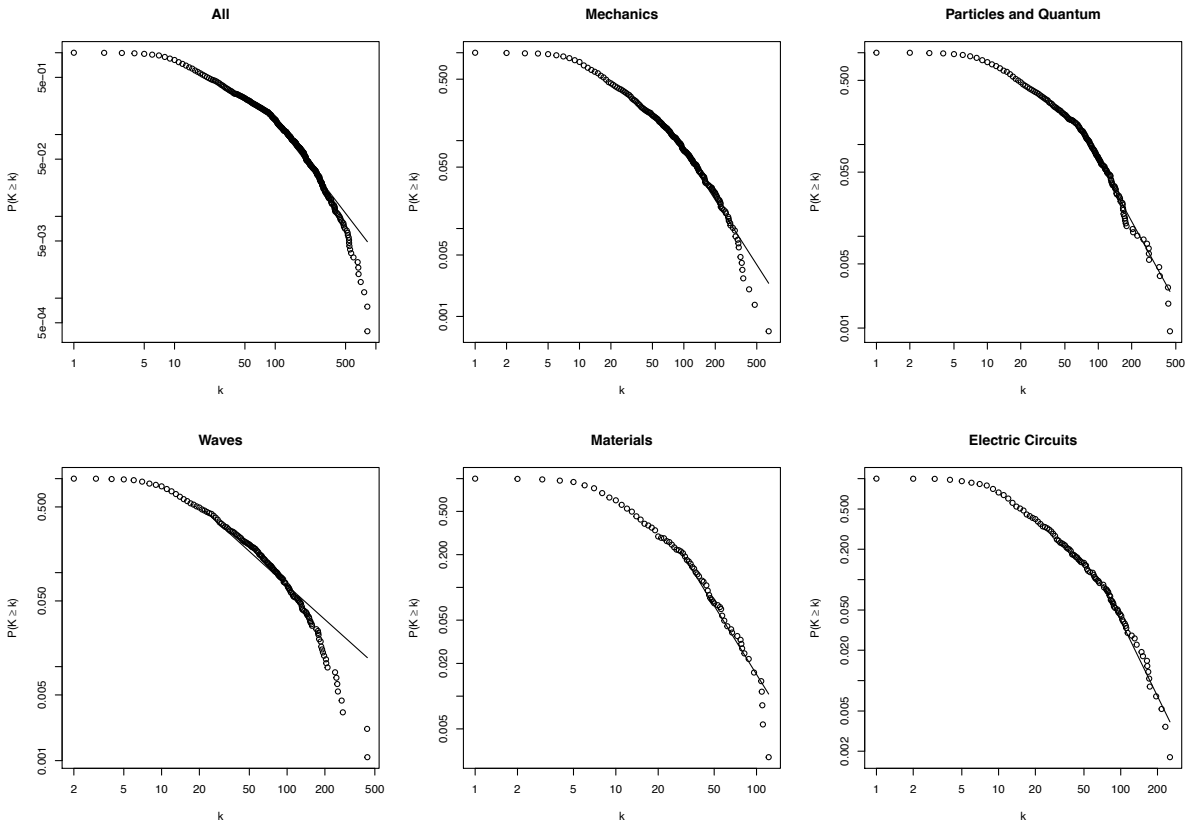


Figure 7.7 The cumulative vertex (sign))degree distributions for sign networks for all the chapters in the A-level physics course textbook (Breithaupt, 2015), and the chapter for each section topic (according to the book contents). The distributions are plotted with log-log axis and the solid line represent a best-fitting power-law model.

*Table 7.3 The parameters and Kolmogorov-Smirnov statistics (KS) for a power-law model fit to the vertex degree distributions for textbook sign networks. Note * indicates that the hypothesis that a power-law model is good fit to the data can be rejected. Note, α is the scaling parameter (or exponent) for each power-law fit, k_{min} is the cut-off value for k in each fit that defines the tail of the distribution, and n_{tail} is the number of vertices tail of the distribution above k_{min}.*

Participant	α	95% Confidence Limits		k_{min}	95% Confidence Limits		KS	P
		α_{lower}	α_{upper}		$k_{min\ lower}$	$k_{min\ upper}$		
All Chapters	2.600	2.420	2.840	85	83	158	0.060	0.000*
Mechanics	2.850	2.360	3.290	83	48	132	0.060	0.047*
Particles and Quantum Mechanics	3.210	2.820	3.780	72	62	107	0.050	0.219
Waves	2.200	2.060	2.380	22	20	49	0.070	0.000*
Materials	3.080	2.150	3.940	29	9	46	0.070	0.147
Electric Circuits	3.480	1.950	4.360	72	8	88	0.060	0.495

Table 7.4 Words with the highest degree (hub-words) in all chapters and the mechanics chapter in the textbook sign networks.

All Chapters			Mechanics		
Rank	Word	k	Rank	Word	k
1	Can	823	1	**Force**	650
2	Use	822	2	Object	478
3	**Force**	765	3	Speed	422
4	**Energy**	707	4	**Mass**	371
5	Light	673	5	Can	364
6	Figure	669	6	**Time**	359
7	Two	659	7	**Energy**	349
8	Object	599	8	Figure	336
9	Direct	568	8	Move	336
10	Example	554	10	Use	333
11	Electron	546	11	Velocity	328
12	**Time**	542	12	Vehicle	313
13	Speed	541	13	Example	308
14	**Mass**	537	13	Two	308
15	Therefore	527	14	Distance	295
16	Move	524	15	Therefore	280
17	Wave	514	16	Weight	278
18	Equal	492	17	Direction	270
19	Show	482	18	Act	269
20	**Particle**	478	19	Equal	266
21	Difference	475	20	Acceleration	253

7.4.4.3 Summary

The two speculative studies in this section suggest that the frequency of word occurrence in instruction resources (e.g., textbooks) may play some role in how individuals identify hub-words, and the sign they connect to. This is a possible area for future study.

7.5 Claim 5: The Sign Networks Contain Some Different Hub-Words

Claim 4 identifies several hub-words that are typically present in the sign networks of physicists and psychologists. Claim 3 stressed how hub-words are responsible for organising the growth of the clusters in the sign networks in the Steyvers and Tenenbaum (2005) model. Hence, hub-words are important features of the placeholder sign structure that the TOOC proposes guides the learning processes that drive conceptual development. Therefore, shared hub-words are an important shared piece of placeholder sign structure that is a feature of both physicists' and psychologists' sign networks, despite their different conceptual understanding of mechanics. However, not all hub-words are shared by the participants. *Mass* is a hub-word in the physicists' network but not in the psychologists' sign networks.

The presence of mass as a hub-word in the physicists' network but not in the psychologists' network suggests that sign structure around *mass* may be an important piece of the placeholder sign structure for developing a Newtonian conceptual understanding of mechanics (CS2). The interview analysis of participants' use of *mass* in their explanation of the motion of the film-canister rocket and falling cup-cake case would appear to confirm this, for several reasons.

First, the content analysis of the interview transcripts showed *mass* is used frequently in physicists' explanations, second only to *force*. This is consistent with the sign *mass* being associated with a concept that the physicists view as crucial to their theory of mechanics. Moreover, this is consistent with the central and fundamental role of mass in Newtonian mechanics. In contrast, *mass* is either not used by psychologists (two of the five psychology participants) or it is used infrequently (only 6 times by three participants, ranked 53rd, 32nd and 108th in most frequently used words) (see Table 6.1). This would suggest *mass* does not play a crucial role in the psychologists' intuitive theory of mechanics.

Second, the LG analysis of the interview explanations showed *mass* primarily functions as a cause when it used by physicists, and when it is infrequently used by psychologists. Like *force*, physicists associated the causal role of *mass* with acceleration, changes in velocity, or momentum. Whereas, in the few incidences when psychologists used *mass*, it was associated with falling or general description of motion. Similar to the interpretation of the concepts the physicists and psychologists associated with *force*, this suggests these groups of participants share narrow conceptual content associated *mass*, given that this content is defined by a concept's most central inferential conceptual role (Carey, 2009, 2011), but have different broad content. Hence, as was

suggested in claim 4, it is possible an early stage in the conceptual development of mass could be mapping the sign *mass,* with its associated placeholder value, onto the core concept of cause. For *force*, this early mapping to a core concept of cause seemed plausible, and indeed beneficial, given the role of force in Newtonian mechanics. However, the role of mass in Newtonian mechanics is not obviously associated with cause. This discrepancy between the role of mass in Newtonian mechanics and participants' theories of mechanics can be understood by taking the interaction between the hub-words *mass* and *force* into account.

This LG analysis of participants use of *mass* demonstrated that in all but one case, *mass* was used as cause is association with another sign, which also had a causal role. Most frequently, this dual causal role was defined by the co-occurrence of *mass* and *force*. Furthermore, the use of these two hub-words was most frequently associated with explanations that involve changes in weight or Newton's 2nd Law. Effectively, *mass* acts as a secondary cause that acts when it mediates or affects a primary cause (i.e., causes more weight, or reduces accelerations). This interpretation is more consistent with mass acting as a property of an object, which is also more consistent with the role of mass in Newtonian mechanics. However, this interpretation does not directly map onto Carey's core concepts of *object*, *number*, *agent*, and *cause*. This is not necessarily an issue for any account of how Quinian bootstrapping might lead to the development of an intuitive theory of mechanics because it is not a requirement of this account to associate the narrow content connected to an initial placeholder sign with a core concept. However, it perhaps suggests Carey has not identified a complete inventory of all core concepts. Alternatively, it is striking that the interaction between mass and force can be characterised in terms of Ohm's p-prim. This raises the possibility that the narrow content could be built from relating core concepts (i.e., causes and agents), which can then be mapped onto important placeholder sign structure, such as hub-words.

The interaction between *mass* and *force* as hub-words may be crucial for the development of the concepts associated with these signs, and participants' intuitive theories of mechanics and matter. Section 2.5.2.2.2 emphasised that Carey's analysis of children's early theories on matter (held until up to the age of approximately 12 years old) did not distinguish between weight and density. Instead, these two concepts were undifferentiated as a single concept of 'degree of heaviness'. As children develop an adult theory of matter, they learn they associate weight with a property of matter and differentiate weight, volume, and density.

In Carey's analysis of the development of these concepts, she does not clearly distinguish between weight as a force and mass as a measure of the quantity of matter in an object. This analysis suggests that development of *mass* hub-word and importance placeholder sign may be involved in

creating this distinction. For example, the merged physicists' sign network shows connection between *mass*, *weight*, and *gravity* forming a cluster. Furthermore, the interview analysis demonstrates that the physicists have successfully differentiated mass from weight and gravity. Perhaps significantly, the sign *mass* is not present in the merged psychologists' network, which only has connections between *gravity* and *weight*.

The interview analysis also suggest that the physicists have developed the concept of mass as inertia through their use of statements that are consistent with Newton's 2nd Law. The physicists' merged sign network show mass connecting the *mas*s | *weight* | *gravity* cluster to the signs *force* (via *weight*) and *momentum*. All the concepts associated with these signs could be seen as essential to develop a concept of inertia. Hence, the role of *mass* in the sign network is consistent with the picture of hub-words organising cluster of sign structure, which guide simultaneous and coordinated changes in key concepts. These changes would appear to be important in the development of physicists' theory of mechanics, but also in their theory of matter. Again, this is consistent with the Quinian bootstrapping process outlined in TOOC.

7.5.1 Implications for Teaching

The implications for teaching that arose from claim 4 stressed were based on the evidence that suggested that psychology students shared significant amounts sign placeholder structure with physicists. This included hub-word, *force*, and some structure around motion. However, claim 4 also stressed that for the psychologists, any representational content associated with this sign structure was still at placeholder stage (i.e., force is recognised as a cause but for different things). Put simply, claim 4 stressed that all participants had made some connection with appropriate signs, but only the physicists had developed a full conceptual understanding.

The implications from claim 5 are different from this observation because this claim demonstrate the psychologists are missing key placeholder structure in terms of the hub-word mass. This suggests that they have not recognised the importance of this sign. It is an open question as to why this the case. In terms of the discussion in section 7.4.4, it could be that *mass* is not frequently used in course textbooks or classroom dialogue at GCSE. However, given the argument outlined in this section about the importance of mass, and the associated sign structure, it is important to work out why *mass* is does not become a hub-word.

7.6 Limitations and Future Work

The findings from this thesis have highlighted several areas for possible future work, which could be some of the limitation of this research. These areas are listed and discussed below.

7.6.1 Further Evidence for Quinian Bootstrapping

The thesis has presented some evidence that supports the Carey's proposal that Quinian bootstrapping drives the development of concepts within intuitive theories. However, this evidence could be strengthened in two ways.

First, a more systematic approach could be taken to establishing whether aspects of placeholder sign structure develop before conceptual understanding. For instance, these findings seem to suggest that sign structure around motion develops before a full conceptual understanding of important kinematic concepts (i.e., velocity, acceleration). This finding is based semi-spontaneous questions that participants were asked in the semi-structured interviews. Hence, any future study could specifically plan to explore participants' understanding of these concepts. Furthermore, a distinction has been made in this thesis between concept and sign networks based on the conditions that are applied to identify vertices and form edges in these networks. Concept networks apply more stringent conditions, which require declarative knowledge, in contrast with sign networks that require less developed knowledge. Hence, it could be possible to investigate whether placeholder sign structure precedes conceptual development by creating, and then comparing, sign and concept networks.

Second, the evidence presented here that supports Quinian bootstrapping relates to the development of an intuitive theory of mechanics. However, Carey's present bootstrapping as a process that is general in the sense that it applied to all conceptual development in intuitive theories. Indeed, Carey cites examples from the development of an intuitive theory of matter to support her arguments for bootstrapping (Carey, 2009). Hence, the evidence presented here in support of Quinian bootstrapping (i.e., the presence and association of hub-words with placeholder values before fully developed concepts) should be present in other domains of knowledge with associated intuitive theories (e.g., thermodynamics). This hypothesis could be tested using the methodology developed in this thesis.

7.6.2 Small Sample Sizes and Using Data Lost Through Attrition

The size of the physicist and psychology samples are small (six and five participants in each group respectively). This implies that some of the findings that involve comparisons between the two groups (e.g., the comparison of the network parameters or in the presence of hub-words in the networks) may be susceptible to type II errors, where the null hypothesis is incorrectly accepted. To avoid this problem, the initial sample size for both groups was higher (13 physicists and 10 psychologists). However, there was significant participant attrition (see section 3.3.3.2.3). Participants were allowed to drop out for several ethical reasons (also see section 3.3.2.4).

Two participants in the original sample did not submit any answer to the on-line data collection for the sign networks. However, the remaining eight participants who eventually dropped out did submit data. A further four only completed one iteration of data collection in the snowball sampling method. Hence, their data could not be used. However, the remaining four participants completed more than one iteration. But at least three data collection iterations were needed for the snowball sampling method to successfully capture a sign network without significant distortion of the network structure (see section 3.3.3). This condition ruled out the use of the data from a further two participants.

The remaining three participants complete three or more iterations but did not complete their final iteration. This creates an issue because the network finishing technique, described in section 3.3.4.1, cannot be used to create sign networks from this data. This technique is a practical necessity when using snowball sampling, but it does distort the network degree distributions. Developing another technique to finish the sign network for the participants with incomplete final iterations could distort the degree distribution in a different way. This would make comparing the distribution from the complete and in-complete networks difficult. However, the addition of three extra networks to the sample may be beneficial. Hence, it is intendent to explore ways to use this data in any future work.

7.6.3 Additional Measures of Network structure

The sign networks in this study are not easy to visualise and, therefore, it is not possible to compare them by qualitative inspection. Hence, this analysis has used measures that capture features of network structure to compare these networks. Initially, the mean shortest path length, mean degree, clustering coefficient have been used to give a basic characterisation of the sign networks. However, other measures have been omitted from this initial analysis that may provide additional perspectives to the comparison of the networks. For instance, there are several measures of

network centrality that have not been applied to the networks (e.g., eigenvector centrality, betweenness centrality).

Centrality attempts to identify the most important vertices in a network (Newman, 2018). The simplest way to measure centrality is using vertex degree by defining the most important vertices as the most highly connected vertices in the network. Hence, by this measure of centrality, the hub-words in this study are the most important signs in the sign networks. However, degree is a crude measure of centrality (Newman, 2018) and measures like eigenvector centrality and closeness centrality have incorporate more sophisticated definitions of vertex importance. For example, Koponen and Nousiainen (2014) use betweenness centrality to identify key concepts that contribute significantly to the connectivity of the concept networks, and therefore the flow of information through these networks. It would be interesting to use measures of betweenness centrality to identify whether hub-words play the same role as the key concepts identified by Koponen and Nousiainen (2014). This would provide another example of placeholder sign structure reflecting conceptual structure.

7.6.4 Communities in the Sign Networks

Modular structure is a feature of individual concept network for electricity and magnetism (Koponen & Nousiainen, 2018b). The findings of this study suggest that placeholder structure (e.g., cluster of signs) in sign networks are growing around hub-words (*force*, *energy*, *mass*, *time*, *particle*), which are associated with key physical concepts. Therefore, sign networks might contain modular structure, like the modular structure found in concept maps, that is associated with hub-words.

Algorithms have been developed to identify sub-groups or communities with networks (Newman, 2004). These will be used to try and identify communities, or modules, within the network. An initial attempt to identify communities in a physics sign network are shown in Figure 7.8, and the words in one community in this network are presented in

Table 7.5. The identification of modular structure in sign network using community detection could be an area of future work.

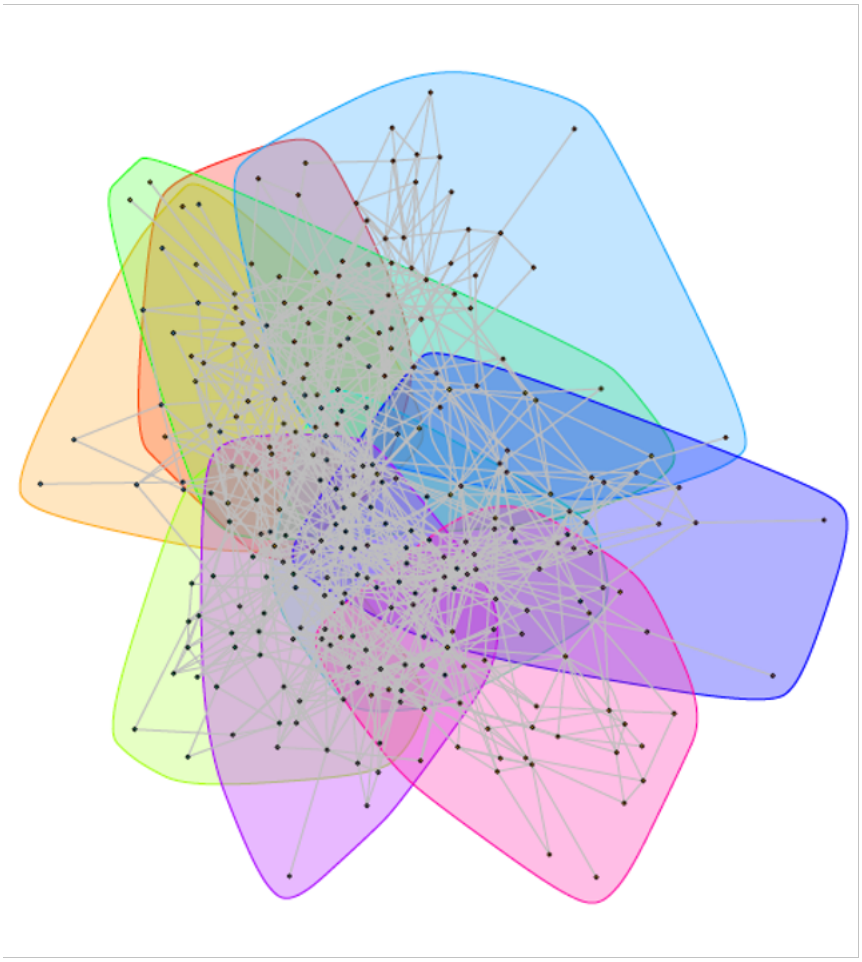


Figure 7.8 The communities in 12P05's sign network identified by the Cluster Louvain function in iGraph (Csardi & Nepusz, 2006), which implements a multi-level modulars algorithm to detect community structures, which Yang, Algesheimer, and Tessone (2016) suggest is appropriate for a network of this size.

Table 7.5 The words in the 'energy' community in the 12P05 sign network.

Vertex (sign)	Degree
Energy	20
Dissipated	12
Heat	11
Wasted	10
Inelastic	9
Efficiency	9
Kinetic	8
lost	8
collision	8
surroundings	7
converted	7
momentum	6
conservation	6
inefficient	6
yield	6
transfer	6

7.6.5 Quinian Bootstrapping in Groups of Student with Different Language Profiles

The participants in this study all had English as a first language and were predominately from higher socio-economic groups (see section 3.3.1.2). Moreover, these students had similar vocabularies but different conceptual understandings of mechanics (claim 1). This is consistent with claim 2 that identified that students have similar macrostructure for their network but different microstructure. However, students whose first language is not English or students from lower-socio-economic groups could have different vocabularies. Hence, the macrostructure of their sign network could be different. This may affect the findings outlined here regarding the nature of the placeholder sign structure that develops during the Quinian bootstrapping process that drives the development of concept in intuitive theories. Again, the methodology developed here could be used to investigate the sign network for students with different language profiles.

7.7 Summary

This discussion has present five claims that highlight features of the placeholder sign structure that could guide conceptual development in individuals' intuitive theory of mechanics. These claims have provided a framework to analyse and breakdown the Quinian bootstrapping process, in a similar way to the count sequence in the development of natural number. These claims are summarised and related to the research question posed in the thesis in the following conclusions chapter. This section also summarises the implications for teaching and identifies areas for future work.

8 Conclusion

Conceptual change has been the focus of this thesis, and how a greater understanding of this process, particularly the mechanism that drives conceptual change, might lead to improvements in science teaching. TOOC was adopted as the theoretical framework for this thesis because this theory provides the most complete account of conceptual change. Specifically, TOOC uses extensive empirical evidence to identify core concepts as the foundations of subsequent conceptual representative (the innate conceptual repertoire). TOOC then uses Quinian bootstrapping to explain how more advanced concepts, including those embedded in intuitive theories, develop from this innate conceptual repertoire. This explanation is supported by a precise and detailed account of how the concept of natural number develops from a core cognition of number. Carey (2009) uses knowledge of the placeholder sign structure the guides this conceptual development to develop this account and to marshal empirical evidence that supports it. This detailed knowledge of placeholder sign structure is not known for any examples of the development of concepts in intuitive theories. Hence, a central aim of this thesis has been to examine the placeholder sign structure that forms when individuals learn mechanics, and to use this knowledge to develop a more detailed and precise account of the Quinian bootstrapping process that the TOOC proposes drives conceptual development.

8.1 Summary of Claims

The discussion in chapter 7 has demonstrated how the findings in this thesis, captured in terms of the claims outlined in this chapter, are consistent with the TOOC and can be interpreted in terms of this framework. Specifically, these claims have described the placeholder sign structure that guides the development of intuitive theories of mechanics in detail. These claims are:

1. The physicists and psychologists (that form the expert and novice groups) have a different conceptual understanding of mechanics.
2. The sign networks of the physicists and psychologists have similar macrostructure but there are differences in the microstructure.
3. All the sign networks contain hub-words.
4. The physics and psychology sign network have shared hub-words (*force*, *energy*, *time*, *particle*).
5. The physics and psychology sign networks contain some different hub-words (*mass*).

The claims listed above have illuminated features of the Quinian bootstrapping process that is responsible for developing intuitive theories in mechanics, and provided some evidence that this process drives conceptual change in this domain of knowledge. These features and evidence are summarised below in terms of this study's research questions.

1. *What are the sign networks of individuals within a scientific domain of knowledge?*

2. *What are the similarities and differences between the networks for individuals who have different levels of scientific understanding (novices and experts)?*

Research questions 1 and 2 are primarily addressed by claims 1 and 2. The evidence presented as part of claim 1 demonstrates that the physicists and psychologists have different conceptual understandings. Broadly, psychologists have what Hestenes et al. (1992) define as an everyday understanding of mechanics (CS1) and physicists a Newtonian understanding of mechanics (CS2). This conclusion is supported by the FCI scores and the interview analysis.

Despite these differences in conceptual understanding, claim 2 asserts that the macrostructure of placeholder sign networks of the physicists and psychologists share several common features. These networks have a similar number of vertices and edges. They are sparse and have a low average degree. Hence, most signs are only connected to a relatively small number of other signs. Despite this, the networks have short average path. Hence, it is possible 'jump' between signs in a short number of steps (~6). Finally, the sign networks contain clusters of signs. These features

suggest that the placeholder structure that guides conceptual change is economical in the sense that only a small number of connections between neighbouring signs can the drive conceptual change that is needed to form new intuitive theories. However, the clusters in the networks suggest that simultaneous and coordinated changes in concepts are required to develop these theories. Moreover, the short path length in the networks suggest that connections between clusters are important in conceptual development.

The difference in the microstructure of the sign networks, which is shown by the aggregate sign networks, support the findings that the placeholders sign structure that guides conceptual development is minimal or economical, forms clusters, and requires connections between different parts of the network. For instance, *weight* | *gravity* | *mass* form a small cluster in the aggregate physicist network, but this cluster is absent in the psychologist network. In both aggregate networks, there is what could be identified as a motion cluster around *speed* | *time* | *distance* | *velocity* | *acceleration*. Furthermore, the aggregate physics network seems to show that there are important shared connections between clusters (i.e., between *force* and *energy* via *work*).

a. Are hub-words responsible for the majority of connections in the networks?

Claim 3 states that both the physicists' and psychologist's' placeholder sign networks contain hub-words. Hence, claims 3 addresses research questions 2a. The presence of hub-words in the networks have several implications that add to the picture of the sign networks shown by claim 2.

First, the hub-words explain the short path lengths that are a shared feature of macrostructure of the networks. Hence, clusters of signs with a small number of connections seem to be a feature of the placeholder sign structure that guides conceptual development, these signs are connected via hub-words to other clusters in the network. Again, the aggregate sign networks seem to emphasise the importance of these connections between clusters. The physicists' aggregate network clearly contains clusters of signs (e.g., *acceleration* | *vector* | *velocity*) but these clusters seem to have shared connections (i.e., the same connections are made by 5 out of the 6 physicists) via hub-words such as *force, mass,* and *energy*. In contrast, the psychologists' aggregate networks show isolated clusters that are not connected via shared edges.

Second, a model based on preferential attachment, such as the model developed by Steyvers and Tenenbaum (2005), can explain the presence of hub-words in the network. In the Steyvers and Tenenbaum (2005) model, any new sign entering the network is likely to connect to a sign with a high degree, and then the neighbours of that sign (which explains the clustering in the network).

Hence, the placeholder sign networks for physicists and psychologists develop by forming small clusters of signs, but these clusters effectively grow around hub-words. This suggests that these words play some sort of crucial role in organising the structural development of the networks. Therefore, hub-words are important features of the placeholder sign structure.

b. If so, are the hub-words the same for different individuals?

Claims 4 and 5 addresses research question 2b. Claim 4 state that physicists and psychologists share the hub-words, *force, energy, time,* and *particle.* Within the context of claim 3, this implies that physicists and psychologists share important sign structure. However, claim 5 demonstrates that not all hub-words are shared by both groups. *Mass* is a hub-word for physicists but not for psychologists. Given that the physicists have a Newtonian (CS2) rather than an everyday (CS1) conceptual understanding of mechanics, this suggests that the hub-word *mass* plays an important role in the development of individuals' intuitive theory of mechanics.

3. Do these individuals use hub-words from their networks when they reason about a scientific experiment or demonstration? If so, how?

Both physicists and psychologists frequently use the hub-word *force* in their explanations of how objects move. Moreover, they identify *force* as a cause. This would suggest that they share the narrow content for the concept associated with this sign. However, physicists associate *force* with causing acceleration or changes in velocity, whilst psychologists associate *force* as cause of more general, undifferentiated motion. Therefore, they do not share broad content. This is consistent with the psychologists associating the sign *force* with a placeholder value and an everyday conceptual understanding of mechanics (CS1). In contrast, the physicists would appear to have a more developed concept of force associated with the hub-word *force*, which explains why they have a Newtonian understanding of mechanics (CS2). Claim 3 demonstrated the importance of hub-words as placeholder structure. Moreover, claim 4 showed hub-words are shared. Hence, the fact that psychologists and physicists share hub-words, and in particular *force*, is evidence that placeholder structure could precede full conceptual understanding. This supports TOOC and specifically Quinian bootstrapping as a mechanism that drives conceptual change.

The sign *mass* is a hub-word in the physicists' networks but not in the psychologists. *Mass* is frequently used in the physicists' explanations but not by the psychologists and, like *force*, is identified as a cause of motion. However, *mass* is usually used as cause in conjunction with *force*. This interaction suggests that the development of *mass* as a hub-word is important for the

development of both the concept of mass and force, most likely in terms of mass being as inertia. This is consistent with TOOC, which argues that simultaneous and coordinated changes in concepts are essential for conceptual development. Evidence for simultaneous changes in concepts, which are initiated by sign connection and coordinated by hub-words, is also present in the aggregate networks. These network show that a cluster of sign *weight* | *gravity* | *mass*, which is connected to *force*, is present in the aggregate physics network but not in the psychology network.

8.1.1 Summary and Contribution to Knowledge

The conclusions listed in the previous section have summarised the research that forms this thesis. These conclusions can be more concisely summarised to emphasise how this thesis has contributed to knowledge. First, this research has shown how snow-ball sampling with considered word associations is a technique that can be used to measure the placeholder sign structure that forms when individuals develop an understanding of concept within their intuitive theory of mechanics. Second, it has demonstrated how network analysis can quantitatively characterise the placeholders sign structure, and the efficacy of using a model of network development as analytical tool to try to understand how this structure develops. Overall, this analysis has been used to build a more detailed picture of the placeholder sign network structure associated with the development of intuitive theories. This picture has illuminated the Quinian bootstrapping process, which is central to the TOOC, and has provided some evidence this process is one mechanism that is driving conceptual change. Hence, this thesis has contributed to the more precise model construction and theory building and evaluation approach to conceptual change called for by Amin and Levrini (2017b).

8.2 Summary of Implications for Teaching

The findings summarised in section 8.1 strongly support the view that language (and more generally, symbol systems) is a fundamental mechanism that drives and organises conceptual change (Amin, 2017). Hence, pedagogical approaches that support language learning should be central to science teaching. More specifically, these findings support Carey's second thesis in TOOC which identifies Quinian bootstrapping as the process that enables conceptual development. This thesis has mapped the placeholder sign structure that develops through this process as individuals develop concepts within an intuitive theory (in this case, mechanics). The nature of this placeholder sign structure has several implications for teaching.

First, placeholder sign networks contain hub-words, and it is possible that sign networks develop around these hub-words. It seems that learners can recognise these hub-words without specific pedagogical approaches, perhaps due to their frequent use (see section 7.4.4). However, it also seems reasonable to suggest that hub-words in specific topics could be identified and then highlighted during teaching, this might help learners to develop any required concepts. Similarly, the sign structure contains clusters of signs, which suggest that conceptual development is driven by simultaneous and coordinated changes in closely related concept. Again, examining the sign structure in different topics could help to identify concepts that need to be simultaneously developed (e.g., mass, weight, and gravity). Using knowledge of placeholder structure (and the Quinian bootstrapping process) to inform curriculum design and then teaching is an approach that has been adopted by Smith (2007) and Smith et al. (1985), which they claim to be beneficial in promoting conceptual development. These findings support this view.

Second, this analysis suggests that the psychologists, who despite having studied mechanics at GCSE and passed this assessment, have acquired some of the required sign structure but do not have fully developed concepts associated with the necessary signs (i.e., *force* and, perhaps, *acceleration*, *velocity*, etc…). This would suggest pedagogies that are specifically aimed to develop the meaning of key scientific terminology could be beneficial in helping more students to develop a conceptual understanding of mechanics and other topics that heavily involve conceptual change. For instance, the teaching strategy identified by Mortimer and Scott (2003) that uses difference patterns of interaction to create different communicative approaches that encourage language development, could promote the modelling activities that would help to bootstrap an understanding of key concepts. This approach has been adopted in the researchers target school

to develop a new mechanics curriculum and pedagogy, and the teachers involved with this project have made encouraging comments about its efficacy and impact on learning.

9 References

Abrahams, I., Homer, M., Sharpe, R., & Zhou, M. (2015). A comparative cross-cultural study of the prevalence and nature of misconceptions in physics amongst English and chinese undergraduate students. *Research in Science & Technological Education, 33*(1), 111-130.

Albert, R., Jeong, H., & Barabási, A.-L. (1999). Diameter of the world-wide web. *Nature, 401*, 130.

Alexander, R. J. (1997). *Policy and practice in primary education : Local initiative, national agenda* (2nd ed.). London: Routledge.

Amin, T. G. (2017). Representations, concepts and concept learning. In T. Amin & O. Levrini (Eds.), *Converging perspectives on conceptual change : Mapping an emerging paradigm in the learning sciences*. London: Routledge.

Amin, T. G., & Levrini, O. (2017a). *Converging perspectives on conceptual change : Mapping an emerging paradigm in the learning sciences* (1st ed.). London: Routledge.

Amin, T. G., & Levrini, O. (2017b). Introduction. In T. G. Amin & O. Levrini (Eds.), *Converging perspectives on conceptual change : Mapping an emerging paradigm in the learning sciences* (1st ed.). London: Routledge.

AQA. (2015a). GCSE combined science: Trilogy. Retrieved from https://www.aqa.org.uk/subjects/science/gcse/combined-science-trilogy-8464/changes-for-2022

AQA. (2015b). GCSE physics. Retrieved from https://www.aqa.org.uk/subjects/science/gcse/physics-8463/changes-for-2022

AQA. (2017). A-level physics. Retrieved from https://www.aqa.org.uk/subjects/science/as-and-a-level/physics-7407-7408/specification-at-a-glance

AQA. (2022). A-level Psychology. Retrieved from https://www.aqa.org.uk/subjects/psychology/as-and-a-level/psychology-7181-7182/specification-at-a-glance

Bagler, G. (2008). Analysis of the airport network of India as a complex weighted network. *Physica A: Statistical Mechanics and its Applications, 387*(12), 2972-2980.

Bani-Salameh, H. N. (2016). Using the method of dominant incorrect answers with the FCI test to diagnose misconceptions held by first year college students. *Physics Education, 52*(1), 015006.

Barabási, A.-L. (2002). *Linked*. New York: Basic Books.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*, 509 LP - 512.

Barabási, A.-L., Jeong, H., & Albert, R. (2000). Error and attack tolerance of complex networks. *Nature, 406*(6794), 378-382.

Barabási, A.-L., & Pósfai, M. (2016). *Network science*. Cambridge: Cambridge University Press.

Baram-Tsabari, A., Sethi, R. J., Bry, L., & Yarden, A. (2006). Using questions sent to an ask-a-scientist site to identify children's interests in science. *Science Education, 90*(6), 1050-1072.

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). Quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software, 3*(30), 774.

BERA. (2018). *Ethical guidelines for educational research*. London: British Educational Research Association.

Berkeley, G. (1910). *A new theory of vision : And other select philosophical writings*. London: Dent.

Biber, D., Reppen, R., & Conrad, S. (1998). Norming frequency counts *Corpus linguistics: Investigating language structure and use* (pp. 263-264). Cambridge: Cambridge University Press.

Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods Instruments & Computers, 33*(1), 73-79.

Black, P. J., & Lucas, A. M. (1993). *Children's informal ideas in science*. London: Routledge.

Block, N. (1986). Advertisement for a semantics for psychology. In P. A. French (Ed.), *Midwest studies in psychology* (pp. 615-678). Minneapolis: University of Minnesota.

Block, N., & Campell, J. (1987). Functional role and truth conditions. *Proceedings of the Aristotelian Society, Supplementary Volumes, 61*, 157-183.

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, Massachussetts: MIT Press.

Braisby, N., & Gellatly, A. (2012). *Cognitive psychology* (Second ed.). Oxford: Oxford University Press.

Breithaupt, J. (2015). *AQA physics: A-level student book* (Second ed.). Oxford: Oxford University Press.

Breithaupt, J., & Calder, G. (2016). *Physics*. Oxford: Oxford University Press.

Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., . . . Wiener, J. (2000). Graph structure in the web. *Computer Networks, 33*(1), 309-320.

Broido, A. D., & Clauset, A. (2019). Scale-free networks are rare. *Nature Communications, 10*(1), 1017.

Brookes, D. T., & Etkina, E. (2009). "Force," ontology, and language. *Physical Review Special Topics-Physics Education Research, 5*(1).

Brookes, D. T., & Etkina, E. (2015). The importance of language in students' reasoning about heat in thermodynamic processes. *International Journal of Science Education, 37*, 759-779.

Brown, C. (2022). Narrow mental content. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2022 ed.). Standford: Metaphysics Research Lab, Stanford University.

Bruun, J., & Brewe, E. (2013). Talking and learning physics: Predicting future grades from network measures and force concept inventory pretest scores. *Physical Review Special Topics - Physics Education Research, 9*, 1-13.

Cachapuz, A. F. C., & Maskill, R. (1987). Detecting changes with learning in the organization of knowledge: Use of word association tests to follow the learning of collision theory. *International Journal of Science Education, 9*, 491-504.

Cachapuz, A. F. C., & Maskill, R. (1989). Using word association in formative classroom tests: Following the learning of Le Chatelier's principle. *International Journal of Science Education, 11*, 235-246.

Cain, M. J. (2002). *Fodor : Language, mind and philosophy*. Cambridge: Polity.

Caldarelli, G. (2007). *Scale-free networks*. Oxford: Oxford University Press.

Cancho, R. F. i., & Sole, R. V. (2001). The small world of human language. *Proceedings of the Royal Society B: Biological Sciences, 268*, 2261-2265.

Carey, S. (1978). The child as a word learner. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), *Linguistic theory and psychological reality*. Cambridge, Massachussetts: MIT Press.

Carey, S. (1985a). Are children fundamentally different thinkers and learners from adults? In S. F. Chipman, J. W. Segal, & R. Glaser (Eds.), *Thinking and learning skills* (Vol. 2). NJ: Hillsdale.

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

Carey, S. (1988a). Conceptual differences between children and adults. *Mind & Language, 3*(3), 167-181.

Carey, S. (1988b). Reorganisation of knowledge in the course of acquistion. In A. Strauss (Ed.), *Ontogeny, phylogeny, and historical development* (pp. 1-27). Norwood, NJ: Ablex.

Carey, S. (1991). Knowledge acquisition of conceptual change? In S. Carey & R. Gelman (Eds.), *The epigenesis of mind : Essays on biology and cognition* (pp. 133-169). Hillsdale, NJ: Erlbaum.

Carey, S. (2000). Science education as conceptual change. *Journal of Applied Developmental Psychology, 21*, 13-19.

Carey, S. (2009). *The origin of concepts*. Oxford: Oxford University Press.

Carey, S. (2011). Précis of the origin of concepts. *Behavioral and Brain Sciences, 34*(3), 113-124.

Carey, S. (2015). Why theories of concepts should not ignore the problem of acquisition. In E. Margolis, H. C. Barrett, S. Laurence, A. Avarguès-Weber, M. Giurfa, J. Plotnik, N. S. Clayton, R. Seyfarth, D. L. Cheney, & B. Mahon (Eds.), *The conceptual mind : New directions in the study of concepts*. Cambridge, Massachussetts: MIT Press.

Cassels, J. R. T., & Johnstone, A. H. (1985). *Words that matter in science - a report of a research exercise*. London: Royal Society of Chemistry.

Chi, M. T. H. (1992). Conceptual change within and across ontological categories: Implcations for learning and discovery in sciences. In R. Giere (Ed.), *Cognitive models of science: Minnesota studies in the philosopy of science* (pp. 129-186). Minneapolis: University of Minnesota Press.

Chi, M. T. H. (2005). Commonsense conceptions of emergent processes: Why some misconceptions are robust. *Journal of the Learning Sciences, 14*, 161-199.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*(2), 121-152.

Chi, M. T. H., & Roscoe, R. D. (2002). The processes and challenges of conceptual change. In M. Limón & L. Mason (Eds.), *Reconsidering conceptual change: Issues in theory and practice* (pp. 3-27). Dordrecht: Springer.

Chi, M. T. H., & Slotta, J. D. (1993). The ontological coherence of intuitive physics. *Cognition and Instruction, 10*, 249-260.

Chi, M. T. H., Slotta, J. D., & De Leeuw, N. (1994). From things to processes: A theory of conceptual change for learning science concepts. *Learning and Instruction, 4*, 27-43.

Chin, C. (2006). Classroom interaction in science: Teacher questioning and feedback to students' responses. *International Journal of Science Education, 28*(11), 1315-1346.

Christianson, N. H., Sizemore Blevins, A., & Bassett, D. S. (2020). Architecture and evolution of semantic networks in mathematics texts. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 476*(2239), 20190741.

Christie, F., & Derewianka, B. (2008). *School discourse: Learning to write across the years of schooling*. London: Continuum.

Clark, A. (2014). *Mindware : An introduction to the philosophy of cognitive science* (2nd ed.). New York: Oxford University Press.

Clark, E. V. (1993). *The lexicon in acquisition*. Cambridge: Cambridge University Press.

Clauset, A. (2022). Power-law distributions in empirical data. Retrieved from https://aaronclauset.github.io/powerlaws/

Clauset, A., Shalizi, C., & Newman, M. (2009). Power-law distributions in empirical data. *SIAM Review, 51*(4), 661-703.

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (Eighth ed.). London: Routledge.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 33*(NOV), 497-505.

Crawley, M. J. (2013). *The R book* (2nd ed.). Oxford: Wiley-Blackwell.

Crystal, D. (2003). *The Cambridge encyclopedia of the English language* (2nd ed.). Cambridge: Cambridge University Press.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695.

Dabney, K. P., Tai, R. H., & Scott, M. R. (2016). Informal science: Family education, experiences, and initial interest in science. *International Journal of Science Education, Part B, 6*(3), 263-282.

Dancy, M. (2000). *Investigating animations for assessment with an animated version of the force concept inventory.* North Carolina State University.

Dasgupta, S. (2016). *Computer science: A very short introduction*. Oxford: Oxford University Press.

De Deyne, S., & Storms, G. (2008a). Word associations: Network and semantic properties. *Behavior Research Methods, 40*(1), 213-231.

De Deyne, S., & Storms, G. (2008b). Word associations: Norms for 1,424 dutch words in a continuous task. *Behavior Research Methods, 40*(1), 198-205.

Department of Education. (2015). National Curriculum in England: Science programmes of study. Retrieved from https://www.gov.uk/government/publications/national-curriculum-in-england-science-programmes-of-study/national-curriculum-in-england-science-programmes-of-study

DfE. (2018). Using pupil premium: Guidance for school leaders. Retrieved from https://www.gov.uk/guidance/pupil-premium-effective-use-and-accountability

DfE. (2020). Main findings: State-funded schools inspections and outcomes as at 31 march 2020. Retrieved from https://www.gov.uk/government/statistics/state-funded-schools-inspections-and-outcomes-as-at-31-march-2020/main-findings-state-funded-schools-inspections-and-outcomes-as-at-31-march-2020

diSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and Instruction, 10*, 105-225.

diSessa, A. A. (2002). Why "conceptual ecology" is a good idea. In M. Limon & L. Mason (Eds.), *Reconsidering conceptual change: Issues in theory and practice* (pp. 29-60). Dordrecht, The Netherlands: Kluwer.

diSessa, A. A. (2007). An interactional analysis of clinical interviewing. *Cognition and Instruction, 25*, 523-565.

diSessa, A. A., Gillespie, N. M., & Esterly, J. B. (2004). Coherence versus fragmentation in the development of the concept of force. *Cognitive Science, 28*, 843-900.

diSessa, A. A., & Sherin, B. L. (1998). What changes in conceptual change? *International Journal of Science Education, 20*, 1155-1191.

Dorogovtsev, S. N., & Mendes, J. F. (2002). Evolution of networks. *Advances in Physics, 51*(4), 1079-1187.

Driver, R., Guesne, E., & Tiberghien, A. (1985). *Children's ideas in science*. Milton Keynes: Open University Press.

Driver, R., Squires, A., Rushworth, P., & Wood-Robinson, V. (1994). *Making sense of secondary science*. London: Routledge.

Duit, R., & Treagust, D. F. (2003). Conceptual change: A powerful framework for improving science teaching and learning. *International Journal of Science Education, 25*, 671-688.

Eaton, P., & Willoughby, S. D. (2018). Confirmatory factor analysis applied to the force concept inventory. *Physical Review Physics Education Research, 14*(1), 010124.

Field, A. P. (2013). *Discovering statistics using IBM SPSS statistics : And sex and drugs and rock 'n' roll* (4th ed.). London: SAGE.

Fodor, J. (1975). *The language of thought*. Cambridge, Massachussetts: Harvard University Press.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports, 486*(3), 75-174.

Garskof, B., & Houston, J. (1963). Measurement of verbal relatedness: An idiographic approach. *Psychological review, 70*(3), 277-288.

Gillespie, C. S. (2015). Fitting heavy tailed distributions: The poweRlaw package. *Journal of Statistical Software, 64*(2), 1-16.

Ginsburg, H. (1997). *Entering the child's mind : The clinical interview in psychological research and practice*. Cambridge: Cambridge University Press.

Goldwater, M. B., & Schalk, L. (2016). Relational categories as a bridge between cognitive and educational research. *Psychol Bull, 142*(7), 729-757.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and bayes nets. *Psychological review, 111*(1), 3-32.

Granello, D. H., & Wheaton, J. E. (2004). Online data collection: Strategies for research. *Journal of Counseling and Development, 82*(4).

Grayling, A. C. (2001). *Wittgenstein : A very short introduction*. Oxford: Oxford University Press.

Grimberg, B. I., & Hand, B. (2009). Cognitive pathways: Analysis of students' written texts for science understanding. *International Journal of Science Education, 31*(4), 503-521.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64-74.

Halford, G. S., Wilson, W. H., & Phillips, S. (2010). Relational knowledge: The foundation of higher cognition. *Trends in Cognitive Sciences, 14*(11), 497-505.

Halliday, M. A. K. (2004). On the language of physical science. In J. Webster (Ed.), *The language of science* (Vol. 5). London: Continuum.

Halliday, M. A. K., & Martin, J. R. (1993). *Writing science: Literacy and discursive power*. London: Falmer Press.

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). *An introduction to functional grammar* (3rd ed.). London: Arnold.

Halloun, I. A., Hake, R., Mosca, E., & Hestenes, D. (1995). Force concept inventory (revised 1995).

Halloun, I. A., & Hestenes, D. (1985a). Common sense concepts about motion. *American Journal of Physics, 53*(11), 1056-1065.

Halloun, I. A., & Hestenes, D. (1985b). The initial knowledge state of college physics students. *American Journal of Physics, 53*(11), 1043-1055.

Hampton, J. A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior, 18*(4), 441-461.

Hanrahan, M. U. (2006). Highlighting hybridity: A critical discourse analysis of teacher talk in science classrooms. *Science Education, 90*(1), 8-43.

Henderson, B. J., Langbeheim, E., & Chi, M. T. H. (2017). Addressing robust misconceptions through the ontological distinction between sequential and emergent processes. In T. Amin & O. Levrini (Eds.), *Converging perspectives on conceptual change : Mapping an emerging paradigm in the learning sciences*. London: Routledge.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher, 30*(3), 141-158.

Hovardas, T., & Korfiatis, K. J. (2006). Word associations as a tool for assessing conceptual change in science education. *Learning & Instruction, 16*, 416-432.

Howe, C. (2013). Scaffolding in context: Peer interaction and abstract learning. *Learning, Culture and Social Interaction, 2*(1), 3-10.

Howe, C., Taylor-Tavares, J., & Devine, A. (2016). Recognition as support for reasoning about horizontal motion: A further resource for school science? *Research in Science & Technological Education, 34*(3), 273-289.

Huffman, D., & Heller, P. (1995). What does the force concept inventory actually measure? *The Physics Teacher, 33*(3), 138-143.

Ioannides, C., & Vosniadou, S. (2002). The changing meanings of force. *Cognitive Science Quarterly, 2*, 5-61.

Jasien, P. G. (2010). You said "neutral", but what do you mean? *Journal of Chemical Education, 87*(1), 33-34.

Jasien, P. G. (2011). What do you mean that "strong" doesn't mean "powerful"? *Journal of Chemical Education, 88*(9), 1247-1249.

Jisc. (2022). Retrieved from jisc.ac.uk

Johnstone, A. H., & Moynihan, T. F. (1985). The relationship between performances in word association tests and achievement in chemistry. *European Journal of Science Education, 7*, 57-66.

Keil, F. (1989). *Concepts, kinds, and cognitive development*. Cambridge, Massachusetts: MIT Press.

Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *Proc Natl Acad Sci U S A, 105*(31), 10687-10692.

Kemp, C., & Tenenbaum, J. B. (2009). Structured statistical models of inductive reasoning. *Psychol Rev, 116*(1), 20-58.

Keys, C. W. (1999). Language as an indicator of meaning generation: An analysis of middle school students' written discourse about scientific investigations. *Journal of Research in Science Teaching, 36*(9), 1044-1061.

Kitcher, P. (1978). Theories, theorists and theoretical change. *The Philosophical Review, 87*(4), 519-547.

Kitcher, P. (1988). The child as parent of the scientist. *Mind & Language, 3*(3), 217-228.

Kittel, C., & Kroemer, H. (1997). *Thermal physics* (2nd ed. Vol. W. H. Freeman and Company): New York.

Kleinberg, J. M., Kumar, R., Raghavan, P., Rajagopalan, S., & Tomkins, A. S. (1999). *The web as a graph: Measurements, models, and methods*. Paper presented at the Computing and Combinatorics, Berlin, Heidelberg.

Koponen, I. T., & Nousiainen, M. (2014). Concept networks in learning: Finding key concepts in learners' representations of the interlinked structure of scientific knowledge. *Journal of Complex Networks, 2*(2), 187-202.

Koponen, I. T., & Nousiainen, M. (2018a). Concept networks of students' knowledge of relationships between physics concepts: Finding key concepts and their epistemic support. *Applied Network Science, 3*(1), 14.

Koponen, I. T., & Nousiainen, M. (2018b). Modelling students' knowledge organisation: Genealogical conceptual networks. *Physica A: Statistical Mechanics and its Applications, 495*, 405-417.

Koponen, I. T., & Nousiainen, M. (2019). *Lexical networks and lexicon profiles in didactical texts for science education*. Paper presented at the Complex Networks and Their Applications VIII.

Koponen, I. T., & Pehkonen, M. (2009). Coherent knowledge structures of physics represented as concept networks in teacher education. *Science & Education, 19*, 259-282.

Kuhn, T. S. (1996). *The structure of scientific revolutions* (3rd ed.). Chicago: University of Chicago Press.

Lappi, O. (2007). *Conceptual change in cognitive science education: Towards understanding and supporting multidisciplinary learning*. Paper presented at the EuroCogSci07: The European Cognitive Science Conference, Delphi.

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing, 16*(1), 33-51.

Laurence, S., & Margolis, E. (2002). Radical concept nativism. *Cognition, 86*(1), 25-55.

Lefever, S., Dal, M., & Matthíasdóttir, Á. (2007). Online data collection in academic research: Advantages and limitations. *British Journal of Educational Technology, 38*(4), 574-582.

Lemke, J. K. (1990). *Talking science*. Norwood, NJ: Ablex.

Levrini, O., & DiSessa, A. A. (2008). How students learn from multiple contexts and definitions: Proper time as a coordination class. *Physical Review Special Topics - Physics Education Research, 4*, 10107.

Lijffijt, J., Nevalainen, T., Säily, T., Papapetrou, P., Puolamäki, K., & Mannila, H. (2014). Significance testing of word frequencies in corpora. *Digital Scholarship in the Humanities, 31*(2), 374-397.

Locke, J. (1975). *An essay concerning human understanding*. Oxford: Clarendon Press.

Lommi, H., & Koponen, I. T. (2019). Network cartography of university students' knowledge landscapes about the history of science: Landmarks and thematic communities. *Applied Network Science, 4*(1), 6.

Lumivero. (2023). Powerful automated transcription software. Retrieved from https://lumivero.com/products/nvivo-transcription/

Madsen, A., & McKagan, S. (2022). Physport implementation guide: Force concept inventory (FCI). Retrieved from https://www.physport.org/assessments/assessment.cfm?A=FCI

Mahajan, S. (2005). Observations on teaching first-year physics, physics/0512158. Retrieved from https://ui.adsabs.harvard.edu/abs/2005physics..12158M

Margolis, E., & Laurence, S. (2022). Concepts. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy*. Stanford: Metaphysics Research Lab, Stanford University.

Maskill, R., & Cachapuz, A. F. C. (1989). Learning about the chemistry topic of equilibrium: The use of word association tests to detect developing conceptualizations. *International Journal of Science Education, 11*, 57-69.

McCloskey, M. E. (1983). Naive theories of motion. In D. Gentner & A. Stevens (Eds.), *Mental models*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets? *Memory & Cognition, 6*(4), 462-472.

McGinn, M. (2013). *The Routledge guidebook to Wittgenstein's philosophical investigations*. London: Routledge.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological review, 85*(3), 207-238.

Medin, D. L., & Smith, E. E. (1984). Concepts and concept formation. *Annual Review of Psychology, 35*, 113-138.

Mercer, N. (2013). The social brain, language, and goal-directed collective thinking: A social conception of cognition and its implications for understanding how we think, teach, and learn. *Educational Psychologist, 48*, 148-168.

Milgram, S. (1967). The small world problem. *Psychology Today, 2*, 60-67.

Morais, A. S., Olsson, H., & Schooler, L. J. (2013). Mapping the structure of semantic memory. *Cognitive Science, 37*(1), 125-145.

Mortimer, E. F., & Scott, P. H. (2003). *Meaning making in secondary science classrooms*. Maidenhead: Open University Press.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, Massachusetts: MIT Press.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological review, 92*(3), 289-316.

Nakiboğlu, C. (2008). Using word associations for assessing non major science students' knowledge structure before and after general chemistry instruction: The case of atomic structure. *Chemistry Education: Research and Practice, 9*(4), 309-322.

Nation, P. (1983). Testing and teaching vocabulary. *Guidelines, 5*, 12-25.

Nation, P. (1990). *Teaching and learning vocabulary*. Boston: Heinle and Heinle.

Nation, P. (2012). The vocabulary size test. Retrieved from https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-tests/the-vocabulary-size-test/Vocabulary-Size-Test-information-and-specifications.pdf

Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9-13.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1999). The university of south florida word association, rhyme, and word fragment norms. Retrieved from http://www.usf.edu/FreeAssociation/

Newman, M. E. J. (2004). Detecting community structure in networks. *The European Physical Journal B, 38*(2), 321-330.

Newman, M. E. J. (2005). Power laws, pareto distributions and zipf's law. *Contemporary Physics, 46*(5), 323-351.

Newman, M. E. J. (2010). *Networks : An introduction*. Oxford: Oxford University Press.

Newman, M. E. J. (2018). *Networks : An introduction* (2nd ed.). Oxford: Oxford University Press.

OED. (2018). *Oxford English dictionary*. Oxford: Oxford University Press.

Ofsted. (2012). *Inspection report for x school (redacted to preserve anonymity)*. Retrieved from https://www.gov.uk/government/organisations/ofsted

Ofsted. (2017). *Short inspection report for x school (redacted to preserve anonymity)*. Retrieved from https://www.gov.uk/government/organisations/ofsted

Ohlsson, S. (2009a). Meaning change, multiple routes, and the role of differentiation in conceptual change: Alternatives to resubsumption? *Educational Psychologist, 44*(1), 64-71.

Ohlsson, S. (2009b). Resubsumption: A possible mechanism for conceptual change and belief revision. *Educational Psychologist, 44*(1), 20-40.

Osborne, J., & Dillon, J. (2008). *Science education in europe: Critical reflections* (Vol. 13): London: The Nuffield Foundation.

Oxford, U. o. (2018). Curec approved procedure 25: Non-invasive research with children. Retrieved from https://researchsupport.admin.ox.ac.uk/sites/default/files/researchsupport/documents/media/ap25_curec_approved_procedure_idrec.pdf

Parnafes, O. (2007). What does "fast" mean? Understanding the physical world through computational representations. *Journal of the Learning Sciences, 16*, 415-450.

Pastor-Satorras, R., & Vespignani, A. (2002). Epidemic dynamics in finite size scale-free networks. *Phys Rev E Stat Nonlin Soft Matter Phys, 65*(3), 035108/035104-035108.

Perez-Granados, D. R., & Callanan, M. A. (1997). Parents and siblings as early resources for young children's learning in mexican-descent families. *Hispanic Journal of Behavioral Sciences, 19*(1), 3-33.

Piaget, J. (1954). *The construction of reality in the child*. London: Routledge and Kegan Paul.

Piaget, J., & Inhelder, B. (1974). *The child's construction of quantities : Conservation and atomism*. London: Routledge and Kegan Paul.

Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychon Bull Rev, 21*(5), 1112-1130.

Pinker, S. (2009). *How the mind works*. New York: W.W. Norton.

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*, 211-227.

Putnam, H. (1975). The meaning of 'meaning' (pp. 215-271): Cambridge University Press.

QSR International Pty Ltd. (2020). Nvivo.

Quine, W. V. (1960). *Word and object*. Cambridge, Massachusetts: MIT Press.

R Development Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Rescorla, M. (2019). The language of thought hypothesis. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2019 ed.). Stanford University: Metaphysics Research Lab.

Rey, G. (2014). Innate and learned: Carey, mad dog nativism, and the poverty of stimuli and analogies (yet again). *Mind & Language, 29*(2), 109-132.

Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior, 12*(1), 1-20.

Rogers, B. (2015). *Classroom behaviour : A practical guide to effective teaching, behaviour management and colleague support* (4th ed.). London: SAGE.

Rosch, E. (1975a). Cognitive representations of semantic categories. *Journal of experimental psychology. General, 104*(3), 192-233.

Rosch, E. (1975b). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104*(3), 192-233.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*(3), 382-439.

Rusanen, A.-M. (2014). Towards to an explanation for conceptual change: A mechanistic alternative. *Science and Education, 23*, 1413-1425.

Rusanen, A.-M., & Pöyhönen, S. (2013). Concepts in change. *Science and Education, 22*, 1389-1403.

Savinainen, A., & Scott, P. (2002). The force concept inventory: A tool for monitoring student learning. *Physics Education, 37*(1), 45-52.

Savinainen, A., & Viiri, J. (2008). The force concept inventory as a measure of students conceptual coherence. *International Journal of Science and Mathematics Education, 6*(4), 719-740.

Seah, L. H. (2015). Understanding the conceptual and language challenges encountered by grade 4 students when writing scientific explanations. *Research in Science Education*, 1-25.

Seah, L. H., Clarke, D., & Hart, C. (2015). Understanding middle school students' difficulties in explaining density differences from a language perspective. *International Journal of Science Education, 37*, 2386-2409.

Seah, L. H., Clarke, D. J., & Hart, C. E. (2011). Understanding students' language use about expansion through analyzing their lexicogrammatical resources. *Science Education, 95*(5), 852-876.

Seah, L. H., Clarke, D. J., & Hart, C. E. (2014). Understanding the language demands on science students from an integrated science and language perspective. *International Journal of Science Education, 36*, 952-973.

Sen, P., Dasgupta, S., Chatterjee, A., Sreeram, P. A., Mukherjee, G., & Manna, S. S. (2003). Small-world properties of the Indian railway network. *Physical Review E, 67*(3), 036106.

Sherin, B. (2017a). Elements, ensembles and dynamic constructions. In T. Amin & O. Levrini (Eds.), *Converging perspectives on conceptual change : Mapping an emerging paradigm in the learning sciences*. London: Routledge.

Sherin, B. (2017b). The nature of concepts and conceptual change. In T. Amin & O. Levrini (Eds.), *Converging perspectives on conceptual change : Mapping an emerging paradigm in the learning sciences*. London: Routledge.

Slotte, V., & Lonka, K. (1999). Spontaneous concept maps aiding the understanding of scientific concepts. *International Journal of Science Education, 21*, 515-531.

Smith, C. (2007). Bootstrapping processes in the development of students' commonsense matter theories: Using analogical mappings, thought experiments, and learning to measure to promote conceptual restructuring. *Cognition and Instruction, 25*, 337-398.

Smith, C., Carey, S., & Wiser, M. (1985). On differentiation: A case study of the development of the concepts of size, weight, and density. *Cognition, 21*(3), 177-237.

Smith, C., Maclin, D., Grosslight, L., & Davis, H. (1997). Teaching for understanding: A study of students' preinstruction theories of matter and a comparison of the effectiveness of two approaches to teaching about matter and density. *Cognition and Instruction, 15*(3), 317-393.

Song, Y., & Carheden, S. (2014). Dual meaning vocabulary (DMV) words in learning chemistry. *Chemistry Education Research and Practice, 15*, 128-141.

Spelke, E. S. (2000). Core knowledge. *American Psychologist, 55*(11), 1233-1243.

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological review, 99*(4), 605-605.

Spelke, E. S., Kestenbaum, R., Simons, D. J., & Wein, D. (1995). Spatiotemporal continuity, smoothness of motion and object identity in infancy. *British Journal of Developmental Psychology, 13*(2), 113-142.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science, 10*(1), 89-96.

Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods, 38*(4), 598-605.

Steinberg, R. N., & Sabella, M. S. (1997). Performance on multiple-choice diagnostics and complementary exam problems. *The Physics Teacher, 35*(3), 150-155.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science, 29*, 41-78.

Strike, K. A., & Posner, G. J. (1992). A revisionist theory of conceptual change. In R. A. Duschl & R. J. Hamilton (Eds.), *Philosophy of science, cognitive psychology, and educational theory and practice* (pp. 147-176). Albany, New York: State University of New York Press.

Swirski, H., Baram-Tsabari, A., & Yarden, A. (2018). Does interest have an expiration date? An analysis of students' questions as resources for context-based learning. *International Journal of Science Education, 40*(10), 1136-1153.

Tan, A. L. (2009). *The discourse of secondary school biology: Inter-relating interactional features and teachers' theories*. (Unpublished doctoral dissertation), Nanyang Technological University, Singapore.

Taylor, R. D. (2012). *Language in the classroom: Exploring students' use and understanding of key scientific terms*. (Unpublished Master's Dissertation), University of Oxford, Oxford.

Taylor, R. D., & Hillier, J. (2018). Developing science explanations in the classroom: The role of the written narrative. In A. Childs & G. J. Stylianides (Eds.), *Classroom-based interventions across subject areas: Research to understand what works in education*. Abingdon, UK: Routledge.

Thagard, P. (2000). *Coherence in thought and action*. Cambridge, Massachusetts: MIT Press.

Topp, N. W., & Pawloski, B. (2002). Online data collection. *11*, 173-178.

Treagust, D. F., & Duit, R. (2008). Conceptual change: A discussion of theoretical, methodological and practical challenges for science education. *Cultural Studies of Science Education, 3*, 297-328.

Van Zele, E., Lenaerts, J., & Wieme, W. (2004). Improving the usefulness of concept maps as a research tool for science education. *International Journal of Science Education, 26*, 1043-1064.

Voitalov, I., van der Hoorn, P., van der Hofstad, R., & Krioukov, D. (2019). Scale-free networks well done. *Physical Review Research, 1*(3), 033034.

Vosniadou, S. (2017). Initial and scientific understandings and the problem of conceptual change. In T. Amin & O. Levrini (Eds.), *Converging perspectives on conceptual change : Mapping an emerging paradigm in the learning sciences*. London: Routledge.

Vosniadou, S., & Skopeliti, I. (2014). Conceptual change from the framework theory side of the fence. *Science & Education, 23*(7), 1427-1445.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica, 57*(2), 307.

Vygotsky, L. S. (1986). *Thought and language* (A. Kozulin, Trans.). Cambridge, Massachusetts: MIT Press.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*, 440-442.

Weiskopf, D. A. (2023). The theory-theory of concepts. Retrieved from https://iep.utm.edu/

Wellington, J. J., & Osborne, J. (2001). *Language and literacy in science education*. Buckingham: Open University Press.

Wertsch, J. V. (1991). *Voices of the mind : A sociocultural approach to mediated action*. Cambridge, Massachusetts: Harvard University Press.

Willinger, W., Alderson, D., & Doyle, J. C. (2009). Mathematics and the internet: A source of enormous confusion and great potential. *Notices of the American Mathematical Society, 56*(5), 586-599.

Wiser, M., & Amin, T. G. (2007). Microworld-based interactions for conceptual change in science. In M. Limon & L. Mason (Eds.), *Reconsidering conceptual change. Issues in theory and practice.* (pp. 357-338). Dordrecht/London: Kluwer Academic Publishers.

Wiser, M., & Carey, S. (1983). When heat and temperature were one. In D. Gentner & A. L. Stevens (Eds.), *Mental models*. Hillsdale, N.J.: Erlbaum.

Yang, Z., Algesheimer, R., & Tessone, C. J. (2016). A comparative analysis of community detection algorithms on artificial networks. *Scientific Reports, 6*(1), 30750.

10 Appendices

10.1 CUREC Approval

The pilot study (described in appendices 1 and 2) had CUREC approval. The proposed methodology for the main study is the same as the pilot study. Therefore, an extension of the original CUREC approval to cover the timeframe of the main study was asked for and approved.

The original CUREC 1A form and original supporting materials are given below.

10.1.1 CUREC 1A Application Form

CENTRAL UNIVERSITY RESEARCH ETHICS COMMITTEE (CUREC)

Form CUREC 1A Checklist for the Social Sciences and Humanities

The University of Oxford places a high value on the knowledge, expertise, and integrity of its members and their ability to conduct research to high standards of scholarship and ethics. The research ethics clearance procedures have been established to ensure that the University is meeting its obligations as a responsible institution.

They start from the presumption that all members of the University will take their responsibilities and obligations seriously and will ensure that their research on human subjects is conducted according to the established principles and good practice in their fields and in accordance, where appropriate, with legal requirements. Since the requirements of research ethics review will vary from field to field and from project to project, the University accepts that different guidelines and procedures will be appropriate.

- Please check the **CUREC flowchart and NHS flowchart** first to see if you need ethics approval.
- Please complete this form using a word processor and email it, together with your supporting documents, to your **Departmental Research Ethics Committee (DREC)** (**if applicable**). If you don't have a DREC please email this form to **ethics@socsci.ox.ac.uk** using your official **ox.ac.uk email** address. **Only emailed applications will be accepted.**

WHAT THIS CHECKLIST IS DESIGNED FOR

This **CUREC 1A checklist** is designed largely for research that falls within the Divisions of Social Sciences and Humanities where ethical issues are relatively few and straightforward. Interviews, field work and oral history are also included in the CUREC process.

The **full CUREC 2 application** is only required where certain project characteristics (e.g. type of participants, or procedures) result in a more complex set of ethical issues. It is expected that only in a limited number of cases will it be necessary for researchers to complete a CUREC 2 application. The checklist below will direct you to a CUREC 2 application if needed.

WHAT THIS CHECKLIST WILL NOT ASSESS

This checklist does not cover research governance, satisfactory methodology, or compliance with the requirements of publishers when administering their tests or questionnaires. As principal researcher, it is your responsibility to ensure that requirements in these areas are met.

CUREC does not review studies classed as **audit** (see Glossary and Decision Flowchart for CUREC on our website).

Further information on the University's research ethics procedures are available from the CUREC website: www.admin.ox.ac.uk/curec.

SECTION A: Filter for CUREC2 application This section determines whether your study raises more complex issues which require the completion of a full application for ethical review, known as the CUREC 2 application. **(Please mark 'X' in the Yes/No column as appropriate to indicate your response.)**		
1. Are research participants classed as **people whose ability to give free and informed consent is in question**? (This may include those under 18 (though see "competent youths" in **FAQ C12**), prisoners, or adults "at risk".) Your attention is drawn to the University's **Safeguarding Code of Practice** and its implications for researchers involving children or adults at risk, including the need for the work to be risk assessed and for researchers to undertake related training. (**Note**: If any of your participants are aged 16 or under, please answer 'Yes' here **and** also answer question 5 below.)	Yes	No X
2. By taking part in the research, will participants be at serious risk of criminal prosecution (e.g. by providing information on drug abuse or child abuse)?	Yes	No X
3. Does the research involve the **deception** of participants?	Yes	No X
4. Does your research raise issues relevant to the **Counter-Terrorism and Security Act (the Prevent duty)**, which seeks to prevent people from being drawn into terrorism? Please see advice on this on our **Best Practice Guidance web page**.	Yes	No X
If you have answered 'No' to all of the questions above please go to **Section B**. If you have answered 'Yes' to any question above continue to question 5 below.		
5. Is your project covered by a CUREC **approved procedure** (formerly known as "CUREC Protocols")?	Yes	No
If yes, please give research procedure number(s):		
If you answered '**Yes'** to **ANY** of questions **1-4**, **and** answered '**No'** to question **5**, **please stop completing this checklist and do not submit it for ethical review**. Instead, please complete the CUREC 2 application form from the CUREC website. Then submit the CUREC 2 form for ethical review. If you answered 'Yes' to ANY of questions 1-3, **and** answered 'Yes' to question 5, please go on to **Section B**.		

SECTION B: Contact details and project description (NB: must be typed not handwritten)	
Contact details:	
1. **Principal researcher/**supervisor (title and name) (if student research):	Dr Judith Hillier (supervisor)
2. Name of student (if student research):	Mr Richard Taylor (student)
3. Degree programme, e.g. DPhil, BA, MPhil, BSc, MSc (if student research):	DPhil
4. Department or Institute name:	Department of Education
5. Address for correspondence (if different from 4 above):	
6. University e-mail (**not** private email) and telephone contact:	richard.taylor@education.ox.ac.uk
7. Name and status of others taking part in the project, e.g. third year undergraduate; postdoctoral research assistant:	Dr Ann Childs (supervisor)

SECTION B continued	
Project description:	
8. Title of research project:	The role of language in developing scientific concepts
9. List of location(s) where project will be conducted:	Gillingham School, Hardings Lane, Gillingham, Dorset, SP8 4QP.
10. If your research involves overseas travel or fieldwork, by the time the research starts, will you have completed and returned a travel risk assessment form? (This has to be approved by your department before you travel. If you are travelling overseas, you are strongly advised to take out University travel insurance.)	Yes No N/A X
11. Anticipated duration of research project overall:	0 months or 1 years (maximum 5)
12. Anticipated start and end dates of the research project involving human participants:	From: (01/11/16) To: (31/10/17) Please note that you will need ethics approval **before** you start your research. CUREC1As may take up to 30 days to process.

13. Title and very brief and simple lay description of **research** (about 150 words), plus description (about 200 words) of the nature of participants.

a) Title and brief lay description of research (150 words). When describing the research, please include your methodology, how you are applying professional guidelines, and the use to which results/data will be put. **Please also declare any conflicts of interest here.**

The role of language in developing scientific concepts

This is a pilot study to investigate the efficacy of using coordination classes (diSessa, 2002) to describe how students' understanding of scientific concepts (e.g. force, acceleration) evolve as a result of tuition. Coordination classes specify that an individual possesses a particular concept when they can describe and explain physical phenomena in given way. Hence, semi-structured interviews will be used to explore how groups of two Year 12 students describe different examples of physical phenomena and the explanations they give for these phenomena. These interviews will be recorded to enable any dialogue to be transcribed to allow for a detailed analysis of any descriptions and explanations.

diSessa (2002) implies the acquisition of a particular coordination class requires an individual to develop a complex network of knowledge which connects relevant ideas. A modified word-association test in a second series of interviews will be used to investigate each student's 'network of knowledge'. This methodology is described in greater detail in the supporting documents section.

Explicit written consent for participation in the interviews and for audio recordings to be made will be obtained.

The data and results from this pilot study will be used to inform the research design for a wider DPhil study of students' understanding of scientific concepts.

As a part-time student, the DPhil researcher is also a teacher at the target school and is therefore a member of the Gillingham School Community. Whilst this will be an advantage in ensuring participants feel at ease during the research process, (BERA Guidelines for Ethical Research, Article 20), this potentially raises ethical issues highlighted within this guidance (Article 12). These issues will be addressed in section 14.

SECTION B continued

b) Description of participants and obtaining informed consent (200 words). When describing participants, please include

- criteria for inclusion/exclusion
- method of recruitment
- processes for consent to participate

Please ensure you attach as separate documents (if applicable, in English translation):

- your recruitment and advertisement material e.g. a poster or brief invitation letter/ email
- information for participants to read (or hear) before they agree to take part e.g. written participant information sheets or (only if applicable) oral information scripts
- a document to record informed consent. Templates for written consent forms and/or oral consent scripts (in case of an oral consent process) are available from the CUREC website
- a guide to interview questions (this may be a list of questions to be asked, or a preliminary scope of questions), or a sample of other instruments (such as a sample questionnaire)
- (if relevant) debriefing document after participants have taken part

Criteria for inclusion/exclusion

Year 12 and 13 students who are developing normally will be the focus of this pilot study. Within this group of students, two further target groups will be identified.

- Year 12 and 13 students who are studying A-level Physics. This group is selected as presumably their understanding of the coordination classes such as *force* and *acceleration* will have developed during their A-level tuition.
- Year 12 and 13 students who are studying A-level Psychology but not A-level Physics. The entry requirements to study Psychology at A-level require a grade C in GCSE Science. Therefore, it is a reasonable assumption they have some understanding of coordination classes such as *force* and *acceleration*, but this understanding will not have been enhanced by A-level tuition. Therefore, this group is selected to provide a contrast with the A-level Physics group.

Both of these groups also represent stakeholders who have the potential to benefit from this research. In terms of possibly identifying misconceptions and improvements to pedagogy (see section 14) for the A-level Physics group and participation in a study investigating learning for the A-level Psychology group. These benefits will be given as possible reasons for students to become participants (see below).

Method of recruitment

Year 12 and 13 students studying A-level Physics and Psychology (and not A-level Physics) will be invited to 'opt-in' to this research. To allow students to make this decision, a short presentation to the students will introduce the study. This presentation (see additional material) will outline the rationale behind the study and the possible reasons for students to participate. However, to allow students to make an informed choice, they will be made aware of the likely time commitments expected of them, given examples of interview scenarios (including exemplar questions) and informed of who their data will be shared with and for what purpose (see section 14). It will be stressed that no student is under obligation to participate in this study and if they do decide to take part, they can withdraw at any time and refuse consent for any data collected to be used (see section 14). This information will also be provided for students in a letter, which also provides students with the opportunity to give their written consent (see attached material). Students will opt-in to this research by returning this letter.

Process for consent to participate

Lorna Lyons, the Head Teacher of Gillingham School, will be asked to give permission for this study to be carried out. A copy of the consent letter that will be sent to Mrs Lyons is attached to this application.

The Year 12 and 13 students being asked to participate in this study will be aged between 16 and 18 years. In addition, this study will use techniques that are likely to be reasonably familiar to the students through the usual monitoring of their understanding by their teachers (for example, student will be familiar with being ask to explain

whether they think an object is accelerating by their teacher). It also fits the guidelines published under "Best practice guidance: research involving competent youths". Hence, only the explicit written consent of the participants will be obtained. However, parents of participants will be informed of their child's involvement in this study and will be given information about the study.

14. What are the ethical issues connected with your research and what steps have you taken to address them? Please do not answer 'none'. The committee needs to see evidence that you have identified potential ethical issues with respect to your research and have taken steps to address them. These issues could relate to:

- your own physical and psychological safety as a researcher (please see the University's and Social Science Division's Safety in Fieldwork guidance
- participant burdens and/or risks, and
- data protection/ confidentiality.

For more guidance on ethical issues, please see www.admin.ox.ac.uk/curec/resources/.

The student researcher's dual role as a teacher at Gillingham School and a DPhil student may introduce explicit tensions (BERA Ethical Guidelines for Educational Research, Article 12) which could raise a number of ethical issues. These issues are anticipated and addressed below.

This study could identify individual or systematic misconceptions in topics taught by the researcher or his colleagues. These misconceptions have the potential to restrict the future attainment of the participants or other students within the community. Therefore, if systematic misunderstandings are discovered, this information will be feedback anonymously to the teachers of the participants. These teachers can then deal appropriately with these misconceptions and make any necessary changes to future teaching of these topics. No judgments about individual teaching will be made. In addition, the general findings of the research will be shared with the teachers of Gillingham School with attention to the limitations of the research (in accordance with BERA Ethical Guidelines for Educational Research, Article 38).

In the case of individual misconceptions, the DPhil researcher (as potentially a teacher of the participants) would have a responsibility to try to correct these misconceptions. Consequently, participants will be debriefed individually at the end of the research process. In this session, one-to-one tuition will be given by the principal researcher to help participants overcome any misconceptions they might have. However, this risks participants being given an advantage over students who do not take part in this research. Therefore, these individual misconceptions will also be feedback anonymously to teachers to raise their awareness of possible misunderstandings within the group. In addition, the participants will be made aware of the findings of the research as a group (in accordance with BERA Ethical Guidelines for Educational Research, Article 31).

In his role as a teacher at Gillingham School, the DPhil researcher is required to assess students to measure their attainment and report this data in various ways (working at grades, UCAS predicted grades, attitude to learning schools). It will be made clear to students who wish to participate that this research process will not contribute to these forms of assessment.

As a member of the Gillingham School community, it would be difficult for the participants to remain anonymous to the student researcher. Moreover, as a teacher at Gillingham School, he would be expected to use students' names in any interaction to establish positive relationships. This is consistent with Article 20 of the BERA Ethical Guidelines for Educational Research which states participants should be made to feel at ease. However, the participants of this study, including students, teachers and the school, will be made anonymous in all subsequent data analysis and research reports. All the data collected will be kept strictly confidential and available only to the student researcher and his supervisors. Any audio recordings will be stored securely on a password-protected computer, which is kept in a single locked location, and deleted at the end of this project.

It is necessary not to place a 'bureaucratic burden' on the participants (BERA Ethical Guidelines for Educational Research, Article 21) or risk impacting on their attainment by significantly altering their normal school timetable (BERA Ethical Guidelines for Educational Research, Article 24). Hence, all interviews will take place within participants 'free sessions' and will not exceed more than four sessions. Participants will be offered a choice for when these sessions take place.

Section B continued		
15. Will you obtain informed consent according to CUREC guidelines and good practice in your discipline before participation?	Yes X	No
If you have marked '**No**', please give a brief explanation and justification for this decision here:		
16. Will your research involve discussing sensitive issues? This could be information relating to race or ethnic origin, political opinions, religious beliefs, physical/mental health, trade union membership, sexual life or criminal activities.	Yes	No X
If you have marked '**Yes**', please make sure that you have included some **supporting information** (as directed in question 13 of this section) showing the range of questions covering these issues.		
17. Will you ensure that **personal data** collected directly from participants or via a **third party** is held and processed in accordance with the provisions of the Data Protection Act?	Yes X	No
18. How will you ensure that any personal or sensitive data are captured, transferred and stored securely? In particular if data are to be captured electronically, please consult with your local IT department and, with respect to University IT security policies, please comment on how you will capture such data in the first instance, how you will transfer them over networks or via portable media and how and where data will be stored. For more information please see the University's web pages on research data management: http://researchdata.ox.ac.uk/university-of-oxford-policy-on-the-management-of-research-data-and-records/		
Data will be captured, transferred and stored securely per Approved Procedure 25. No reference to individuals will be made within the final report to uphold anonymity. All consent forms and individual paper records will be kept in a locked filing cabinet. Anonymity data keys will be stored separately from datasets. The researchers will store electronic anonymized data on password encrypted storage media. The researcher's supervisor will have copies of this data, also password encrypted. Electronic data will be stored indefinitely, while hard copy data will be destroyed five years after the project. There is no perceived harm, risk, or possible injury anticipated from this proposed study.		

SECTION C: Methods and procedures to be used	
Method used: Please ensure you have addressed any potential ethical issues related to these methods in Section 14 and in your Participant Information Sheet	**Please mark 'X'**
1. Analysis of existing records	
2. Snowball sampling (recruiting through contacts of existing participants)	
3. Use of casual or local workers e.g. interpreters	
4. Participant observation	
5. Covert observation	
6. Observation of specific organisational practices	
7. Participant completes questionnaire in hard copy	
8. Participant completes online questionnaire or other online task	
9. Using social media	
10. Participant performs paper and pencil task	X
11. Participant performs verbal or aural task (e.g. for linguistic study)	
12. Focus group	
13. Interview	X
14. Audio recording of participant (you will generally need specific consent from participants for this)	X
15. Video recording of participant (you will generally need specific consent from participants for this)	
16. Photography of participant (you will generally need specific consent from participants for this)	
17. Others (please specify):	

SECTION D: Professional guidelines and training		
In this section, please mark 'X' against at least one of the following professional guidelines you aim to adhere to. You should use the principles listed in your chosen guideline(s) in conducting your own research. **Note**: this is not an exhaustive list.		**Please mark 'X'**
Research specialism/ methodology	**Association and guidance document**	
Anthropology	Association of Social Anthropologists of the UK and Commonwealth	
Criminology	British Society of Criminology: Code of Ethics for Researchers in the Field of Criminology	
Education	British Educational Research Association Ethical Guidelines for Educational Research	X
Geography	Association of American Geographers Statement on Professional Ethics Royal Geographical Society: Research Ethics and Code of Practice	
History	Oral History Society of the UK Ethical Guidelines	
Internet-based Research	British Psychological Society: Conducting Research on the Internet Association of Internet Researchers Ethics Guide Also see our Best Practice Guidance on internet-based research	
Law (Socio-Legal)	Socio-Legal Studies Association: Statement of Principles of Ethical Research	
Management	Academy of Management's Professional Code of Ethics	
Political Science	American Political Science Association (APSA) Guide to Professional Ethics in Political Science	
Politics	Political Studies Association. Guidelines for Good Professional Conduct	
Psychology	British Psychological Society Code of Ethics and Conduct British Psychological Society: Conducting Research on the Internet Also see "Internet-based Research" guidance above	
Social Research	Social Research Association: Ethical Guidelines	
Sociology	The British Sociological Association: Statement of Ethical Practice	
Visual Research	ESRC National Centre for Research Methods Review Paper: Visual Ethics: Ethical Issues in Visual Research	
Other professional guidelines. Please specify the other guidelines used here*:*		

Please indicate what training in research ethics the researchers involved with this study have received, e.g. the title of the course and date completed (online training available at www.admin.ox.ac.uk/curec/training).

If no formal training has been undertaken, please indicate any discussions of research methodology between researchers and supervisors here.

The DPhil researcher (Richard Taylor) has undertaken research ethic training during his MSc and DPhil studies.

MSc in Learning and Teaching

Masters in learning and teaching seminar taught by May Cheng and Anne Edwards (Summer 2011).

Masters in learning and teaching seminar taught by Nigel Fancourt and Anne Edwards (October 2011).

DPhil

Foundation in Educational Research (FER) Research Ethics Session (Michaelmas Term 2015)

In addition, the DPhil researcher and has supervisors have discussed the ethical dimensions of this research (e.g. appropriate timing for interviews to avoid bureaucratic burden and participant selection) during supervisions which have taken place in Trinity term and Michaelmas term 2016.

SECTION E: Signatures

- 'Electronic signatures' sent as email confirmations from a University of Oxford email address can be accepted. Separate emails should come from each of the relevant signatories as outlined below, indicating acceptance of the relevant responsibilities.
- If you have obtained handwritten (wet-ink) signatures, please scan them and the rest of the checklist pages to create a single PDF document and email through.

Please ensure this checklist is signed by:

For staff research:	**For student research**:
1. **Principal researcher**	1. **Principal researcher (project supervisor)**
2. **Head of Department (or nominee)**	2. **Head of Department (or nominee)**
	3. **Student researcher**

1. **Principal researcher signature/supervisor signature** (if student research)

I understand my responsibilities as **principal researcher** as outlined in the CUREC glossary and guidance on the CUREC website.

I declare that the answers above accurately describe the research as presently designed, and that a new checklist will be submitted should the research design change in a way which would alter any of the above responses so as to require completion of CUREC 2 (involving full scrutiny by an IDREC). I will inform the relevant IDREC if I cease to be the principal researcher on this project and supply the name and contact details of my successor if appropriate.

Signature: ……

Print name (block capitals): …….. **Date**: ……..

2. **Departmental endorsement signature**

I have read the research project application named above. On the basis of the information available to me, I:

(i) consider the principal researcher to be aware of her/his ethical responsibilities in regard to this research;
(ii) consider that any ethical issues raised have been satisfactorily resolved or are covered by relevant professional guidelines and/or CUREC approved procedures, and that it is appropriate for the research to proceed (noting the principal researcher's obligation to report should the design of the research change in a way which would alter any of the above responses so as to require completion of a CUREC 2 full application);
(iii) am satisfied that: the proposed project design and scientific methodology is sound; the project has been/will be subject to appropriate **peer review**; and is likely to contribute to existing knowledge and/or to the education and training of the researcher(s) and that it is in the **public interest**.

Signed by Head of Department or nominee *(example nominees for student research include the Director of Graduate Studies/ Director of Undergraduate Studies)*:

Signature: ……

Print name (block capitals): …….. **Date**: ……..

3. **Student signature** (if student research)

I understand the questions and answers that have been entered above describing the research, and I will ensure that my practice in this research complies with these answers, subject to any modifications made by the principal researcher properly authorised by the CUREC system.

Signed by student: Richard Taylor (electronic signature)

Date: 10th October 2016

Print name (block capitals): Richard Taylor

SECTION F: SUBMITTING THE COMPLETED CHECKLIST	**Please mark 'X'**
1. Check you have completed all sections (A-E)	X
2. Ensure your application is signed by you, your supervisor (if student) and department	X
3. **Please attach all supporting documents (see section B, question 13b for details).** If the appropriate supporting documentation is not included with your application, you will then be asked to provide this separately. **This may well delay the ethical review process, and thus the start of your research.**	X
4. Ensure you have declared conflicts of interest (if any) in Section B, question 13.	X
5. If your department has a Departmental Research Ethics Committee (DREC), submit this checklist and supporting information to the appropriate departmental officer.	X
6. If your department does not have a DREC, submit the checklist and supporting information to the SSH IDREC (email ethics@socsci.ox.ac.uk).	
7. Applications must be sent by email from your official ox.ac.uk email account. Please do not send applications by post.	X

10.1.2 Headteacher Consent Letter

University of Oxford
Department of Education
15 Norham Gardens, Oxford OX2 6PY
Director: Jo-Anne Baird
Tel: +44 (0)1865 274024 Fax: +44 (0)1865 274027

[redacted]

I write to ask for permission to conduct research at [redacted] part of my DPhil in Education at the University of Oxford. My doctoral research project, *The role of language in the development of scientific concepts,* explores how a student's understanding of scientific concepts influences the descriptions and explanations they give of simple natural phenomena. My supervisors for this project are Dr. Judith Hillier and Dr. Ann Childs.

The research will take place with Year 12 and 13 Physics and Psychology students. These students have been selected as students who potentially benefit from involvement in this research, either through a greater understanding of the process of learning Physics or from participating in a study concerned with learning. Also, these two groups also provide a contrast between two different levels of expertise in Physics. Students will be given the opportunity to opt-in to this research after being given information about the study.

I intend to use interviews with groups of two students to assess how the students describe and explain simple observable phenomena (e.g. a falling lump of plasticine or a dynamic trolley moving on a ramp). I would like to take notes and audio recordings of of these interviews. In addition, I would like to ask students to list words that help them to understand scientific concepts. Then students will be asked to list words they associate with their first words to create a 'snowball' sample. In this way, I am hoping to map the network of terminology students use to help them understand scientific concepts. Of course, both of these activities would be subject to each student's consent, which they can withdraw at any time and without reason. Their parents or guardians will also be informed of their child's decision to participate and given information about the study.

By participating in the research, our school would help to inform our understanding of how our students learn Physics. I would disseminate this understanding to other members of the Science Department, which would contribute to our on-going goal of finding ways to improve the attainment of our students.

The University of Oxford has strict procedures on conducting ethical research with teachers and young people, consistent with current British Educational Research Association guidelines. All participants, including students and the school, would be made anonymous in all research reports. The data collected would be kept strictly confidential, available only to my supervisors and myself, and not used for any purpose that has not been specified without obtaining the further consent of all participants. All audio recordings would be destroyed at the end of the research period, and kept in locked conditions until then.
If you feel you would like me to carry out this research in Gillingham School, I would be grateful if you could indicate your consent by ticking the approval box below.

Thank you for your time and attention.

Yours sincerely,

Richard Taylor.

The role of language in learning scientific concepts

Richard Taylor
Gillingham School and University of Oxford, Department of Education

I give permission for Gillingham School students to take part in the above study.

Signature: _____ **Date:** _____

10.1.3 Participant Consent Form

University of Oxford
Department of Education
15 Norham Gardens, Oxford OX2 6PY
Director: Jo-Anne Baird
Tel: +44 (0)1865 274024 Fax: +44 (0)1865 274027

Researcher: Richard Taylor
E-Mail: richard.taylor@education.ox.ac.uk

VOLUNTEER CONSENT FORM

CUREC Approval Reference:

The role of language in developing scientific concepts

- Your school has agreed to take part in a study run by Oxford University looking at how students learn Physics.
- If you take part, Mr Taylor will interview you with another student about how you describe and explain what happens in some simple demonstration. He will also ask you about words that help you understand scientific concepts, like *force*. Audio recordings of these interviews will be made.
- **If you are happy take part, please fill in the form below and return it to Mr Taylor as soon as possible.**
- To find out more about the study, please read the attached information sheet. You can also e-mail us at richard.taylor@education.ox.ac.uk, or Judith Hillier (Associate Professor in Science Education, judith.hillier@education.ox.ac.uk) if you have any questions.

Name: _____ _____
Forename Surname

I have read and understood the details of the above study, and have had the opportunity to ask questions and discuss the study with others. I have received satisfactory answers to my questions. I understand that the project has received ethics clearance through the University of Oxford's ethical approval process for research involving human participants, and understand who will have access to the data, how it will be stored and what will happen to the data at the end of the study. I understand that participation is voluntary and that I am free to withdraw at any time, without giving any reason and without my education being affected in any way. I understand how to raise a concern or make a complaint.

I give permission to take part in the above study.

Signature: _____ **Date:** _____

Name of researcher: _____

Signature: _____ **Date:** _____

10.1.4 Participant Information Sheet

University of Oxford

Department of Education

15 Norham Gardens, Oxford OX2 6PY

Director: Jo-Anne Baird

Tel: +44 (0)1865 274024 Fax: +44 (0)1865 274027

Student Researcher: Richard Taylor

E-Mail: richard.taylor@education.ox.ac.uk

http://www.education.ox.ac.uk/about-us/directory/richard-taylor/

The role of language in developing scientific concepts

PARTICIPANT INFORMATION SHEET

In partnership with researchers at the University of Oxford, your school has agreed to take part in a study to investigate the development of scientific concepts (e.g. force, acceleration). We would like to invite you to be part of this study. We very much hope you would like to take part, but before you decide, it is important that you understand why the study is being done and what it will involve.

What are we trying to find out?

This research is part of a project is looking at how secondary school students learn Physics. We are especially interested in how students learn to describe and explain things they see in the natural world (e.g. a falling ball or a car speeding up) in the same way as expert scientists. Many researchers suspect our understanding of scientific concepts (e.g. force, acceleration) determines what descriptions and explanations we give of things we see in the world. We also know that an expert scientist's understanding of these concepts is different from other peoples' understanding. Unfortunately, we do not know in detail how someone's understanding of these concepts changes as they learn scientific ways to describe and explain. Researchers also disagree about the best way to measure these changes. We hope our findings will allow us to find a method for studying how someone's understanding of scientific concepts changes when they learn. We will then be able to use this method to help us understand the role scientific concepts play in learning scientific descriptions and explanations.

You can obtain more information about this project by contacting the research team (contact details overleaf)

What will happen if you take part?

You will be asked to take part in two types of interview.

- Interview 1
  You will be shown a demonstration (e.g. a toy car rolling down a ramp). You will then be asked to describe and then explain what you have seen. It is likely you will be asked some questions by the interviewer to help them understand your thinking.

- Interview 2
  You will be asked to list words that help you understand a particular scientific concept (e.g. chemical). You will then be asked to think of words that help you understand these words and so on.

Mr Taylor will conduct these interviews. Another student will be interviewed with you. You will be asked to give up a number of your private study sessions for these interviews. However, you will be free to choose when these sessions will be and you will not be expected to give up more than four. Audio recordings of these interviews will be made.

What happens to the results of the study?

Individual results and recordings are kept strictly confidential and only used for the purposes outlined in this information sheet. A code number will be used to identify you to members of the research team, and all information and results are kept in a locked filing cabinet in the University. All audio recordings will be destroyed at the end of the project. A summary of our findings will be given to you and the school. We also aim to publish our findings in scientific journals, but this may be two to three years from the end of the study.

What are the any benefits or risks of participating in this research?

By participating in this research, you will contribute to our understanding of how students of your age learn Physics. This understanding will be shared with your teachers. Hence, this research has the potential to help other students in your school community learn Physics by giving their teachers a better understanding of this learning process. In addition, this research has the potential to reveal misconceptions you might have about aspects of Physics. If this is the case, you will be given tuition to help you correct these misconceptions if you wish.

If you are a Psychology student, this is an opportunity to take part in a study which looks at aspects of learning. What we are trying to find out about learning, why we have chosen this methodology and the findings of this research will be shared with you after the study has taken place.

The risks involved with this study are low as the questions you will be asked are similar to the questions you would be asked by your teacher as part of everyday lessons. You will not be asked any personal questions or asked to make judgements about your teachers. However, it is important to understand that you might be asked searching questions about your understanding of Physics. This can be uncomfortable for some people. So you have the right to withdraw from the research at any time and without reason. Also, it is important to stress that information from this study will not be used to inform any formal assessment your teachers may make about your attainment in your A-level subjects (e.g. working at grades, UCAS predicted grades).

Who is conducting this research?

This research project is organized by Richard Taylor, who is a DPhil student at the University of Oxford (supervised by Judith Hillier and Ann Childs), and a Physics teacher at Gillingham School. The research is funded by the Economics and Social Science research council. This study has received ethics clearance through the University of Oxford's Central University Research Ethics Committee.

What if there is a problem?

If you have a concern about any aspect of this project, please speak to the relevant researcher (Richard Taylor, richard.taylor@education.ox.ac.uk) or his supervisor, Judith Hillier (judith.hillier@education.ox.ac.uk), who will do her best to answer your query. The researcher should acknowledge your concern within 10 working days and give you an indication of how he intends to deal with it. If you remain unhappy or wish to make a formal complaint, please contact the chair of the Research Ethics Committee at the University of Oxford (Chair, Social Sciences and Humanities, Inter-Divisional Research Ethics Committee; Email: ethics@socsci.ox.ac.uk; Address: Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD). The chair will seek to resolve the matter in a reasonably expeditious manner.

What should I do next?

Please fill in the enclosed form and return it to Mr Taylor if you would like to take part in this study. Please remember that you may withdraw at any time, without penalty and without giving a reason by notifying Mr Taylor or Dr Hillier. If you decide to take part, please take home a copy of this information sheet to your parents, guardians or carers to inform them of your choice and provide them with details of the research you have chosen to participate in.

If you would like to discuss the research with someone other than Mr Taylor beforehand (or if you have questions afterwards), please contact:

Dr Judith Hillier,
Associate Professor in Science Education (Physics),
Oxford University Department of Education,
15 Norham Gardens,
Oxford,
OX2 6PY,
Email: judith.hillier@education.ox.ac.uk

10.2 Merged Word Tables

Table 10.1 Merged words for each participant (plural).

Participant	Merged Word	Vertex (Sign)	Frequency of Merge	Stimulus Words
12P02	degrees	degree	1	scale
12P02	dimensions	dimension	6	cubic deformation size space structure volume
12P02	downwards	downward	2	falling weight
12P02	edges	edge	1	shape
12P02	events	event	2	process time
12P02	forces	force	1	intermolecular
12P02	happens	happen	1	occurrence
12P02	lines	line	1	axis
12P02	molecules	molecule	1	intermolecular
12P02	numbers	number	3	counting mathematical scale
12P02	particles	particle	4	fundamental gas liquid temperature
12P02	points	point	1	length
12P02	products	product	1	reaction
12P02	readings	reading	1	observer
12P02	seconds	second	1	time
12P02	shapes	shape	1	dimension
12P02	steps	step	2	logical quantised
12P02	stores	store	1	transfer
12P02	towards	toward	3	attract closer pull
12P02	trends	trend	1	related
12P02	units	unit	3	counting discrete quantised
12P02	upwards	upward	1	height
12P02	vertices	vertex	1	shape
12P02	vibrations	vibration	1	temperature
12P02	waves	wave	3	interfere superpose superposition
12P03	atoms	atom	1	mass
12P03	baryons	baryon	1	strange
12P03	bosons	boson	2	effect exchange
12P03	changes	change	1	awareness
12P03	charges	charge	1	attraction
12P03	circumstances	circumstance	2	conditional luck
12P03	collisions	collision	2	energy momentum
12P03	colours	colour	1	dispersion

12P03	combusts	combust	1	alcohol
12P03	components	component	2	circuitry incline
12P03	constants	constant	1	calculation
12P03	currents	current	1	convection
12P03	differences	difference	1	comparison
12P03	dimensions	dimension	10	beyond density flat function measurement minus size unit unquantifiable volume
12P03	downwards	downward	1	weight
12P03	drugs	drug	3	driver reaction recreational
12P03	dynamic	dynamics	1	equilibrium
12P03	electrons	electron	20	atom build-up charge conductor discharge distribution electromotive element flow function ground leftover loss minimum overcome Planck potential shock static voltage
12P03	fermions	fermion	1	fundamental
12P03	forces	force	1	perpendicular
12P03	gluons	gluon	1	radioactive
12P03	hadrons	hadron	1	weak
12P03	homosapiens	homosapien	1	race
12P03	joules	joule	1	energy
12P03	leptons	lepton	1	weak
12P03	levels	level	1	excitation
12P03	masses	mass	1	attraction
12P03	mesons	meson	1	strange
12P03	metres	metre	1	highest
12P03	moments	moment	8	balance centre cradle principle rocking rotation tipping weight
12P03	neanderthal	neanderthal	1	human
12P03	neutrons	neutron	2	nucleus strong
12P03	nucleons	nucleon	3	binding radioactive strong
12P03	numbers	number	1	phase
12P03	objects	object	2	gravity relative
12P03	optics	optic	3	multipath refractive index transmission
12P03	packets	packet	4	catastrophe discrete Planck quantised
12P03	particles	particle	1	electromagnetic
12P03	patterns	pattern	1	matterwave
12P03	photons	photon	15	alcohol annihilation catastrophe combust disappear discrete electromagnetic energy excitation incident negative quantum threshold ultraviolet

12P03	planes	plane	4	displacement opposition perpendicular resultant
12P03	properties	property	2	comparison duality
12P03	protons	proton	5	element iron56 mass nucleus strong
12P03	pupils	pupil	1	dilation
12P03	quarks	quark	5	hadron mass proton radiation strong
12P03	results	result	1	perfect
12P03	slits	slit	1	diffraction
12P03	sources	source	1	double
12P03	springs	spring	1	mechanical
12P03	stores	store	8	conservation dissipation drop elastic energy joule stationary transfer
12P03	theories	theory	4	beyond breaking conventional difference
12P03	units	unit	5	calculation converting dimension small
12P03	vectors	vector	1	bearing
12P03	wavelengths	wavelength	2	dispersion refractive index
12P03	waves	wave	3	energy gamma phase
12P05	amps	amp	1	current
12P05	angles	angle	1	trigonometry
12P05	arms	arm	1	hand
12P05	asymptotes	asymptote	1	graph
12P05	atoms	atom	2	body combustion
12P05	bosons	boson	2	fundamental subatomic
12P05	calculations	calculation	2	applied trigonometry
12P05	cells	cell	1	tissue
12P05	centimetres	centimetre	1	ruler
12P05	circuits	circuit	1	resistance
12P05	collisions	collision	4	heat light particle threshold
12P05	components	component	2	circuit electricity
12P05	computers	computer	1	data
12P05	degrees	degree	1	angle
12P05	dials	dial	1	analogue
12P05	electrons	electron	7	carbon current electricity metal photoelectric potential difference water
12P05	equations	equation	7	coefficient equipment mathematics multiply Pythagoras trigonometry workings
12P05	fingers	finger	1	hand
12P05	forces	force	2	magnitude terminal
12P05	fringes	fringe	4	light slit superposition young
12P05	galaxies	galaxy	1	space

12P05	graphs	graph	3	analysis calculus statistic
12P05	hands	hand	1	clock
12P05	interactions	interaction	1	similar
12P05	joules	joule	4	elastic energy inelastic work
12P05	kilograms	kilogram	1	mass
12P05	lengths	length	1	trigonometry
12P05	masses	mass	1	space
12P05	measurements	measurement	1	trigonometry
12P05	metres	metre	3	displacement travelled wavelength
12P05	millimetres	millimetre	1	ruler
12P05	ohms	ohm	1	resistance
12P05	orbits	orbit	1	space
12P05	ores	ore	1	mining
12P05	oscillations	oscillation	2	medium wavelength
12P05	particles	particle	15	body Brownian collision combustion fundamental heat interaction natural reaction space subatomic terminal thermal tissue transfer
12P05	photons	photon	3	collision dark threshold
12P05	products	product	2	reaction yield
12P05	properties	property	2	same similar
12P05	radians	radian	2	angle phase
12P05	reactions	reaction	1	collision
12P05	results	resultant	7	analysis answer calculus experiment mean median practical
12P05	rocks	rock	3	geology geothermal ore
12P05	seconds	second	4	clock speed stopwatch time
12P05	sides	side	1	trigonometry
12P05	slits	slit	4	light superposition wave young
12P05	solutions	solution	1	multiply
12P05	statistics	statistic	1	analysis
12P05	times	time	1	multiply
12P05	triangles	triangle	2	Pythagoras trigonometry
12P05	units	unit	1	convention
12P05	vibrations	vibration	1	collision
12P05	wavefronts	wavefront	1	wave
12P05	waves	wave	5	bright frequency interference phase vibration
12P05	Youngs	young	1	slit
12P08	angles	angle	1	resolving
12P08	antiparticles	antiparticle	1	particle

12P08	antiquarks	antiquark	1	fundamental
12P08	areas	area	1	inhabitable
12P08	atoms	atom	3	decomposition density molecule
12P08	baryons	baryon	1	strong
12P08	bosons	boson	1	interaction
12P08	calculations	calculation	4	Pythagoras radius series suvat
12P08	carriers	carrier	2	power thermistor
12P08	circuits	circuit	1	electricity
12P08	coils	coil	3	anticlockwise magnetic opposing
12P08	collisions	collision	5	chain compression mechanics momentum thermistor
12P08	components	component	15	angle circuit composition electricity hypotenuse magnitude ohm perpendicular power Pythagoras resistor series slope total triangle
12P08	compressions	compression	1	shift
12P08	constants	constant	1	equation
12P08	degrees	degree	2	heat triangle
12P08	dimensions	dimension	4	area cubed hypotenuse volume
12P08	diodes	diode	1	directional
12P08	downwards	downward	5	elevator gravitational gravity heavy weight
12P08	electrons	electron	19	atom attraction current diffraction electricity electromagnetic energy gain ion ionisation magnetised oxidation release repulsion resistivity resistor shell thermistor voltage
12P08	equations	equation	3	numerical radius suvat
12P08	factors	factor	2	rate variable
12P08	fields	field	3	charged core electricity
12P08	forces	force	12	angle elevator equal equilibrium hypotenuse law sine standing terminal trigonometry unbalanced work
12P08	gluons	gluon	2	interaction strong
12P08	grams	gram	1	mole
12P08	gravitons	graviton	2	gravity interaction
12P08	hadrons	hadron	2	interaction small
12P08	interactions	interaction	1	gluon
12P08	ions	ion	1	thermistor
12P08	isotopes	isotope	2	decay uranium
12P08	kinetics	kinetic	1	radius
12P08	laws	law	2	conserved newton
12P08	lengths	length	1	triangle

12P08	leptons	lepton	3	fundamental particle small
12P08	levels	level	6	excitation ground ionisation photoelectric Planck shell
12P08	metals	metal	1	electromagnet
12P08	metres	metre	6	area deceleration radius stiffness unit wrench
12P08	molecules	molecule	2	density ozone
12P08	moles	mole	3	Avogadro concentration volume
12P08	moments	moment	9	balance clockwise core distance equal equilibrium lever resolving rope
12P08	neutrons	neutron	3	atom isotope mass
12P08	newtons	newton	1	law
12P08	ohms	ohm	3	resistance resistivity resistor
12P08	particles	particle	24	Avogadro boiling charge contamination decay electromagnet electromagnetic fundamental gas gravitational gravity hadron interaction internal liquid mole nuclear quark ratio release small solid state strangeness
12P08	photons	photon	6	absorbed emitted energy gain release small
12P08	planets	planet	12	doppler extra-terrestrial gravitational gravity inhabitable moon redshift relativity shift space universe weight
12P08	protons	proton	6	atom attraction electromagnetic isotope mass repulsion
12P08	quarks	quark	11	down fundamental gluon interaction neutron nucleus particle proton small strangeness up
12P08	radians	radian	6	angle clockwise cosine phase pi wave
12P08	rates	rate	1	concentration
12P08	reactions	reaction	3	catalyst equilibrium ratio
12P08	satellites	satellite	2	field gravitational
12P08	scale	scales	1	proportional
12P08	seconds	second	4	rate stopwatch time unit
12P08	shells	shell	3	electron ground level
12P08	springs	spring	1	energy
12P08	stars	star	1	universe
12P08	states	state	1	equilibrium
12P08	terminals	terminal	2	attraction opposite
12P08	units	unit	3	metric ohm watt
12P08	upwards	upward	2	elevator normal
12P08	values	value	1	measuring

12P08	variables	variable	3	change equation proportional
12P08	vectors	vector	4	hypotenuse resultant triangle trigonometry
12P08	watts	watt	1	power
12P08	wavelengths	wavelength	1	doppler
12P08	waves	wave	20	annihilation constructive critical diffraction directional emitted energy frequency interference light nuclear oscillate photoelectric polarise radio shift sound superimpose ultraviolet vibration
12P12	actions	action	2	activity behaviour
12P12	afterwards	afterward	1	before
12P12	aspects	aspect	1	identification
12P12	atoms	atom	1	molecule
12P12	behaviours	behaviour	1	conditions
12P12	bits	bit	1	information
12P12	borders	border	1	international
12P12	causes	cause	1	property
12P12	communications	communication	1	remote
12P12	conforms	conform	2	fluid liquid
12P12	controls	controlled	1	possesses
12P12	countries	country	1	international
12P12	differences	difference	1	generic
12P12	electrons	electron	2	nitrogen oxygen
12P12	holds	hold	1	possesses
12P12	items	item	2	set two
12P12	limits	limit	1	normal
12P12	liquids	liquid	1	stream
12P12	metres	metre	1	distance
12P12	molecules	molecule	1	liquid
12P12	neutrons	neutron	1	oxygen
12P12	objects	object	4	direction interaction separation system
12P12	operations	operation	1	base
12P12	oscillates	oscillate	1	wave
12P12	owns	own	1	possesses
12P12	particles	particle	4	gas matter ray uncountable
12P12	properties	property	2	conditions opposite
12P12	protons	proton	2	nitrogen oxygen
12P12	qualities	quality	1	object
12P12	records	record	1	memory
12P12	retains	retain	1	elastic
12P12	shows	show	1	pointer

12P12	sides	side	2	Pythagoras stereo
12P12	verifies	verify	1	identification
12P12	waves	wave	2	frequency ray
12P12	wheels	wheel	1	bicycle
12S01	activities	activity	1	procedural
12S01	atoms	atom	2	charge proton
12S01	brakes	brake	1	car
12S01	capitals	capital	2	country earth
12S01	careers	career	2	future occupation
12S01	coins	coin	1	pay
12S01	colours	colour	4	calm eye look sight
12S01	countries	country	4	capital earth geography world
12S01	days	day	1	occurring
12S01	degrees	degree	1	temperature
12S01	differences	different	1	compare
12S01	digits	digit	1	number
12S01	directions	direction	2	compass north
12S01	electrons	electron	4	charge electrical particle voltage
12S01	emotions	emotion	2	colour feeling
12S01	events	event	3	episodic timeline unpredictable
12S01	examinations	examination	1	test
12S01	eyes	eye	2	look sight
12S01	favours	favour	1	pay
12S01	feelings	feeling	1	emotion
12S01	forces	force	6	acceleration balance momentum movement opposite weight
12S01	forwards	forward	1	front
12S01	graphs	graph	1	maths
12S01	homes	home	1	world
12S01	hours	hour	3	day occurring time
12S01	instructions	instruction	2	obedience rule
12S01	jobs	job	2	career work
12S01	joules	joule	1	energy
12S01	languages	language	2	sentence word
12S01	letters	letter	2	capital word
12S01	lights	light	1	electricity
12S01	links	link	1	schema
12S01	memories	memory	1	event
12S01	minutes	minute	4	day hour occurring time
12S01	neutrons	neutron	2	electron particle

12S01	newtons	newton	1	law
12S01	numbers	number	8	graph maths measurement order pattern sequence test timeline
12S01	paragraphs	paragraph	1	capital
12S01	particles	particle	10	air atom collision friction kinetic movement neutron proton solid thermal
12S01	plans	plan	1	career
12S01	protons	proton	4	charge electrical electron particle
12S01	rules	rule	1	law
12S01	schemas	schema	1	memory
12S01	seconds	second	4	day minute occurring time
12S01	sentences	sentence	4	capital English language paragraph
12S01	separate	separates	1	parallel
12S01	shapes	shape	2	maths pattern
12S01	subjects	subject	2	examination GCSE
12S01	switches	switch	1	electrical
12S01	tests	test	1	examination
12S01	units	unit	1	measurement
12S01	values	value	1	scale
12S01	words	word	8	capital English language letter order paragraph sentence test
12S02	atoms	atom	2	hydrogen nitrogen
12S02	balls	ball	1	newtoncradle
12S02	bonds	bond	1	vibration
12S02	bricks	brick	1	build
12S02	cells	cell	2	muscle skin
12S02	centimetres	centimetre	1	ruler
12S02	degrees	degree	1	thermometer
12S02	electrons	electron	2	ion transfer
12S02	fingers	finger	2	count hand
12S02	grams	gram	1	weight
12S02	hours	hour	1	clock
12S02	legs	leg	5	chair desk stand table walk
12S02	letters	letter	1	print
12S02	limbs	limb	1	body
12S02	lips	lip	2	mouth tongue
12S02	materials	material	1	construct
12S02	measurement	measure	1	degree
12S02	millimetres	millimetre	1	ruler
12S02	minutes	minute	2	clock time

12S02	molecules	molecule	1	synthesis
12S02	muscles	muscle	2	limb press
12S02	nails	nail	1	hand
12S02	particles	particle	16	air bond chemical compact energy expand gas heat kinetic liquid oxygen solid state temperature thermal vibration
12S02	poles	pole	1	attraction
12S02	seconds	second	3	clock sixty time
12S02	sense	senses	2	feel tongue
12S02	shapes	shape	1	print
12S02	toes	toe	1	foot
12S02	vertices	vertex	1	corner
12S02	waves	wave	1	ocean
12S05	angles	angle	1	geometry
12S05	appliances	appliance	1	electricity
12S05	atoms	atom	2	charge nucleus
12S05	backwards	backward	1	behind
12S05	birds	bird	1	morning
12S05	brackets	bracket	2	algebra bidmas
12S05	branches	branch	1	root
12S05	buds	bud	2	taste tree
12S05	cells	cell	1	biology
12S05	chemicals	chemical	1	chemistry
12S05	droplets	droplet	1	snow
12S05	eyes	eye	2	colour face
12S05	faces	face	1	surface
12S05	features	feature	1	face
12S05	forces	force	1	physics
12S05	grams	gram	2	law weight
12S05	hours	hour	2	clock time
12S05	joules	joule	1	energy
12S05	kilos	kilo	1	weight
12S05	laws	law	2	bidmas newtons
12S05	leaves	leaf	3	chlorophyll photosynthesis tree
12S05	lips	lip	1	face
12S05	metres	metre	1	measurement
12S05	minutes	minute	2	clock time
12S05	neutrons	neutron	2	atomic nucleus
12S05	onwards	onward	1	ahead
12S05	organs	organ	1	biology

12S05	particles	particle	3	liquid physics snow
12S05	planets	planet	3	night space star
12S05	plants	plant	4	biology green photosynthesis slug
12S05	protons	proton	2	atomic nucleus
12S05	seasons	season	1	weather
12S05	seeds	seed	2	apple sow
12S05	sports	sport	1	weights
12S05	stars	star	3	celestial night space
12S05	triangles	triangle	1	trigonometry
12S05	waves	wave	1	boat
12S09	aims	aim	2	ambitious optimistic
12S09	alevels	alevel	3	content school work
12S09	animals	animal	1	cute
12S09	backwards	backward	1	direction
12S09	books	book	1	roalddahl
12S09	cars	car	1	traffic
12S09	centimetres	centimetre	2	metre ruler
12S09	errors	error	1	backward
12S09	exams	exam	2	stress syllabus
12S09	eyes	eye	2	looking see
12S09	films	film	1	up
12S09	fingers	finger	1	hand
12S09	forwards	forward	1	backward
12S09	friends	friendly	2	alone family
12S09	goals	goal	1	optimistic
12S09	grams	gram	1	scales
12S09	kilograms	kilogram	2	heavy weight
12S09	knees	knee	1	head
12S09	lanes	lane	1	motorway
12S09	lessons	lesson	2	boring maths
12S09	measurements	measurement	1	mass
12S09	metres	metre	2	centimetre distance
12S09	mistakes	mistake	1	backward
12S09	mocks	mock	1	stress
12S09	numbers	number	3	centimetre counting weight
12S09	outdoors	outdoor	2	nature outside
12S09	parents	parent	1	family
12S09	planets	planet	1	rocket
12S09	roads	road	1	traffic
12S09	rockets	rocket	1	space

12S09	skies	sky	1	earth
12S09	stars	star	6	dreaming galaxy moon night solar space
12S09	toes	toe	2	foot head
12S09	words	word	1	essay

Table 10.2 Merged words for each participant (tense).

Participant	Merged Word	Vertex (Sign)	Frequency of Merge	Stimulus Words
12P02	altered	alter	1	different
12P02	bonded	bonding	1	connected
12P02	caused	cause	1	determined
12P02	changed	change	1	different
12P02	decrease	decreasing	5	less lose multiplier reduce step
12P02	decreased	decreasing	2	smaller squashed
12P02	expand	expanding	1	stretch
12P02	happened	happen	2	already occurred
12P02	increase	increasing	4	enlarging gain multiplier step
12P02	increased	increasing	6	bigger expanding extra greater larger more
12P02	influenced	influence	2	determined related
12P02	link	linked	1	relate
12P02	moving	move	2	falling work
12P02	pair	paired	1	relate
12P02	reduced	reduce	1	lower
12P02	representing	represent	1	quantifying
12P03	acting	acts	4	centre comparison integration pivot
12P03	bending	bend	2	density refraction
12P03	breaks	breaking	1	duality
12P03	changing	change	1	orbit
12P03	charging	charge	1	fixed
12P03	disprove	disproving	1	same
12P03	orbiting	orbit	2	centripetal dark
12P03	rock	rocking	1	cradle
12P03	transferred	transfer	1	gpe
12P03	varying	varies	4	fibre leftover multipath refractive index
12P05	change	changing	3	field rate steep
12P05	changes	changing	5	gradient gravitational gravity relative variable
12P05	controlled	control	1	manned
12P05	ending	end	1	infinite
12P05	finished	finish	2	closed complete
12P05	formulated	formulating	1	counted
12P05	spacing	space	1	fringe
12P05	use	used	2	converted purpose
12P05	waste	wasted	1	efficiency
12P08	cancelling	cancel	1	opposing

12P08	close	closed	3	mars solid Venus
12P08	emit	emitted	1	supplied
12P08	fall	falling	1	ball
12P08	measure	measuring	3	amount meter stopwatch
12P08	measured	measuring	1	effect
12P08	polarised	polarise	2	light transverse
12P08	released	release	4	emitted heater loss photon
12P08	stopped	stopping	1	stationary
12P12	achieve	achieved	1	overcome
12P12	block	blocking	1	atom
12P12	blocked	blocking	2	barred hidden
12P12	bordered	border	2	bounded region
12P12	bound	bounded	1	limit
12P12	changed	change	3	offshoot processed transferred
12P12	changing	change	1	altering
12P12	contracted	contract	1	short
12P12	control	controlled	3	autonomous own remote
12P12	covering	cover	1	surface
12P12	escaped	escape	1	passed
12P12	exchange	exchanged	1	own
12P12	fall	falling	1	down
12P12	follows	following	1	then
12P12	gifted	gift	1	presented
12P12	handed	hand	1	left
12P12	head	heading	1	ear
12P12	holding	hold	1	bearing
12P12	limited	limit	1	choice
12P12	move	moving	3	rocket stable transport
12P12	moved	moving	2	gone transferred
12P12	multiply	multiplied	1	multiple
12P12	orbiting	orbit	1	planet
12P12	pass	passed	1	transmission
12P12	process	processed	2	start unending
12P12	processed	processed	1	information
12P12	protected	protect	2	enclosed inside
12P12	receive	receiving	1	transmitting
12P12	removed	remove	1	abstract
12P12	rest	resting	1	still
12P12	rotate	rotating	1	pedal
12P12	send	sending	2	transmission transmitting

12P12	shaped	shape	1	ear
12P12	spin	spinning	1	turning
12P12	store	stored	2	constant elastic
12P12	surfaced	surface	1	floor
12P12	surround	surrounding	2	cover stereo
12P12	surrounded	surrounding	2	bounded unbound
12P12	transfer	transferred	2	resistance transmission
12P12	transferring	transferred	1	working
12S01	balanced	balance	3	equal equilibrium law
12S01	follow	followed	1	obedience
12S01	learning	learn	1	recall
12S01	linked	link	1	together
12S01	timed	time	1	examination
12S02	building	build	3	brick exterior inside
12S02	condensed	condensed	1	compact
12S02	point	pointed	2	peak vertex
12S02	stopped	stop	1	still
12S02	suspend	suspended	3	hang rope string
12S02	tissue	tissue	2	muscle wipe
12S09	bored	boring	1	tiring
12S09	cry	crying	1	depressed
12S09	relax	relaxed	1	break
12S09	repeat	repeating	1	again
12S09	sleeping	sleep	1	dreaming
12S09	stressed	stress	1	exhausted
12S09	studying	study	1	reading
12S09	travel	travelling	1	road
12S09	watching	watch	1	looking
12S09	writing	write	2	essay word

Table 10.3 Merged words for each participant (word class).

Participant	Merged Word	Vertex (Sign)	Frequency of Merge	Stimulus Words
12P02	accelerate	acceleration	5	force gravity inertia turn weight
12P02	accelerating	acceleration	1	falling
12P02	algebra	algebraic	4	calculation equation mathematical symbolise
12P02	attraction	attract	5	bonding closer gravitational gravity intermolecular
12P02	compressible	compressed	1	fluid
12P02	deform	deformation	3	bend distort manipulate
12P02	displacement	displace	1	return
12P02	elasticity	elastic	3	compressed spring stiffness
12P02	indeterminate	indeterminism	1	probable
12P02	interact	interaction	5	affect collision influence interfere reaction
12P02	interference	interfere	1	superposition
12P02	logic	logical	1	reason
12P02	measure	measurement	4	charge coulomb degree observer
12P02	measured	measurement	1	relative
12P02	measuring	measurement	1	number
12P02	occur	occurrence	1	happen
12P02	opposition	opposite	2	drag friction
12P02	previously	previous	1	already
12P02	proportional	proportionality	1	ratio
12P02	reverse	reversible	1	undo
12P02	separated	separation	1	apart
12P02	similar	similarity	3	accurate comparison linked
12P02	subtract	subtraction	2	deduct minus
12P02	unpredictability	unpredictable	1	indeterminism
12P03	combustion	combust	2	destruction ethanol
12P03	conditions	conditional	1	circumstance
12P03	conserved	conservation	3	kaon strangeness weak
12P03	construction	construct	3	antinode slit source
12P03	constructive	construct	1	pattern
12P03	destruct	destruction	1	superposition
12P03	destructive	destruction	1	pattern
12P03	different	difference	2	dispersion refractive index
12P03	dimensional	dimension	2	human length

12P03	direction	directional	10	amplitude anticlockwise bearing conventional displacement down north orbit point vector
12P03	dissipated	dissipation	1	charge
12P03	electromagnetically	electromagnetic	1	cloud
12P03	electromagnetism	electromagnetic	1	negative
12P03	environment	environmental	1	fittest
12P03	gravitation	gravitational	3	attraction centripetal gain
12P03	interact	interaction	1	charged
12P03	measure	measurement	3	number range stopping
12P03	measuring	measurement	2	gauge small
12P03	opposite	opposition	1	inverse
12P03	oppositional	opposition	1	newton
12P03	optical	optic	1	fibre
12P03	polarised	polarisation	1	source
12P03	propagation	propagate	1	origin
12P03	repulsive	repulsion	2	alpha radioactive
12P05	conserved	conservation	1	elastic
12P05	cost	costly	2	economy yield
12P05	destruction	destructive	1	superposition
12P05	directed	direction	1	push
12P05	directional	direction	1	acceleration
12P05	dissipation	dissipated	1	kinetic
12P05	efficient	efficiency	3	direct system yield
12P05	geological	geology	1	ore
12P05	initiate	initial	1	start
12P05	intersection	intersect	2	asymptote origin
12P05	mathematical	mathematics	1	formulating
12P05	measure	measurement	3	degree dial unit
12P05	measured	measurement	2	amount counted
12P05	measures	measurement	1	preventing
12P05	measuring	measurement	2	apparatus stopwatch
12P05	oscillating	oscillation	1	vibration
12P05	physically	physical	1	same
12P05	physics	physical	1	natural
12P05	precision	precise	1	practical
12P05	terminate	terminal	2	end finish
12P05	termination	terminal	1	complete
12P05	vibrating	vibration	2	heat particle
12P08	accelerate	acceleration	4	faster gain increase momentum
12P08	attract	attraction	1	opposite

12P08	calculate	calculation	1	proportional
12P08	calculating	calculation	1	cubed
12P08	cancellation	cancel	1	frequency
12P08	coherent	coherence	4	constructive directional ripple superimpose
12P08	concentrated	concentration	1	compression
12P08	conservation	conserved	9	coalesce collision difference interaction law strangeness transfer weak work
12P08	different	difference	4	initial isotope opposite ratio
12P08	direct	directional	1	linear
12P08	direction	directional	25	anticlockwise bearing coherence component destructive diffraction dimension displacement field horizontal interference negative opposite parallel polarise refraction resolving resultant superimpose torque triangle trigonometry vector vertical wave
12P08	disproportionation	disproportionality	1	spring
12P08	effervescence	effervescing	1	rise
12P08	insulate	insulator	1	surround
12P08	insulation	insulator	1	greenhouse
12P08	ionising	ionisation	1	gamma
12P08	magnetism	magnetic	2	attraction repulsion
12P08	mathematical	mathematics	3	cosine cubed trigonometry
12P08	measurement	measuring	5	depth kelvin length metre scalar
12P08	measurements	measuring	2	metric numerical
12P08	oscillation	oscillate	8	bungee microwave parallel peak perpendicular polarise transverse vibration
12P08	oscillations	oscillate	1	sound
12P08	progress	progression	1	rate
12P08	proportion	proportional	4	concentration density numerical spring
12P08	proportionality	proportional	5	constant cubed deformation equation string
12P08	resist	resistance	1	overcome
12P08	strange	strangeness	2	particle quark
12P08	toxic	toxicity	2	alpha inhabitable
12P08	vibrating	vibration	1	oscillate
12P12	conserved	conservation	1	system
12P12	different	difference	7	absence contrast distinguishable indistinguishable offshoot same unique
12P12	direct	direction	2	purposeful rocket
12P12	directional	direction	2	flow stream
12P12	efficiency	efficient	1	regulation

12P12	intentional	intent	1	Purposeful
12P12	interact	interaction	1	contained
12P12	locate	location	1	find
12P12	measured	measure	1	data
12P12	move	moving	3	rocket stable transport
12P12	normally	normal	1	bottom
12P12	operator	operation	1	multiplied
12P12	orbitals	orbit	1	atom
12P12	oriented	orientation	1	facing
12P12	protective	protect	1	surface
12P12	push	pushes	2	pedal thrust
12P12	pushing	pushes	1	working
12P12	quantifiable	quantify	1	measurable
12P12	relative	relatively	11	down fast left location near resting right smaller stationary still up
12P12	relativity	relatively	1	classical
12P12	rotation	rotating	3	pivot spinning turning
12P12	separate	separation	2	cut remove
12S01	difference	different	1	comparison
12S01	measure	measurement	2	coin scale
12S01	natural	nature	1	nature
12S01	quickly	quick	1	replicate
12S01	similarities	similar	1	comparison
12S02	accelerate	acceleration	1	forward
12S02	attract	attraction	3	ion pole proton
12S02	construction	construct	2	nail structure
12S02	significant	significance	1	powerful
12S05	advancement	advance	1	technology
12S05	ill	illness	1	pale
12S05	measure	measurement	2	gage mean
12S09	measuring	measurement	1	scales
12S09	omnipresence	omnipresent	1	god

Table 10.4 Response words that are not merged with plural inflections of an existing stimulus word. Reasons for this are given. The number of times these words could be merged (count of occurrence) is also listed.

Response Word	Possible Merge to Create Single Vertex	Participants	Count of Occurrence	Stimulus Words	Reason for Not Merging
newtons	newton	12P08	4	pull push stiffness wrench	*newtons* (unit of force) versus *newton* (Isaac)
scale	scales	12P08	3	cubed numerical small	*scales* (used to measure mass) versus *scale* (to change size)
newton	newtons	12P05 12S02	6	constant friction gravity heavy scientist gravity	*newtons* (unit of force) versus *newton* (Isaac)
newtons	newton	12S05	1	weights	*newtons* (unit of force) versus *newton* (Isaac)
weights	weight	12S05	2	gym newton	*weight* (force due to mass) versus *weights* (masses for exercise)

Table 10.5 Response words that are not merged with tense inflections of an existing stimulus word. Reasons for this are given. The number of times these words could be merged (count of occurrence) is also listed.

Response Word	Possible Merge to Create Single Vertex	Participants	Count of Occurrence	Stimulus Words	Reason for Not Merging
circle	circling	12P12	1	circular	*circling* | *circle* implies *circle* (noun) rather than *circle* (verb).
close	closed	12P12 12P08	4	adjacent near neighbouring together Mars solid Venus	*close* is used in the sense of 'close to' rather than 'close the door'.
face	facing	12P12	1	side	*side* | *face* implies *face* (noun) rather than *face* (verb).
mean	meaning	12P12	1	average	*average* | *mean* implies *mean* (noun) rather than *mean* (verb).
present	presented	12P12	4	absence gift gone receiving	Stimulus words imply *present* (noun) rather than *present* (verb).
repeat	repeating	12S01	5	everyday followed memorising repeating replicate	*repeat* | *repeating* is a word association, which implies *repeat* (noun) versus *repeat* (verb).
spacing	space	12P03 12P05	1	duality fringe	*space* | *universe* is a previous word association, which implies *space* (as the earth is in space) rather than *spacing* (as in the space between the fringes).
surroundings	surround	12P08	5	atmosphere endothermic exothermic room surround	*surrounding* | *surround* is a word association, which implies *surround* (noun) versus *surround* (verb).
working	work	12P12	2	effective together	*working* is used in the sense of 'I work in a team', rather than 'work done'.
workings	work	12P05	2	answer formula	*working* is used in the sense of 'I need to show my workings', rather than 'work done'.

Table 10.6 Response words that are not merged with word class inflections of an existing stimulus word. Reasons for this are given. The number of times these words could be merged (count of occurrence) is also listed.

Response Word	Possible Merge to Create Single Vertex	Participants	Count of Occurrence	Stimulus Words	Reason for Not Merging
alternatives	alternating	12P08	1	pollution	*pollution* | *alternative* implies alternative in the sense of 'there is no alternative', rather than the 'alternating current'.
bars	barred	12P12	1	barred	*bars* is a response word for *barred*
container	contained	12P12	2	fluid liquid	*fluid* | *container* implies the 'container is full', rather than the 'liquid is contained'.
electromagnetic	electromagnet	12P08	18	annihilation constructive destructive energy excitation infrared interaction lepton magnetic microwave photoelectric photon polarise radiation radio transverse ultraviolet wave	Stimulus words imply *electromagnetic* in the sense of 'electromagnetic effect', rather than 'this is an electromagnet'.
fitting	fitness	12S05	1	loose	*loose* | *fitting* implies fitting in the sense of 'this jumper doesn't fit', rather than 'my fitness was great'.
function	functional	12P03	7	animal binomial dimension function inverse plus rate	*Functional* is a response word for *function*.
inhibit	inhibition	12S01	1	inhibition	*Inhibit* is a response word for *inhibition*.
physical	physics	12S05	1	sport	*sport* | *physical* implies *physical* 'as in that was a physical game', rather than 'the physics lesson was fun'.
present	presentation	12P05	1	presentation	*Present* is a response word for *presentation*.
presentation	presented	12P12	1	presented	*Presentation* is a response word for *presented*.

strangeness	strange	12P03	5	kaon meson neutral strange weak	*strangeness* response word for *strange*
subjective	subject	12S01	1	art	*subject* is response to several other words (*English, maths, gcse, science, examination, geography, physics, literature, test, homework*). So, the use of *subjective* for *art* seems to be deliberate.

10.3 Compound Word and Phrase Tables

Table 10.7 Changes to response words that have been presented as a combined compound word. The reason codes are explained in Figure 3.11.

Participant	Presentation	Seed Word	Response Words			Vertex Decision	Reason	Notes
			1	2	3			
12P03	combined	race	down	force		down force	IS-0	
12P04	combined	bending	path	altering		path / altering	IS-0	
12P04	combined	motion	kinetic	energy		kinetic / energy	IS-0	
12P04	combined	radiation	electron	scattering		electron / scattering	IS-0	
12P05	combined	metre	standard	unit		standard / unit	IS-0	
12P08	combined	molecule	lattice	simple	giant	lattice / simple / giant	IS-0	
12P04	combined	molar	molar	mass		mass	NIS-0	Molar is the stimulus word and response word. Hence, mass is kept as a response and molar is removed.
12P05	combined	combustion	CO_2			carbon-dioxide	NIS-R	CO_2 is kept as a compound response but merged with carbon-dioxide, which is a previous response to a stimulus word.
12P03	combined	gpe	work	done		work	Precedent	Work is a previous response and stimulus is a synonym for work-done.
12P03	combined	potential	work	done		work	Precedent	Work is a previous response and stimulus is a synonym for work-done.
12P04	combined	power	work	done		work	Precedent	Work is a previous response and stimulus is a synonym for work-done.
12P03	combined	varies	energy	transfer		energy / transfer	Precedent	Energy and transfer have both been previously given as a sequential response.

Table 10.8 Changes to response words that have been presented as a sequential compound word. The reason codes are explained in Figure 3.11.

Participant	Presentation	Seed Word	Response Words				Vertex Decision	Reason	Notes
			1	2	3	4			
12P03	sequential	beegees	staying	alive			staying alive	NIS-A	
12P03	sequential	conditional	bayes	theorem			bayes theorem	NIS-A	
12P03	sequential	density	refractive	index			refractive index	NIS-A	
12P03	sequential	dispersion	refractive	index			refractive index	NIS-A	
12P03	sequential	environmental	power	station			power station	NIS-A	
12P03	sequential	galaxy	milky	way			milky way	NIS-A	
12P03	sequential	material	youngs	modulus			youngs modulus	NIS-A	
12P03	sequential	material	self	healing			self healing	NIS-A	
12P03	sequential	medium	refractive	index			refractive index	NIS-A	
12P03	sequential	point	black	hole			black hole	NIS-A	
12P03	sequential	refraction	refractive	index			refractive index	NIS-A	
12P03	sequential	refractive index	snells	law			snell's law	NIS-A	
12P03	sequential	space	refractive	index			refractive index	NIS-A	
12P03	sequential	spring	youngs	modulus			youngs modulus	NIS-A	
12P03	sequential	spring	hookes	law			hookes law	NIS-A	

12P03	sequential	statistics	standard	deviation			standard deviation	NIS-A	
12P03	sequential	tightness	youngs	modulus			youngs modulus	NIS-A	
12P03	sequential	triangle	refractive	index			refractive index	NIS-A	
12P04	sequential	distribution	monte	carlo			monte-carlo	NIS-A	
12P08	sequential	sun	main	stage			main-stage	NIS-A	
12S03	sequential	atom	spread	out			spread-out	NIS-A	
12S03	sequential	exam	a	level			a-level	NIS-A	
12S03	sequential	exam	check	up			check-up	NIS-A	
12S03	sequential	film	rom	com			rom-com	NIS-A	
12S03	sequential	film	sci	fi			sci-fi	NIS-A	
12S03	sequential	kidney	pelvic	floor			pelvic-floor	NIS-A	
12S03	sequential	problem	hold	up			hold-up	NIS-A	
12S03	sequential	tread	tip	toeing			tip-toeing	NIS-A	
12P05	sequential	gravity	general	relativity			general relativity	NIS-R	
12P05	sequential	space	general	relativity			general relativity	NIS-R	
12S03	sequential	noise	high	pitched	low	pitched	high / pitched / low	IS-0	Pitched is a repeated response word. Hence, one pitched is removed.
12P04	sequential	light	slit	double	slit		slit / double	IS-1	Slit is a repeated response word. Hence, one slit response was removed.
12P08	sequential	refraction	critical	angle			critical-angle	IS-1	Angle is give twice as a response word to refraction.
12S03	sequential	anxious	panic	panic	attack		panic / attack	NIS-0	Panic is a repeated response word. Hence, one panic was removed.
12S03	sequential	cooked	cook	through			through	NIS-0	Cook is a tense inflection of cooked. Hence, cook is removed.

12S03	sequential	leave	left	behind		left / behind	NIS-0	Leave is a tense inflection of leave. Hence, left is removed.
12S03	sequential	saddle	horse	horse	riding	horse / riding	NIS-0	Horse is a repeated response word. Hence, one horse is removed.
12S03	sequential	sports	sports	kit		sports / kit	NIS-0	Sports is stimulus and response word. Hence, sports is removed.
12S03	sequential	steer	steering	wheel		steer / wheel	NIS-0	Steer is a tense inflection of steering. Hence, steering is removed.
12P04	sequential	activation	starts	starting	off	start /off	NIS-1	Start and starting are repeated response words. Hence, start and starting were merged according to the word class inflection procedure.
12P04	sequential	closed	closed	off		off	NIS-1	Closed is both stimulus and response word. Hence, closed was removed.
12P04	sequential	travel	mass	mass	point	mass / point	NIS-1	Mass is a repeated response word. Hence, one mass is removed.
12S03	sequential	emergency	fire	fire	engine	fire-engine	NIS-1	Fire is a repeated response word. Hence, one fire is removed.
12P03	sequential	deBroglie	matter	wave		matterwave	Precedent	Matterwave is a previous response word presented as a combined response.
12P03	sequential	element	work	function		workfunction	Precedent	Workfunction is a previous response word presented as a combined response.
12P03	sequential	height	work	done		work	Precedent	Work done is synonym for work, which is a previous response.
12P03	sequential	incident	work	function		workfunction	Precedent	Workfunction is a previous a response word presented as a combined response.
12P03	sequential	overcome	work	function		workfunction	Precedent	Workfunction is a previous response word presented as a combined response.
12P03	sequential	surface	work	function		workfunction	Precedent	Workfunction is a previous response word presented as a combined response.
12P03	sequential	threshold	work	function		workfunction	Precedent	Workfunction is a previous response word presented as a combined response.

12P03	sequential	ultraviolet	work	function	workfunction	Precedent	Workfunction is a previous response word presented as a combined response.
12P05	sequential	voltage	potential	difference	potential difference	Precedent	Potential-difference is a previous response word presented as a combined response.
12P05	sequential	voltmeter	potential	difference	potential difference	Precedent	Potential-difference is a previous response word presented as a combined response.

Table 10.9 Changes to response words that have been presented as phrases. The reason codes are explained in Figure 3.11.

Participant	Presentation	Seed Word	Response Words				Vertex Decision	Reason
			1	2	3	4		
12P03	combined	location	twenty	four	hours		twenty four / hours	p-crw
12P03	combined	potential	work	done			work	p-crw
12P03	sequential	kaon	created	via	strong	decay	created / strong / decay	p-crw
12P04	combined	example	case	of			case	p-crw
12P04	combined	kilojoule	a	thousand	joules		thousand / joules	p-crw
12P04	combined	space	speed	of	light		speed / light	p-crw
12P04	combined	wasted	not	with			with	p-crw
12P04	combined	wasted	not	useful			useful	p-crw
12P04	sequential	amount	measure	of	mass		measure / mass	p-crw
12P04	sequential	graph	results	from	table		results / table	p-crw
12P04	sequential	law	not	breakable			breakable	p-crw
12P04	sequential	show	display	and	tell		display / tell	p-crw
12P04	sequential	tread	line	of	best	fit	line / best-fit	p-crw
12S03	sequential	synagogue	place	of	worship		place / worship	p-crw
12P05	sequential	impossible	cannot	happen			-	p-rm

10.4 Strict Test Criteria Tables for the Sign Network Degree Distributions

*Table 10.10 Summary of whether each in-degree (k_{in}) sign network has met the criteria for the strict test for a scale-free structure in a network developed by Broido and Clauset (2019). * Indicates that the null hypothesis that the data cannot distinguish between the two distributions can be rejected (P<0.1). # indicates that a power-law model can be rejected as a plausible fit to this distribution.*

		Alternative Distribution is a Better Model				Super-Weak	Weakest	Weak	Strong
Cohort	Participant	Poisson	Power-Law with Exponential Cut-Off	Exponential	Log-Normal	Alternative Distribution is a Better Model	Power-Law Distribution Model Cannot Be Rejected (P>0.1)	Power-law Region Contains at Least 50 nodes	2 < **α**< 3
Physics	12P02		✓		✓			✓#	
	12P03		✓		✓		✓	✓	✓
	12P04		✓		✓		✓		✓
	12P05		✓	✓	✓		✓		
	12P08		✓		✓		✓	✓	
	12P12		✓*		✓	✓	✓	✓	✓
Psychology	12S01		✓		✓		✓	✓	
	12S02		✓*	✓	✓	✓	✓	✓	✓
	12S03		✓		✓		✓	✓	
	12S05		✓*	✓	✓	✓		✓#	✓#
	12S09		✓*	✓	✓	✓	✓	✓	✓

*Table 10.11 Summary of whether each out-degree (k_{out}) sign network has met the criteria for the strict test for a scale-free structure in a network developed by Broido and Clauset (2019). * indicates that the null hypothesis that the data cannot distinguish between the two distributions can be rejected (P<0.1). # indicates that a power-law model can be rejected as a plausible fit to this distribution.*

		Alternative Distribution is a Better Model				Super-Weak	Weakest	Weak	Strong
Cohort	Participant	Poisson	Power-Law with Exponential Cut-Off	Exponential	Log-Normal	Alternative Distribution is a Better Model	Power-Law Distribution Model Cannot Be Rejected (P>0.1)	Power-law Region Contains at Least 50 nodes	$2 < \alpha < 3$
Physics	12P02		✓	✓	✓		✓	✓	
	12P03		✓*	✓	✓	✓	✓	✓	
	12P04		✓	✓	✓		✓		
	12P05	✓	✓	✓	✓		✓		
	12P08		✓	✓	✓			✓	
	12P12	✓	✓	✓	✓		✓	✓	
Psychology	12S01	✓	✓	✓	✓		✓		
	12S02				✓		✓		
	12S03	✓	✓	✓	✓		✓		
	12S05		✓	✓	✓		✓	✓	
	12S09				✓		✓	✓	

10.5 Hub Word Frequency Tables

Table 10.12 Hub-Words Threshold 0.015

Word	Physics	Psychology	Total	p
force	6	2	8	0.061*
energy	4	2	6	0.567
mass	4	0	4	0.061*
measurement	1	2	3	0.545
particle	2	1	3	1.000
time	1	2	3	0.545
movement	1	2	3	0.545
change	2	0	2	0.455
direction	2	0	2	0.455
distance	1	1	2	1.000
motion	2	0	2	0.455
small	0	2	2	0.182
light	1	1	2	1.000

Table 10.13 Hub-Words Threshold 0.01

Word	Physics	Psychology	Total	p
force	6	3	9	0.182
energy	5	3	8	0.545
time	4	3	7	1.000
particle	4	2	6	0.567
mass	5	0	5	0.015**
direction	2	2	4	1.000
measurement	1	3	4	0.242
unit	3	1	4	0.545
light	1	3	4	0.242
acceleration	3	0	3	0.182
change	3	0	3	0.182
charge	2	1	3	1.000
small	1	2	3	0.545
speed	1	2	3	0.545
wave	3	0	3	0.182
electron	3	0	3	0.182
movement	1	2	3	0.545
water	0	3	3	0.061*
difference	2	0	2	0.455
distance	1	1	2	1.000
gravity	2	0	2	0.455
kinetic	2	0	2	0.455
motion	2	0	2	0.455
number	0	2	2	0.182
shape	0	2	2	0.182
transfer	2	0	2	0.455
velocity	2	0	2	0.455
work	1	1	2	1.000
conservation	2	0	2	0.455
resistance	2	0	2	0.455
measure	0	2	2	0.182
happy	0	2	2	0.182
maths	0	2	2	0.182

Table 10.14 Hub-Words Threshold 0.005

Word	Physics	Psychology	Total	p
time	6	5	11	1.000
force	6	4	10	0.455
distance	5	4	9	1.000
energy	6	3	9	0.182
gravity	6	3	9	0.182
movement	5	4	9	1.000
particle	6	2	8	0.061*
speed	4	4	8	1.000
direction	3	4	7	0.545
mass	6	1	7	0.015*
work	4	3	7	1.000
constant	5	1	6	0.080
measurement	3	3	6	1.000
unit	4	2	6	0.567
weight	3	3	6	1.000
light	3	3	6	1.000
acceleration	5	0	5	0.015**
change	4	1	5	0.242
current	3	2	5	1.000
kinetic	4	1	5	0.242
small	2	3	5	0.567
electron	4	1	5	0.242
electricity	1	4	5	0.080*
angle	4	0	4	0.061*
charge	3	1	4	0.545
heat	1	3	4	0.242
length	2	2	4	1.000
number	2	2	4	1.000
shape	1	3	4	0.242
space	3	1	4	0.545
temperature	1	3	4	0.242
transfer	4	0	4	0.061*
vector	4	0	4	0.061*
velocity	4	0	4	0.061*
proton	3	1	4	0.545
strong	1	3	4	0.242
water	0	4	4	0.015**

10.6 Agree 'Ideal' Explanations of the Motion of the Objects in the Demonstrations with Force Concept Inventory (FCI) Cross References

The demonstration shown to participants in the interviews are described below. The 'idea' explanation of the motion of each of the objects agreed by the physics teachers in the target school, referring to the class textbook, are given in the following tables, with the FCI cross references.

10.6.1 Film Canister Rocket

Participants are shown a film canister and the reaction of a fizzy table with water. The canister has a coin (2 pence payload stuck to the body of the canister). During the demonstration, the canister is filled with about a third of water. A fizzy tablet is added to the canister and the canister is left to 'pop'. For the second demonstration, the canister is reversed, so the lid is launched and not the canister.

Section of Demonstration	Agreed Explanation	FCI Cross-Reference
Film canister launch	When the film canister 'pops', gas escapes in a downward direction and exerts an equal and opposite force on the canister.	Newton's Third Law. Impulsive forces.
	This force is greater than the weight of the canister.	Superposition Principle. Vector sum and cancelling forces.
	Therefore, there is a resultant force upwards which causes the canister to accelerate upwards. Hence, the canister is given a large velocity in this direction.	Newton's Second Law. Constant force implies constant acceleration.
Film canister flight	The canister is now travelling upwards. Air resistance and weight are the only forces acting on the canister and both of these forces act in a downwards direction.	Kinds of force. Fluid contact (air resistance) and gravitation.
	Hence, there is a resultant force acting in the opposite direction to the velocity of the canister.	Superposition Principle. Vector sum and cancelling forces.
	Therefore, the canister accelerates and the velocity of the canister decreases.	Newton's Second Law. Constant force implies constant acceleration.
	Eventually, the canister reaches zero velocity. However, weight is still acting on the canister. Hence, there is still a resultant force acting on the canister. This resultant force continues to accelerate the canister but now the speed of the canister increases, and it moves towards the ground. It now has increasing negative velocity. The canister might reach a terminal velocity similar to the cupcake case.	Kinematics. Velocity distinguished from speed.
Film canister landing	When the canister hits the desk, the desk exerts a force on the canister.	Newton's Third Law. Impulsive forces.
	This is greater than the weight. Hence, there is a resultant force which decelerates the canister.	Newton's Second Law. Constant force implies constant acceleration.
Lid as Projectile	When the film canister 'pops', gas escapes in a downward direction and exerts an equal and opposite force on the lid. The lid has less mass than the canister with payload. Hence, if the force of the 'pop' is roughly similar, then the initial velocity of the lid will be greater.	Newton's Second Law. F=ma, acceleration inversely proportional to mass.
	Therefore, it should reach a higher height as the acceleration of the will be independent of weight.	Gravitation. Acceleration independent of weight.

10.6.2 Falling Cupcake Case

During an interview, each participant was shown demonstration of three different types of cupcake case falling from 1m to a desk.

1. Cupcake case only.
2. Cupcake case with paperclips attached.
3. Screwed-up cupcake case.

A metre rule was used to show the height of a 1m.

Section of Demonstration	Agreed Explanation	FCI Cross-Reference
Falling cupcake case only	The cupcake case initially accelerates but then reaches a terminal velocity. This is because the weight of the object is initially greater than any air resistance (which is zero when the case is at rest). Therefore, there is a resultant force acting downwards.	Superposition Principle. Vector sum and cancelling forces.
	Hence, the case accelerates towards the floor.	Newton's Second Law. Constant force implies constant acceleration.
	However, as the velocity of the case increases, the air resistance acting in the opposite direction to the weight increases.	Kinds of force. Fluid contact (air resistance).
	Eventually, the air resistance and weight are equal but acting in opposite directions. Therefore, there is no longer a resultant force acting on the case and, hence, the case is no longer accelerating. It has reached a terminal velocity.	Newton's First Law Cancelling forces.
Falling cupcake case with paper clips	The cupcake case with paperclips will hit the ground first because it reaches a higher terminal velocity. This is because the weight of the paperclip case is greater because more mass has been added. Hence, the air resistance must be greater before it is equal and opposite to the weight, consequently, there is no resultant force acting on the case and, therefore, the case stops accelerating.	Newton's First Law Cancelling forces.
	The air resistance depends on velocity. So, the velocity is higher before the air resistance balances the weight.	Kinds of force. Fluid contact (air resistance).
Falling screwed-up cupcake case.	The cupcake case with paperclips will hit the ground first because it reaches a higher terminal velocity. This is because the screwed-up case has less surface area. Hence, the air resistance is less for a given velocity.	Kinds of force. Fluid contact (air resistance).
	So, the velocity is higher before the air resistance balances the weigh and there is no resultant force acting on the case. Therefore, the case stops accelerating at a higher velocity.	Newton's First Law Cancelling forces.

10.7 Demonstration coverage of the conceptual dimensions and sub-dimensions identified in the Force Conceptual Inventory (FCI).

Inventory Section		Inventory Item	Further explanation	Falling cupcake case	Film Canister Pop	Trolleys and Ramps
0	Kinematics	Velocity discriminated from position	20E	Ticker tape timer style question - use dot at regular intervals to determine velocities and acceleration.		
		Acceleration discriminated from velocity	21D	Ticker tape timer style question - use dot at regular intervals to determine and compare acceleration. Recognise that equally spaced dots indicate zero acceleration.	Can be assessed by asking; 'Does the cake cup case fall at the same speed as it falls?' 'Is it accelerating all the way to the floor?'	Acceleration of the canister in the air versus the velocity of the canister in the air (constant acceleration versus changing velocity).	Acceleration of the trolleys on the ramps and on the flat using the distance markers.
		Constant acceleration implies parabolic orbit	23D, 24E	Projectile motion - constant horizontal velocity and vertical acceleration equal to g.			
		Constant acceleration implies changing speed	25B	Similar to projectile motion but with a rocket. X velocity remains constant but Y velocity increase - acceleration at right angles to X direction.		Acceleration for the canister in the air.	Acceleration of the trolleys down on the ramps and then on the flat surface.
		Vector addition of velocities	(7E)	Distinguishing between a component of velocity in a single direction (i.e. x) versus the vector sum (resultant) of two velocities in x and y directions. This question is not clear.			

Inventory Section			Inventory Item	Further explanation	Falling cupcake case	Film Canister Pop	Trolleys and Ramps
1	First Law	Velocity direction constant with no force	4B, (6B), 10B, 26B	The effect of a resultant force on the direction an object is travelling in. Frequent use of circular motion examples.			Change in velocity (direction and not speed) when a trolley changes direction when it moves from the inclined plane to the flat. Prediction for what would happen to the trolley if the surface was ice or if there was more friction.
		Speed constant with no force	8A, 27A	No resultant force means no acceleration (change in speed).			
		With cancelling forces	18D, 28C	Again, no resultant force means no acceleration. However, 'balanced forces' lead to the lack of a resultant force.	This is a terminal velocity problem. Hence, 'balanced' forces (air resistance and weight) are a key part of the explanation.		
2	Second Law	Impulsive force	(6B), (7E)	The hockey puck example again. No contact implies no force - the force doesn't die away.	Acceleration of the cupcake case when it hits the desk.	Acceleration of the canister or lid when the gas escapes from the canister and acceleration of the canister and lid when they hit the desk.	
		Constant force implies constant acceleration	24E, 25B	Force applied a right angle to constant velocity in x direction gives rise to constant acceleration in y direction (change in direction and speed).		Acceleration of the canister by its weight when it has left the desk.	Acceleration of the trolley by the pulley masses and by a component of their weight down the slope.

Inventory Section			Inventory Item	Further explanation	Falling cupcake case	Film Canister Pop	Trolleys and Ramps
3	Third Law	For impulsive forces	2E, 11E	Forces exerted in a collision and with different size objects pushing on each other - momentum ideas.	Force exerted by the desk when the film canister 'lands'.	Force exerted on the canister by the escaping gas and by the desk when the canister hits 'lands'.	
		For continuous forces	13A, 14A	Car and lorry pushing on each other - accelerating and then moving at constant speed.			Tension in the wire between the pulley masses and the trolley
4	Superposition Principle	Vector sum	19B	Boy and man pulling at different angle on a block - resultant force question.			Different components of the weight for different inclines.
		Cancelling forces	(9D), 18B, 28C	Balanced' forces.	Balancing' of air resistance and weight	Force exerted by the gas is greater than the weight of the canister.	

Inventory Section			Inventory Item	Further explanation	Falling cupcake case	Film Canister Pop	Trolleys and Ramps
		Solid contact - passive	(9D), (12B, D)	Forces acting on a book at rest on a table - 'balanced' forces at rest - and balanced forces in the vertical plane for an object moving at constant velocity in the horizontal plane (ice puck).	Cup case stationary on the table.	Canister stationary on the table.	
		Solid contact - impulsive	15C	Why does ball bounce?			
5	Kinds of Force	Solid contact - friction opposes motion	29C	Block slowing down and coming to rest when a force pulling the block forward at a constant velocity is removed.			The trolley slowing down on the flat surface in different scenarios.
		Fluid contact - air resistance	22D	Forces acting a golf ball during flight - air resistance and gravity.	Air resistance on cupcake case and how velocity and surface area affect air resistance.	Air resistance acting on the canister during flight.	
		Fluid contact - buoyant	12D	Forces acting on a book at rest on a table - 'balanced' forces at rest.	Can be assessed by asking; 'Can you explain why the cupcake case is just sitting there stationary on the table?'.	Can be assessed by asking; 'Can you explain why the canister is just sitting there stationary on the table?'.	

Inventory Section		Inventory Item	Further explanation	Falling cupcake case	Film Canister Pop	Trolleys and Ramps
5	Kinds of Force			This demonstration reinforces the intuitive idea that weight is proportional to acceleration when falling because the cupcake case with the added paperclips will fall faster. However, this is a terminal velocity affect and not a confirmation of the initiative explanation.	Acceleration of the canister in the air and the comparison with the lid of the canister. The acceleration of the lid is not greater than the acceleration of the canister. The difference in the height reached by the lid and the canister is a result of the different initial velocities.	
		Gravitation - acceleration independent of weight	5D, 9D, (12B, D), 17C, 18B, 22D, 1C, 3A	When and how gravitational forces are acting on falling and thrown objects (falling metal balls, ball thrown into air, ball rolling off table) and also acceleration of these objects.			
		Gravitation - parabolic trajectory	16B, 23D	Projectile motion - constant horizontal velocity and vertical acceleration equal to g.			

10.8 Interview Questioning Protocol

10.8.1 Falling cupcake case

10.8.1.1 Cupcake Case Only

Demonstrate cake-cup falling from approximately 1m.

- Can you describe what just happened to the cake-cup?

  It's falling.
  It fell to the floor slowly.

- Ok, it falls to the floor.

  → Can you give me a bit more detail about how the cake cup moves when it is falling?

  → Can you tell me what you mean by slowly?

- OR

- When I drop the cake-cup from about 1m it falls to the floor. Can you describe how the cake-cup moves when it is falling?

- Ok, so you said – DESCRIBE MOTION.

  → Is there anything you want to add/clarify before we move on?

  → It the case always moving in the same way?

  → Does the case always fall to the floor at the same speed?

  → What about when the cake-cup hits the floor, how would you describe what happens then?

- Can you explain why the cake-cup moves like that?

  → Can you draw that for me?

  → What is weight?

→ Are there any other forces acting on the case?

→ What is air resistance?

- What happens to the case WHEN IT HITS THE DESK?

→ Does its speed change? How?

→ Does it accelerate?

→ How can you explain that?

10.8.1.2 With Paperclips Attached

Drop case with paperclips and, again, demonstrate the cake-cup falling from 1m. Compare with 'normal' case.

- Is there anything different now about the way the cake-cup moves when it falls?
- So, ok – The cake cup DESCRIBE MOTION and HIGHLIGHT DIFFERENCES.
- Can you explain why the cake-cup moves like that?
    - → Can you draw that for me?
    - → What about when it hits the floor? Is there anything different happening?

10.8.1.3 Screwed-Up Cupcake Case

Take a case, screw-up it up and let it fall again. Compare with 'normal' case.

- Is there anything different now about the way the cake-cup moves when it falls?
- So, ok – The cake cup DESCRIBE MOTION and HIGHLIGHT DIFFERENCES.
- Can you explain why the cake-cup moves like that?
    - → Can you draw that for me?
    - → What about when it hits the floor? Is there anything different happening?

10.8.2 Film Canister Rocket

10.8.2.1 Film Canister as Payload

Film canister 'pops' with canister and 2p coin as payload.

'I've got a film canister rocket here with a 2p glued to the top as a payload. I'm going to put a bit of water into the canister and add a tablet. Let participant see the reaction. *Then I'm going to put the lid on, turn it upside down and wait.'*

- Can you describe what happened there?

The tablet reacted.

- Ok, so the tablet reacted with the water. Why is that important?

Gas is released. The pressure pushes the canister into the air.

- Ok, so can we unpick a bit what happens when the canister goes up.

  → If we start, with the canister on the desk before it pops.

    - How is it moving then?
    - Why is it stationary?

  → Now what about when it pops?

    - How does its movement change?
    - Why does it move/speed up?
      - Does it accelerate?

- Ok, so now the canister is in the air.

  → Can you describe how the canister moves when it is in the air.

  It goes up, reaches a maximum point, and then comes down.

  → Ok, REFLECT – why does it do that?

- Ok, so can we talk in a bit more detail about the motion of the canister in the air. REFER TO DRAWING. We've said it does up and then comes back down.

  → Can you give me any more detail about how it moves in the air?

    - What about the speed of the canister?

      – Is the speed the same?

    - What about the velocity of the canister?

      – Is the velocity the same?

    - So, is the canister accelerating at any point?

      – Is the acceleration the same?

      – Is the canister accelerating at the maximum point?

- So now we've put a bit more detail onto our description of the motion. We've said REFLECT BACK. Is there anything else you want to add to your description?

  → Why does the canister accelerate here (up)? Why does the canister accelerate here (down)?

  → We've mentioned gravity. Are there any other forces acting on the canister?

  → Are the forces the same when the canister is going up compared with going down?

  → At the maximum point, we said the canister was REFLECT BACK. What are the forces acting on the canister at the maximum point?

10.8.2.2 Lid as Payload

Film canister 'pops' with lid as payload.

'Ok, so I'm going to do the same thing as before but this time with the lid up.'

- Ok, so is there any difference between the first and the second pop?

  → In the movement of the lid compared with the movement of the canister?

  The lid reached a higher maximum height.

- Can you explain REFLECT THE DIFFERENCE.

  → Can we look at the motion again in more detail?

  - At the pop:
    - Is the speed of the lid the same?
    - Is the velocity the lid the same?
    - Is the acceleration of the lid the same?
  - In the air:
    - Is the speed of the lid the same?
    - Is the velocity the lid the same?
    - Is the acceleration of the lid the same?

  → So, we've said the lid reaches a higher velocity/speed when the canister pops. Why is that?

  → Are the forces acting on the lid the same as the forces acting on the canister?

  - Force from the pop/weight/air resistance.

10.9 Additional Tables from the Interview Analysis

Table 10.15 The frequency counts for hub-words in each participant's interview. T is the total number of words in the interview, C is the raw frequency count, C/T is the fractional frequency count (raw count/total number of words in an interview), N is the normalised work frequency count (hub-word occurrence per 500 words), and R is the hub-word rank in the list of most frequently occurring words in an interview. Note the normalised word frequency count is given per 500 words and not per 1000 words as most interview are ~ 2000 to 3000 words (Biber et al., 1998). These descriptive statistics include the interviewer's speech.

		T	Force				Mass				Energy				Time				Particle			
Cohort	Participant		C	$\frac{C}{T}$	N	R	C	$\frac{C}{T}$	N	R	C	$\frac{C}{T}$	N	R	C	$\frac{C}{T}$	N	R	C	$\frac{C}{T}$	N	R
Physics	12P02	1831	82	0.045	22.392	2	8	0.004	2.185	53	0	0.000	0.000	-	11	0.006	3.004	35	0	0.000	0.000	-
	12P03	2326	86	0.037	18.487	1	15	0.006	3.224	35	6	0.003	1.290	86	16	0.007	3.439	33	0	0.000	0.000	-
	12P04	2773	141	0.052	25.796	1	12	0.004	2.195	57	53	0.019	9.696	6	10	0.004	1.829	70	4	0.001	0.732	130
	12P05	2449	86	0.035	17.558	3	17	0.007	3.471	31	0	0.000	0.000	-	10	0.004	2.042	51	3	0.001	0.612	137
	12P08	2082	71	0.034	17.051	3	12	0.006	2.882	36	1	0.000	0.240	231	8	0.004	1.921	59	12	0.006	2.882	36
	12P12	2027	72	0.036	17.760	2	19	0.009	4.687	24	0	0.000	0.000	-	8	0.004	1.973	57	1	0.000	0.247	234
Psychology	12S01	2546	36	0.014	7.070	9	0	0.000	0.000	-	0	0.000	0.000	-	2	0.001	0.393	179	0	0.000	0.000	-
	12S02	1640	27	0.016	8.232	12	6	0.004	1.829	67	0	0.000	0.000	-	11	0.007	3.354	40	1	0.001	0.305	191
	12S03	2046	47	0.023	11.486	6	6	0.003	1.466	71	0	0.000	0.000	-	8	0.004	1.955	54	5	0.002	1.222	84
	12S05	2524	57	0.023	11.292	4	0	0.000	0.000	-	1	0.000	0.198	240	16	0.006	3.170	39	1	0.000	0.198	240
	12S09	2190	13	0.006	2.968	34	5	0.002	1.142	94	0	0.000	0.000	-	9	0.004	2.055	53	0	0.000	0.000	-