UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Taxonomic Organization and Evolutionary Relatedness Understanding in Experts and Novices

Permalink

https://escholarship.org/uc/item/2pn86860

Author

Marcroft, Tina Ashley

Publication Date

2024

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA SAN DIEGO

SAN DIEGO STATE UNIVERSITY

Taxonomic Organization and Evolutionary Relatedness Understanding in Experts and Novices

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Math and Science Education

by

Tina Ashley Marcroft

Committee in charge:

University of California San Diego

Stanley Lo, Chair
Thomas Bussey
Chris Halter

San Diego State University

Regis Komperda, Co-Chair
Scott Kelley
Joanne Lobato

2024

Copyright

Tina Ashley Marcroft, 2024
All rights reserved

The dissertation of Tina Ashley Marcroft is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

_____

_____

_____

_____

_____
Chair

University of California San Diego

San Diego State University

2024

DEDICATION

This dissertation is a tribute to those who have steadfastly supported me throughout this journey. Your belief in me has been unwavering, especially during the countless challenging moments that "struggle" barely begins to describe. I am profoundly grateful for your enduring support. To Carrie Umetsu, Janet Buckner, Princess Gilbert, Amisha Gadani, and Stephanie Steele: your wisdom, sharp wit, and patience have seen me through many tears and tribulations. My gratitude also extends to my committee, advisors, and cohortmates who I thank for their invaluable guidance. I am deeply thankful to my mentors, Regis Komperda, and Stanley Lo, for their extensive guidance and mentorship through a journey that was undoubtedly demanding; their steadfast support has not gone unnoticed. To Isabel White, Kristin Tenney, and Talia LaTona-Tequida, my gratitude knows no bounds for your kindness and encouragement. Most importantly, I dedicate this dissertation to my partner, Sam Kahn, whose boundless love and support made the completion of this work possible. My heartfelt thanks to all of you.

TABLE OF CONTENTS

Dissertation Approval Page....iii

Dedication....iv

Table of Contents....v

List of Tables....x

List of Figures....xi

List of Abbreviations....xiii

Acknowledgements....xiv

Vita....xv

Abstract of the Dissertation....xvi

Chapter 1 – Introduction....1
- Phylogenetic Trees in Biology....1
- Tree-Thinking Education....3
- Relatedness Understanding: The Foundations of Tree-Thinking?....7
- Motivations for This Research....8
- My Study....10
- Research Questions....11
- Contribution to the Literature....12
- Dissertation Summary....13

Chapter 2 – Literature Review....15
- Tree-Thinking....15
  - Summary....15
  - What is Tree-Thinking?....15
  - Tree-Thinking: A Problem?....16
  - The Misconception Literature....18
    - Inappropriate Mapping of Time....20
    - Progress is Depicted....20
    - Backbone is the "Main Line" of Evolution....21
    - Complexity Increases with Time....23
    - Similarity Indicates Relatedness....24
    - "Reading Across Tips"....26
    - Node Counting Determines Relatedness....27
    - Evolutionary Change Only Occurs at Splits....28
    - Unbroken Lineages Have Longer Evolutionary Histories....29
    - Unbroken Branches Have Ancestral Traits and Have Not Changed....30
    - Other Misconceptions or Misinterpretations....30

Summary 31
Instructional Strategies 32
Tree Analysis versus Tree Construction 32
Use of Alternative Metaphors 34
Framing Tree-Thinking and Relatedness 36
Tree-Thinking 36
What About Relatedness? 46
Biology Education and Relatedness 47
Organismal Classification Outside of Biology Education 48
What Do We Need to Know About Relatedness? 50
This Study 50
What Work is Needed? 50
Research Questions 52
Frameworks Employed in this Study 54
Embodied Cognition: Metaphor and Classification 54
Cognitive Metaphor (Gesture) 54
Categorization: Classical Theory, Exemplar Theory, and Prototype Theory 55
Classical Theory 56
Prototype Theory 56
Exemplar Theory 59
Prototype Theory and Exemplar Theory as a Continuum 60
Criticisms of the Dominance of Prototype, Classical, and Exemplar Theories 60
What Does All of This Mean? 63
Expert-Novice Comparison 64
Representations 65
What Are Representations? 66
What Are The Functions of Representations? 68
Chapter Summary 69

Chapter 3 – Methods 71
Summary 71
Positionality 71
Frameworks Employed by this Study 72
Expert-Novice Framework 73
Constructivist Perspective 75
Cognitive Perspective 76
Embodied Cognition 77
Contextualized Learning 78
Representations as Tools versus Targets 79
Prototype Theory 81
Procedure Overview 81
Rationale for Qualitative Methods 82
Tasks in Detail 84
Task Summary 84
Task Development 86
Rationale for Segmenting Participants into Experts and Novices 94
Undergraduate Participant as Novices 94

Faculty and Post-Doctoral Researchers as Experts 96
Sampling Strategy and Recruitment 98
Sample Composition 101
Clinical Interviews 102
Data Summary 103
Analytical Methods 104
Unit of Research 105
Grounded Theory 106
Background and Implementation 106
Why Grounded Theory 108
Coding in Grounded Theory 110
Establishing Validity 111
Coding Procedure 114
Primary Description of Procedure 114
Expert and Novice Separation 118
Data Reduction 118
Definitions and Operationalization of Terms 121
Defining Evolutionary Relatedness 121
Operationalizing Relatedness 122
Internalization 123
Category, Subcategory, Codes 123
Chapter Summary 124

Chapter 4 – Relatedness Understanding: Structure and Strategy 125
Summary 125
Introduction 125
Overview of Results 127
Structural Metaphors 130
Vertical 133
Description 133
Top-Down and Bottom-Up 136
Scientific Normativeness 140
Experts versus Novices 141
Flat 142
Description 142
In-Out and Out-In 147
Scientific Normativeness 148
Experts versus Novices 149
Tree-Like 153
Description 153
Twisted Lineages and Traditional 159
Scientific Normativeness 162
Experts versus Novices 163
Anthropocentrism 166
Description 166
Scientific Normativeness 170
Experts versus Novices 170

Family 172
Description 172
Scientific Normativeness 175
Experts versus Novices 175
Storytelling 177
Description 177
Scientific Normativeness 182
Experts versus Novices 183
Siloed 186
Description 186
Scientific Normativeness 190
Experts versus Novices 191
Grouping Strategies and Grouping Criteria 193
Grouping Strategy 195
The Strategies 195
Normativeness 208
Experts versus Novices 208
Grouping Criteria 213
The Criteria 214
Normativeness 220
Experts versus Novices 221
Authority and Essentialism: Special Cases of Criteria and Strategies 225
Chapter Summary 226

Chapter 5 – Relatedness: What Do These Findings Mean 228
Summary 228
Research Question 1: The Nature of Taxonomic Reasoning Structure 228
Commonalities Among Multiple Structures: Group Size, Hierarchicalness, and Time 229
Hierarchy 229
Directionality (Group Size) 232
Time 234
Research Question 2: Taxonomic Reasoning Strategies and Criteria 236
Similarity: Prototype Theory, and Isotype Strategies 237
Essentialism and Its Possible Effects 239
The Ubiquity of Relatedness Understanding 241
Essential Components of Relatedness Understanding 241
Dynamic Taxonomic Reasoning 242
Implications for Research and Teaching 243
Prevalence of Metaphors in the Field 244
Impact of Metaphorical Speech on Learning 246
Structure, Strategy, and Criteria as Diagnostics 248
Backtracing as Indicators of Progress 248
Targets of Learning 249
Grouping Criteria Proximity 252
Being Explicit About Methods and Cognition 253
Future Directions 255

Toward a Synthetic Model of Taxonomic Thinking....256
Limitations....259
Lack of Gesture Documentation....259
Inauthentic Settings....261
Expert-Novice Comparison....262
Task Coverage....262
Final Takeaways....263

Appendix A. Primer on Phylogenetic Trees....266

Appendix B. Final Interview Protocol....280

Appendix C. Final Task Slides....283

Appendix D. Contact Summary Sheet....286

Appendix E. Codebook....287

References....308

LIST OF TABLES

Table 1: Summary of Common (Mis)Conceptions in the Tree-Thinking Education Literature 19

Table 2: Taxa Included in the Task 87

Table 3: Probing Questions Correspondence to Research Questions 93

Table 4: Table of Participant Demographic Information. Bolded pseudonyms were given by the author. 102

Table 5: Coding Steps 117

Table 6: Summary of the Highest-Order Categories Comprising Relatedness Understanding 128

Table 7: Subcategories of Structural Metaphor 132

Table 8: Number and Frequency of Structural Metaphor Responses 133

Table 9: Number of Individuals Displaying Overlapping Structural Metaphors (Transcripts Only) 167

Table 10: Number and Frequency of Grouping Strategy Responses 195

Table 11: Number and Frequency of Grouping Criteria Responses 214

LIST OF FIGURES

Figure 1: An ultrametric phylogenetic tree. 2

Figure 2: Two phylogenetic trees with same topology communicating the same relationships. 3

Figure 3: Equivalent trees – left is unladderized, right is ladderized. 21

Figure 4: Bracketed tree and diagonal tree with highlighted “backbones.” 22

Figure 5: Author's representation of the vertical structural metaphor. 134

Figure 6: Gardner's relatedness diagram (expert). 138

Figure 7: Turtle's relatedness diagram (novice). 139

Figure 8: Louise's relatedness diagram (novice). 140

Figure 9: Author’s representation of flat structural metaphor. 143

Figure 10: Sita’s relatedness diagram (expert). 146

Figure 11: Elisa's relatedness diagram (novice). 146

Figure 12: Peppermint's relatedness diagram (novice). 147

Figure 13: Author's representation of the tree-like structural metaphor. 153

Figure 14: Ella's relatedness diagram (novice). 157

Figure 15: Witch's relatedness diagram (expert). 157

Figure 16: Mika's relatedness diagram (expert). 158

Figure 17: Author's representation of the anthropocentric structural metaphor. 167

Figure 18: Author's representation of the family structural metaphor. 173

Figure 19: Author's representation of the storytelling structural metaphor. 178

Figure 20: John's relatedness diagram (novice). 179

Figure 21: Victoria's relatedness diagram (novice). 180

Figure 22: Author's representation of the siloed structural metaphor. 187

Figure 23: Seline's relatedness diagram (novice). 189

Figure 24: Seline's relatedness diagram (novice) when explaining her diagram. 189

Figure 25: Tentative proposed model of relatedness understanding. 256

Figure 26: An example of a topophylogeny 267

Figure 27: Tree featuring a collapsed clade. 268

Figure 28: Tree containing a polytomy. 269

Figure 29: Bracketed/squared tree (left) and diagonal/triangular tree (left) with identical tree topologies. 270

Figure 30: Rounded bracket tree 270

Figure 31: Bracket tree fanned out into a circle 270

Figure 32: An unrooted tree. 272

Figure 33: Tree with ingroup and outgroup labeled. 272

Figure 34: Cladogram showing relative time where branch lengths are meaningless (Left). Chronogram showing absolute time where branch lengths refer to time (Right). 273

Figure 35: Example of issues with sequence alignment. 277

LIST OF ABBREVIATIONS

MENA	Middle Eastern North African
MRCA	Most Recent Common Ancestor
PAD	Plant Awareness Disparity
SABER	Society for the Advancement of Biology Education Research
SACNAS	Society for the Advancement of Chicanos/Native Americans in Science
STEM	Science, Technology, Engineering and Mathematics

ACKNOWLEDGMENTS

I would like to acknowledge my chairs, Regis Komperda and Stanley Lo. Their help in motivating me to conduct this study and write this dissertation has been invaluable. I thank them for all of the hard work they have done.

I would also like to acknowledge my fellow graduate students who are part of the Math and Science Education Doctoral Program at SDSU and UCSD. I appreciate the community we have worked hard to cultivate, and I love that we look out for one another. I thank the MSED community for all their support.

All chapters contain unpublished material coauthored with Komperda, Regis and Lo, Stanley. The dissertation author was the primary researcher and author of these chapters.

VITA

2005 – 2009 Bachelor of Science in Biology, University of California, Los Angeles

2010 – 2015 Master of Science in Biology, University of California, Los Angeles

2010 – 2017 Teaching Assistant, University of California, Los Angeles

2017 – 2024 Graduate Research Assistant at San Diego State University

2017 – 2024 Doctor of Philosophy in Math and Science Education, University of California San Diego & San Diego State University

PUBLICATIONS

Marcroft, T. A., Rasmussen, C., & Kelley, S. T. (2022). Computing in Bioinformatics and Engaged Student Learning: Student Perspectives on Anticipatory Activities and Innovative Apps. *Journal of College Science Teaching*, *52*(2).

Lazenby, K., Tenney, K., Marcroft, T. A., & Komperda, R. (2023). Practices in instrument use and development in *Chemistry Education Research and Practice* 2010–2021. *Chemistry Education Research and Practice*, *24*(3), 882–895. https://doi.org/10.1039/D2RP00275B

Santini, F., Sorenson, L., Marcroft, T., Dornburg, A., & Alfaro, M. E. (2013). A multilocus molecular phylogeny of boxfishes (Aracanidae, Ostraciidae; Tetraodontiformes). *Molecular Phylogenetics and Evolution*, *66*(1), 153–160. https://doi.org/10.1016/j.ympev.2012.09.022

Wassenbergh, S. V., van Manen, K., Marcroft, T. A., Alfaro, M. E., & Stamhuis, E. J. (2015). Boxfish swimming paradox resolved: Forces by the flow of water around the body promote manoeuvrability. *Journal of The Royal Society Interface*, *12*(103), 20141146. https://doi.org/10.1098/rsif.2014.1146

ABSTRACT OF THE DISSERTATION

Taxonomic Organization and Evolutionary Relatedness Understanding in Experts and Novices

by

Tina Ashley Marcroft

Doctor of Philosophy in Math and Science Education

University of California San Diego, 2024
San Diego State University, 2024

Stanley Lo, Chair
Regis Komperda, Co-Chair

Phylogenetic trees, diagrams depicting hypothesized evolutionary relationships between taxa, are used widely across multiple biology subdisciplines. Being able to read these diagrams is therefore a crucial step in achieving expertise in biology. Tree-thinking, however, has been found to be quite difficult, especially given the abstract nature of the diagram and its reference to directly unobservable macroscopic biological phenomena.

The biology education literature, therefore, has dedicated some effort into investigating the ways in which tree-thinking might be improved in undergraduates. While these investigations have been fruitful, they have not yet included a systematic examination of expertise in

comparison to novices in order to fully characterize the skills and conceptions required for tree-thinking. They also have not yet examined the concept at the root of tree-thinking: relatedness understanding. To remedy these issues, I conducted a study employing expert and novice framing of relatedness understanding. I also draw upon cognitive metaphor and Prototype Theory to inform this study which has not yet been employed by the tree-thinking community. I attempted to characterize the everyday ways individuals categorize and identify organisms and how they might differ by experience.

I conducted a series of clinical interviews with 29 individuals: 11 experts and 18 novice undergraduate students. Participants grouped organisms from photos and were asked to explain and draw their reasoning. I analyzed the resultant transcripts and artefacts using Grounded Theory.

Three primary components of relatedness understanding were apparent in all participants: *structural metaphor*, *grouping strategy*, and *grouping criteria*. *Structural metaphor* refers largely to spatial conceptual metaphors describing the arrangement of taxa. *Grouping strategy* refers to group membership determination while *grouping criteria* refers to the properties used with strategies to determine group identity. Experts exhibited understandings consistent with novices and partially non-normative, suggesting that some repression of nascent relatedness understandings may be occurring, or heuristics may be used in certain conditions. Further investigation of the impact of these understandings on tree-interpretation is necessary as well as whether or not learner generated representations of these understandings can be leveraged into normative interpretation of phylogenetic trees.

CHAPTER 1 – INTRODUCTION

Phylogenetic Trees in Biology

Phylogenetic trees (or simply *trees*) are diagrams integral to research in several biology subdisciplines including epidemiology, evolutionary medicine, population genetics, biogeography, and systematics (e.g., Boni et al., 2020; Buckner et al., 2015; McPeek et al., 2008; Santini et al., 2013). Typically, they depict the evolutionary history of organisms (organismal relatedness) but are also used to understand a variety of complex micro and macroevolutionary processes such as speciation, trait evolution, the emergence of zoonoses, and dispersal. Thus, trees are intrinsically tied to evolution which is often recognized as an organizing concept in biology (Baum et al., 2005; Dobzhansky, 1973).

Given trees' importance and pervasiveness in biology, many have argued that students (and the public) should be trained in the interpretation of these trees, also referred to as *tree-thinking* (Baum et al., 2005; Baum & Offner, 2008; Catley, 2006; Catley & Novick, 2008; MacDonald & Wiley, 2012; O'Hara, 1997). In addition, many calls to improve undergraduate biology education have focused on improving representational competence—the interpretation and usage of diagrams in science (Bauerle et al., 2009). Unfortunately, phylogenetic trees are often difficult for students and the public to understand (Halverson, 2011; Kummer, 2017; Meir et al., 2007; Phillips et al., 2012; Young et al., 2013) likely because they are symbolic, abstract diagrams (Pierce, 1931, as cited in Chandler, 2019) sharing few features with the biological phenomena they represent.

Tree representations are variable, but they all contain edges (lines) and points at which lines converge or diverge (nodes; see Figure 1; also see Appendix). These lines represent organismal lineages through time, and nodes represent splitting events (e.g., speciation). All

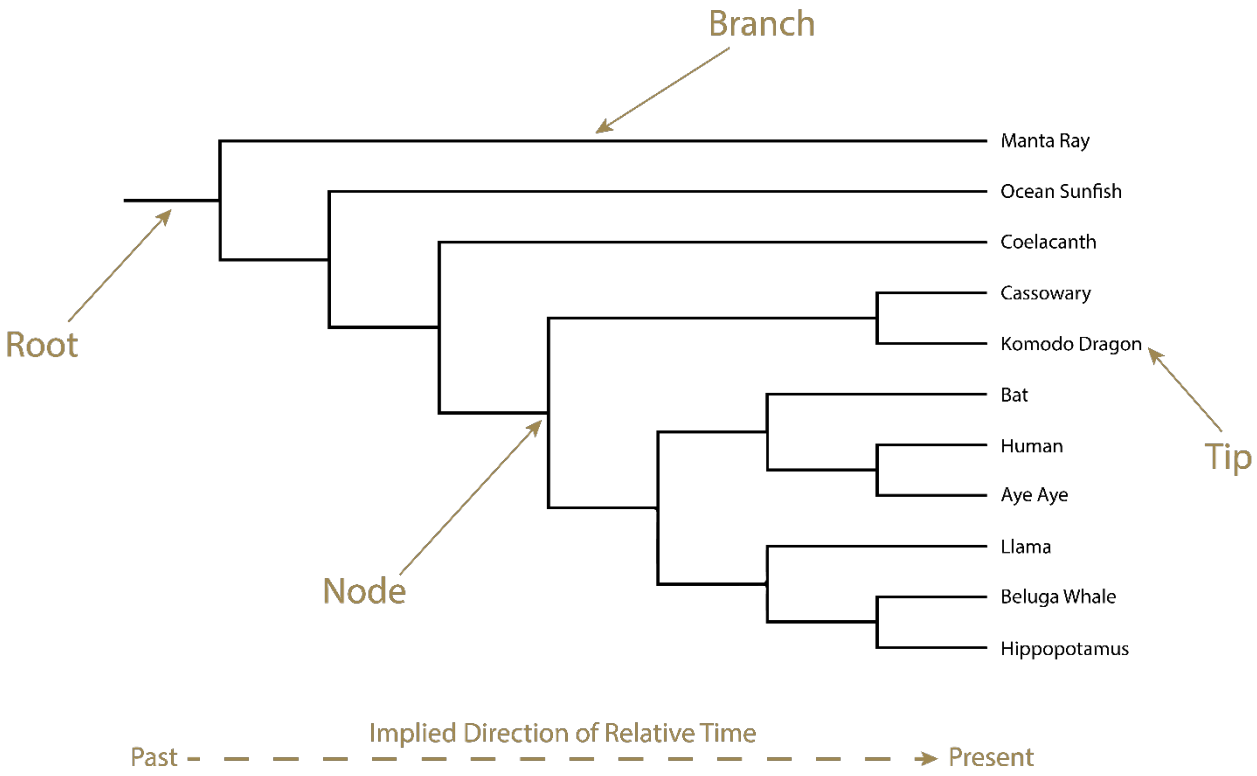


Figure 1: An ultrametric phylogenetic tree.

lineages culminate at the tips of the tree, typically representing extant organisms. These simplistic two-dimensional diagrams, however, often fail to adequately communicate the hierarchical relationships they are meant to represent. The comb-like structure of trees often leads to sequential interpretations of organismal relationships, misleading novices to believe that adjacent tip lineages are closely related when the order of splitting events (tree topology) indicates otherwise (Gregory, 2008; Kortz et al., 2012; Meir et al., 2007). Correctly interpreting trees requires understanding tree shape; most importantly, lineages that share the same node are more closely related to each other than to any other lineages, regardless of the lineages' ordering at the tips. It is this property, best illustrated by the diagrams' ability to be rotated about nodes without losing meaning (Figure 2), that is likely difficult for students to understand.

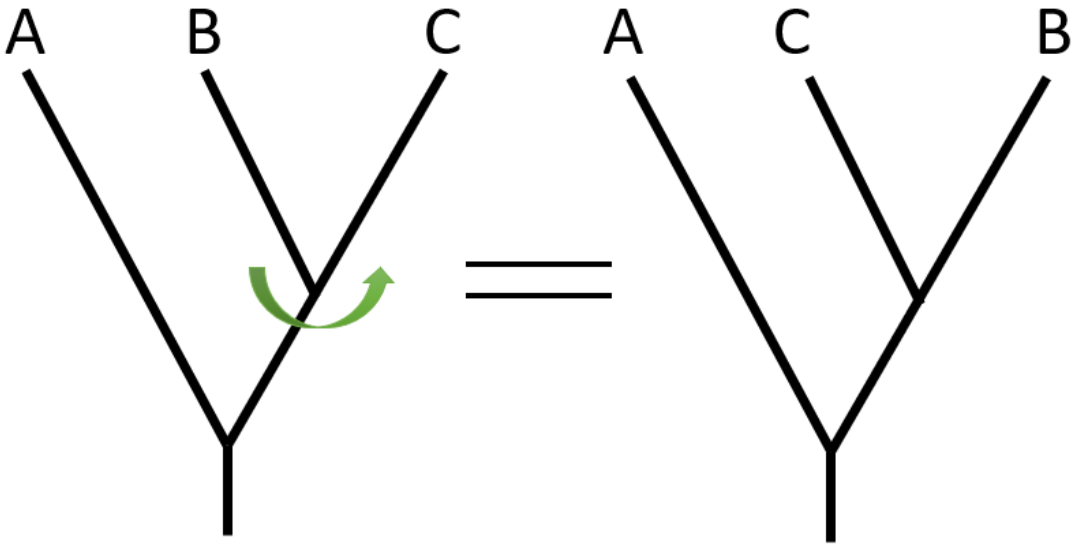


Figure 2: Two phylogenetic trees with same topology communicating the same relationships.

Additionally, the assumptions associated with the tree structure are often not clearly articulated, which some recent education researchers have implied is a problem consistent with other diagrams in science (Wright et al., 2018).

Tree-Thinking Education

As a result of the difficulties in interpreting tree diagrams, the biology education community has directed some attention towards tree-thinking. Much of the literature consists of interventions intended to help students construct and understand phylogenetic trees (e.g., Crisci et al., 1993; Davenport et al., 2015; Eddy et al., 2013). A large proportion of the remaining literature focuses on misconception work (e.g., Gregory, 2008; Halverson et al., 2011; Kummer et al., 2016; Meir et al., 2007; Meisel, 2010), some of which adopting a cognitive psychology perspective focused on innate (biological or common developmental) perceptual biases that may lead to tree misinterpretation (Liu & Novick, 2022; Novick & Catley, 2007; Novick & Fuselier, 2019; Novick, Stull, & Catley, 2012).

The aforementioned misconception work that was common early in the field's development provided no student utterances as evidence for their assertions and therefore only presented the instructors' interpretations of students' (mis)understandings (e.g., Baum et al., 2005; Baum & Offner, 2008; Meisel, 2010). Those misconceptions were then adopted, mostly unaltered and a priori, into other applied studies (e.g., Meir et al., 2007; Perry et al., 2008). As these hypotheses were not originally subject to inductive and generative methods, they may not adequately describe the realities of tree-interpretation. While recent researchers have investigated these conceptions in additional detail (Dees & Momsen, 2016; Dees et al., 2014; Halverson et al., 2011; MacDonald, 2014; Matuk & Uttal, 2020), more work is required. Sampling and contexts are still limited. Expertise, for example, has been only briefly examined in two studies (Halverson et al., 2011; Novick & Catley, 2007) so it is unknown how novice conceptions compare to expert conceptions, one of which only investigated the differences in translation between tree-types and did not investigate any other elements of expertise. Examining expertise in detail may elucidate different components of expert tree-thinking that exist, thereby clarifying the goals of tree-thinking education or heretofore undocumented variants of normative and non-normative reasoning may exhibit alternative avenues by which tree-thinking could be supported. In addition, supporting those findings with evidence in the form of utterances or artifacts will further validate those findings and provide opportunities for researchers to form new questions given the more detailed data provided to them.

Similarly, much of the literature on interventions also lacks compelling evidence on their effectiveness with many studies providing general estimates of performance in the form of grades or assessments with limited validity evidence, if any metrics are given at all (Davenport et al., 2015; Firneno et al., 2023; Gendron, 2000; Gibson & Cooper, 2017; Huang & Whittall, 2018;

Kumala, 2010b; Perry et al., 2008). Typically, evidence-based measures of the construct to be assessed are recommended to ensure the validity and reliability of the results (American Educational Research Association et al., 2014), and as general performance measures may be inadequate for capturing true performance differences within the domain of interest. The field has, however, very recently begun to recognize and address this need with the recent influx of well validated instruments with good support (Jenkins et al., 2022; Kong et al., 2022; Kummer et al., 2019; Schramm et al., 2021a). While these assessments often did include the input of experts as advised by assessment validation recommendations (American Educational Research Association et al., 2014; Arjoon et al., 2013; Barbera & VandenPlas, 2011), a comprehensive and generative examination of expertise and a comparison with novices has still not been conducted to my knowledge. Such a targeted examination may lead to clarification of the goals of representational competence in tree-thinking; indeed, I believe there is some discourse to be had around some of the misconceptions in the literature and the way they were framed. "Similarity equals relatedness," for example, is variably defined (Baum et al., 2005; Blacquiere et al., 2020; Dees et al., 2014; Gregory, 2008; Halverson, 2011; Halverson et al., 2011) and likely refers both to the revelation that traits can be analogous or to specific operationalizations of "similarity" based methods (distance-based methods) that are less favored such as neighbor-joining (Blair & Murphy, 2011). Such a definition requires reevaluation should we consider the ways in which such an understanding was arrived at and the basal understanding of relatedness upon which that understanding is made.

In addition to this general lack of evidence in the literature, much of the newer work which does provide evidence is concerned with finding the single best type of diagram for biology instruction (MacDonald, 2014). This work has been productive and has included surveys

of the types of diagrams that are common in educational texts (Catley & Novick, 2008; Schramm et al., 2021b) as well as the ways that diagonal and bracketed trees differ in the ways they are interpreted (Dees et al., 2017; MacDonald, 2014; Novick & Catley, 2007; Novick & Fuselier, 2019; Novick et al., 2011), with a strong majority of the literature favoring bracketed trees. This work, however, does not often consider findings in lieu of the broader representational literature in science education. While some tree-thinking researchers have begun to draw from more productive domains of the representational literature (Dees & Momsen, 2016; Dees et al., 2014; Halverson et al., 2011; MacDonald, 2014; Matuk & Uttal, 2020), other phenomena such as multiple representations, dual-channel processing, etc. (e.g., Kelly et al., 2017; Opfermann et al., 2017; Terwel et al., 2009) and their relationship to tree-thinking have remained mostly unaddressed. Only recently have learner-generated representations and representational competence been referenced in the tree-thinking literature (Halverson & Friedrichsen, 2013; Matuk & Uttal, 2020; Schramm et al., 2019).

The ways in which context may impact tree interpretation also has not been fully investigated. While some attempts have been made to examine the impact of "prior knowledge" on tree thinking (Halverson et al., 2011; Novick & Catley, 2014), they are limited in scope and do not address the breadth of contexts that might impact tree-thinking such as plants. Plant blindness, also known as plant awareness disparity (PAD) (Parsley, 2020), for example, has been widely documented in the biology education literature (Achurra, 2022; Wandersee & Schussler, 1999) where simply replacing animal organisms with plants in evolution assessments cause students be more likely to choose incorrect answers (Heredia et al., 2016). Students do not recognize that plants are subject to the many of the same evolutionary processes as animals, likely because they are so unfamiliar. Specific contexts, e.g., specific taxa or diagram additions,

may therefore encourage different behaviors during tree-thinking. As this context varies by discipline and taxa of interest, this aspect of tree-interpretation warrants further investigation as it may variably impact tree-thinking.

Relatedness Understanding: The Foundations of Tree-Thinking?

Amid these varying gaps that demand attention, one is of significant importance due to its capacity to help the field establish a more cohesive foundational understanding of tree-thinking: relatedness. Relatedness here refers to the categorization of organisms and the understanding of their connectedness, i.e., how organisms are understood and relate to one another. The tree-thinking literature is, quite obviously and very appropriately, predominantly concerned with the interpretation of these normative representations. However, the literature does not engage with these foundational understandings of relatedness and classification which may offer insight into the ways in which tree-thinking may be supported. Indeed, other domains of science education have examined underlying conceptions to normative representations indirectly through the examination of students' modeling and representation of natural phenomena in the classroom (Margolin et al., 2022; Park et al., 2019; Windschitl et al., 2018). Especially in K-12 contexts, some interventions recommend that students create their own understanding of a phenomena and represent it, present it to other students, and subsequently refine their ideas and accompanying representation. The hope is that by both simulating the authentic process by which authentic problem solving, negotiation, and solicitation of critique is used by researchers to generate knowledge and by having students call upon their own resources to construct these understandings, students can be scaffolded toward normative understanding which implicitly includes normative representations.

In contrast, much of the tree-thinking literature is preoccupied with determining the "best" tree to provide learners (e.g., Dees et al., 2017; Novick & Catley, 2007), and almost all tree-thinking interventions (e.g., Gendron, 2000; Goldsmith, 2003; Novick & Fuselier, 2019; Novick et al., 2011; Novick, Stull, & Catley, 2012), even those that attempt to authentically replicate modern tree-building processes (Bokor et al., 2014; Firneno et al., 2023; Julius & Schoenfuss, 2006; Karimi et al., 2017), always assume and recommend that students couch their reasoning in the form of a tree, as if the tree were the only model or representation that could conceivably depict relatedness. While instruction on and about the formalisms of normative representations is critical and necessary (Florence & Yore, 2004), the existence of those exercises in other domains suggest that a similar exploration of everyday student understanding (resources) is required. In short, though trees depict hypotheses of relatedness, relatedness understanding outside the context of a tree, including everyday understandings of relatedness, have not been investigated. This is despite our understanding in science education that everyday understandings of natural phenomena can sometimes be used to scaffold toward normative understanding (Pea, 2004; Sherin et al., 2004). This is a potentially productive area of research as it may elucidate the precursor ideas that modulate or are used to construct nascent relatedness, and potentially by extension tree-thinking, understanding.

Motivations for This Research

In conceiving this study, I was motivated to investigate the underlying understandings of taxonomic thinking among adult learners, i.e., the everyday understandings of what organisms are and how they relate to one another without the context of a phylogenetic tree. I particularly was interested in adult learners as tree-thinking and tree-building are very often taught in introductory biology courses (Bokor et al., 2014; Firneno et al., 2023; Karimi et al., 2017). While

some literature does exist in psychology and linguistics on how young children think about broad taxonomic groups, i.e., essentialist versus non-essentialist conceptions of taxonomic groups (Gelman & Coley, 1990; Gelman & Markman, 1987; Johnson & Solomon, 1997; Springer, 1996; M. G. Taylor et al., 2009), to my knowledge none of such literature exists for undergraduate or other higher-level students, let alone in biology education or tree-thinking and only one such study exists for K-12 students in science education (Trowbridge & Mintzes, 1988). While biology education has focused on the learning of specific taxonomic groups, specifically as it pertains to the proverbial "walk through the tree of life" where students learn researchers' current consensus on the evolutionary relationships between organisms (i.e., what are lobe-finned fishes and what fishes are currently considered part of the lobe-finned fish group) (Ballen & Greene, 2017), the structure and process of said groupings as well as everyday taxonomic thinking and their potential impact on tree-interpretation has not been examined save for one recent study (Matuk & Uttal, 2020). My goal, therefore, was to document and describe different components of potentially generalizable taxonomic thinking in undergraduates and experts such that we can understand how best to support different ways of learning and how these understandings may or may not translate into better tree-reading and tree-construction by undergraduate biology students.

This study was further motivated by the other aforementioned aspects of the tree-thinking education literature that are also wanting, specifically the lack of deep qualitative evidence to support "(mis)conception" work as well as a lack of examination of experts and how they compare to novices. While much of the literature does describe tree-related understandings of taxa and their understanding or misunderstandings (e.g., Baum et al., 2005; Gregory, 2008; Kortz et al., 2012; Meir et al., 2007), most of it lacks evidence in the form of utterances provided to

readers or did not examine their data using exacting qualitative methods. I hope, therefore, not only to provide such evidence but also more rigorously interrogate such evidence with the use of qualitative tools. Similarly, expertise outside of a handful of studies (Halverson et al., 2011; Novick & Catley, 2007), has not yet been fully characterized. Formal examinations of expertise can elucidate viable avenues of potential support in the form of heretofore undocumented variants of normative understanding constructed from different basal understandings as well as clarify the goals of instruction. While Daniel (nee Halverson) and Friedrichsen have formally enumerated a series of competencies required for tree-thinking based on the representational literature in other domains of science education (Halverson & Friedrichsen, 2013), an examination of expertise in this specific domain can account for previously unaccounted for learning goals; experts may not necessarily be cognizant of all aspects of a concept that are of pedagogical use (Fisher & Keil, 2016). Indeed, as the reader shall soon observe, some of the misconceptions cited in the literature, especially those such as similarity equals relatedness, may need reexamination in light of these results, with greater attention paid to the ways in which normative understanding is communicated to and scaffolded for students.

My Study

To fill these gaps in understanding of tree-thinking, I conducted a study examining the ways in which experts and novices understand relatedness and ultimately group taxa by analyzing both utterances and diagrams generated by participants. Using a primarily expert-novice lens informed by the literature on embodied cognition and constructivism, I interviewed undergraduates and faculty, with the former broadly representing novices and the latter experts. I conducted one primary task that probed conceptions of relatedness by having participants view a variety of photographs of macroscopic taxa. They were probed for the ways in which they

thought about grouping organisms and the ways in which components of these organizing ideas interact.

Research Questions

This study was intended to address the overarching research question of "What are the ways in which individuals with varying experiences in biology understand relatedness?" In attempting to obtain a clearer picture of this question, two more detailed research questions were formulated. These questions are:

1. *Understanding Relatedness* – In what ways do experts and novices conceive of organism relatedness?

This question interrogates the ways in which experts and novices group organisms, either in an everyday sense or a scientifically normative, i.e., evolutionary, sense. This question is intended to partially scrutinize the nature of these understandings which, outside of the child psychology literature, have not yet been investigated, especially not in the undergraduate biology and tertiary biology context. These groupings may be equivalent or analogous to scientific classifications and thus may or may not be scientifically normative. This question gets at the shape of their understanding of relatedness: what are the ways in which taxa are structured or ordered and what are the ways in which they are described. I also discuss how these ideas differ by experience, i.e., the variation in understandings between novices and experts and how they may or may not differ, in order to better understand the goals of tree-thinking education as well as identify potential areas upon which nascent ideas might be scaffolded.

2. *Investigating Ways in Which Taxa are Grouped* – How do experts and novices group (i.e., classify) organisms?

This question investigates the methods by which individuals assign taxa to groups in a lay setting. This question gets at the nature of these strategies and the components of this understanding. Probing questions and the task discussed were intended to illuminate the ways in which participants decided to sort taxa. These strategies for doing so similarly may or may not be normative. I further investigated how these strategies might compare across varying experience levels and what that might mean for instruction.

Contribution to the Literature

This study expands and improves on the existing literature evaluating students' understanding of evolutionary relatedness and, presumably, interpretation of formal phylogenetic trees. First, this qualitative study was inductive and generative. It provides evidence for each construct generated in the form of utterances and artifacts for each concept examined. By comparing conceptions across groups with differing experience with phylogenetic trees, I hoped to characterize experts' and novices' everyday understanding of relatedness with the assumption that said understanding can lead us towards better characterization of relatedness understanding, the possible concepts that underly tree-thinking, and the potential everyday understandings instructors could leverage to support tree-thinking in the future. I expected to document some interpretations not yet described in the literature, given that such a study has not yet been conducted (i.e., a pointed examination of expertise in comparison to novice responses). I also assumed that such a study could result in data that may help the field construct a rigorous model of relatedness understanding which is not so strongly tied to normative understanding, i.e., trees.

Secondly, this study employed perspectives that appear infrequently in the tree-thinking literature. Instead of the typical cognitive psychology perspective, I adopted a primarily constructivist perspective which assumes that prior experiences influence perception as well as

an embodied perspective that seeks to understand some of these understandings in terms of the spatial metaphors employed to describe such diagrams. In addition, I explicitly recognized and formally recast tree-thinking as two distinct phenomena, i.e., conceptions of evolutionary relatedness and tree-interpretation. While some recently have recast tree-thinking as an understanding of evolution and interpretation separately (Kummer, 2017), this perspective might miss the ways students conceive of taxa which may or may not align with biologically and evolutionarily determined taxa. In addition, this study assumes that both experts and novices can conceive of relatedness and group organisms accordingly utilizing different resources, and that several reasonable strategies or conceptions for doing so may exist.

Dissertation Summary

In the following chapters, I contextualize my study beginning with a literature review (Chapter 2) that briefly describes phylogenetic trees with additional details provided in Appendix A. I will discuss the general areas of research that have been conducted in these areas and the perspectives by which these studies have been implemented, i.e., the frameworks which they employed. I do this for both relatedness and tree-thinking. I then articulate the gaps in the literature and the frameworks from other literature such as the cognitive psychology literature on organismal grouping and philosophy.

In Chapter 3, I describe my methods followed by a discussion of my background and how it may have affected my analyses. I then discuss the broad frameworks I employ and specifically how they impact my methods. I follow with a description of the general procedures I followed as well as other information such as sampling composition. I then discuss the tasks and summarize the types of data I obtained from the study. I follow with a discussion of my analytical

methodology, including the use of clinical interviews and analysis using Grounded Theory (Glaser & Strauss, 1967; Strauss & Corbin, 1994).

In Chapter 4, I present the results of my study. I begin with the analysis of the relatedness task and the different components of relatedness understanding that emerged from the data. I discuss how they related to my two primary research questions, and I provide evidence for those assertions. I also provide a cursory discussion of each of the subcategories observed as well as a comparison of expert and novice expression of said categories. A codebook containing the categories, subcategories, and so on, as well as the definitions of each code is provided in the appendix.

In Chapter 5, I begin with an in-depth discussion of the ramifications of these results in the context of the literature. I discuss what potential pedagogical innovations these results might inspire as well as the relevant next steps in relatedness understanding research. In addition, I discuss the future of tree-thinking research given these results. I follow that discussion with a detailed examination of the limitations of this study. I close by briefly summarizing the major takeaways of this study.

Chapter 1 contains unpublished material coauthored with Komperda, Regis and Lo, Stanley. The dissertation author was the primary researcher and author of this chapter.

CHAPTER 2 – LITERATURE REVIEW

In this section I discuss the work that has been done in tree-thinking education. I also discuss the limited work that has been done with regards to animal classifications in the psychology literature and how little that work has been applied to tree-thinking. I then detail the frameworks by which the aforementioned work has been done. I follow with a discussion of those perspectives that remain unaddressed in the literature and how they tie into my research questions. I follow with a discussion of the frameworks that this study employs.

Tree-Thinking

Summary

In the following sections, I discuss the main ways in which tree-thinking has been investigated in the biology education literature. I first discuss the largest portion of the literature that focuses on misconception work, i.e., largely what "incorrect" non-scientifically normative ideas students have. I close with a discussion of the interventions that have been created to help high school and undergraduate students improve their tree-thinking abilities.

What is Tree-Thinking?

Tree-thinking refers to any of the practices relating to the reading of phylogenetic trees (Baum & Smith, 2013). Phylogenetic trees are diagrams in biology that use lines and points at which lines join to represent differing levels of taxonomic groups (see Tree Primer in the Appendix for details). The relative ordering of the points where these lines join indicates the evolutionary history of these organisms. Tree-thinking therefore involves the interpretation of patterns such that the relevant information being communicated by its creator can be correctly understood by readers. This information includes the larger taxonomic groups that organisms belong to, the relative positioning of common ancestors, and the relatedness among lineages.

Tree-thinking may also refer to the utilization of trees in addition to their interpretation. Trees are often used to interpret patterns of trait evolution or other biological phenomena and appear in combination with other diagrams such as maps or morphospaces. A student who excels at tree-thinking, for example, would be able to determine what the relationships between organisms are and also coordinate that understanding with the other information provided by the diagram. In other words, they could describe what evolutionary patterns are occurring and at some level how they are related to the other biological phenomena depicted.

Expert tree-thinking may also potentially entail a deeper understanding of the processes underlying macroevolution as well as systematics (Novick, Schreiber, & Catley, 2014). Experts, as in other studies regarding expert representational competence (Kozma et al., 2000; Novick & Catley, 2007), are capable of flexibly translating between diagram types, indicating they have a deeper understanding of said representation and the concepts associated with it such that context (i.e., certain features) does not matter and they are able to focus on and translate the most salient components. Students, being novices, even if able to interpret phylogenetic trees, may struggle with such a task as they presumably lack such an understanding.

Tree-Thinking: A Problem?

Substantial scholarship has been dedicated to tree-thinking since the inception of the domain of biology education. Tree-thinking is recognized as a particularly difficult activity, with many novices and even experts very commonly being observed as struggling to interpret trees (Phillips et al., 2012). Tree-thinking has been split, at times, into tree-interpretation and tree-construction with separate efforts attempting to support the acquisition of those skills (Baum et al., 2005; Eddy et al., 2013; Halverson et al., 2009; Halverson & Friedrichsen, 2013). However, a portion of the community has also assumed that support in tree-construction may support tree-

interpretation (Eddy et al., 2013; Julius & Schoenfuss, 2006), so much of the literature includes tree-construction interventions intended to support tree-interpretation (Crisci et al., 1993; Dees & Momsen, 2016; Karimi et al., 2017; Kumala, 2010b; Singer et al., 2001).

Regardless of its definition, it appears help is needed in this area. Calls have been made in biology education to radically improve both undergraduate evolution education and representational competence (Bauerle et al., 2009). While tree-thinking has not explicitly been singled out as an area requiring improvement by the broad biology education community, some in evolutionary biology have explicitly advocated for such (Baum et al., 2005; Phillips et al., 2012; Sandvik, 2008). In addition, given the huge breadth of the field which depends on the usage of trees (e.g., Boni et al., 2020; Buckner et al., 2015; McPeek et al., 2008; Santini et al., 2013) and given that trees both are part of evolution and are extremely common representations that biologists are presumably required to understand, tree-thinking is of particular importance to biology education.

Indeed, much of the early literature refers to misconceptions that instructors have encountered in the classroom or even among colleagues; however, many of them lack evidence in the form of student assertions (Baum et al., 2005; Gregory, 2008; Meisel, 2010; Sandvik, 2008). Later literature became focused on reducing the appearance of such misconceptions by working extensively on what types of phylogenetic trees are the most pertinent for new learners (Catley & Novick, 2008; Dees et al., 2018; Liu & Novick, 2022; MacDonald & Wiley, 2012; Novick et al., 2011; Novick, Stull, & Catley, 2012; Novick, Pickering, et al., 2014), the assumption being that these external representations have features that are more easily or innately understood by novice viewers. However, this perspective ignores some of the other representational and modeling literature that promotes the utility of learner generated models and

representations which may help support student learning (Ainsworth & Scheiter, 2021; Matuk & Uttal, 2020; Van Meter & Garner, 2005) given that any natural phenomena can be represented "correctly" in a number of ways. In addition, the literature lacks any substantive comparisons between experts and novices which could impact the ways in which we conceive of learning goals. Experts may not be self-cognizant enough without the help of exhaustive qualitative analysis methods to thoroughly assess and provide evidence for such learning goals.

In the following sections, I discuss this misconception literature and the extensive literature on interventions. I then follow with a discussion of the broad areas of work that have been conducted and the frameworks they employ. I then discuss the major gap I wish to address by this study and the state of the literature in terms of that topic. I follow with a discussion of the study I conducted and the appropriate research questions. I end with a discussion of the frameworks I intend to employ which do not necessarily come from the tree-thinking literature.

The Misconception Literature

Much of the literature on tree-thinking focuses on conceptions or misconceptions. In the following paragraphs I will briefly discuss how these ideas have been discussed in the literature and their relationship to one another. A summary of these misconceptions appears in (Table 1). I discuss these misconceptions as they are related to some of the concepts I expected to observe in the study (see Results). Note that many of these misconceptions are overlapping, either due to the connectedness between each of the underlying conceptions or because of methodological differences.

Table 1: Summary of Common (Mis)Conceptions in the Tree-Thinking Education Literature

Misconception (Names Vary by Author)	**Description**	**Comment**	**References**
Inappropriate Mapping of Time	Axis of time placed in an incorrect direction.		(Dodick & Aharonson, 2009; Gregory, 2008; Meir et al., 2007).
Progress is Depicted	Progression or linearity is assumed to be shown in the tree.	Relates to hierarchical or non-hierarchical forms of grouping in this study.	(Crisp & Cook, 2005; Giusti & Scott, n.d.; Gregory, 2008; Halverson et al., 2011; Matuk & Uttal, 2012).
Backbone is the Main Line of Evolution	Typically observed in diagonal trees, greater importance and directionality is imposed upon the tree backbone.		(Ainsworth & Saffer, 2013; Halverson, 2011; Matuk & Uttal, 2012; Novick, Stull, & Catley, 2012).
Complexity Increases with Time	Complexity of organisms going down the tree (typically down tips) increases.	Relates to *Reading Across Tips* and *Progress is Depicted.*	(Gregory, 2008; Nee, 2005; O'Hara, 1992, 1997).
Similarity Indicates Relatedness	Organisms that are similar are more related.	Typically interpreted with the context of inappropriately broad morphological traits or phylogenetically uninformative traits (and disfavored tree-building methods) to determine relatedness. However, this conception may be retained and modified by biologists and is currently unrecognized.	(Baum et al., 2005; Gregory, 2008; Halverson et al., 2011). Also (Dees et al., 2014; Halverson, 2011; Halverson et al., 2011; Kummer et al., 2016).
"Reading Across Tips"	Individuals will read the tips sequentially from top to bottom or left to right, depending on the tree's orientation. The ordering of the tips indicates relatedness relationships.	Relates to *Complexity Increases with Time.*	(Baum et al., 2005; Kummer, 2017; Kummer et al., 2016; Meisel, 2010). Also, potentially (Dees et al., 2014; Gregory, 2008; Novick, Stull, & Catley, 2012).
Node Counting Determines Relatedness	The fewer nodes that separate two tip taxa, the more closely related they are.		(Dees et al., 2014; Halverson, 2011; Kummer, 2017; Kummer et al., 2016, 2019; Stenlund et al., 2021)
Evolutionary Change Only Occurs at Splits	Morphological and other change is only assumed to occur at nodes.	Complicated by the fact that gradual and punctuated equilibrium are recognized as normative hypotheses.	(Baum et al., 2005; Gregory, 2008; Meir et al., 2007)
Unbroken Lineages Have Longer Evolutionary Histories	Lineages that appear as longer lines on the diagram are assumed to have experienced more evolution.	Experienced the same amount—lack of splits may be due to sampling issues. Also, evolution can occur after a split has occurred.	(Gregory, 2008)
Unbroken Branches Have Ancestral Traits and Have Not Changed	Lineages that appear long are assumed to have the traits of the estimated ancestor.	Anagenesis may have occurred, or lack of sampling causes the appearance no splits occurring.	(Crisp & Cook, 2005; Gregory, 2008; Novick & Catley, 2007)

Inappropriate Mapping of Time

One misconception or misinterpretation commonly discussed in the literature is the inappropriate mapping of time (Dodick & Aharonson, 2009; Gregory, 2008; Meir et al., 2007). Trees can be oriented such that relative time begins from any direction. However, many of the trees in the literature are oriented such that the root is on the bottom or left side of the diagram. Trees oriented with the root at the top or right of the diagram are still technically correct; they are simply interpreted differently. The researcher viewing the tree would mentally rotate the diagram, were they not used to seeing those versions. Novices, however, might consistently suggest time is occurring on the same axis, i.e., perhaps assume that time is depicted as left to right but not rotating their interpreted axis of time with the diagram. This description of "rotation" should not be confused with the rotations about nodes that are commonly discussed in the literature—rotated trees simply refer to rooted trees that differ in terms of the root starting point rather than the direction that they are facing on a page.

Progress is Depicted

Another misconception ties into a misconception observed in evolution, i.e., the perception of progress and linearity (Crisp & Cook, 2005; Giusti & Scott, n.d.; Gregory, 2008; Halverson et al., 2011; Matuk & Uttal, 2012). Students will often be given a tree like the one pictured (Figure 3) in the biodiversity portion of an introductory biology class and simply memorize the order of the tips. Therefore, in Figure 3, though both trees depict identical evolutionary relationships, a student might interpret that humans are equally closely related to the aye-aye and hippo as they are located on either side of the human in the diagram. Conventionally, the diagram suggests the aye aye and human are actually more closely related to each other than either are to the hippo. Ladderization (the orientation of the tree such that the

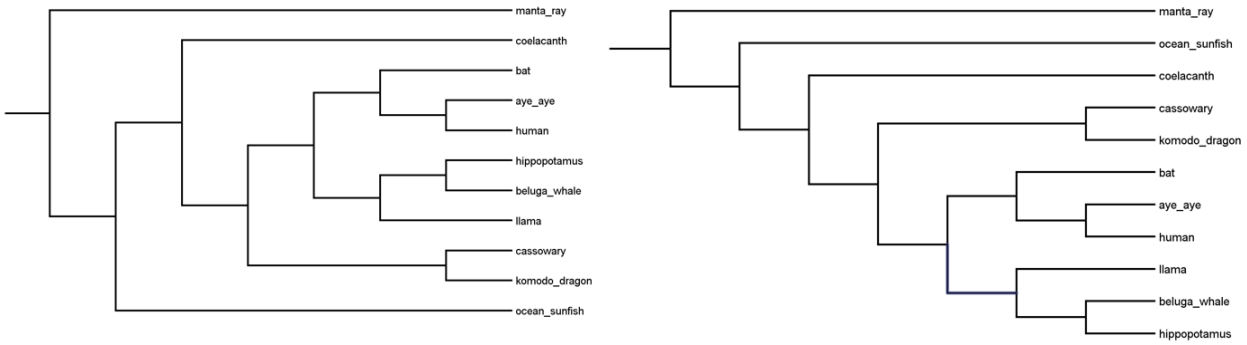


Figure 3: Equivalent trees – left is unladderized, right is ladderized.

branches are in order of length) impact this incorrect interpretation with an individual suggesting that the human is more closely related to the llama than the hippo as they are in closer proximity. In short, essentially interpreting the sequence of the tips (i.e., reading across the tips) suggests inappropriate evolutionary relationships to students and ladderized trees' linearity and orderedness may imply that some level of directional progress or sequence is represented by the tree when that is not the case. The literature is unclear about how this "progress" differs from the idea of increasing complexity, as discussed below.

Backbone is the "Main Line" of Evolution

In this conception, students interpret the backbone of a tree as being the "main line" or primary "path" or "ladder" of evolution (Ainsworth & Saffer, 2013; Halverson, 2011; Kummer et al., 2016; Matuk & Uttal, 2012; Novick, Stull, & Catley, 2012). The backbone of a tree typically refers to the longest line observable from the root – backbones are more easily visible in diagonal type trees (Figure 4). Novices will generally assign importance to that backbone and interpret it as a single directional unit (Novick & Catley, 2007). This interpretation is considered nonnormative because the backbone is not considered special or different from any other part of the tree and it does not necessarily imply a sequence of events. Lineages are considered equal to one another after a split, and the backbone is not considered a special "ancestor" or lineage. Instead, it is a visual byproduct of the ladderization of a tree as well as sampling issues and

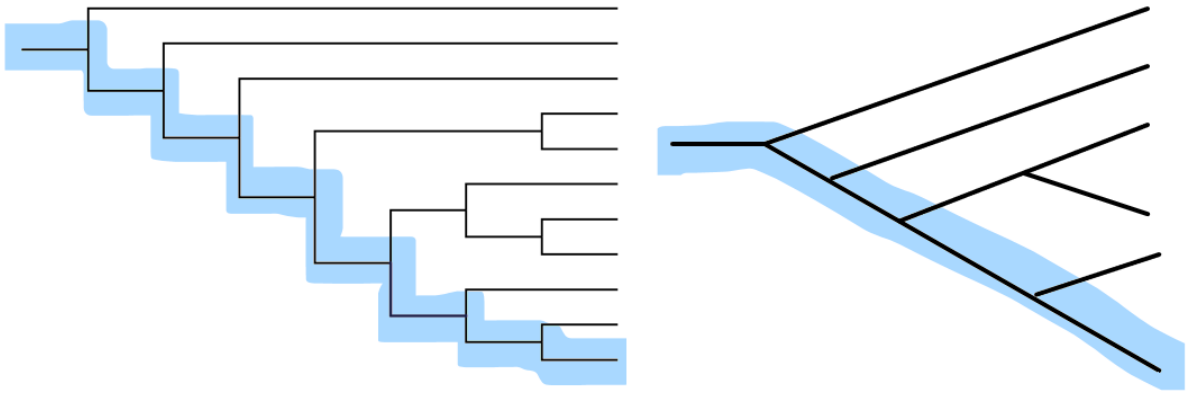


Figure 4: Bracketed tree and diagonal tree with highlighted “backbones.”

indeed some imply giving students unladderized trees might deter this behavior (Crisp & Cook, 2005; Meisel, 2010; Omland et al., 2008) though they may increase cognitive load. Note that for the unladderized tree in Figure 3 it is more difficult to intuit a single “line” unit or series of events.

Ainsworth and Saffer (2013) have suggested students interpret the backbone as a "path" of evolution (note that this definition using directionality is difficult to disentangle from the aforementioned "progress depicted" misconception). Novick, Catley, and Schreiber (2012) and Novick and Catley (2007) have observed that when drawing trees students will draw a solid line representing the backbone and subsequently draw branches coming off the backbone.

Matuk (2010) has suggested that this interpretation of trees is related to our storytelling tradition, where the backbone provides a time component to the story and the branches off the main line are characters. Matuk, however, does not only examine the backbone but instead attributes all features of the diagram to mental models related to storytelling. Novick and Catley (2007) alternatively suggested that this tendency to assign greater importance to this line is a byproduct of our innate (biological or developmental) perceptual orientation towards long unbroken lines, referred to as the Gestalt principle of Good Continuation. However, Novick, Stull, and Catley (2012) confusingly also relate this behavior to the misconception "reading

across tips" and explain the scanning of such lines as a byproduct of our native language reading order experience.

As a result of this work examining this misconception, a very large proportion of the literature is dedicated to comparing students' interpretation of bracketed trees versus diagonal trees (Dees et al., 2017; Novick & Fuselier, 2019). As diagonal trees' backbones are more noticeable, many argue that bracketed trees should be used more often (Dees et al., 2017, 2018; Novick & Catley, 2007). These trees are often compared, as both bracketed and diagonal trees are the most common trees observed in modern textbooks and recent literature in empirical phylogenetics. As a result, some textbooks appear to have changed as a result (Schramm et al., 2021b) of this finding and to coincide with the systematics literature which tends to nowadays examine huge numbers of taxa for which bracketed trees are easier to interpret (note that this is mostly restricted to systematics, many other tree types are used in other disciplines or in in-person communication with colleagues).

Complexity Increases with Time

This misconception is related to a similar misconception in evolution, i.e., that organismal complexity increases with time (Gregory, 2008; Nee, 2005; O'Hara, 1992, 1997). Students do not suggest that single lineages evolve into more complex ones, but rather, given more time, increasingly complex organisms will appear. Students will examine a tree, often ladderized, and suggest that newer lineages or shorter branches will be more complex. This conception might be related to common representations of evolution such as the ascent of man showing evolution occurring sequentially and culminating in the appearance of humans (Zallinger, 1965, as cited in Blake, 2018). The figure is not available for reproduction here but it

is common enough that the reader may have seen it showing a series of primate and humanoid figures that gradually begin to walk upright, culminating in the last image representing a human.

This misconception is nonnormative because trait evolution is dependent on the circumstances of the lineage in question. Certain parasites, for example, have lost complexity in structure, having a streamlined gut, given their highly preprocessed food source—a complex gut is costly to maintain if unnecessary. Alternatively, the circumstances in which a lineage has experienced could lead to great complexity if that is what was selected for.

It is notable that this particular misconception in part overlaps with the "progress is depicted" and "backbone is main line" misconceptions in that all three have qualities of linearity and sequential events.

Similarity Indicates Relatedness

Similarity indicates relatedness is one of the most prevalent misconceptions in the literature, as I shall argue presently, and in my opinion needs some reevaluation given its vagueness. It typically describes when students will interpret two organisms as being closely related because they are similar in some way (Baum et al., 2005; Gregory, 2008; Halverson et al., 2011; Kummer, 2017; Kummer et al., 2016). For example, a student might interpret a whale to be closely related to fish when they are more closely related to other mammals because they share the same body shape. Unfortunately, this mistake is a result of the convergent evolution, a process by which many disparate species that share a very old ancestor converge on the same evolutionary traits because they were subjected to the same selective pressures.

Historically these analogous or convergent traits have posed a problem for algorithms used to determine relatedness—simply counting and comparing the number of differences between species for example might be problematic if these traits are present in the dataset

(Philippe et al., 2011). Other methods such as parsimony attempt to circumvent this problem by instead searching for trees with the fewest evolutionary events that could explain the data (thereby allowing for multiple appearances of traits, i.e., events, in different trees). Given the importance of this paradigm shift in evolutionary biology and its subsequent spawning of newer methods, it follows that the biology education community has spent a lot of time and effort in the literature investigating it.

What is additionally unclear is that the literature (and my personal experience) also implies that the rejection of "similarity" as being a tool for understanding relatedness is that it may be referring to specific phylogenetic reconstruction methods that are now generally disfavored, primarily distance-related methods where essentially the number of trait difference refers to the level of similarity. This is in contrast to other methods where differences are assumed to occur at different rates and that differences are used to infer evolutionary events which in turn are used to estimate relatedness.

Regardless of the vagueness of its definition, as will be discussed in the results of this dissertation, the framing of this conception as erroneous is potentially problematic. Though processes like convergent evolution do impair our ability to determine relatedness, the general principle could be considered the same regardless of the way the trees are constructed. Though we might use molecular data, the underlying concept could still be considered generally the same, i.e., that two sequences are likely to be more similar if they are related. The only difference is that this conception has been modulated to be more probabilistic, taking into consideration nuances such as differences in substitution rates. In other words, we use similarity in some sense to estimate relatedness.

A few researchers in the literature have treated the formulation of this conception more carefully as a result. Dees et al. (2014) referred to this conception as "external insights" and as using disconnected outside information. Halverson (2011) and Halverson et al. (2011) observed a specific variant of this conception that used organisms' environment to group species. While not discussed by Halverson et al. (2011) and Dees et al. (2014) explicitly, I believe both are examples of contextualization and knowledge-in-pieces (diSessa, 2018). Different prior knowledge or conceptions of relatedness based on overall physical appearance might be prompted depending on the organisms that appear on the tree. Indeed, evidence for this contextualization has been documented in natural selection assessments; students score lower on questions featuring plants when compared to identical questions featuring animals (Heredia et al., 2016).

This is one of the few (mis)conceptions that may be apparent in both novices and experts suppressed, as a heuristic, or slightly modified (Smith et al., 1994). If so, it and other conceptions may serve as a potential cognitive resource that educators can leverage for scaffolding students' understanding of tree-thinking, relatedness, and evolution.

"Reading Across Tips"

The most discussed misconception in the literature is "Reading Across Tips." "Reading across tips" is the practice of interpreting phylogenetic trees by interpreting the tips from left to right or top to bottom to determine evolutionary relatedness (Baum et al., 2005; Kummer, 2017; Kummer et al., 2016; Meisel, 2010). Students might, for example, interpret a tree of vertebrates such that the relationships are in order or close to each other. As mentioned, in the progress is depicted misconception, an individual might perceive the human to be equally closely related to the hippo and aye aye as they are placed before and after humans in the sequence (Figure 3).

However, appropriate reading of the tree reveals that humans are actually more closely related to aye ayes and distantly related to hippos.

Exactly why people are observed doing so is unclear and being investigated. One such explanation could be that it relates to the progress is depicted misconception and therefore is an outcome of having that particular interpretation or underlying understanding of the tree. Alternatively, it could refer to reading order.

The literature itself appears to not have reached consensus either on what is actually meant by this misconception or how to accurately characterize and differentiate it from other misconceptions. Kummer (2017) and Kummer et al. (2016) document this misconception as "reading across tips" but defines it as the "tip distance." Dees et al. (2014) referred to this misconception as "branch tip proximity." Baum et al. (2005) and Novick, Catley, and Schreiber (2012) suggested it is due to reading order. Gregory (2008) confusingly suggests students “read meaning into the order” of the tips. There is conflation about this misconception's formulation; is this referring to reading order, absolute tip distance, or relative distance (ordering)?

Node Counting Determines Relatedness

In this conception, students count the number of nodes on the path leading from one organism to another (Dees et al., 2014; Halverson, 2011; Kummer, 2017; Kummer et al., 2016, 2019; Stenlund et al., 2021). If taxon A and taxon B have 3 nodes between them and taxon A and taxon C have 2 nodes between them, a student will interpret taxon A and taxon C as being more closely related. Halverson (2011) referred to this conception as "node emphasis" while Dees et al. (2014) and Kummer (2017) and Kummer et al. (2019) referred to it as "node counting."

This misconception is considered nonnormative because in trees that have differential splits across the tree, individuals may interpret certain pairs of lineages to be more closely related

when they are not. Figure 3 can illustrate such a misconception. A student using this misconception might want to compare the following pairs of organisms, the bat and the llama and the bat and the whale. The bat and the llama have three nodes between them while the bat and the whale have four nodes between them. A student using this conception would therefore conclude that the bat is more closely related to the llama than to the whale. However, normative reading of the tree would suggest that the bat is most closely related to the ancestor leading to those three taxa and that the llama is actually more closely related to the whale, specifically the ancestor of the whale and hippo.

Evolutionary Change Only Occurs at Splits

In this misconception, students assume that morphological change and speciation only occur at nodes and no change occurs along lineages (Baum et al., 2005; Gregory, 2008; Meir et al., 2007). In other words, the shape or features of organisms do not change at any time points during the tree except at the moment of a split. This may not be true for a variety of reasons.

First, sampling issues may make the recovery and reconstruction of some splits impossible. If a species is missing due to extinction or unavailability, the splits including that lineage may not be recovered and the reconstructed tree will exhibit fewer nodes when compared to real evolutionary history. This explanation only holds if evolution is assumed to occur near splitting events.

Second, evolutionary change can appear anywhere on a lineage. Rather than only occurring at the tips, lineages can change slowly via anagenesis, i.e., evolutionary change that does not involve splitting into multiple species. This explanation only holds, however, if evolution is assumed to be gradual.

This misconception is problematic in that it overlaps with a legitimate debate in biology about rates of trait evolution, i.e., the existence of punctuated equilibrium versus gradual evolution (aka phyletic gradualism). Punctuated equilibrium is the hypothesis that most evolutionary change occurs around nodes; lineages generally experience stasis. Gradual evolution—what Darwin believed was most common—assumes evolution is slow, incremental, and generally occurs along lineages. Literature suggests that both phenomena do occur, but that punctuated evolution may be more common (Gould & Eldredge, 1977).

Only one author, Gregory (2008), acknowledges this debate. However, he suggested this understanding occurs without an understanding of the aforementioned theory and should therefore be considered wrong. Given the literature on the use of everyday understanding, I believe this might not be a productive take as instead it could be a source for supporting that understanding with the appropriate theory—indeed a lot of the authenticity and problem-solving literature implicitly leans into the ideas that the processes by which researchers came up with knowledge should be replicated by students and presumably researchers at some point shared these ideas. In addition, these ideas may have not included an understanding of the underpinnings of their hypotheses, as for example, Darwin understandably inappropriately modeling the units of heredity as gemmules rather than genes or alleles.

Unbroken Lineages Have Longer Evolutionary Histories

In this misconception, long branches in trees are perceived to have experienced longer evolutionary histories (Gregory, 2008). This conception is incorrect as every lineage has experienced the same amount of time if the tips of the tree all represent extant (living) organisms. This conception is therefore related to the backbone misconception in that it assumes longer units are important and presumably have experienced "more evolution."

This misconception may stem from two sources. First, its appearance as one long unbroken line simply makes students interpret it as such. A second explanation for this sampling bias discussed previously, lines that may appear as single units may actually have more splitting events than are visible in the tree. Neither of these explanations have been discussed in the literature.

Unbroken Branches Have Ancestral Traits and Have Not Changed

Students exhibiting this misconception typically will assume long branches near the root have ancestral traits; species poor groups near the root are considered more "basal" (Crisp & Cook, 2005; Gregory, 2008; Novick & Catley, 2007). If a trait appears on the lineage just before the split leading to the long branch, a student might assume that the trait appeared and has remained morphologically unchanged in the long branch. While this is sometimes true, as in the case of living fossils – organisms whose morphology has remained unchanged for longer than most other lineages as evidenced by the fossil record – this is often not the case. Lineages may undergo anagenesis—though some argue it is uncommon (Gould & Eldredge, 1977). Alternatively, these lineages may have split but their splits have been obscured by poor sampling—this observation assumes that evolution is punctuated and typically occurs at nodes.

Other Misconceptions or Misinterpretations

Stenlund et al. (2021) examined secondary school students' conceptions after using an interactive Tree of Life display. This display allowed students to zoom in to view different hierarchical groups. They found that before and after zooming, attendees did not detect that time scales visible had changed. Alternatively, they perceived the time during which the zooming animation occurred was proportional to the differences in time observed between splits.

Essentially the majority of said misinterpretations found were unique to the presentation of said tree, i.e. an interactive tree animation.

Matuk (2010) found similar misleading artifacts from animating trees. When given trees that animate in different directions, i.e., from tips to root or the reverse, students will tend to interpret trees as either "converging" or "diverging." Matuk (2010) also offered a unique, more semiotic perspective to our understanding of tree-thinking. She believes that humans' gravitation towards storytelling best explains many of the misconceptions observed in the literature. She argued that the linearity, increasing complexity, and other features common to many of the described conceptions correspond to traditional elements of storytelling such as character and arcs.

Summary

The primary takeaway from this literature is that some of it is preoccupied with the interpretation of the tree itself, such as in the case of the backbone. However, some of the literature may not necessarily be tied to the tree—progress is depicted is also considered a type of misconception in the evolution education literature and is referred to (or related to, depending on your interpretation of the misconception) as teleological interpretation of evolution (Hammann & Nehm, 2020; Phillips et al., 2010; Schramm & Schmiemann, 2019).

Though not discussed in explicit detail in this section, these misconception papers very often lack evidence in the form of student utterances or artifacts—many of them were written at the time by an expert and presumably learned from their own interactions with students or colleagues. While many of them are often true, the lack of rigorous qualitative examination means that it may be more difficult to find consensus in the literature regarding appropriate more or more accurate models of such behavior. Indeed, many of the misconceptions are overlapping

and descriptions of them in many of the papers would employ descriptions also attributable to other misconceptions.

A few of the more promising misconceptions to take note of when reading this dissertation, given the eventual direction this study took, are the similarity equals relatedness, progress is depicted, and backbone is the main line of evolution—these will be discussed in greater detail with reference to my results in Chapter 5.

Instructional Strategies

Tree Analysis versus Tree Construction

A large segment of the literature is also dedicated to the presentation and/or evaluation of instructional strategies, ranging from direct instruction and tree-analysis (Davenport et al., 2015; Novick & Catley, 2016; Novick, Stull, & Catley, 2012; Offner, 2016) to more student-centered approaches often involving group-work (Eddy et al., 2013; Firneno et al., 2023; Kosinski, 2006; Novick & Catley, 2018; Singer et al., 2001). Direct instruction techniques typically involve explicit instruction about trees' components and their conventions as well as the analysis of existing trees (Davenport et al., 2015; Offner, 2016). Student-centered approaches vary, but many of them involve guided problem-based learning in the form of tree construction activities (Eddy et al., 2013; Singer et al., 2001). Many of these activities were designed for undergraduates, but a handful were designed for K-12 students (Crisci et al., 1993).

Tree-construction lessons typically involve giving students a dataset and directing them to construct a tree, typically in groups (Singer et al., 2001). Some of these tree construction tasks include real data from real organisms (Bokor et al., 2014; Firneno et al., 2023; Huang & Whittall, 2018; Kuzoff et al., 2009; Lents et al., 2010; Schneider et al., 2012; Singer et al., 2001) or fake data about real organisms (Crisci et al., 1993), while others include fake data about

fantasy organisms (Crisci et al., 1993; Gendron, 2000). Still others require students to determine their own trait data to classify fake organisms represented by real objects such as gummy bears or machine screws (Crisci et al., 1993; Kosinski, 2006; Kumala, 2010a). A handful of tree construction lessons have students use authentic tools such as GenBank and MEGA to construct trees (Bokor et al., 2014; Kuzoff et al., 2009; Lents et al., 2010). Those lessons are typically found in the context of more molecular evolution-based courses and provide so much trait data that students cannot determine relatedness without the help of algorithms. Some even employ a combination of these strategies (McCullough et al., 2020).

While not explicitly discussed in the literature, the tree construction lessons that involve generating or using a smaller dataset (and thus do not require algorithms, at least initially) may better help students understand the conceptions underlying tree construction compared to those that do not. Those lessons force students to wrestle with individual traits such that they must develop a strategy for tree construction, presumably drawing upon theory regarding "authenticity" in lab design. However, much like some "authentic" labs, the bioinformatics-based tree construction lessons potentially risk being cookbook labs, where instruction is more about the procedure than underlying conceptual understanding; students plug data into a computer program without understanding what is occurring within them (Brownell et al., 2012; Osborne, 2014; Singer et al., 2001). Indeed, several of the lessons present in the literature appear just that (Huang & Whittall, 2018; Kuzoff et al., 2009; Singer et al., 2001) and it is unclear the extent to which relevance of the labs is being emphasized in each of them (Oliver et al., 2002).

There is some argument over the effectiveness of tree-construction on student understanding of evolutionary relatedness and tree interpretation. Most of the literature suggests that tree construction improves student understanding of formal trees (Eddy et al., 2013).

However, there is some evidence that it may not (Young et al., 2013) and some even suggest that tree construction might inhibit tree interpretation (Halverson, 2011). In addition, Baum et al. (2005) interestingly argued that knowing how to construct trees is unnecessary for interpreting them. Note, however, that all these tree-construction activities focus solely on determining evolutionary relatedness using tree diagrams, not other possible diagrams of evolutionary relatedness.

Use of Alternative Metaphors

Of additional note is the usage of metaphors in these lessons to scaffold students' understanding of phylogenetic trees. (This is of particular import given the role of metaphor which I discuss in subsequent chapters.) These metaphors may give insight into the resources instructors can recruit to support students' understanding of evolutionary relatedness and interpretation of trees, and as metaphors (and any form of representation, be they external or internal) are by definition imperfect approximations of the phenomena they represent, they also may unintentionally incur misunderstandings when students hone in on features that the instructor does not intend them to draw parallels with (Catley et al., 2010; Pearce et al., 2023; Talanquer, 2022).

Goldsmith (2003) for example used the metaphor of a race to scaffold students' understanding of trees. He implicitly used body movement to represent time and stamps to represent the accumulation of new traits. Students were asked to imagine a group of students splitting into groups in the woods and at certain points in time obtaining stamps on their cards until all reach the finish line. Each individual thus will have the resulting notepads with stamps and will be asked to reconstruct the racecourse.

Davenport et al. (2015) alternatively used the evolution and organization of languages to scaffold students' understanding of evolution. Their method involved using trees with the same topology and introducing more familiar topics to organize (like languages and families) to transition into instruction about organismal evolution. It is notable that unlike the former metaphor, it more explicitly and extensively ties the tree metaphor to language use which may increase cognitive load.

Boyce (2015) and Davenport et al. (2015) both used human ancestry and/or pedigrees to scaffold students' understanding of evolution. Pedigrees are formal diagrams that superficially resemble trees. They are arguably significantly different than phylogenetic trees in the sense that they depict pairing and resultant offspring. These parings are represented by unions of lines rather than divergences—unions are not often observed in phylogenies except at the microbial level (Kunin, 2005). The utility of pedigrees is of additional note, given their resemblance to phylogenetic trees. On the one hand, they may induce incorrect interpretations of trees based on pedigree conventions, especially as pedigrees are arguably more common in media. On the other hand, they may be useful resources upon which to scaffold, assuming that the history of science reflects the learning trajectory some might undergo. Indeed, several science historians and cultural studies researchers have suggested that genealogical trees (which are not quite the same as pedigrees) were precursors to trees (C. R. Miller & Hartzog, 2020).

I encourage the reader to take note of these metaphors as I discuss them in relation to my findings later in this study as well as my usage of cognitive metaphor. Cognitive metaphor is not quite the same as the use of complex metaphor for explanation, as the former typically assumes more fundamental embodied metaphors for constructing new ideas. Nevertheless, metaphor is of

note as it can be considered the usage of a type of cognitive resource upon which students can scaffold their new knowledge of the normative understanding they are attempting to learn.

Framing Tree-Thinking and Relatedness

As suggested in Chapter 1, there is a dearth of investigations into various aspects of tree-thinking which also include framing studies in particular ways to focus on particular results. I discuss the various frameworks by which tree-thinking, representation, and evolutionary relatedness have been approached. Given the historically vague usages of the term and some differences in the ways that they have been described (Bodner & Orgill, 2007; Bussey et al., 2020; Imenda, 2014; Luft et al., 2022; Rodriguez et al., 2023), I use the term framework here loosely to refer to any major bodies of work that represent sets of assumptions, theoretical perspectives, and their associated methodologies from which the aforementioned topics were approached and which may or may not be explicitly named or identified. Note that these assumptions, perspectives, etc., may overlap as some theoretical perspectives also often co-occur with and inform certain methodological approaches, and these users from frameworks often influence each other during their development.

Tree-Thinking

A sizable portion of the tree-thinking literature does not explicitly invoke a specific framework (Baum et al., 2005; Baum & Offner, 2008; Gregory, 2008; Meir et al., 2007) which has been common in biology education until recently (Bussey et al., 2020). A lot of early work in biology education and science education more broadly consisted of studies that were, like many fields, more practitioner focused (National Research Council, 2012) and largely consisted of broad observations of misconceptions (Baum et al., 2005; Baum & Offner, 2008; Gregory, 2008; Meir et al., 2007) or recommended interventions (Crisci et al., 1993; Gendron, 2000; Goldsmith,

2003) both of which were typically not investigated with the types of rigorous methods and frameworks now often employed in science education.

As the discipline developed, however, some areas of the field implicitly or explicitly began to use familiar frameworks. These frameworks largely consisted of constructivist, representational competence, Gestalt theory, expert-novice framing, and cognitivist (information processing) which I will discuss presently. These perspectives largely align with the major players in tree-thinking, with certain research groups favoring certain subsets of these frameworks.

Frameworks in the Literature. I begin here with a brief description of the frameworks employed as they exist outside of the biology education literature and then describe how they have been employed in the literature. I follow with an analysis of the broad patterns of framework use in the tree-thinking literature.

Constructivism. Constructivism is a large body of work stemming from the work of child psychologist Jean Piaget (Bodner & Orgill, 2007; Glasersfeld, 1995; Piaget, 2001). This work is predominantly predicated on the idea that learners actively generate knowledge based on prior experience. In other words, new knowledge structures are created out of pre-existing knowledge structures. This perspective is opposed to the typical folk ideas of intelligence in that it presupposes that human ideas, thoughts, etc., are not illogical even when they do not match normative "correct" thought; rather, they have internal logic. This work typically is associated with qualitative analysis of clinical interviews where the researchers formulate a hypothesis of the meaning that the learner is making in the form of a schema.

A constructivist perspective has been adopted in a number of studies in the tree-thinking literature, though less than anticipated. As previously mentioned, much of the literature consists

of practitioner focused opinion pieces or interventions that do not substantiate their assertions with evidence like student utterances (Schramm et al., 2019) and perhaps therefore cannot be considered constructivist. Later studies, especially those by Daniel (nee Halverson), use a constructivist framework to investigate tree-thinking deeply and qualitatively (Halverson, 2009; Halverson et al., 2009, 2011).

Information Processing (Cognitivist) Perspectives. Information processing is a learning theory that broadly consists of two main subfields: strong and weak information processing. Strong information processing represents the initial work in the field that is more cognitive in nature and uses the structure and function of computers as a metaphor and model for cognition. Studies utilizing this framework typically generate algorithm-like models of cognition that help explain behavior. This perspective tends to be more procedural and typically focuses on modeling human cognition based on lines of computer code or documenting the often inarticulable (by the user) procedural steps taken during problem solving. This is in contrast to constructivist work that is typically focused more on meaning-making or meaning assigned to concepts rather than procedural knowledge.

The second major body of work in this area is typically focused on higher order cognition such as metacognition and knowledge structures used to understand the problems—this is sometimes referred to as weak information processing. This body of work is preoccupied with the heuristics used in problem-solving as well as the noticing of problem or object features over others. This perspective also is interested in characterizing how people approach problems and manage their cognitive processes and additionally can also refer to the ways in which they organize their knowledge. This latter aspect of IP is often referred to as metacognition.

In the tree-thinking literature, derivatives of early information processing that resulted in the broad cognitivist traditions we see today such as cognitive psychology are apparent in some of the more perceptually focused tree-thinking studies. Some studies use Gestalt models to understand the ways in which students parse trees (Novick & Catley, 2007; Novick & Fuselier, 2019) while others are interested in determining the features of particular tree variants that might aid student understanding (Catley & Novick, 2008; Dees et al., 2017, 2018; Novick et al., 2011).

Gestalt Theory. Gestalt Theory is a body of work from the cognitive psychology tradition, originally stemming from information processing, that essentially describes how humans parse the visual field. Gestalt Theory consists of a series of principles that predominantly describe how humans innately or near innately interpret visual patterns (Wagemans et al., 2012) but is also referred to as describing the "dynamics of cognition" (Mungan, 2023). Gestalt theory is particularly prevalent in the psychology literature but is sometimes employed in other domains such as education as in tree-thinking (Novick & Catley, 2007; Novick & Fuselier, 2019).

Gestalt Theory comes from the more cognitivist information processing traditions including cognitive psychology and is therefore preoccupied with the perception of objects in the human visual field. Often Gestalt theory related studies will involve examining what is essentially innate or near innate (i.e., learned early) tendencies of humans to interpret certain shapes or patterns in predictable ways (Mungan, 2023; Wagemans et al., 2012). Gestalt theory, when used in the education literature, often refers to a series of principles that are hypothesized to guide how humans parse the visual field. For example, unbroken lines tend to be conceived of or depict constant or stable processes—this tendency is specifically referred to as the Gestalt principle of Good Continuation (Novick & Catley, 2007).

Tree-thinking literature that draws upon Gestalt theory often relates to the ways in which trees are perceived. For example, some studies examine how lines are perceived as "units" of a diagram based on their arrangement (Novick & Catley, 2007), and this information is used to determine what forms of trees are better suited for novice learners to interpret (Novick & Catley, 2007; Novick & Fuselier, 2019).

Expert-Novice Framework. The expert-novice framework is closely related to and arose out of the information processing tradition where comparisons between experts and novices were used to understand larger metacognitive behaviors and constructs. In short, the expert-novice comparisons examined the heuristics experts used in doing work related to their expertise, the organization of their knowledge, the strategies they use to deploy such knowledge, and the aspects of the problem framework which they notice and use to determine appropriate strategy deployment. Expert-novice framework has also been implicitly used in the strong information processing tradition, i.e., more procedurally based contexts where the types and usages of certain procedures were compared. This procedural understanding of expert-novice cognition I will not be referring to in this study.

The expert-novice framework is also often borrowed for some constructivist studies where they are used to compare the meaning of concepts between experts and novices (Smith et al., 1994). Expert-novice frameworks have also been broadly used in the representational literature, i.e., the comparison of how experts and novice examine, interpret, and utilize scientifically normative diagrams in chemistry, and other sciences (Clement et al., 2005; Fisher & Keil, 2016; Goodwin, 1995; Hay et al., 2013; Kozma & Russell, 1997). These usages typically implicitly draw upon either constructivist or information processing framework works (or both) when being used.

In tree-thinking this framework has been very rarely employed. I could only find a handful of studies that did so (Halverson et al., 2011; Novick & Catley, 2007; Phillips et al., 2012), and even then they left readers wanting; one in particular used only a single expert for comparison and another examined partial expertise (students with biology experience).

Representational Competence. Representational competence refers to the phenomenon of experts' successful interpretation and use of scientific external representations, i.e., diagrams or symbols. Representational competence generically has been modeled as a linear process involving the acquisition of a variety of representation-related competencies. However, Daniel et al. (2018) and others (Halverson & Friedrichsen, 2013; Schramm et al., 2021a) have argued that representational competence does not necessarily have an agreed upon definition in biology education and implied it may not necessarily be as simple. She argued that representational competence on some level entails representational fluency, which is the ability to translate between diagrams (Daniel et al., 2018; Halverson & Friedrichsen, 2013), the ability to reason within the formal diagram of interest, a shift from external to internal representations (Clement et al., 2005; Gilbert, 2007), as well as other skills such as the usage and analysis of said diagrams. Whether or not these skills constitute a general progression or pattern of learning is unknown, at least in the tree-thinking literature.

Given this definition, the entire field of tree-thinking could therefore be argued as being in some sense concerned with representational competence. However, only new tree-thinking researchers and Daniel have explicitly framed it as such (Halverson et al., 2009; Halverson & Friedrichsen, 2013; Schramm et al., 2019, 2021a). Daniel has codified her own interpretation of representational competence learning goals based on the representational competence literature in other scientific domains (Halverson & Friedrichsen, 2013), and it is based on the normative

behaviors of experts as well as generalized aspects of those other representational competence frameworks (e.g., experts are able to use diagrams to generate more research questions—assumed regardless of the domain of interest). Similarly, these researchers have produced competing instruments intended to measure tree-thinking (Blacquiere et al., 2020; Jenkins et al., 2022; Kong et al., 2022; Kummer et al., 2019; Schramm et al., 2021a). As assessment implies learning goals and therefore representational competence goals, it can be argued that these researchers have thus engaged in the examination and determination of representational competence skills as they should be applied to this field.

How These Frameworks Are Employed in the Literature

Meaning-Making Orientation. Much of the deep qualitative work investigating many of the tree thinking misconceptions the aforementioned scholars have identified was spearheaded by Kristy Daniel (née Halverson) (Bucklin & Daniel, 2017; Daniel et al., 2018; Halverson, 2009, 2011; Halverson et al., 2009, 2011; Jenkins et al., 2022). Her work largely utilized constructivist, representational competence, and expert-novice frameworks. Her work could broadly thus be considered as focusing primarily on meaning-making, and to a lesser extent, the impact of other factors on tree-interpretation.

Daniel has used the constructivist frameworks in the qualitative sense (Halverson, 2009), typical of much qualitative work used broadly in other education subdisciplines. Constructivist axioms such as the influence of prior knowledge on new knowledge being formed were assumed, and the conceptions she hypothesized about were analogous to schemas. Methods such as clinical interviews and Grounded Theory were also used as is common in constructivist studies (Ginsburg, 1997; Mann & MacLeod, 2015; Posner & Gertzog, 1982).

Daniel also occasionally employs an expert-novice framework (Halverson et al., 2011), which is unsurprising given that constructivist frameworks sometimes co-occur with expert-novice frameworks in addition to information processing or cognitivist frameworks. One of her studies examined students' conceptions about phylogenetic trees in the context of the classroom (Halverson et al., 2011) related to her dissertation work which explicitly cites constructivist framing (Halverson, 2009). Expert students and the instructor were classified as experts in comparison to the main group. The implication was that the expert instructor supplied scientifically normative ideas about phylogenetic trees, and she used this information to compare those ideas to novice ideas. Expert-novice frameworks had widely been used in the broader representational literature in great detail but had not been used extensively with respect to tree-thinking. Her study is one of the few to employ such a perspective in the tree-thinking literature.

Daniel, however, most frequently and most explicitly references representational competence frameworks (Halverson et al., 2009; Halverson & Friedrichsen, 2013), wherein she applied aspects of representational competence as it pertains to chemistry and other fields to phylogenetic trees. Representational competence is a framework that is more commonly present in other domains of science education that essentially aims to map out the skills and knowledge necessary to interpret formal representations, usually referring to normative representations found in math and science such as graphs, phylogenetic trees, molecular models, etc. She created a representational competence scale specifically for tree-thinking (Halverson & Friedrichsen, 2013) based on her prior work (Halverson, 2009, 2010, 2011; Halverson et al., 2009, 2011) as well as the representational competence literature. In doing so she essentially codified her recommendation of the learning goals of tree-thinking.

More recently Daniel has begun to incorporate elements of more cognitivist frameworks. She and her collaborators have begun to examine how students engage in pattern recognition using eye tracking equipment and has stated her interests in investigating their relationship to spatial reasoning (Daniel et al., 2018). She also participated in the creation of a validated assessment intended to measure tree-thinking (Jenkins et al., 2022). She has also investigated the impact of some methods on the assessment of representational competence (Halverson et al., 2013). Her work, therefore, has begun to shift to that of more quantitative approaches intent on measuring the meaning she has investigated.

Indeed, this quantitative turn, with a focus on learning goals (i.e., cast as representational competence here), has also been reflected in some of the newer faces in tree-thinking. Tyler Kummer, Yi Kong, and Thilo Schramm (Kong et al., 2022; Kummer et al., 2019; Schramm et al., 2021a) have each recently begun to publish in the field and have similarly published competing validated assessments which together imply a trend towards determining the appropriate targets for learning in tree-thinking. Some of these authors have also begun to use both her work and the early misconception work to investigate the prevalence of some misconceptions (Kummer et al., 2016) as well as reinterpret the literature's findings on misconceptions (Schramm et al., 2019) in addition to offering alternative models of tree-thinking (Schramm et al., 2021a).

Perception and Other Perspectives. The other significant presence in the literature is collaborators Kefyn Cately and Laura Novick. Novick and Cately are likely the most prolific writers in the tree-thinking literature (e.g., Catley et al., 2010, 2013; Novick et al., 2011; Novick & Catley, 2007, 2018; Novick & Fuselier, 2019). Novick has a background in educational psychology and Catley in biology. The pair have explored a broad range of topics which span

several frameworks. While they are very often not explicit about the frameworks they use (insofar as they explicitly do not often include "framework" sections at times), they do implicitly and explicitly draw from recognizable bodies of work, the most obvious of which is Gestalt Theory.

Cately and Novick have suggested it is this tendency that may explain why novices tend to interpret the backbone as a single unit, i.e., "main line" and that subsequent lineages "branch off" this primary linage (Novick & Catley, 2007; Novick, Stull, & Catley, 2012). They and their collaborators have also suggested that other Gestalt principles may impact the way individuals interpret phylogenetic trees. Novick and Catley (2014) also have used frameworks such as Chinn and Brewer's taxonomy to categorize how prior knowledge of taxa can impact their acceptance of scientifically normative taxonomic groups.

With a few exceptions (Catley et al., 2010; Novick & Catley, 2013, 2014), Novick and Catley's work is essentially part of a subset of the literature predominantly preoccupied with determining what types of trees are most optimal for tree-interpretation (Catley & Novick, 2008; Catley et al., 2012, 2013; Dees et al., 2017, 2018; Dees & Momsen, 2016; MacDonald & Wiley, 2012; Novick & Catley, 2007; Novick, Pickering, et al., 2014; Novick, Stull, & Catley, 2012). This method has been fruitful and has led to changes in textbooks as a result of their findings (Schramm et al., 2021b). These studies are typically more quantitative in nature and focus primarily on more cognitivist and perception-based perspectives. Others in this domain have similarly examined the relative utility of different types of diagrams in different populations such as among museum attendees and high school students (Bokor et al., 2014; Catley et al., 2013; MacDonald & Wiley, 2012; Novick, Pickering, et al., 2014; Stenlund et al., 2021).

This perspective, however, does ignore some factors that may impact tree thinking such as the underlying meaning of said trees (which I discuss momentarily) as well as some factors not part of the visual field that may also impact tree-thinking. While Novick and Catley (2007) have noticed that experts are able to flexibly translate between tree types, their perspective in some sense neglects some of the underlying reasoning that might drive this ability to flexibly translate between structures, namely an understanding of the underlying conception that is not so rigidly context bound. Such a perspective, therefore, would not necessarily be overly concerned with one particular optimal tree (though it is very useful in cases of limited time as in a museum or where the content to be understood has nothing to do with the tree itself as in instruction about organismal diversity) but rather developing an understanding that is robust enough to allow for interpretation of said diagram regardless of diagram type.

What About Relatedness?

One crucial aspect of the aforementioned work that appears to be underappreciated is the underlying understanding of tree-thinking. What is essentially tree-thinking but an attempt to make sense of a series of taxa based on a set of rules? In addition, such categorization of taxa results in the possible identification of taxa. Tree-thinking, therefore, is an operationalized model and external representation of how we know what an organism is and what group of organisms they belong to. In short, it is a representation that implicitly relates to how we know what organisms are and how we come to determine what they are and how they are grouped.

The answer to this question, I argue, is our understanding of relatedness. Relatedness here refers to the ways in which we categorize organisms presumably with different processes. Each organism is its own "concept" or unit of knowledge and either unique or generalizable processes have occurred to generate such concepts.

The tree-thinking literature is woefully bereft in that regard with the majority of the literature being dedicated to representational competence as it pertains mostly to the diagrammatic formalisms, and to a lesser extent, the evolutionary processes implied by it. However, Daniel et al. (2018) do suggest to some extent that this is an area of interest—they make a distinction between "model competence" and "representational competence," referring to the former as the sort of conceptual underpinning of the representations in question. Despite this recognition, to my knowledge no one outside of Matuk and Uttal (2020) has yet to investigate this aspect of taxonomic understanding and perhaps its effect (or lack thereof) on tree-thinking. While some have examined specific "prior knowledge" that affects tree thinking (Halverson et al., 2011; Novick & Catley, 2014), no one has framed their study as examining the concepts that are to be represented by trees or any other related diagram.

Biology Education and Relatedness

Relatedness, in a normative sense, refers to evolutionary relatedness. Evolutionary relatedness refers to the ways in which taxonomic groups are related to one another according to their evolutionary history. Taxonomy refers to the ways in which researchers determine taxonomic groups which are essentially identifications of groups, and within the last century or so the norms of taxonomy have changed such that systematists' goals are to reconstruct the evolutionary history of organisms and change taxonomic classification to reflect such relationships.

The literature reveals a dearth of work related to organismal relatedness understanding. While a few authors have highlighted the importance of relatedness understanding, only a handful of papers investigate the nature and strategies employed for grouping organisms as well as how those behaviors might translate to tree-interpretation or tree-building. Matuk and Uttal

(2020) investigated the grouping behaviors of non-majors as well as their representative diagrams and discussed the possible implications of those behaviors on tree-thinking. Novick and Catley (2014) similarly discussed specific taxonomic group contexts (i.e., identity of taxonomic groups) on the interpretation of phylogenies. They explicitly examined the understanding of reptiles and how that impacted the interpretation of trees with misconceptions about taxonomic identity indeed impacting tree interpretation. Relatedness therefore has been largely uninvestigated in the tree-thinking literature.

Organismal Classification Outside of Biology Education

Despite this dearth of data on organismal understanding in biology education, other fields have dedicated some energy towards their investigation. A subset of the child developmental psychology literature is interested in characterizing the nature of children's biological knowledge as well as how they come to be (Gelman & Coley, 1990; Rhodes & Gelman, 2009; Tarłowski, 2019; M. G. Taylor et al., 2009; Tsou, 2022). This knowledge is often in the form of grouping and identification of organisms as well as the properties they assign to such organisms.

Natural kinds are often described in this literature and refer to the types of generic categories people might make of quite literally any potentially "natural" thing, including taxonomic groups, which are assumed to be representative of reality. For example, someone might argue that hydrogen is a "thing" while another might argue that specific morphs of *Enstatina eschscholtzii*, a type of Californian salamander, are each distinct species. For chemists, some of these natural kinds have clear demarcations: what constitutes a specific element is clearly defined and bound and has very clearly different properties than other elements. In that sense it could be argued that hydrogen is a "thing" that is demarcated by the natural world. The aforementioned salamander, though, is much more difficult. We might see that morph as visually

different and may suggest it is its own "thing," i.e., species. However, we also see that it is reproductively isolated from one group and not from another which is itself not isolated from other *E. eschscholtzii* morphs. Is that morph therefore a "thing" constituted and determined by the environment? It is arguable then that in these cases, these types of natural kinds are fuzzy or do not exist at all.

It follows then that there are different formulations of said natural kinds. Some natural kinds are attached to an essentialist belief, i.e., that certain kinds have some sort of inherent essence that determines said properties. Other disciplines have other philosophies. Biology might therefore assume that "things" like these things exist and operate independently in the environment. For example, someone who is an essentialist might suggest that trees have leaves "because they are trees." They suggest the kind of a "tree" has some sort of inherent essence that bestows on it the properties. In other words, the property of having leaves, therefore, is an outcome of its essence.

The natural kind literature is varied and includes many modifications to different philosophical interpretations of the term (both in the lay sense and philosophy of science sense) but it also refers to the types and kinds of organismal categories children create as well as the sources of these categories and the processes by which these categories come to exist (Gelman & Markman, 1987; Tsou, 2022). The latter portion of the field is tied to more strongly cognitive traditions and therefore is typically aligned with perspectives such as information processing and embodied cognition.

There have also been a few linguistics studies investigating word classes which include some animals (Shapiro & Palermo, 1970, as cited in Catley et al., 2010; Mervis et al., 1976). However, these are not all groups are of interest to undergraduate biology courses and referred to

exceptionally large taxonomic groups (i.e., animals and plants rather than something like birds versus reptiles) and their findings were not examined with respect to tree-thinking.

What Do We Need to Know About Relatedness?

This dearth of investigation in tree-thinking education regarding organismal relatedness requires closer examination of relatedness both in the normative and non-normative sense. As such we should likely use literature from these other domains which have begun to look at them but have not necessarily connected them to undergraduate and expert understandings of tree-thinking as most of them are predominantly focused on young children.

As with other representations discussed in the field, it may be that this "prior knowledge" in the form of models of taxonomic identity may impact tree-thinking. While this study does not necessarily investigate the impact of specific taxonomic groups on tree-thinking, it seeks to at least characterize the ways in which this relatedness understanding might appear across experience levels. In other words, our everyday understanding of organisms, how we identify and determine what organisms are what, may impact our interpretation of the trees in addition to the features of the tree which may or may not be inherently distracting.

This Study

What Work is Needed?

In accordance with the aforementioned literature, I therefore proposed a study that would address some of these gaps. As the most pressing issue was that of the underlying understanding of trees, I conducted a study that sought to probe at everyday understandings of relatedness in hopes that such understandings could be compared to tree-thinking behaviors and investigated as potential supports for normative tree-thinking. In other words, rather than focusing on the

interpretation of trees, I sought to determine how people understand what an organism is and how they determine what organisms belonged to what categories.

In an attempt to provide the tree-thinking literature with more robust evidence for any (mis)conceptions, I decided that this study should be qualitative involving clinical interviews. These interviews would more deeply get at the meaning students gave to these categories as well as the strategies employed to generate these categories. In addition, as the literature is so sparse—outside the context of early childhood psychology and linguistic studies be (Gelman & Coley, 1990; Rhodes & Gelman, 2009; Tarłowski, 2019; M. G. Taylor et al., 2009; Tsou, 2022)—regarding taxonomic understanding, qualitative methods were better suited for what is essentially pointed formal observation and theory generation (Kuczynski & Daly, 2003).

I additionally sought to address the dearth of expert-novice evidence in the literature. To my knowledge, only a handful of studies involve the examination of expertise (Halverson et al., 2011; Novick & Catley, 2007; Phillips et al., 2012). Those that do only examine a small subset of expert behavior (translation, for example) or employ an exceedingly small comparison group. While the aforementioned misconception literature was, quite obviously, created by experts, and their viewpoints therefore could be considered as a type of "expertise" discussed in the literature, the field is likely better aided by having outsiders with different perspectives (education researchers, for example) to more robustly examine and document such behavior as experts are not immune to blind spots either content-wise (Fisher & Keil, 2016) or pedagogically given that it requires a different skillset (K. K. H. Chan & Hume, 2019; Cochran et al., 1993). Much of the other representational competence literature in other domains such as chemistry (Kozma et al., 2000; Kozma & Russell, 2005) have provided more of this evidence which can implicitly function as both a tool for determining precise areas where novices might require support and

identifying heretofore undescribed components of said understandings which might be leveraged for interventions.

I additionally wish for this study to help the field orient itself toward other types of interventions that involve the underlying conceptions of representations, which could be referred to as model conceptions (Daniel et al., 2018), and are helpful in other areas of science education. Those interventions prioritize discussion, argumentation, and communal creation of representations of the phenomena they are investigating and thereby presumably better scaffold students' everyday understanding toward normative understanding (Windschitl et al., 2018). This position is a more flexible in that it supports those alternative representations which may themselves be valid methods of sensemaking. I argue that the systematists might have been so reliant on trees—as evidenced by the pervasiveness of tree-based jargon—that we may have essentially forgotten about alternative learner generated diagrams (Ainsworth & Scheiter, 2021; Matuk & Uttal, 2020; Van Meter & Garner, 2005) that may be useful teaching aids. While the literature on the "best tree" is useful for cases where individuals need to parse the diagrams quickly (e.g., museums) or students require learning about the information contained within the tree only (e.g., in a textbook section on snakes), students might also be better served by being able to engage with the underlying understanding of said diagrams such that, presumably, other expert behaviors such as translation might be better supported (in addition to simply having practice equating one feature to another among different diagrams).

Research Questions

As a result of these desires, I therefore generated two primary research questions as components to my larger desire to investigate organismal relatedness, the concept presumably

underlying any phylogenetic tree or any other related equivalent diagram that attempts to communicate the proposed relationships between taxa.

1. Understanding Relatedness -- In what ways do experts and novices conceive of organismal relatedness?

My first question, therefore, was intended to get at the ways in which organismal relatedness might be constituted or conceived. This question essentially gets at the shape of the mental structures as well as some of the spatial metaphors that I expected individuals to employ which presumably would be related to external representations. This question was intended to help determine the ways in which individuals might understand relatedness and how that might differ with experience. Given my reservations on some of the aforementioned misconceptions in tree-thinking, I intended to determine if these understandings were consistent across both experience groups, or not.

2. Investigating Ways In Which Taxa are Grouped -- How do experts and novices group (i.e., classify) organisms?

My second question was intended to interrogate the ways in which individuals would group organisms. I assumed that this question would get at the methods or strategies employed to create such taxonomic groups and be used to identify taxa. I also desired to look at the ways in which they may or may not vary across experience; again, given some of my misgivings about the ways in which some of the misconceptions have been formulated in the literature, I expected that there may be some similarities across experience groups.

Frameworks Employed in this Study

Embodied Cognition: Metaphor and Classification

The first primary framework upon which I draw is embodied cognition. Embodied cognition is a broad learning theory that comes out of cognitivist and information processing traditions, especially cognitive linguistics. Embodied cognition broadly posits that cognition is composed of borrowed mental constructs which typically represent basal understandings of the world. Embodied cognition is unique in that it recognizes the role of the body's placement in the world as a source of understanding the world; the space around a person's body, for example, might be used to understand the concept of time (Núñez et al., 2012). Stronger variants of the theory even suggest that the body and mind are so tightly intertwined that some posit that "simulations" of said activities, etc., being borrowed are "ran" in the mind (Fodor, as cited in Adams, 2010). Linguists were originally most involved in this area given their cognizance of the pervasive use of metaphor in everyday language (see below). Later, however, other cognitive scientists began to investigate these ideas using less conventionalized measures of thought, i.e., spontaneous co-speech gesture. I only know of one study in the tree-thinking literature that uses such a framework (Matuk & Uttal, 2012).

Cognitive Metaphor (Gesture)

Cognitive metaphor is a component of the embodied framework that describes the borrowing of mental structures to create new mental structures. It is similar to constructivism in that sense but it more explicitly suggests that aspects of the new knowledge to be created are mapped on to features of the old knowledge that is being borrowed. Individuals, for example, might map their understanding of time egocentrically with the future implied to be in front of

them, the past behind them, and the present as their body. Typically, these metaphors are often more primitive in nature and represent fundamental ways that our bodies interact with the world.

These were observed initially by the occurrence of pervasive metaphors in every aspect of our language. Affable people, for example, are often considered "warm" perhaps suggesting that people may be mapping friendliness onto their understanding of temperature. However, language is highly conventionalized; most people who speak my language would understand what I meant if I said a person was "warm." That, however, could suggest that my usage of the term "warm" came from my learning of my language's conventions. In other words, perhaps the first person who used the term "warm" drew parallels between friendliness and temperature but subsequent usages do not guarantee that. Someone may have simply copied them, but that may not necessarily represent the mental structures they are calling upon.

Therefore, the field investigated spontaneous co-speech gesture which are essentially unplanned, mostly unconscious gestures that occur during everyday speech (this does not include highly conventionalized gestures such as the "thumbs up"). The implication is that these gestures give insight into understanding, especially as it is assumed that the body and mind are closely connected.

Categorization: Classical Theory, Exemplar Theory, and Prototype Theory

One important set of frameworks upon which I will draw for a sizable portion of this dissertation is the hypothesized theories that describe what categories consist of. Do note that as I describe these that the literature itself sometimes conflates, and also criticizes the conflation of, categories that exist in the real world versus categories that exist in the mind and some combination of the two. I argue for the sake of this study that the real world contains phenomena we can approximate with internal conceptions and focus primarily on the mind.

Classical Theory

Classical theory is one of the earlier hypotheses of category formulation that is typically associated with Aristotle (Lakoff, 1987; Studtmann, 2023) and is not part of the embodied cognition tradition; however, I discuss it here as it is often compared to embodied hypotheses. Classical theory suggests that a category comprises of a group of things that all share common traits. The implication of classical theory is that these set of traits must be harbored by all individuals that comprise said category. So, an example might be that the category ***TREE*** contains individual things that trees all have, such as leaves, trunks, etc. Aristotle (Lakoff, 1987; Studtmann, 2023) and others do not indicate that these categories are necessarily cognitive or representations of the external world that might not constitute reality, rather that they likely are reality. Aristotelian understandings of categories have been described as describing artifacts and natural kinds, where natural kinds are assumed to be ontologically real categories that exist in the world. This theory of categorization has been dominant, at least in Western scholarship, until very recently.

One major property of classical theory's categories is its uniformity and lack of centrality (Lakoff, 1987). Given that all members of a set or category must have all required properties or characteristics, it suggests that potential members are evaluated on an all-or-nothing basis. Either they are part of the category, or they are not. That assumes that no gradations of membership occur (as opposed to being on a gradient which I will discuss in the prototype theory section).

Prototype Theory

However, in the 1960s and 70s during the explosion of work in the cognitive and other sciences, linguists like Lakoff (1987, 2012) have suggested a different, more cognitivist framework, for the ways categories are composed. Hampton (1979, 1981) first described it in the

literature as polymorphous theory while later Rosch (1973) and Lakoff (1987) and others (Wittgenstein, 1986) independently characterized and subsequently popularized it as *prototype theory*. Rather than individuals of a group all harboring a set of traits, they suggest that categories consist of members that harbor a subset of possible traits. They imply that cognitively we have some idea of a "prototype" or "typical" hypothetical organism that has a set of traits that have differing probabilities of appearing in practice. Wittgenstein (1986) observed that most members of a category do not neatly fit into a category and that categories lack "clear boundaries," yet we still describe them as being a member of said category. To use the previously used example, the category ***TREE*** might include green leaves, leaves, presence of a trunk, presence of bark, etc. Some individuals like palm trees do not have bark, yet we still characterize them as trees. Alternatively, deciduous trees lose their leaves for part of the year, yet we also consider them trees.

The idea of traits, however, is complicated. Lakoff (1987) and others (Wittgenstein, 1986) are quick to specify that they do not need to be traits or component parts of the object or phenomenon in question. They can simply relate to, and therefore traits can also refer to relationships between concepts (of objects, activities, etc.). They refer to the class of "cricket" as potentially containing anything relating to the game of "cricket." In that sense, those objects like a wicket and ball might be related to "cricket" but not necessarily a property of "cricket." Similarly, an evented category of "things to take on a camping trip" may not include items that have comparable properties. I do suggest, however, that this take might be semantic in nature given one could argue that a property of said items are their higher likelihood to be taken on a camping trip or to be used in cricket. It may be that these authors make a distinction between contextual activity and internal properties inherent to said object, activity, etc. (For the sake of

this study, I make no distinction between the two and do not investigate the potential impact of the distinction between the two.)

Categories in prototype theory have a number of qualities associated with them such as centrality and gradedness. Centrality refers to the quality of centeredness; one could analogize this quality to a normal distribution in two or three dimensions. Members of a category closer to the center are thought to be better examples of said category; someone might suggest that a pine tree is more of a ***TREE*** than a cycad which has unusual fuzzy orange cones, is short, and somewhat palm-tree like. Centrality suggests that members exist across a gradient and therefore the categories exhibit gradedness. Individuals can rank members based on their conception of what is more or less a better example of said category.

Another property of said categories described by Wittgenstein (1986) was that boundaries were unclear (given the probabilistic nature of criteria presence in members) as well as unstable (J. R. Taylor, 2015). The fuzzy boundaries that are present do not seem to be fixed over time—which makes sense given the changeable nature of language and concepts culturally—but also within an individual. Individuals can encounter new potential members of a category and modify the category accordingly. Perhaps an individual encounters a new object X that has A, B, C properties consistent with Class Alpha, but also has D properties inconsistent with Alpha. However, the individual notices that two other members of said group also have D property and amends their understanding of the concept to include A, B, C, and D properties.

Some have criticized this formulation of categorization in that it presupposes that each possible characteristic or trait involved in membership to a category is equally weighted, implying that having a certain number of said traits constitutes membership and therefore there is some cut-off for non-inclusion based on that number (Lakoff, 1987). This formulation does

indeed have issues as we can imagine a case where two potential members of a category might have the same number of traits, but we might suggest that one is more representative or more likely to be a member because those traits vary in importance to the characterization of said category. Therefore, modifications to this theory have suggested that each of these traits have some sort of weight associated with them and that it is the combination of the number and importance of said traits that determines whether someone determines whether that thing being observed is a member of said category.

One additional wrinkle in the concept of similarity in determining a category is that the use of cognitive metaphors can draw similarities between two concepts or classify two concepts even though they might appear particularly distinct. Cognitive metaphors often employ bodily metaphors projected onto new concepts. In the case of a mountain, the base might be referred to as the "foot of a mountain" where the bottom of a list might be referred to as the "foot of the list." A list and mountain may at first seem like there are not that many similarities but the use of the same metaphor to understand the two might cause individuals to group these objects together.

Exemplar Theory

Yet another theory is offered by the literature: the exemplar theory. Exemplar theory differs from prototype theory in that it suggests that an individual member of said category serves as a point of comparison for other potential members of a category. This is in contrast to prototype theory which posits that individuals harbor a hypothetical, likely non-existent, prototype which in colloquial terms could be considered an "average" of the members which is used for comparison with the set of weighted traits previously described. This particular formulation of categorization is often contrasted to prototype theory with suggestions that they

may occur in different circumstances or that they may simply model slightly different aspects of categorization processes (Gradoville, 2023; Murphy, 2016).

Prototype Theory and Exemplar Theory as a Continuum

While I have not yet observed this in the literature, whether due to my inadequate perusal of the literature or a lack of discussion about the topic, the latter two theories could potentially be considered somewhat overlapping. As you shall observe in the dataset I discuss later, prototype formulation of categories might masquerade as exemplar in the case where the individual has little experience with the group but knows that the individual is part of a distinct group. In the dataset I discuss later, one individual, for example, would compare certain organisms with a deer with which they are presumably familiar. It might be that they know that deer are distinct from other groups but are only experienced with deer, so the set of characteristics they associate with this hypothesized deer-related group only relate to that specific member. In this case, the "exemplar" would be the deer in question but in terms of prototype theory, the ***DEER*** category only consists of a hypothetical deer that closely resembles or is the same as the deer in question. For the sake of this study, I do not strongly differentiate between them. I simply take from them the commonality that they both reject the major assumptions of classical theory and suggest that not all criteria may be satisfied when categorizing a new object.

Criticisms of the Dominance of Prototype, Classical, and Exemplar Theories

While these theories in sum appear to be popular descriptions of how we categorize things, some researchers are critical of all of these formulations. One researcher in particular, Eric Margolis (1994), suggested that these theories are inadequately framed, especially given our philosophical goals in determining how the mind generically functions.

The Correspondence between "True" Categories and Perceived Categories.

Margolis (1994) first argued that there is conflation between ontological truth and cognitive categories; in other words, these theories either take no stance or weak stances on the separation between what our minds see versus what actually exists and whether or not there is a correspondence between the two. While I agree this is the case for classical theory as classical theory presupposes that the categories we observed are real natural phenomena rather than a representation of said phenomena and implies any differences between individuals is due to lack of experience or incorrectness, this may not be as true of fuzzier categorization perspectives. Some philosophers have clarified that natural kinds ("true" categories exist), especially in biology, but rather that these categories are simply collections of outcomes that arise from functional or stable natural processes rather than the objects themselves. In other words, real categories are actually functional groups – the unit of a bird species, for example, is functional in that as a whole it might be impacted by the effects of individual-oriented selective processes and that it functions as a unit because individuals within that group are able to mate with each other. That unit also occupies a niche and has a particular function and therefore can be considered a real "thing" that exists (note function here does not necessarily mean utility in service of the environment, rather that it is a unit that is operated on by other processes and may operate on other functional units in the environment).

Mental Process versus Concept. Margolis (1994) also suggested that prototype theory is not necessarily better than classical theory (and by implication, exemplar theory) because of the way both theories are attempting to describe the process of categorization. He suggested that the focus on criteria does not actually hypothesize about innate mental processes involved in categorization, rather that these theories themselves actually are themselves pieces of knowledge.

In other words, individuals can "know" that a member of a category must have all (classical) or some (prototype) of the criteria relevant to the category. He argued that the actual generalizable process or mental construct that cognitive scientists can focus on is actually the same in both, i.e., that knowing about how to categorize things is required for both and that knowledge isn't actually a valid generalization. He has cited that people are typically unaware of the criteria used to categorize and likely search for criteria in the moment to answer any questions regarding how individuals categorize things.

This criticism could pose a problem in my and similar studies, but this particular criticism is not cause for concern given cognitivist models of cognition account for such awareness in the form of procedural knowledge. It may just be that the criteria required for categorization may not be describable by individuals as in physical actions such as riding a bike which require procedural knowledge of limb movement as well as even procedural knowledge of how to solve a math problem. Though expertise is marked by enhanced metacognitive awareness of their pertinent cognitive processes, experts are not always fully aware of the heuristic shortcuts or biases they employ in certain situations as they are susceptible to some of the same biases as students (Fisher & Keil, 2016; Kelemen et al., 2013).

I additionally suggest that this criticism is likely unwarranted given that it is a model of cognition. While true that frameworks and schools of thought should narrow the ways in which we describe phenomena in science, it is also true that words can change over time, either in a community, in general, or even within a particular interaction. It is therefore important for individuals to thoroughly describe their understanding of some particular core construct which may or may not coincide with Margolis' formulation of how cognition regarding categorization occurs.

Gradedness versus All-Or-Nothing Membership. Margolis (1994) also criticized prototype theory for suggesting that all-or-nothing membership does not exist and cannot be described by prototype theory. He argued that given his formulation of the (lack of) difference between classical and prototype theory that both can be true in certain circumstances. Margolis (1994), however, may not have been familiar with Lakoff's book *Women, Fire, and Dangerous Things* where he explicitly mentions some culturally operationalized categories such as "U.S. Senator" which are unarguably well-defined—an individual is either a senator or not. Though Lakoff implied that prototype theory might describe a dominant process for category creation and may not be the be-all-end-all. In addition, though not explicitly stated by Lakoff, it is arguable that classical categorization could be considered a variant of prototype theory. Where prototype theory hypothesizes that members of a category have certain probabilities of having particular traits, these "probabilities" could reasonably set at 1, which would thereby result in what is essentially classical theory of categorization.

This theme of gradedness and probabilities of trait evolution will be one of the major foci of this study (Chapter 5). As you shall soon read, the usage of criteria and the relative appearance of said criteria seem to be a potential area of confusion from both a normative sense and an everyday sense.

What Does All of This Mean?

The primary takeaway from all of this literature is that essentially there are several ways in which categorization has been hypothesized to occur: classical, prototype, etc. There is some contention on whether these refer to innate mental processes—Margolis suggested it actually represents knowledge rather than a procedure or mental process given the shape of said assumptions and the fact that we use both of them. Regardless of these formulations, I referenced

this literature when analyzing my data given that they are dominant models of categorization. In addition, given the spatial qualities of trees (at least on the external representation level) I thought conceptions of relatedness therefore might have spatial qualities. The examination of those elements presumes an embodied perspective so I have therefore discussed it here.

I am also inclined to lean toward Margolis' argument that they simply represent knowledge structures that are deployed as strategies for categorization, though that ultimately would not make a difference in the final analysis. I also note that while some aspects of the literature seem to conflate their understanding of these mental structures with reality, I do very clearly distinguish between mental representations and the real natural phenomena that we are attempting to represent.

Expert-Novice Comparison

Finally, I employ a broad expert-novice framework. This framework has been very widely applied to a variety of areas (e.g., representation, conceptual understanding) and domains (e.g., chemistry, math), and paired with varying theories (e.g., constructivism and information processing).

This framework is generally more methods focused in that it mostly encompasses the comparison of the two groups. Implicit in this framework is the obvious notion that experts likely know more about the subject to be learned than novices (National Research Council, 2000, 2012). Comparison of these two groups is intended to elucidate the areas in which students struggle to achieve expertise. Alternatively, this comparison can also highlight aspects of expertise, either process wise (i.e., the use of certain heuristics) or perhaps in terms of knowledge structures employed (specific knowledge used) that may experts themselves may not necessarily be cognizant of as requiring instruction; indeed, pedagogical knowledge is considered itself a

domain separate from that of the one the expert is attempting to communicate (K. K. H. Chan & Hume, 2019; Cochran et al., 1993). This perspective can also be considered a type of formal analysis of what a field or topic's learning goals should be. If experts are engaged in certain activities and we expect students to achieve expertise, it follows that some of those activities we expect students to perform. It may have been, however, that prior characterization of said learning goals did not adequately account for such skills. Lastly, this framework can be thought of as a first step in understanding the progression of learning that might occur in that it shows the start and endpoints of said learning process. As such, a comparison could, for example, elucidate the mental resources students come in with as candidates for scaffolding toward expertise, especially those to which experts share similar variants.

In this study, I essentially employ this framework using all of the aforementioned assumptions. I also use it as the chemistry representation literature often uses this framework (Kohl & Finkelstein, 2008; Kozma & Russell, 1997; Pande & Chandrasekharan, 2021). As the chemistry representation literature appears a bit more robust than the biology representation literature given the importance of representation to chemistry and biology does not often utilize this framework, I chose to employ this framework, in addition to the aforementioned reason that very few pointed examinations of expertise exist in the literature at all.

Representations

Another set of literature that I briefly engage in is that of representations in science education. There is a particularly broad set of work, especially in the sphere of chemistry education, which is primarily interested in documenting the ways in which representations coincide or are a part of cognition as well as the ways in which we can support the learning and use of formal external representations (scientific diagrams: graphs, figures, etc.). However, the

extent to which I use this literature is limited given that I discuss the content outside the context of these trees. Nevertheless, I also examine participants' drawings of their understanding which partially draws upon the representational literature. Additionally, as the intent of this study is to ultimately support the interpretation of formal phylogenetic tree interpretation, it is pertinent to discuss some aspects of this literature.

What Are Representations?

Representations typically refer to external representations that exist outside of the body such as diagrams or models (Kozma, 2020; Pande & Chandrasekharan, 2021). These diagrams can include things such as cartesian coordinate graphs, path diagrams, tree diagrams and molecular structure diagrams. These diagrams essentially function to represent certain phenomena, whether they are real things such as natural phenomena or fictional or representational things as in illustrations. These diagrams may include inscriptions, which essentially are any type of orthographic representation (language). These inscriptions can come in the form of notes or captions on the external representations. These inscriptions may also use more abstract orthography as in mathematics where letters are abstract stand-ins for a phenomenon or variable. These external diagrams, however, are not limited to paper or screen-based two-dimensional representations. Statues, museum displays, manipulative models and the like can also be considered external representations.

Representations, however, have also been described in the literature as being internal; mental constructs, therefore, are a type of representation (Harrison & Treagust, 2000; Rapp & Kurby, 2008). Such an understanding of representations aligns with the literature framing cognition as a symbol manipulation system where symbols are implied to be either external or internal (National Research Council, 2000; Sherin, 2000). Representations in this formulation are

therefore imperfect models of a phenomena or concept; any mental approximation of said phenomena or concept therefore could be considered a type of representation. Considering internal mental states as a type of representation is thus the widest definition of representation, i.e., that representation can include any sort of symbol whether they are external or internal.

In this study and throughout the following review, I refer to representation as referring to external representation but not the aforementioned internal representation unless explicitly stated as such. While external representations, as I will later discuss, can constitute thought given its role as a mediation tool, I will characterize this role as being distinct from internal representations. While I acknowledge that they play a role in cognition, i.e., in that they can feedback, trigger, and sustain thought (De Lenos Meira, 2002), I do not necessarily consider them as constituting or itself being cognition. Alternatively, my framework of representation usage and existence could be thought of as an individual having knowledge of or being capable of representing a concept in a variety of ways, be they external or internal and that the process of using, modifying, or creating said diagrams can in turn impact internal representations (conceptions).

I additionally make the distinction between generated representations and conventionalized representations. Generated representations refer to representations that are less conventionalized and made by an individual or group. While generated representations most obviously can and are used as tools for thinking (Kozma, 2020; Kozma et al., 2000), conventionalized premade diagrams can also be used as such, even though they might assign meaning to and notice features that are agreed upon by and salient to the expert community (National Research Council, 2000). In this study, I primarily focus on generated representations and recognize that these may or may not resemble conventionalized diagrams and that part of

this study involves the comparison of the two to make statements on the utility of these representations for scaffolding toward normative tree-thinking.

What Are The Functions of Representations?

Representations typically are considered as having two major related functions, i.e., that they are used to either represent the individual's understanding of a concept (for the purposes of thinking) or communicate it (Kozma, 2020; Kozma et al., 2000). Indeed, the representational literature characterizing expert representational practices has done just that. Experts sometimes want to communicate with peers, for example. Scientists may use graphs or diagrams to communicate to other scientists their results of a study in a journal article and they may use these representations to, for example, summarize the current consensus on the topic at hand. They may also generate their own representations in order to process a topic either alone or with a colleague. They might also use diagrams to communicate with individuals with less expertise such as students or lay audiences. Museums represent the most extreme case where they display associated external diagrams that function to communicate complex topics quickly to a lay audience.

However, others in the literature suggest there is more nuance to these two usages of representations. De Lenos Meira (2002) and Latour (1990) have suggested that the older literature is primarily concerned with mapping representations onto internal representations of said phenomena and that it ignores the external, but cognitive, usages as a manipulatable cognitive tool as with other mediation tools such as manipulatives or writing systems. He argued not only for mediation tools' typical function, i.e., to offload some cognition to the environment, but also that its manipulation can also trigger and support new insights. Latour (1990) also suggested that there are additional features typically unaccounted for in the literature that relate

to their function. Latour suggested that external representations are unique in their size: they can be changed without changing their contents, they are often transportable through space (e.g., paper diagram or model), they are in some sense stable in that once created (at least officially) they stay that way when distributed, they can be combined with other diagrams or inscriptions and made sense of, and that they can represent and simulate manipulation in other dimensions. De Lenos Miera (2002), however, pushes back on the immutability and suggested that its role as a mediation tool suggests that meaning of representations change over the course of their usage, not only in a larger sense (community refining a normative model over long periods of time), but in the smaller sense where individuals read, mark, and create diagrams.

Chapter Summary

In this chapter, I have discussed what the tree-thinking literature has currently achieved. The tree-thinking education literature has primarily investigated the perceptual understanding of phylogenetic trees, as well as the misconceptions commonly associated with it. Much of the literature also continues a series of interventions which seek to support undergraduate tree-thinking.

Despite these advances in our understanding of tree-thinking and how to best support student tree interpretation and tree building, several areas still require investigation. While the literature has a lot of information on misconceptions, a lot of the misconception literature does not substantiate its assertions with data in the form of student utterances with the exception of a few researchers. In addition, no systematic investigation of expertise exists outside of two studies that examine a few narrow aspects of expertise. As a result, we may not necessarily have a good understanding of the goals of systematics education, and we do not have a baseline comparison for determining what types of student understandings can be leveraged for tree-thinking.

Additionally, given the lack of qualitative theory building investigations into these misconceptions, the field as a whole has not appropriately differentiated between these misconceptions or investigated the reasons for such misconceptions let alone come to an agreement on said differences. In addition, representations are linked to underlying models of understanding which have not yet been investigated in this area; this underlying understanding we could consider as an understanding of organismal relatedness, i.e., the ways in which individuals identify and group taxa.

This study therefore is intended to address these gaps in the fields of understanding of relatedness and how they might be related to and presumably support tree-thinking. This study essentially investigates the structure and ways in which experts and novices understand organismal relatedness. It therefore uses several frameworks, i.e., the expert novice framework for comparing different experiential groups and embodied cognition for its coverage of bodily and spatial metaphors as well as its theories on categorization. These frameworks have rarely been employed in this particular field with most studies; if they provide any framework, they are cognitivist frameworks and occasionally constructivist frameworks.

This study therefore is intended to investigate two research questions.

1. In what ways do experts and novices conceive of organismal relatedness?
2. How do experts and novices group (i.e., classify) organisms?

Chapter 2 contains unpublished material coauthored with Komperda, Regis and Lo, Stanley. The dissertation author was the primary researcher and author of this chapter.

CHAPTER 3 – METHODS

Summary

In the following paragraphs, I discuss my methods in great detail. I begin by briefly discussing my positionality. I then broadly describe the procedure for the study. I follow with details on other aspects of the study such as the sampling strategy, task structure, and sample composition. I then summarize my analytical procedures followed by a detailed discussion of the methods chosen and any operationalized definitions that were lacking in the literature necessary for this study. I then discuss the frameworks employed during both the construction and analysis.

Positionality

Given that I later discuss the understandings of nature and taxonomic relationships in diverse cultural contexts, it is prudent that I discuss my background in the form of a positionality statement (Collins & Stockton, 2018; Creswell, 2013; Yilmaz, 2013). I identify as a biracial cis woman with one White parent and one Vietnamese parent. My history is complex but, in short, I have experienced and lived in both extremely privileged and extremely unprivileged situations, with the majority of my life in the latter situation. One parent was also college-educated while the other was not. My Vietnamese parent is a first-generation immigrant to the United States. Both parents were bilingual though I was raised monolingual. I have experience with other languages (French, Japanese, and Vietnamese) though my knowledge is insubstantial; I am not fluent in any other language. I have, however, taken a significant amount of linguistics coursework (I intended to minor), so I have some limited knowledge of the social issues surrounding language research as well as broad knowledge about the field. I raise these issues given the sensitive nature of discussing other cultural and linguistic experiences to which I may or may not identify.

I also would like to mention my religious background. I am not religious and identify as an atheist. This point is relevant to this dissertation given the sensitivity of the religious discourse in the United States regarding evolution (J. D. Miller et al., 2022). Despite biology being thought of as more diverse than other STEM fields, evolutionary biology, in particular, is not diverse (O'Brien et al., 2020). There is evidence indicating that the lack of religiosity in the field might be contributing to the lack of diversity given that some minoritized students generally identify as more religious (Barnes et al., 2020). Additionally, some have cited concerns about the behavior of instructors as being potentially exclusive (Barnes & Brownell, 2017). Given this sensitivity and the fact that I will be discussing understandings of relatedness that may relate to some of this contentious discourse, I highlight my background here.

To provide additional context to as my decision to study this subject, I provide my professional background, especially given the perspective that researchers interact with and affect, both negatively and positively, data collection and analysis (Collins & Stockton, 2018). I hold a Masters in Biology and have previously conducted work in systematics and comparative morphology (Santini et al., 2013; Wassenbergh et al., 2015). During this time, I was a teaching assistant for seven years for a variety of upper and lower division biology courses. These experiences sparked my interest in teaching and tree-thinking. I therefore drew upon these experiences when devising and conducting this study.

Frameworks Employed by this Study

I begin with a brief discussion of the various frameworks, i.e., major bodies of work, upon which I drew to construct the study as well as conduct the subsequent analyses. I discuss these here to explicitly connect the components of this study to the frameworks mentioned in Chapter 2.

Expert-Novice Framework

The expert-novice framework is common in the science education representational literature and is typically associated with the information processing and cognitivist learning theories as well as some constructivist studies (National Research Council, 2000, 2012). As previously discussed in Chapter 2, this framework involves the comparison of experts and novices for understanding what aspects of their cognition differ between the two within a particular domain. Such a comparison can make relevant the learning goals for a particular domain as well as elucidate potential avenues by which learning might occur. Experts, for example, may apply different heuristics to a different set of circumstances than novices. Such an insight can suggest additional attention to be made to those heuristic strategies in interventions such that they are communicated and scaffolded appropriately.

This framework, however, has been criticized in that it can encourage a deficit framing; comparing the two groups implies that novices know less about a subject and such an assertion can encourage criticism of novices by users of data in this framework. However, I believe special precautions can be taken to reduce the likelihood of that happening. Indeed, some assessment of expertise is required to know what it is people should learn, i.e., what their learning goals should be; we should examine expertise and compare it to novice conceptions. As experts are not always cognizant of every aspect of their own cognition, formal examinations of their expertise are necessary. We also can reduce deficit framing by suggesting to readers that expertise is not necessarily holistic; an individual can be an expert in one domain and a novice in another. By casting the individuals in this study as experts and novices, I can caution that I am not adding any value assessment to either of those groups or labels. Experts simply have achieved a goal but

may not necessarily be cognizant of that expertise and appropriately communicating that to novices and therefore their experience requires examination.

Similarly, these studies have been criticized for also not being interested in application and therefore deficit in that sense. In other words, they only point out the "mistakes" without being interested in helping students' learning. I believe this perspective can be dealt with by casting this study as a first step in the understanding of a learning trajectory or progression. In other words, expert-novice comparison studies can function as "endpoints" of a potential hypothesis of progression. Of course, there are other issues to contend with, i.e., that experts may not necessarily fully constitute the "endpoint" given that they are a different population of individuals. Indeed, some critical perspectives suggest that constituent conceptions of a "final" expert conception may be different. Different populations, having different experiences and therefore different cognitive resources, may construct their own normative understanding of a phenomenon that may not necessarily share the same constituent parts as the current set of experts but are still nonetheless normative. This concern can be alleviated in that future studies will take these into account and by framing this study as a cursory examination of these end goals. Indeed, though there are likely to be differences between the novice group and the expert group, they are likely not so different that they do not share any similarities whatsoever. Some of the takeaways from this study should therefore still be applicable to a subset of the population and be valuable to the reader. Future researchers, however, simply need to implement precautions to ensure that the full swath of phenomena is ultimately accounted for.

Constructivist Perspective

As alluded to in Chapter 2, the expert-novice framework in general is often employed in two senses: the constructivist perspective and the cognitive perspective. I refer to both in this study.

As part of the constructivist perspective, I assume that learners use prior knowledge to construct their understanding of new concepts. I also assume that this prior knowledge varies from person to person given our differing life experiences. I therefore went into this study expecting that while experts express scientifically normative ideas, the ways in which they construct these ideas may differ both in the professional and everyday sense. A prior unpublished study suggested that population geneticists when asked to illustrate evolution tend to focus on the individuals and microevolution while phylogeneticists used to investigating large timescales focus more on macroevolutionary aspects of evolution (Marcroft, 2019).

Similarly, I assumed that understandings of evolution from an everyday perspective might differ; some individuals may have constructed their understanding of evolution initially from a teleological perspective, as is common in a lot of novices and suppress, mask, or modify those mental structures when presenting scientifically normative ideas, though glimpses of such novice ideas are observable in certain situations (Hammann & Nehm, 2020; Kampourakis, 2020; Smith et al., 1994; Trommler & Hammann, 2020; Werth & Allchin, 2020). In other words, though I expected experts to express normative ideas, I assumed that the ways in which they were to be shaped may differ according to their experiences as in novices.

It is additionally arguable that I adopt a social constructivist perspective (Bloomberg & Volpe, 2019; Creswell, 2013). Social constructivists adopt elements of constructivism and additionally recognize the role of other external social factors in learning such as external social

cues, cultural tool use in thought mediation, etc. As my methods are primarily inductive, I allowed for the appearance of these external factors. Indeed, I recognize that conceptions are socially derived though I primarily focus on the internal state (conceptions, skills, etc.). I do not draw upon this perspective strongly, however, as this study does not explicitly address external factors insofar as they relate to those internal states. I additionally engaged in data reduction methods that eliminated placeholder codes for some of these other elements that fall under this social constructivist perspective but were beyond the scope of this study.

Cognitive Perspective

The second manner in which I employed the expert-novice framework is in the information processing sense. Much of the representation of literature that exists, especially in chemistry education, uses this framework (Kozma et al., 2000; Kozma & Russell, 2005). This work tends to include assumptions of cognitive load and dual channel processing as well as focusing on identifying features of these diagrams that novices and experts tend to differ on (Stenlund & Tibell, 2019; Sweller, 2005; Talanquer, 2022). Some of the literature, for example, discusses how novices will tend to focus on "superficial" features of the diagram instead of the salient features that experts attend to (Talanquer, 2022).

However, as this study does not interrogate interpretation of formal diagrams, its use is somewhat different. The information processing and cognitivist perspectives (National Research Council, 2000) also focus on differences in phenomena such as connectedness and arrangement of concept networks, procedural knowledge uses and implementation, flexibility in strategy deployment, etc. I addressed some of these elements but not as strongly as the constructivist variant. I did not, for example, formally examine concept networks or construct algorithms that model procedural thought (Anderson, 1996; National Research Council, 2000). Most of the

assumptions of this area were employed and therefore are more applicable to the methods involved, for example the use of retrospective questions rather than talk-alouds. Concern for inhibition of aspects of cognition given that tree-thinking and, by extension, taxonomic reasoning are likely cognitively challenging. I also addressed other aspects of cognitive load in design such as through the number of taxa given to individuals as well as the types of data (photos) given. I additionally was sensitive to both what I was asking of the participants and the framing. Formal methods like card sort tasks or interventions that rigidly recommend the usage of a phylogeny can increase load, given individuals must fit their understanding to the conditions of those tasks.

Embodied Cognition

This study also uses aspects of the embodied cognition perspective. The embodied perspective includes a focus on metaphors recruited to structure new ideas (see Chapter 2). Affable people, for example, might be described as warm, indicating people map their understanding of temperature onto their understanding of social closeness, possibly from the experience of hugging or being close to another person. These metaphors in turn may therefore be used for description of higher-order understandings and may not necessarily be related to the experiences of a person, rather they are just used to structure or understand new ideas. This theory was initially constructed based on the usage of metaphors in language as previously illustrated (Lakoff, 2012; Lakoff & Johnson, 1999). Linguists and other cognitive scientists occupy much of the space in this literature, documenting these types of metaphors used (Gottwald et al., 2015; Lakoff, 2012; Núñez et al., 2012). I similarly expected to see evidence of these metaphors in the ways that they describe features and interpret the entire diagram. As I consider these metaphors part of the meaning associated with aspects of their reasoning, I attended to those appearances. In addition, as a function employing Grounded Theory which

suggests that the results arise from an interaction between the researcher and the subject, I found it difficult not to notice such references given I have some knowledge of the perspective.

Do note, however, that I do not consider this perspective to be heavily applied to the study. Rigorous embodied perspectives typically consider spontaneous co-speech gesture as better evidence of such metaphors (Núñez, 2006; Núñez et al., 2012; Núñez & Sweetser, 2006). Language is also highly conventionalized, meaning that some metaphors are simply used because others have been observed using it (Kövecses, 2008). To use the previously cited example, people may describe others as "warm" simply because others have been observed doing so. Their exposure to that usage is what predicts the usage of the metaphor rather than their mental structure being composed of that metaphor. Despite this failure, this study can provide some of the first evidence, albeit weak, for these metaphors. For more on this limitation, see Chapter 7.

Finally, I also pull from the embodied perspective some theory on categorization. In particular I pull from the literature on prototype theory and its comparison to other formulations of categorical cognition such as classical theory (Lakoff, 1987; Margolis, 1994; Storms et al., 2000; Studtmann, 2023). My results suggest usage of categories in a manner adequately described by prototype theory and as such they are referenced in the framework portions of Chapter 2 as well as the discussions in Chapter 5.

Contextualized Learning

This study also assumes context has an effect on understanding and learning. Context can broadly refer to features of a task, for example, or the physical or social environment in which a learner inhabits, or even the psychological context (relevance) of an activity (Brown et al., 1989; Foster, 2008; Kember et al., 2008; National Research Council, 2000). Several frameworks more or less refer to learning as being contextually bound with some variation, such as knowledge-in-

pieces (diSessa, 2018). I, for the purposes of this study, do not distinguish between them when using this assumption. These suggest that knowledge is prompted in different ways by different contexts and that general or abstracted knowledge arises later, likely after experiences of learning with different contexts.

This study, to an extent, centers this contextualization insofar as it involves multiple types of taxa which could be considered a sort of context. As I previously argued, context has not been well investigated in reference to trees or taxonomic reasoning; only one such study does so (Catley et al., 2010). This study was constructed such that the ways in which individuals' understanding may vary in each of these contexts could be observed. These contexts include the appearance of familiar and unfamiliar taxa as well as the misconceptions associated with such specific taxa.

Context is additionally invoked in the sense that I do not necessarily assume that conceptions are stable. I expected to find conceptions that may or may not be easily changeable based on context. I am simply interested in documenting these conceptions. As previously mentioned, I accounted for changeability in ideas, whether due to learning or context, by attempting to notice signs of change and asking participants directly if they had changed their mind during the span of the task.

Representations as Tools versus Targets

While the representation literature is often used in those studies relating to normative understandings of representations, I do not employ them here strongly, given I am more interested in describing the conceptual underpinnings of tree-thinking which is relatedness understanding. However, as it is tangentially related to tree diagrams and I employ the analysis

of artifacts which may or may not resemble normative diagrams, I do employ some aspects of the representation literature.

Representations in the literature have been characterized in a number of ways, all of which are pertinent to tree-thinking. Of course, one of the goals of evolution education is to have students interpret trees in a scientifically normative way. In that sense, representations are cast as pre-existing scientifically normative tools and learners are attempting to interpret them (Kozma, 2020; Kozma et al., 2000; Kozma & Russell, 2005; Talanquer, 2022). However, another way to conceive of representations is their usage by learners to make sense of concepts. An individual can make an illustration in order to make sense of a new concept (Van Meter & Garner, 2005). Representations in that sense are then are created by individuals and function as mediations tools to aid their learning. These inscriptions and diagrams can even occur on top of preexisting representations---a student might make inscriptions or diagrams on top of a pre-existing representation to make sense of the representation itself or to communicate something.

This study employed both of these ways of understanding representations to differing degrees. On a broader scale, this study is, of course, centered on the constituent precursor concepts that may inform the scientifically normative interpretation of phylogenetic trees; however, direct interpretation was not investigated by this study. This study did however require students to draw their understanding of relatedness. I therefore used their generated representations to gain insight into their understanding of a concept and compared them to each other and to the normative diagrams (i.e., trees). It is therefore arguable that I used representations primarily as a tool for understanding thought (their inscriptions and diagrams) but also recognize that these results will be eventually used to investigate their use as a target for their understanding (interpretation of given diagrams).

However, I also recognize that the act of diagram construction may itself impact reasoning as observed in students and experts (Kozma, 2020). I accommodated this feature of representation use by having individuals give their initial explanation in words and then ask them to use a diagram so the diagram could be considered somewhat "retrospective." Regardless, this assumption was not rigid and were there indications of changes in reasoning, given the drawing portion of the task, I asked individuals to clarify their thought processes.

Prototype Theory

The final and most prominent framework discussed in the analysis and therefore can be considered used by this study is prototype theory. Prototype theory, as previously described in Chapter 2, refers to a model of categorization by cognitive scientists employing an embodied framework wherein individuals will often have categories consisting of a hypothesized prototype that may or may not exist and a set of criteria associated with it with varying likelihoods of appearing in members of said category. As you shall observe in Chapters 4 and 5, I find categories analogous to this theory but I also find categories of responses analogous to competing theories of categorization like classical theory. I also discuss the lack of responses that fit under other competing theories such as exemplar theory.

Procedure Overview

This study was conducted with two groups for comparison, experts and novices. Twenty-five participants were interviewed with the final version of the tasks (Appendix B. Final Interview Protocol and Appendix C. Final Task Slides), 16 of which were novices and 9 of which were experts. Ultimately, a total of 29 participants were interviewed using standard clinical interview methodology (diSessa, 2007; Ginsburg, 1997). The participants completed two sets of questions in a single interview, of each which I refer to ask tasks. The first task and

associated set of prompts probed for ideas about relatedness. The second set of tasks were intended to probe for tree-thinking across multiple diagrams. This study, however, was ultimately pared back due to lack of time and resources. The only methods, therefore, that I discuss in detail in this dissertation are in relation to the first task: the grouping task.

These interviews were conducted online using Zoom, as the study began during the COVID-19 pandemic (zoom.us). Each interview was allotted 60 minutes; however, only the first half of each interview is pertinent to this dissertation. Tasks were visible to participants using Explain Everything, an online resource that records audio as well as mouse movements, drawings, etc. and allows viewers to see and write on slides (explaineverything.com). The interviews were recorded from both modalities resulting in two videos per interviewee.

Each interview was followed by the completion of a contact summary sheet and review of the interview. Recordings were transcribed by an AI tool and then cleaned. Resultant transcripts, contact summary sheets, and videos were used in the final analysis, with the videos in particular containing subjects' diagrams associating their explanations. Final analysis was conducted using Grounded Theory (Glaser & Strauss, 1967) using MAXQDA (maxqda.com). Analyses utilized contact summary sheets for triangulation and orientation to the dataset while the transcripts and artifacts were the primary foci of analysis (M. B. Miles & Huberman, 1994). The two videos were edited together into a single video and attached to the transcripts. These videos were consulted when ambiguities were encountered.

Rationale for Qualitative Methods

Qualitative methods were employed for this study given the nature of the questions asked as well as the dearth of literature available on the subject I was interested in. The two research questions I posed asked about the nature of relatedness thinking and the shape they take in both

experts and novices. Questions about the form of thought typically take the form of qualitative methods where researchers can observe the variants of behavior as well as provide a rich description for them (M. B. Miles & Huberman, 1994; Yilmaz, 2013). This is particularly appropriate given the lack of data on evolutionary relatedness in the field; to my knowledge no one has performed any qualitative, let alone quantitative, examinations of relatedness understanding outside of the context of tree-thinking but within biology education. All taxonomic understanding studies tend to fall under other domains such as psychology and emphasize nascent childhood understandings of organisms (Gelman & Coley, 1990; Gelman & Markman, 1987; Rhodes & Gelman, 2009). Those that do appear in the science education literature focus on the learning of normative taxonomic groups (i.e., what is a fish, what is a mammal, rather than how are these grouped) rather than general relatedness understanding (Ballen & Greene, 2017).

In addition, the tree-thinking literature greatly lacks evidence for its assertions, especially qualitative evidence. Much of the early literature (Baum et al., 2005; Gregory, 2008; Meisel, 2010) outlines common tree-thinking conceptions but provides little to no student utterances to substantiate these assertions. While these misconceptions are likely real phenomena, it behooves the field to have more qualitative data to help researchers develop theory and have a better sense of the phenomena that are occurring. Some researchers believe quantitative work to be better suited for understanding quantitative relationships or properties of phenomena, i.e., how often something occurs or how often a phenomenon co-occurs with other phenomena and is incapable of ascribing causal patterns to human phenomena. While I personally disagree and believe that with care causality can be assessed using quantitative methods, I do agree that qualitative methods give researchers a better sense of what is going on in a phenomenon that complements

quantitative work and is better suited for hypothesis and theory generation (Bernard, 2017; M. B. Miles & Huberman, 1994; Yilmaz, 2013). As there is little to no work in this area, qualitative work is, therefore, best suited for these research questions.

Tasks in Detail

Task Summary

To investigate my research questions, I asked subjects to complete two general tasks, only one of which ultimately was analyzed in the final version of this study. This primary task was intended to investigate participants' understanding of relatedness. The second task I do not discuss further in this dissertation but mention for context. A series of associated probing questions were then optionally used depending on the way the interviews unfolded. While the pilot yielded changes to the second task, it did not yield changes to the first task, so I provide only one copy of the task in Appendix C. Final Task Slides. The probing questions were pre-prepared to anticipate a variety of situations but additional probing questions were generated in the moment for unanticipated occurrences. As previously mentioned, these questions help the interviewer prepare for as many subject responses as possible, as generating these probing questions in-the-moment is difficult. This approach is common in semi-structured interviews (Bernard, 2017).

The probing questions associated with these tasks were immediately retrospective, meaning that they probed for subject's thought processes after a task has been completed (Ericsson & Simon, 1993). I chose this approach because I was concerned about the potential for inhibition by talk-alouds. While talk-alouds yield higher-quality information about in-the-moment thought processes that might be later forgotten, talk-alouds can inhibit sudden insight-based thought, also known as spreading cognition (Schooler et al., 1993). As tree-thinking can be

cognitively challenging (Sandvik, 2008), I was concerned that this would adversely affect data collection. To keep things consistent and therefore simpler across all tasks, I extended these retrospective questions to the relatedness task which is now the focus of this dissertation; the tree-thinking tasks were not analyzed for this study. However, I did not foresee that I would be paring down the scope of this study and were the study repeated for this purpose only, some talk-alouds would have likely been implemented. Reasoning about relatedness understanding might not be as cognitively challenging as tree-thinking as it does not require the individual, presumably, to reconcile the structure of a tree with their reasoning process about relatedness, whereas individuals completing this task are freer to express relatedness in any way they would prefer.

I initially believed that retrospective questions occurring immediately after the task would mitigate data loss due to forgetfulness. However, I do recognize that according to some cognitive models of the mind, reflective and metacognitive thinking are themselves problematic (Ericsson & Simon, 1993). As with any other cognitive processes, some features of a phenomenon are perceived by the subject and some are not. If a subject does not perceive something, then they do not know to comment on it. This is true of thinking itself; the subject would have to perceive and be aware of components of their cognition. If subjects have never been prompted to examine their thought in this way, they may be less practiced in noticing cognitive phenomena that would be valuable to this study. In addition, these metacognitive skills may be unique to each domain and not always transferable (Stebner et al., 2022).

Though I initially planned for the study to be primarily retrospective, I did allow participants to talk while completing the task if they so desired, anticipating that some aspects of the tasks I gave might not induce spreading cognition or other phenomena that might use many

cognitive resources and would therefore be inhibited by talk-alouds. That being said, some conflicting cognitive functions may have occurred if you assume participants' awareness of being able to do so might constitute vigilance which presumably uses some cognitive resources albeit less (Cisler & Koster, 2010; Kimble et al., 2014). The interview protocol was flexible and allowed participants to do what they desired—bringing up too many of these rules for what they should monitor during the task might make participants self-conscious and create a negative environment which could alter or prevent the appearance of certain actions (Zsidó et al., 2022). Doing so might increase their cognitive load as they attempt to monitor their speech and other behaviors during the task (Bong et al., 2016). Alternatively, having a negative rapport with participants more hesitant to share their ideas (West & Blom, 2016) while a positive rapport can make participants hesitant to share socially unacceptable ideas (Horsfall et al., 2021; Tourangeau & Yan, 2007).

Task Development

The first set of prompts were designed to investigate the ways in which individuals might make organismal groups as well as the types of organismal groups they generate. I therefore gave them a set of photos of common organisms (Table 2; Appendix C. Final Task Slides). I focused on macroscopic organisms for brevity and their overall familiarity. However, some organisms I chose were specifically chosen for their likelihood of being not recognized by the participants for reasons of plant awareness disparity (Jose et al., 2019; Wandersee & Schussler, 1999, 2001) or other reasons such as their nonoccurrence in the US (e.g., musk deer). After piloting the study once with four participants, two experts and two novices, I did not ultimately change my probing questions or task structure. While the other portions of the protocol were indeed changed, they are not pertinent to the study discussed in this dissertation.

Table 2: Taxa Included in the Task

Taxon Included	Reasoning for Inclusion
Pine tree	• Plant awareness disparity • Needed >1 plant for grouping • Likely familiar/known • Wanted >1 gymnosperm • Atypical leaf morphology • Macroscopic • In North America
Cycad	• Plant awareness disparity • Needed >1 plant for grouping • Likely unfamiliar/unknown • Wanted >1 gymnosperm • Leaf morphology similar to familiar angiosperm: palm tree • Macroscopic
Palm tree	• Plant awareness disparity • Needed >1 plant for grouping • Likely familiar/known • Wanted at least 1 angiosperm • Leaf morphology similar to cycad • Macroscopic • In North America
Great White shark	• Include aquatic vertebrates • Wanted to include chondrichthyans as well as bony fish due to morphological differences in fins, etc. • Likely familiar/known • Macroscopic • Wanted >1 shark for grouping purposes
Whale shark	• Include aquatic vertebrates • Wanted to include chondrichthyans as well as bony fish due to morphological differences in fins, etc. • Likely unfamiliar/unknown • Macroscopic • Wanted >1 shark for grouping purposes
Tuna	• Include aquatic vertebrates • Wanted to include chondrichthyans as well as bony fish due to morphological differences in fins, etc. • Likely familiar/known • Macroscopic • Wanted possible outgroup for sharks • Wanted at least one “typical” fish
Alpaca	• Include terrestrial vertebrates • Likely less familiar/known (in North America) • Macroscopic • Wanted >1 camelid • Wanted several ungulates since morphology distinct • Very different environmental conditions compared to other ungulates in this dataset

Table 2: Taxa Included in the Task, Continued

Taxon Included	Reasoning for Inclusion
Camel	• Include terrestrial vertebrates • Likely familiar/known • Macroscopic • Wanted >1 camelid • Wanted several ungulates since morphology distinct • Very different environmental conditions compared to other ungulates in this dataset
White Tailed Deer	• Include terrestrial vertebrates • Likely very familiar/known • Macroscopic • Wanted several ungulates since morphology distinct • Very different environmental conditions compared to other ungulates in this dataset • In North America
Musk Deer	• Include terrestrial vertebrates • Likely unfamiliar/unknown • Macroscopic • Wanted several ungulates since morphology distinct • Very different environmental conditions compared to other ungulates in this dataset • Distinct morphology likely unfamiliar with: enlarged canines
Domestic Pig	• Include terrestrial vertebrates • Likely very familiar/known • Domesticated/example of human intervention • Macroscopic • Wanted several ungulates since morphology distinct • Very different environmental conditions compared to other ungulates in this dataset • In North America
Blue Whale	• Include aquatic mammals (for comparison to other aquatic vertebrates that are not mammals) • Relationship with artiodactyls like pigs, hippos • "whales are fish" misconception • Likely familiar/known • >1 cetacean for grouping • Very different environmental conditions compared to other ungulates in this dataset
Bottlenose Dolphin	• Include aquatic mammals (for comparison to other aquatic vertebrates that are not mammals) • Relationship with artiodactyls like pigs, hippos • "whales are fish" misconception • Likely very familiar/known • >1 cetacean for grouping • Very different environmental conditions compared to other ungulates in this dataset
Hippopotamus	• Include semiaquatic mammals/semiaquatic animals • Relationship with artiodactyls like cetaceans • "whales are fish" misconception • Likely familiar/known • Very different environmental conditions compared to other ungulates in this dataset

Rationale for Choice of Task. Card-sort tasks were initially recommended (Hewson & Hewson, 1989) as I essentially wanted to have participant group organisms, but I decided against

them because of their limitations. As I anticipated that participant responses may contain some spatial understanding, I believed the card-sort task was too restrictive. In-person and online card-sort tasks restrict individuals to grouping within a single plane using physical movement. While this may reduce the cognitive load as it offloads some of the sorting to the environment and movement (Scaife & Rogers, 1996), I believed this would constrain the type of thought to one that the task could accommodate and paradoxically increase cognitive load (Van Bruggen et al., 2002). In other words, were I to compare the card-sort task to a diagram, it would resemble a Venn diagram in its use. While some card-sort activities allow individuals to use drawings in combination with the cards, I felt the physicality of moving the very large cards might still restrict them to making certain types of representations. Individuals might have more complex innate understandings than what could be gleaned from this task. Alternatively, from a prompting and context perspective, one could argue that the card-sort task resembles the "real" authentic task less than a more abstract context where words or terms could be moved around though they may or may not be constrained by 2D diagram. I did not want participants to be prompted by the card-sort task in that way—though it is notable while I did not use it for research, it still may be useful as a pedagogical tool and a few have tried this (Gibson & Cooper, 2017) though, again, constrained within a tree context as mentioned earlier in this dissertation. As I wanted to get a better understanding of participants' ideas about taxonomic relatedness without the confines of this context (and I did not want them to engage in the additional cognitive load of trying to fit their ideas to an implied diagram or structure as in most of the tree-thinking education literature) I thought the task chosen (i.e., initially verbal response followed by a mostly free representation of said response) would be most appropriate.

While the task does eventually ask them to represent the taxonomic groups on a two-dimensional plane, they are still afforded more freedom than the card-sort task in that they can make their own representation free from the restrictions of movement and form. In addition, the ordering of the subtasks allows for the creation of a representation that resembles their thinking more than a typical task that integrates diagram creation, asking them to describe their thinking in words, giving individuals the ability to potentially reveal their nascent understanding without the context of a normative diagram or task. This of course assumes that "nascent" understanding is represented better or consists of words or something akin to words rather than an iconic "picture" but I am also assuming that being directly directed to create an external diagram is still more confining than being asked to use words. Asking for their explanation first would give insight into their initial mental representation, should they ultimately decide to change the way they explain it in the diagram—indications like utterances about their confidence in the way to represent it or pedagogical views could signal that they have changed their reasoning type. That being said, tasks that are quite open can still increase the cognitive load of the task as individuals may have too much freedom to depict their understanding any way they wish (Ashman et al., 2020; Martin, 2016). Still, I thought this is preferable to the too strongly context-bound card-sort task.

Rationale for Broad Choices of Taxa. The first task, i.e., the organism grouping task, included mammals, fish, and plants (Table 2). I chose specific groups because of common taxonomic misconceptions, some of which have been illustrated in the literature and others which have not. Plants were included because of plant awareness disparity (Achurra, 2022; Jose et al., 2019; Wandersee & Schussler, 1999, 2001); plant awareness disparity has not been examined in coordination with interpretation of tree-thinking. How those two phenomena might

interact has not been examined. Mammals and other macroscopic organisms were included for their high likelihood of being recognized by the participants (Trimble & Van Aarde, 2010). Most specific taxa were chosen as being possibly recognizable by individuals who have lived in the US (e.g., white-tailed deer).

Microscopic organisms were not considered for their unfamiliarity as well as their distinctness from other organisms. Microscopic organisms, typically prokaryotic organisms, tend to have their own set of complexities when being examined by researchers and their pattern of evolutionary history is likely not tree-like (Kunin, 2005). In other words, trees are likely poor models of microbial evolution given their complex genetic histories. This study focuses more on the underpinning of tree-thinking which therefore lends itself more to macroscopic eukaryotic organisms.

I did omit arthropods and other invertebrates for space. I did not, however, mean to omit reptiles and birds as I was interested in some of the misconceptions associated with those organisms (Novick & Catley, 2014). That was an oversight. That being said, the organisms chosen did have to fit on one slide and the photos had to be large enough to adequately show their features. In a non-online context, I could have chosen more taxa but as I was restricted to Zoom, I had to cut down the number of taxa shown. In addition, the original version of the interview had several other tasks to complete and additional taxa would make the interview too long.

Rationale for Specific Taxa. Plants were considered necessary for this particular study (Table 2) because plant awareness disparity is well documented in the literature (Achurra, 2022; Jose et al., 2019; Wandersee & Schussler, 1999, 2001) but not necessarily investigated in the context of tree-thinking. I expected there to be some sort of interaction between the context and

their understanding of phylogenetic trees. Multiple plant taxa were chosen for the organism grouping task because I thought that they might perceive plants to be quite different than animals and set that aside—I wanted to see how they would deal with multiple plants and be forced to create taxonomic groups within plants. I also included plants that might be unfamiliar to typical US citizens such as cycads as they resemble palm trees yet they have cones as they are gymnosperms.

Whales were included amongst other mammals as well as fish, given the common misconception that whales are fish. Likely people are grouping them for morphological reasons—including this example in the organism grouping task would allow me to observer reasons why they grouped them this way as well as how they might contrast to other strategies for grouping or reading phylogenetic trees later. I additionally included hippos as they are the closest extant relative of whales and dolphins. I included pigs which, while an artiodactyl as well, might be grouped by participants monophyletically with hippos outside of whales and dolphins given their morphological similarity. In addition, as I included fish and was looking at the placement of whales, I included whale sharks which are sometimes mistaken as whales given their shape and feeding strategy.

Organisms that I perceived as being moderately to largely unfamiliar to participants were the musk deer, llama, whale shark, and cycad. I included these organisms to see how participants would manage unfamiliar organisms and fit them into or change their preexisting animal categorization structures. This also means they were chosen due to their relationship with other groups. Of course, this assumption was US-centric. Some students with experience in other parts of the world may be familiar with them.

Table 3 indicates the correspondence between each probing question or task and research question. To remind the reader, research question one referred to the ways in which experts and novices conceived of relatedness and thereby implicitly investigated the structure of said understanding. Research question two referred to the ways in which participants grouped organisms.

Research Questions:

1. Understanding Relatedness

   In what ways do experts and novices conceive of relatedness?

2. Investigating Ways in which Taxa are Grouped

   How do experts and novices group (i.e., classify) organisms?

Table 3: Probing Questions Correspondence to Research Questions

Probing/Task Question	**Research Question**
2a. Here are some different organisms. Have you seen any of these before?	2
2b. Are these organisms related or similar to one another at all? How so and how do you know?	1,2
2c. Let's say you were trying to explain to your friend how those organisms you mentioned were related. How would you explain how they are related? Your friend, let's say, really likes looking at diagrams when trying to understand things and asks you to draw a diagram to help you explain. How would you draw it? Go ahead and use the drawing board if you'd like.	1,2
APQ2a. Have you ever thought about how organisms are related to one another? Like, "Oh gee that animal must be related to that animal" for whatever reason?	1,2
APQ2b. How do you know that some organisms are related?	2
APQ2c. What kinds of organisms have you thought about as being closely related or not closely related?	1
APQ2d. What does this [*cursor points to part of the diagram*] mean?	1,2
APQ2e. [*If they as the interviewer if their friend knows about biology or not*] Let's assume they don't have a lot of biology experience. Also, how would it look if they did have a lot of experience?	1,2
APQ2f. What kind of information would you need to figure out how they are related?	2
APQ2g. Are they all equally related to one another?	2
APQ2h. What does it mean for them to be "related" – what does that mean to you?	1,2
APQ2i. What does it mean for them to be "similar" – what does that mean to you?	1,2

Rationale for Segmenting Participants into Experts and Novices

Experts and novices were considered the main comparison groups as very little literature on experts exists in tree-thinking education (e.g., Halverson et al., 2011; Novick & Catley, 2007). Solely charactering expertise, however, was of little interest to me as the comparison between the groups can spark the field's interest in describing the learning trajectory for tree-interpretation. While much of the literature is composed of tree-thinking misconception work as well as practitioner-oriented tree-thinking interventions, only a handful compare start points to end points (Halverson et al., 2011; Novick & Catley, 2007). While expert-novice framing has been criticized for its preoccupation with what is "wrong" with students, its deficit orientation, and not lower likelihood of encouraging the examination of positive resources or strategies that can aid students' learning (Robbins, 2005), I perceive this study and future studies including expertise and richer description of student experiences as a first step towards understanding the process of learning for a subject. While this study does not attempt to formally document a learning trajectory or progression (Lobato & Walters, 2017), it at a minimum gives us insight into the two endpoints of the learning process (Kozma, 2020). There are, however, some limitations to this perspective which you may review in Chapter 5.

This section therefore discusses the groups that were used and the rationale for their inclusion and the specifics around their sampling.

Undergraduate Participant as Novices

I wanted individuals to not have experience with tree-thinking but be biology majors as I wanted to capture nascent understanding of taxonomic relatedness without the confines of tree-thinking. I therefore solicited undergraduates from introductory biology courses as novice participants. These students were unlikely to have been exposed to instruction on phylogeny

conventions as they were presumably starting coursework in their major, as biology majors are assumed to be the primary consumers of these courses. There is a minor problem with this assumption which I address in the limitations section of Chapter 5. Of course, these students are more likely to have biology experience or interest rather than non-biology students, given they chose to take these classes so they may not be "true" novices. However, they are considered novice "enough" as the results from this study would be likely used to construct interventions in the future for this specific group of students. Knowing more about the resources these undergraduate biology students use to construct their understanding of evolutionary relatedness and phylogenies can help researchers and educators design effective and appropriate interventions.

I expected that students in this group would have seen or interacted with phylogenies as they are often part of museum displays (e.g., MacDonald & Wiley, 2012) and appear in high school textbooks (e.g., Catley & Novick, 2008), though the latter typically depicts trees without discussing tree-thinking or interpretation. Some of the tree-thinking literature does involve high school students' (Catley et al., 2013; Gibson & Cooper, 2017) and young children's understanding of trees (Ainsworth & Saffer, 2013). However, it is unclear the extent to which K-12 educators have adopted instruction of representational conventions as trees are not explicitly mentioned by the Next Generation Science Standards (NGSS; Davenport et al., 2015; NGSS Lead States, 2013). NGSS does, however, mention learning about diversity and relatedness broadly as well as communication using diagrams (University of California Museum of Paleontology, 2021a), of which tree-thinking could be considered a part. Regardless, there is no guarantee that they have fully understood or are thoroughly practiced in tree-thinking even if students were exposed to such instruction, especially since tree-thinking is considered difficult

(Phillips et al., 2012; Sandvik, 2008), and deep learning of complex topics likely require a substantial amount of time to occur (Lobato & Walters, 2017).

Faculty and Post-Doctoral Researchers as Experts

The second group consisted of faculty and/or post-doctoral researchers, with faculty of any type from any type of higher education institution. This group could be considered "expert" and is used for both comparison and also an accounting of expert conceptions. Faculty are assumed to have the most normative knowledge (knowledge here I am using in the broadest sense) of a subject as the goal of education is generally to get students to learn this knowledge and adopt experts' practices. This therefore was considered a point of comparison.

Characterization of a "target" for education can essentially clarify our goals. Often educators and education researchers have designated "targets" in the form of learning goals or competency guidelines (Jiang & Elen, 2011), but these are sometimes limited. While the individuals who make these learning goals and guidelines are undoubtedly experts, experts themselves may not necessarily have the immediate self-awareness to identify what is different about their own thinking and student thinking and what exactly constitutes the skills and knowledge necessary to be successful. Having metaknowledge about the subject or even about pedagogy is itself a different set of skills than the implementation of said knowledge in domain specific areas (K. K. H. Chan & Hume, 2019; Cochran et al., 1993). Having formal studies where experts' behaviors and cognitive processes are critically and deeply analyzed can spawn new observations and avenues of research and, indeed, I believe this occurred several times in this study which I elaborate in future chapters.

However, designating these individuals as experts does not necessarily suggest that experts' ideas are superior and that the construction of their understanding of phylogenetics

should be exactly modeled to students (Robbins, 2005). Recruitment of diverse STEM faculty has been difficult (Kaplan et al., 2018) so current STEM faculty often do not resemble students. Students may therefore not be drawing on the same types of cognitive resources when constructing their understanding of relatedness. In addition, as with any members of a community of practice, there exist different opinions though some convergence has occurred, and new members may add or change the community's knowledge once they become a part of it (Mørk et al., 2010). As students are future members of said communities, I recognize that their ideas that are inconsistent with the current normative knowledge of the group may not necessarily be "wrong." I also recognize that casting them as such in a deficit manner is actively detrimental to the learning process (Canning et al., 2019; Davis & Museus, 2019; Painchaud-Nash, 2020), not to mention it being uncharitable.

I also chose to include a wide range of biology experts as tree-thinking expertise is not a monolith; trees are widely but disparately used in a variety of subdisciplines. For example, researchers in evolutionary medicine are often interested in tracking the evolutionary history of cancer cells—tumor cells will compete for resources, and those cells that accumulate mutations that help them reproduce faster will edge out others (e.g., Schwartz & Schäffer, 2017). Other macroevolutionary biologists are interested in reconstructing patterns of species dispersal (e.g., Buckner et al., 2015). Epidemiologists and virologists are alternatively interested in determining which strains were responsible for zoonotic events (e.g., Boni et al., 2020). Population geneticists use trees to track genetic changes on a smaller timescale among individuals and populations (e.g., Themudo et al., 2020). In addition, teaching faculty appear to translate among and coordinate between different subdisciplines more deftly than others in more research-oriented disciplines, implying they have a stronger organizing cognitive structure of the content

to be taught (Marcroft, 2019). I believed that sampling from this diversity would give us a better understanding of the different potential ways in which expertise could appear. While experts typically have more convergence on a subject given the normative nature of the knowledge in their community, I also think there is some divergence given its use in so many fields. In an unpublished pilot, I found that individuals' diagrams of evolution, for example, seemed to vary widely and according to the discipline in which they work.

Sampling Strategy and Recruitment

My initial sampling strategy consisted of convenience sampling, purposeful randomized sampling, and stratified sampling, while my final sampling strategy consisted of convenience sampling, stratified sampling, and snowball sampling (Patton, 1990). All students were sampled from either two-year or four-year degree-granting institutions in the United States and associated territories. They include community colleges, research-intensive universities, and teaching-focused universities. While students were not asked what institution they came from, insight into these institutions was assumed from the institutional email address they used. All students were enrolled in or had been recently enrolled in an introductory biology course which may or may not have included a phylogeny instruction component.

Attempts were initially made to primarily sample from specific institutions, including an HSI, for students, but students were not available in the time span required for the study. In addition, an additional stratified randomization method for choosing experts was employed (where individuals were chosen to represent a breadth of subdisciplines) involving cold-emailing faculty and post-doctoral scholars from the Carnegie Classification of Higher Education and the Directory of Community Colleges (A2Z Colleges, n.d.; Carnegie Classification of Institutions of Higher Education, n.d.). These efforts were ultimately fruitless. After approximately forty

tailored emails and no responses, the sampling method was changed to that of snowball sampling and convenience sampling for both experts and novices.

In the second attempt to recruit participants, I posted twice on the Society for the Advancement of Biology Education Research (SABER) listserv, an email group consisting of biology educators and biology education researchers (https://saberbio.wildapricot.org/). SABER was chosen for their partiality to education—they were assumed to have a higher response rate. Of course, this population would be more attuned to science education and likely use evidence based instructional strategies so they may not be wholly representative of all faculty and/or students. The SABER pool was used for both faculty and student recruitment, with faculty either participating themselves or providing study recruitment information to their undergraduate students in their introductory biology courses.

I also reached out to former colleagues in various biology disciplines to solicit their colleagues for participants for my study. Several individuals from this part of the recruitment process came from this method, but none of them offered the recruitment survey in the class.

I lastly used the Society for the Advancement of Chicano/Native Americans in Science (SACNAS) listserv to solicit student participants (https://www.sacnas.org). I recruited from SACNAS as they have a larger proportion of undergraduates on their listserv compared to other biology related societies. The population also tends to be more diverse so sampling from that population might solicit a wider range of responses. Not all students from SACNAS were currently enrolled in an introductory course. I only asked if they were currently undergraduates. I however believe this deviation from anticipated recruitment is acceptable given that tree-thinking is generally difficult (Phillips et al., 2012; Sandvik, 2008) and conceptual change usually takes considerable time (Lobato & Walters, 2017) so even students who were not currently enrolled in

a course were likely to be comparable to other students in the sample. In addition, some of the individuals in the aforementioned introductory courses were upperclassmen because their institution allowed them to take certain courses out of order.

Snowball sampling with colleagues provided faculty participants, and snowball sampling from SACNAS participants provided other student participants.

Individuals from EvolDir listserv—an email group consisting of evolutionary biologists and other researchers (https://evol.mcmaster.ca/evoldir.html)—were also considered for expert recruitment but my solicitation email was ultimately rejected as the moderators perceived it as not applicable to their users.

Interested participants completed a Qualtrics online form to show interest and when interviewed were asked to complete a consent form. The aforementioned form asked for contact information as well as basic demographic information. Demographic information such as gender and race/ethnicity were included in the initial recruitment survey to ensure that participants were diverse, maximizing the likelihood of capturing a wide variety of experiences and viewpoints. However, these questions were indicated as "optional," so participants did not assume that the study focused on these identities (Pennington et al., 2016). I also asked for this information at this stage rather than the interview so as to avoid stereotype threat (Pennington et al., 2016). This strategy created a considerable time gap between initial solicitation and the interview such that stereotype threat would not be an issue.

Ultimately, time and resource constraints required that I contact all respondents to the initial recruitment; I could not be as selective with participants as I had initially anticipated. Participants were limited to English speakers as I was the only interviewer. Participants were enticed by a compensation of $15 through Venmo as well as by altruistic appeals. As people are

often motivated to participate in studies for altruistic reasons (Williams et al., 2008), I advertised that my study may help future educators better help undergraduate students.

Sample Composition

I initially intended the two groups to be investigated to have 15 participants each for a total of 30 participants. Due to resource and time restrictions, the study ultimately involved 11 experts and 18 novices. The entire study, therefore, contained 29 participants. Of the 29 participants, only three were men, two of which were novices. One participant identified as non-binary and was an expert. Four individuals declined to state their gender, three of which were novices. Given open-ended demographic prompts, 10 participants identified as White, seven of which were experts. Two individuals identified as White Latinx, one of which was an expert and one a student. One novice identified as White and Middle Eastern North African (MENA). Three individuals identified as South Asian, with two being experts. One novice identified as exclusively Hispanic. Three novices each identified as East Asian, Asian (broadly) and MENA. Nine individuals declined to state their race ethnicity.

A summary of these participants and their information is included in Table 4. Participants were asked to create their own pseudonym to avoid misrepresentation and other ethical issues (Brear, 2018) which is why some of them may seem unusual. This practice, however, does not address all potential issues of pseudonym use (Brunger & Russell, 2015; Grinyer, 2002). At the time of procedure creation, I was unaware of the option to discuss all potential ethical implications of pseudonym choice with the pseudonym, so I did not do so. However, I did mention that they should pick something that does not resemble their own name (a few individuals initially picked a pseudonym that closely resembled their first name).

Table 4: Table of Participant Demographic Information. Bolded pseudonyms were given by the author.

Pseudonym	Gender	Race/Ethnicity	Major/Area of Expertise	Academic Year	Status
Alice	Woman	*Decline to State*	Biology	1	Novice
Bhakti	Woman	Asian Indian	Biology	4	Novice
Elisa	*Decline to State*	*Decline to State*	Undecided	1	Novice
Ella	Woman	White	Neuroscience	4	Novice
Esther	*Decline to State*	*Decline to State*	Applied Mathematics	2	Novice
Ida	Woman	Latina/White	Zoology	4	Novice
John	Woman	Hispanic	Biology	4	Novice
King	Woman	White/MENA	Biology	4	Novice
Louise	Woman	North African American	Biology (Undecided)	1	Novice
Maira	Woman	*Decline to State*	Biopsychology	1	Novice
Paradise	Man	Asian	Environmental Science	3	Novice
Peppermint	Woman	White	Environmental Science	2	Novice
Sasha	Woman	East Asian	Biology	4	Novice
Seline	Woman	*Decline to State*	*Decline to State*	*Decline to State*	Novice
Theodore	Man	White	Undecided	2	Novice
Turtle	*Decline to State*	*Decline to State*	Computer Science	2	Novice
Veronica	Woman	*Decline to State*	*Biochemistry*	3	Novice
Victoria	Woman	*Decline to State*	*Decline to State*	*Decline to State*	Novice
Gardener	Woman	White	Biology Education	N/A	Expert
Jean	Woman	White	Biology Education		Expert
Luna	Woman	White	Biology Education		Expert
Mika	*Decline to State*	*Decline to State*	Biogeography		Expert
Professor X	Man	South Asian	Systematics		Expert
Queen	Woman	White	Evolutionary Biology		Expert
Ridley	Non-Binary	White	Evolutionary Ecology		Expert
Sita	Woman	Asian Indian	Evolutionary Biology		Expert
Turaco	Woman	White	Ecology		Expert
Witch	Woman	White, Latinx	Evolutionary Biology		Expert
Zoey	Woman	White	Biology Education		Expert

Clinical Interviews

I conducted clinical interviews as is common in both education research broadly as well as those with a primarily constructivist orientation (Clement, 2000; Ginsburg, 1997). Clinical interviews were chosen in order to capture rich qualitative data (Bernard, 1988; Clement, 2000) about the diversity of ideas surrounding evolutionary relatedness and the interpretation of formal phylogenetic trees and to provide insight into how conceptions and interpretations differ among groups with different experience levels to hypothesize how conceptual change might occur. In other words, as is common in qualitative research (Yilmaz, 2013), I am interested in

characterizing the “what” and “how” of conceptions of relatedness and the interpretation of formal phylogenetic trees.

Clinical interviews are typically semi-structured and thus less prescriptive than fully structured interviews as they allow for on-the-fly changes to prepared questions (Bernard, 1988). I created an interview guide (Appendix B. Final Interview Protocol) to organize my tasks and prepare for potential responses by interviewees (Bernard, 2017). Through this process I engaged in hypothesis testing (Ginsburg, 1997), where I developed hypotheses in-the-moment of participants’ mental processes and used prepared or spontaneous probing questions to assess for these hypotheses. If, for example, I suspected a student was using a node-counting strategy to determine which pairs of lineages are most closely related, I posed a hypothetical situation placing additional lineages between the two the subject picked and asked them to repeat the same task. If I observed that the subject switched to a pair with fewer nodes, my hypothesis was supported. If they chose another pair that does not have the fewest nodes, I knew to pursue other possibilities. Semi-structured clinical interviews were thus well suited for this study, as they were intended to get at the participants’ understanding of concepts and perception of features as they interacted with the tasks.

Data Summary

Ultimately, I collected data in the form of video recordings, transcripts, contact summary sheets, and artifacts from Explain Everything. I recorded video of the interviews using Zoom's built-in video recording function. I also recorded video of the participants’ inscriptions in Explain Everything. Recordings from each interview session were initially transcribed using Sonix.AI and were initially edited by me using their in-house transcript editing tool (sonix.ai). A

subset of the deidentified transcripts were transcribed by Zoom's AI transcription and edited by a third-party transcriptionist.

The resultant videos (Zoom plus Explain Everything) were edited together such that the audio could be paired with both the video of the participant and their drawings or writings on Explain Everything. The stripped and anonymized audio was used by the aforementioned third-party transcriptionist to edit the transcripts.

Contact summary sheets were filled out immediately after interviews. Contact summary sheets capture in-the-moment researcher inferences about student thought or other phenomena which may be lost upon reflection as well as help with data reduction (M. B. Miles & Huberman, 1994). They also better capture contextual information that might later be forgotten upon viewing the videos. Additionally, intangible interaction experiences like body language or interaction norms are better understood in the moment (e.g., O'Malley et al., 1996).

Artifacts from the Explain Everything videos were also examined. These artifacts included drawings made by participants while they worked through their tasks. Screenshots of the final diagrams made were included in the analysis software.

Analytical Methods

In this study, I used a primarily qualitative approach. Qualitative methods allow you to richly describe the variation in phenomena observed in a study group and the processes by which any observed changes may occur (Yilmaz, 2013). As I am interested in adequately characterizing and comparing the breadth of conceptions of evolutionary relatedness and tree interpretation, qualitative methodology was most appropriate. Furthermore, I am interested in how novice ideas may differ from expert ones and if experts still engage in some of these novice ideas as well as breadth of ideas expressed. While the complex relationships between these factors may be

examined by quantitative methods such as structural equation modeling, these approaches are less inductive and disallow the emergence and characterization of previously unaccounted for phenomena; qualitative work is still required to develop new models of mental processes or thoughts (Clement, 2000). Additionally, some have argued that isolating candidate variables in quantitative work may lead to oversimplified modeling of phenomena (Anderson et al., 1999, p. 199). Indeed, Clement (2000) has argued that quantitative approaches are insufficient for studying higher-order thinking. He has argued that controlled studies of higher-order thinking are difficult to conduct due to the broad assortment of factors involved. Qualitative methods may be better suited to investigate behaviors as they relate to complex contexts than quantitative methods.

In the following paragraphs, I discuss my methodology for conducting my analyses. The resultant data from the clinical interviews were analyzed using Grounded Theory (Glaser & Strauss, 1967). I discuss these methods and my reasons for choosing them below. I also discuss the frameworks upon which I designed this study and conducted the analyses. I end with additional assumptions and methods that do not fall under the aforementioned categories.

Unit of Research

The unit of research used in my analyses was one concept or an idea conveyed within a turn or over multiple turns, also known as a complete thought unit (Keyton, 2018). A turn refers to the segment of speech that a single individual conveys without being interrupted in a significant way. Units of analysis often vary from single words in more discourse focused studies to units larger than turns (Roller & Lavrakas, 2015). The grain size of complete thought was chosen because I anticipated that participants would convey multiple concepts or ideas within a single turn. Therefore, to adequately capture all of this data, the unit would have to be smaller

than a turn. Even smaller units of research (such as words) are more appropriate for discourse analysis, so they were also not considered. It is notable, however, that at times, multiple turns were considered part of one idea. This addendum was made when it was clear that after small exchanges, such as after a small clarifying question by me, that a single idea was being communicated over multiple turns and so they were coded as such.

A drawback of this grain size is that there may be some overrepresentation of certain ideas in the case that an individual might interrupt their own thoughts, for example, to bring up another point and return to the same thought. This might be coded as two instances of the same concept when they should be coded as one. However, as this study is not concerned with the frequency with which individuals utilize these ideas, I believe that these would not impact the analysis. As you will see in subsequent results chapters, I typically report the number of individuals who were observed displaying a code or category in order to get a sense of breadth instead of pervasiveness.

Grounded Theory

Background and Implementation

Interview transcripts, and artifacts from Explain Everything were analyzed using Grounded Theory. Grounded Theory is a framework that guides the collection and analysis of qualitative data. Grounded Theory encourages researchers to infer patterns from the data directly, rather than look for patterns determined by outside sources such as the literature or their prior experiences (Glaser & Strauss, 1967; Strauss & Corbin, 1994). Therefore, the process involves the creation of codes or categories after examining some or all the data, rather than using existing codes.

Grounded Theory is appropriate because some variants accommodate post-positivist and social constructivist perspectives, recognizing the researchers' role in qualitative analysis (Charmaz, 2008; Lomborg & Kirkevold, 2003). Philosophically, I maintain that researchers' life experiences and their resultant cognitive structures (or however you choose to characterize them) will affect the interpretation and analysis of data. Indeed, this is a feature of qualitative methods in general (Clement, 2000; Yilmaz, 2013), but some versions of Grounded Theory account for this explicitly (Charmaz, 2008; Lomborg & Kirkevold, 2003).

However, Grounded Theory also has strong emphasis on data-based inference attempts to avoid misinterpretation of the data. Strauss and Corbin (1994) implied that researchers may erroneously analyze qualitative data because their preexisting hypotheses strongly bias the data's coding; researchers will unintentionally and improperly force data into preexisting categories when those categories should have been redefined or remade. Inductive approaches like Grounded Theory can reduce the likelihood of this problem occurring and better represent the participants and the situations described by the data. In addition, Grounded Theory has very explicitly prescribed methods which presumably help achieve validity of the analysis. In other words, Grounded Theory has many concrete methods built into it such as constant comparison that can help researchers ensure that the data has been thoroughly examined. One could argue that sitting with the data can also induce overanalysis but Grounded Theory has some validity guidelines that can minimize that as well (Corbin & Strauss, 1990; Glaser & Strauss, 1967).

Grounded Theory involves a wide array of practices, a subset of which pertinent to this study I describe here. The most prominent feature of these is its usage of hierarchical codes. Typically, coding occurs at three levels: open coding, axial coding, and selective coding (Corbin & Strauss, 1990). These codes generally differ by the degree to which they hierarchically

organize the data, the attention given to the relationships between categories within a level, and the depth of description for each category created.

Why Grounded Theory

Straussian Grounded Theory was my primary mode of analysis for several reasons. This study consisted of examining students' utterances about relationships as well as their diagrams intended to illustrate this understanding. While relatedness understanding has not been investigated in biology education, interpretation of diagrams has been and has been characterized as being perceptual or cognitive. Much of the work in tree-thinking education has adopted this perspective, such as those that have investigated how predispositions towards particular tree interpretations may develop based on our innate parsing of the visual field (Novick & Catley, 2007; Novick & Fuselier, 2019). Interpretation, however, can also be operationalized in terms of meaning assignment or metaphorical connection. If I explicitly interpret a line to be a lineage, for example, I am assigning that symbol the meaning of "evolutionary group." This association of meaning is not necessarily a perceptual tendency towards a particular interpretation influenced by innate biological processes or early development. I believe Grounded Theory's methods can account for this meaning-based operationalization of interpretation (Glaser & Strauss, 1967) as well as coinciding with the frequent use of Grounded Theory in other types of educational and cognitive qualitative research (Lobato et al., 2011; Martinovic & Manizade, 2017).

Grounded Theory was also more appropriate in that it is more inductive than other a priori methods (M. B. Miles & Huberman, 1994). Also referred to as generative approaches (Clement, 2000), inductive approaches essentially allow for categorizations or conclusions to be made from the data (M. B. Miles & Huberman, 1994). A priori approaches, also referred to as convergent (Clement, 2000), center more on coding using preexisting theory or coding schemes

(M. B. Miles & Huberman, 1994). As previously mentioned, the tree-thinking literature for a time referred to misconceptions wholesale without the student utterances to support it (Baum et al., 2005; Gregory, 2008; Meisel, 2010). The use of a priori coding systems in tree-thinking education may have hampered our ability to critically reexamine the applicability of the conceptions currently in the literature. In addition, the lack of data in the field on relatedness understanding necessitates a more generative qualitative approach that can richly describe the phenomena in question in order to better aid theory generation (Kuczynski & Daly, 2003).

Straussian Grounded Theory was chosen specifically over other forms of Grounded Theory as it acknowledges the existence of preexisting theory and allows the researcher to use this theory to understand data and develop categories (Corbin & Strauss, 1990). While many of the misconceptions in the literature are not substantiated by the student utterances, I did expect to see some of the misconceptions described in the literature in the data set, especially if individuals chose to represent their understanding in the form of a phylogenetic tree.

Grounded Theory in general was additionally chosen because of its rigor. Grounded Theory has a variety of built-in concretely described systematic methods—such as constant comparison—that help researchers create viable categories (Clement, 2000; Glaser & Strauss, 1967; Lomborg & Kirkevold, 2003). Viable categories are those that have essentially satisfied a set of criteria that ensure their quality—criteria that differ from the criteria used to evaluate quantitative work. The original formulation of validity criteria included four major themes—fit, generality, understanding, and control (Glaser & Strauss, 1967). To the best of my ability, I adhered to these criteria in an attempt to increase the validity and generalizability (in terms of theory, not quantitative persistence in the population) of the results. The techniques that Grounded Theory describe such as constant comparison, or the attention paid to the

aforementioned categories helped ensure that the codes or categories generated have been exhaustively contemplated.

Coding in Grounded Theory

One of the ways in which Grounded Theory is distinct from other qualitative work is its framing of codes in addition to its methods and background philosophy. Grounded Theory typically has three primary code or category components which I describe here.

Open coding involves the sorting of elementary codes into larger, well-described inferential categories using a process called constant comparison (Bloomberg & Volpe, 2019; Corbin & Strauss, 1990). Constant comparison involves exhaustively and repeatedly comparing each detected instance of a code to each other to refine better and further develop the boundaries of open-coding categories.

Axial coding and selective coding further sort these open-coding categories into broader ones (Bloomberg & Volpe, 2019; Corbin & Strauss, 1990). These are less like open-coding as they involve the sorting and characterization of those inferential categories developed by open-coding. Codes at these levels are formulated with consideration with other categories as well as the processes and contexts observed in the dataset.

Selective coding is of particular note because researchers rarely employ this feature of Grounded Theory. Broad categories created at this selective coding stage are linked together in terms of context and process; a robust theory is developed to explain all phenomena observed in the dataset (Bloomberg & Volpe, 2019; Corbin & Strauss, 1990). Selective coding essentially is where the "theory" in Grounded Theory is developed.

My analyses could be considered as using at a minimum open and axial codes. As shown in the codebook (Appendix E. Codebook) and Chapter 4, smaller codes than at open level were

generated just to get a sense of the transcripts. These were examined to produce open codes which are the smallest units of interesting and interpretable results but in this case were not reported. These were then assembled into larger hierarchical codes which could be considered axial (e.g., grouping strategies). These axial codes included several levels of codes and subcodes. It is arguable whether or not I achieve selective coding congruent with strict Grounded Theory, but I do discuss how these axial codes interact briefly and deep discussion of the interactions of these codes, while initially planned, were scrapped due to lack of time and resources (Chapter 4). I discuss the actual process of coding in detail in the "coding procedure" section of this chapter.

Establishing Validity

One of the most important aspects of Grounded Theory (Glaser & Strauss, 1967) as well as many other qualitative methods (Franklin et al., 2010; M. B. Miles & Huberman, 1994), is the importance of establishing validity. Validity can be simply defined as support for your assertions in the form of the categories, connections, and other results. Validity essentially helps the reader of said results feel confident about an author's assertions.

To ensure that these categories have some level of validity, I both adhered to the aforementioned criteria of validity—fit, generality, understanding, and control. In terms of fit, I attempted to ensure that the evidence in the form of utterances and artifacts closely matched the definitions given. Methods like constant comparison helped me refine the categories and associated definitions to ensure that the definitions were robust enough to account for all of the utterances and artifacts grouped together as part of each category. Many times, for example, I would revise the codes, either by recoding the full set of documents, or by going through each set of utterances per code to ensure that the quotes matched the codes. In some cases, for example, I found that after revision, the codes no longer fit the newest variant of the code so I changed the

code accordingly, removed it as an instance of the code, fit it into another code, or created a new code.

In addition, as is common in Grounded Theory, I spent significantly more time contemplating the codes. While I am unable to provide a precise number of hours with which I coded the data, I did look over the utterances and refined the codes at least five times (more information on this is in the Coding Procedure portion of this chapter). I conducted the analysis of these codes over the course of a year and spent a lot of time poring over them to get a better sense of the dataset.

The understanding criterion was likely the least adhered to criterion of the four criteria. Understanding broadly refers to the criteria's accessibility to the public. While I do generally believe that science should be communicated in the most accessible manner regardless of method, it is perhaps arguable that these codes might be less easily understandable. The structure codes in particular, I could imagine, would be quite difficult to understand without context. I did attempt to make it accessible as possible by providing simpler, streamlined examples which do not necessarily occur in the dataset specifically but give the reader an idea of the general target of said categories (especially did so in the criteria section). I also attempted to relate each category to known phenomena, as they are indeed similar, but of course that may also be confusing as they are not exactly the same (e.g., "appeal to authority" is similar to but not the same as the appeal to authority "fallacy").

In an attempt to make the theory more accessible, I also attempted to pare down the number of phenomena observed as I included many more categories than what is reported here. I attempted to carefully look for commonalities among the utterances and artifacts as well as across codes to see if there were any better ways to describe these phenomena in fewer, more

useful categories (of course utility here refers to the control or utility criterion). This was particularly necessary for the criteria category of which there were many types of answers individuals could give. I decided to collapse these codes into large units, though I kept the smaller codes as subcategories. These subcategories were still left distinct as I suspected them to be of use to people; not only were they distinct in the dataset but also they differed in their normativity which I thought practitioners would be keen to keep track of were they to use these categories in some applicative manner.

Regarding the generality criterion, I attempted to generate codes that are of sufficient breadth as to be usable for practitioners but not so context laden that they are taxing and too numerous. All of the major categories here were observed in all participants' utterances so in that sense they are generalizable in that they seem to be components of relatedness reasoning and that the subcategories are different facets or expressions of each component of relatedness reasoning. I initially started the analysis with hundreds of codes which I realized later were somewhat too context bound and during the coding process I was able to reduce to the categories given. Also note that some of those codes were also removed due to scope and not just collapsed into the current codes (the codes removed due to scope were those referring to phenomena that are important for but not necessarily the focus of this study, e.g., affect or epistemology).

The final criterion of utility/control refers to its utility in context. It is still not clear at this point the level to which this criterion is filled, especially given its more exploratory nature and the lack of theory available in the biology education literature. As this study was a bit more inductive and exploratory in nature given the lack of data on the subject in evolution education, I would therefore expect that some elements of it to be less "useful" to practitioners than others but such details will not be clear until the completed analysis is disseminated. For example, the

structure criterion, while informative, seems at this time less impactful than I initially believed—see my brief discussion of my tentative model of relatedness understanding based on these data in Chapter 5. I initially believed that these structures might be more static across contexts and might impact strategy choice. However, the cursory examination of the data suggests this might not be true and individuals might be employing several structures based on strategies and, rather, that other components of relatedness understanding might impact the choice of structures, if they impact them at all. Such a result would suggest that structures be examined more in the lens of an outcome rather than as a precursor to other components of relatedness understanding and perhaps tree-thinking.

That being said, I do not think that the formulation of said categories is without its use. It appears to explain a much of the dataset and therefore might be generalizable. Future quantitative studies can substantiate the extent to which this might be true and in what proportions. In addition, I offer some tangible takeaways that these results could inspire in greater detail in Chapter 5.

Coding Procedure

Primary Description of Procedure

I began the analysis for this study by examining my contact summary sheets in order to orient myself to each of the transcripts. I used them to remind myself of what had transpired and any other major themes I had observed at the time which may or may not be recognized during the analysis. Due to time constraints, most of the analysis was conducted after primary data collection, unlike what is typically prescribed in Grounded Theory (Glaser & Strauss, 1967), though analysis of pilot data was conducted immediately after.

Once all of the data was collected and transcribed, I began the process of coding using MAXQDA. I used the zigzag method to analyze the data wherein I consulted the literature multiple times during the coding process (Bruce, 2007; Creswell & Poth, 2018). This was done in a nonsystematic manner: sometimes I looked for things related to what I saw, sometimes I generally read about related subjects. I also consulted the edited videos of the interactions was I confused about any portion of the transcript.

I conducted five rounds of coding where codes were added, changed, recoded, or modified (Table 5). The first round of coding was the open coding stage of Grounded Theory as prescribed by the original variant of Grounded Theory (Glaser & Strauss, 1967). This stage of coding was the messiest and entailed looking through the transcripts and coding anything that I observed as potential phenomena of interest. These codes were extremely granular in nature and I ultimately arrived at hundreds of codes—these codes were not reported in this study. In other words, this phase was more about noticing items rather than trying to make connections between any and all of the items. Later stages would be where I modified, moved around, and recoded these codes. During this stage, some of these granular codes were early versions of some of the later codes and categories but most were changed during the coding process.

During phase one, I noticed that I really wanted for the sake of clarity to bin the taxonomic groups that people generated which was particularly taxing. I found that this was distracting so in the second stage of coding, I went through all the transcripts again specifically looking to finalize these sets of taxonomic codes so that I would not necessarily be compelled to think about them and increase the cognitive load as I coded the rest of the document. These taxonomic group codes were straightforward and did not require much analysis and they also were not of primary interest in this study. During this stage, I also coded very obvious versions

of the final codes if they did not require a lot of thought. These obvious codes were rearrangements and reformulations of the previously documented open codes. I also began in this phase sorting out some of the ancillary codes that were irrelevant to my research questions as a type of data reduction.

During the third and fourth rounds of coding, I coded all of the primary and secondary codes which did subsequently change on successive rounds of coding. During this time, greater efforts were made to constantly compare codes. It was during the fourth round of coding I that very rigorously compared codes that were already generated to ensure validity. It was during the fourth round of coding that I finalized the arrangement of the primary, secondary, and tertiary code structure. Note that while constant comparison was employed during most of these phases, they were not exhaustive, especially given the number of codes that I had employed, doing so would require approximately 10^{260} formal comparisons given the number of codes. Instead, I primarily focused on those codes that were at greater risk of describing the same phenomena and not being distinct, i.e., typically those that were facets or variants of the same phenomena (e.g., comparing different grouping structures to one another or comparing different grouping criteria to one another). However, greater attention to comparison was paid to the higher-order codes that described all participants given that they were the primary interest of this study.

During the fifth and final round of coding, I reviewed the codes, their associated utterances, and the definitions for validity. While I went through and recoded all of the transcripts, I primarily focused on recoding the structural metaphor codes and the grouping strategy codes which were more taxing and required more inference by me (as opposed to the criteria codes which were much more phenomenological and obvious). It was also during this time I removed all other codes that were not pertinent to this study. This round of coding also

consisted of several passes where I interrogated each major primary category of responses (strategy, criteria, and structure) so that utterances more easily comparable and other large categories did not serve as distractions to the coding process.

For what could be considered a half round of analysis, I additionally contemplated the content of each primary code and looked for evidence, as well as used deductive reasoning to determine the connections between the primary codes. This was employed for the purposes of reporting the theory portion of Grounded Theory as well as to provide readers with a model of the interactions between the primary components of taxonomic reasoning. This, however, was not completed given a lack of time. A hypothesized model of how these factors interact with limited evidence from the dataset, i.e., a very cursory examination of these connections, is provided in Chapter 5, however.

Table 5: Coding Steps

Coding Round	Codes Established/ Coding Step	Targets	Other Notes
1	Open Coding	Everything, including topics irrelevant to primary research questions	Messy, used for organization
2	Axial Coding	Group Membership (Taxon Identity: e.g., "deer") primarily but other primary codes if obvious	Data reduction, improved ease of analysis, constant comparison employed
3	Axial Coding	All primary codes, secondary codes, and some other codes	Constant comparison employed
4	Axial Coding	All primary codes, secondary codes, and some other codes	Constant comparison employed
5	Axial Coding (possibly Selective)	Multiple passes for each primary category and associated subcategories for ease of comparison.	Constant comparison employed, final data reduction step employed (categories irrelevant to research questions deleted), validity addressed more systematically
5.5	Selective Coding	Connections Between Three Primary Categories (Strategies, Criteria, Structure) and Two Ancillary Categories (Membership, Metacognition) interrogated	Not Completed

During this entire process, multiple memos, notes, comments, etc., were compiled in the pertinent MAXQDA file. These were consulted at various moments during the analysis, either to

remind myself of my prior reasoning for inclusion of a code, or reminders to reformulate codes. Additionally, a codebook using features built into MAXQDA was repeatedly refined at every step of the process. The final codebook is provided in the appendix.

Expert and Novice Separation

Experts and novices were not separated during the analysis. All of the data for each research question were analyzed together as experts were suspected to exhibit glimpses of novice structures (Hammann & Nehm, 2020; Smith et al., 1994). To ensure that these groups were therefore comparable, they were analyzed together and subsequent categories were then sorted by experience for comparison. Doing so ensured that the categories observed were comparable as they were created under the same sets of conditions and data and it additionally makes it easier for the reader to identify common response types.

Data Reduction

Typically, the original variant of Grounded Theory requires cataloging, organizing, and coding essentially everything the researcher notices; the original version of Grounded Theory was supposed to be highly inductive, though its results were assumed to arise in the interaction of the researcher (their background, expertise, etc.) with the data. As a result, I initially attempted to document everything that I noticed. However, given the scope of this research, the final categories reported only represent a subset of the codes and categories initially noticed. While each larger category does indeed describe the entire sample of participants and satisfies part of the generalization criterion, the overarching topics that were not pertinent to the research questions I posed were not analyzed in great detail or reported in this dissertation.

For example, given my limited training on the effect of affect on learning as well as my general attentiveness toward emotion (my positionality includes a disorder which includes

greater attention toward emotion), I did notice and note participants' affect during the interview. A few individuals, for example, appeared to be especially worried and attuned to what I was looking for in the interview, repeatedly asking and clarifying if their answer was what I was looking for. However, given the general scope of these studies, I decided not to analyze them or refine them in great detail; I removed these ill-formed codes and categories approximately halfway into the analysis. While the constructs they describe (e.g., affect) likely do have an effect on the particular participant behaviors and thinking (e.g., tree-thinking or relatedness understanding), it is one factor that is likely very generalizable (e.g., low belonging, high anxiety might affect most learning environments) and it likely does not have much specialized contextual effect (e.g., being anxious might not result in a specific type of tree-thinking behavior, rather than a type of behavior that is general to many contexts, such as quickly switching between a range of possibilities, is likely a result). This is not to suggest this topic is not important, however. It is just that it was too much to analyze given the current scope of this dissertation. This process I undertook can be considered a data reduction strategy (M. B. Miles & Huberman, 1994).

The reader might also therefore ask themselves why these were not simply omitted or not coded at all in the beginning of the analysis. I did not do so for two reasons. First, I did so for ease of analysis. Putting irrelevant information in a bin better helped me isolate and differentiate between the categories I was most interested in. I found myself repeatedly asking myself whether something counted as a category and I found it easier to remember to isolate the utterance if I had a formal code or category the utterance could be placed in. Second, while the original variant of Grounded Theory implies that researchers should document everything and create a model that explains all aspects of the phenomena described, Straussian Grounded Theory—which I

employed—could be argued as allowing for such an omission given the assumption that you have a lens with which you are examining the data coming into a study. Finally, Grounded Theory's validity recommendations on making categories functional, i.e., understandability and usability among other things, suggest that additional organization and data reduction occur to help make better sense of the data. This study only aims to make sense of the primary, i.e., most proximal, reasoning of organismal relatedness. While phenomena like metacognitive amendments to such reasoning and affect should also affect this reasoning, again it was not the primary focus of this study. Future studies can examine these factors in greater detail and amend the model I provide to account for their impact on reasoning. In addition, each of those factors, i.e., more proximal reasoning, metacognitive strategies, and affect, each have their own distinct sets of interventions they could potentially recommend. Affect based results could, for example, impact the way an intervention is delivered to students but perhaps not affect the content nearly as strongly.

I also would like to briefly note that I used the contact summary sheets to identify the portions of the transcripts that would be useful for addressing the research questions I posed. Again, note that the original version of this study included other tasks not described in this study. They were used to omit the relevant tasks. They were also used to highlight certain elements of the interviews that I noticed at the time and perhaps did not notice later on during the first pass of the analysis. In other words, they served as additional highlights to reexamine portions of the transcript, though the entire transcript spanning the task relevant to my research questions were examined overall at least five times.

Definitions and Operationalization of Terms

In this section, I briefly discuss the terms that I often refer to in this text and how they may or may not be operationalized in the context of the study or discussed in the results portion of this dissertation.

Defining Evolutionary Relatedness

One important consideration of this study is my formulation of evolutionary relatedness. To my knowledge, no one in the literature has operationalized this definition outside of the scientifically normative definition used in the phylogenetic context. Therefore, I here provide a definition for evolutionary relatedness as well as relatedness in general, in hopes of clarifying what I expect to see in the results. Said results in concert with said definition could lay the groundwork for future operationalization of said term.

In this study I interrogate the ideas surrounding evolutionary relatedness as it pertains to tree-thinking. It follows that I frame evolutionary relatedness or general relatedness ideas as distinct from tree-interpretation. Framing it this way allows for the recognition that everyday ideas about the categorization and organization of organisms may or may not be organized like a tree. I argue that some of these everyday ideas may be used to construct expert mental structures in a modified way (Hammann & Nehm, 2020; Smith et al., 1994). Instructors can exploit these ideas to scaffold students' scientifically normative understanding of relatedness and tree-thinking. Of course, as I will discuss later, these ideas may be complicated by the fact that scientifically normative thought may involve a conflation of the two sets of ideas. Indeed, it is now the goal of systematics to closely tie evolutionary relatedness to categorization. While taxonomic groups were originally created to describe broad similarity between organisms,

current systematics' standards emphasize that organisms be named and classified in a way that reflects evolutionary history.

In this study, I extend the definition of evolutionary relatedness to relatedness in general, where relatedness refers to, as we shall later see in the results of this study, ideas about (A) group membership, (B) grouping strategies, and (C) other ideas that influence (A) and (B). These ideas may be evolutionary or scientifically normative in nature, but they may also not be. All of these ideas, however, may possibly be recruited into learners' understanding of evolutionary relatedness, and subsequently, tree-thinking.

I note that while these ideas were devised ahead of the analysis, they are not strict. As Straussian Grounded Theory assumes a flexible conception of categories, I assumed these ideas may not hold and change. As the reader will observe later, this definition was later expanded to include other ideas that influence the core ideas of group membership and grouping strategies as well as the shape of these grouping structures (Chapters 4 & 5).

Operationalizing Relatedness

Relatedness here refers to the relationships between taxonomic organisms, whatever that relationship might be, which allows for the grouping of organisms beyond simple description. In other words, I avoided, unless implied by the participant, considering descriptive groups as being understood as related by participants. A participant may, for example, refer to "swimming animals" when discussing things that swim, where the topic of said discussion is the grouping factor, i.e., the fact that they swim. Such a group would not necessarily be termed as exemplifying relatedness though they may be related in the fact that organisms implied by such a term are implied to share the same property, i.e., that they swim. This would be different than saying "swimming animals" where I might suggest that beyond the fact that they swim, they are

related in some way to one another. The implication is that they might share some similarity among them, and that swimming is not necessarily the main part of that discussion.

Evolutionary relatedness, therefore, would then refer to the relatedness of organisms based on their evolutionary history. Presumably, this type of relatedness would be normative. This study is intended to examine multiple types of relatedness of which evolutionary relatedness might be one.

Internalization

I often refer to internalization several times within this study. Though it is a term that is often used in sociocultural frameworks (Damianova & Sullivan, 2011), I did not explicitly use a sociocultural framework here for this study. I do, however, recognize elements of this framework in that part of learning can include internalization of external representations, as in, for example, usage of an abacus to perform calculations. When I refer to internalization, I am referring to a less strict usage of the term, simply referring to learning that involves the replication (albeit imperfect) of some external phenomenon, action, representation, or what have you, internally.

Category, Subcategory, Codes

In this study, I often refer to analytical units like categories, subcategories, and codes. Most of the time I am simply referring to different hierarchical levels of codes, whose type of content can vary. I do not strictly identify which is which unlike some methods such as Grounded Theory, given that the distinction between each can be difficult to match to every context. Some methodologies have specific, albeit somewhat ambiguous, criteria as to the types of information that constitute a category, subcategory, and code; therefore, I do not attempt to differentiate between them except note that there are essentially different levels of organization with the lowest levels of codes as in open coding not reported in this study—they were used to

organize, group, and compare higher-order categories. Subcategories of course are any categories that are beneath others. All of these organizations therefore can be considered types of codes given that they can be assigned to a particular utterance or artifact assigned to a participant. Throughout this dissertation I refer to them casually except in this methods section where I explicitly connect the different levels of codes, categories, etc., to the specific analytical method and framework I am drawing upon.

Chapter Summary

In this chapter, I presented the methods by which I conducted the study. I interviewed 29 individuals, 11 of whom were experts, and 18 of whom were novices. These individuals were asked to complete a task where they were asked how organisms were related and asked to describe their understanding of this relatedness. Transcripts and artifacts were analyzed using Grounded Theory over five phases. The frameworks I employed included the expert-novice framework, embodied cognition framework, contextualized learning, representation as a tool, and prototype theory.

Chapter 3 contains unpublished material coauthored with Komperda, Regis and Lo, Stanley. The dissertation author was the primary researcher and author of this chapter.

CHAPTER 4 – RELATEDNESS UNDERSTANDING: STRUCTURE AND STRATEGY

Summary

In the following chapter I first review my research questions and summarize my reasoning for investigating this question. I then briefly summarize my methods and how they addressed each research question. In the *Structural Metaphor* section, I begin with an overview of my results as well as the general coding criteria used to determine each category. I discuss each subcategory in great detail and provide evidence from the interviews for each. I then discuss the other two major categories of responses in the *Grouping Strategies and Grouping Criteria* section. I end with a brief summary of the results I have presented in this chapter.

Introduction

While investigating how to improve phylogenetic tree and evolution education, it became apparent that one area in need of investigation was student understanding of relatedness. The majority of the tree-thinking literature is primarily filled with papers on either the Gestalt (near innate) perception of trees, common tree misconceptions or misinterpretations, or tree-thinking interventions with little support for their effectiveness. While a few tree-thinking instruments have been recently proposed indicating greater scrutiny of and interest in tree-thinking education scholarship (Blacquiere et al., 2020; Jenkins et al., 2022; Kummer et al., 2019; Schramm et al., 2021a), the vast majority of the literature is as described.

While prior work in tree-thinking and tree-interpretation has been quite fruitful, the field neglects some areas in which the broader representational literature has addressed such as multiple representations (Munfaridah et al., 2021), the influence of spatial understanding (Stieff et al., 2012), and cognitive load (Cook, 2006). In an effort to identify ways to help progress the field, I thought critically about phylogenies and their foundational understanding. I realized that

beyond the tree diagram, researchers are organizing taxa based on their inferred evolutionary history. While this insight may not surprise the reader, it illuminated for me just how profoundly the tree-thinking and systematics community is wedded to the tree metaphor when conceptualizing evolutionary history (which some small parts of the community have criticized (Jenner, 2022)), so much so that even I did not initially consider investigating innate understandings of taxonomic groupings. Much of this metaphor is embedded in much of the field's discourse which uses a lot of tree-based and family-tree-based terminology such as "sister group," "branching," and "bushiness." I argue the entrenchment of said methods is one of the reasons very few in the tree-thinking literature have sought to examine everyday (innate or lay) understandings of taxonomic organization that lie outside the confines of a tree metaphor or diagram (Matuk & Uttal, 2020).

In addition, methodologically the literature leaves much to be desired. As many of the studies were conducted when biology education was a developing field, many lack techniques that could help support their assertions with appropriately valid evidence. Much of the misconception literature (save those conducted by Kristy Daniel nee Halverson) provides little to no student utterances to substantiate their assertions (Baum et al., 2005; Meir et al., 2007; Omland et al., 2008). While likely the authors' observations of misconceptions are valid given their experience as educators as well as experience conversing with their peers, evidence would help the field move forward and more precisely document tree-thinking and associated reasoning so that appropriate interventions can be created.

The field is also deficient in terms of framing. As it is relatively new, much of the literature also lacks references to educational and cognitive frameworks that can help the field make sense of their results and ensure that their studies' assumptions remain theoretically

consistent. In particular, the literature has very few examinations of tree-thinking expertise (outside of Catley, Novick, and Daniel), let alone substantial ones, which can function to orient the field toward the presumably normative understanding and thinking that education is attempting to get students to achieve. While there has been discussion of what these goals should be (Halverson & Friedrichsen, 2013) as well as implicit discussion of these goals through the creation of assessments (Blacquiere et al., 2020; Jenkins et al., 2022; Kong et al., 2022; Kummer et al., 2019; Schramm et al., 2021a), systematic examinations of expertise have not been conducted.

I therefore proposed a study which would address two primary research questions which I enumerated in detail in the first chapter. This study involved interviewing 11 professors, post-docs, and educators over Zoom who I cast as "experts." I also interviewed 18 undergraduate students who were taking introductory biology courses who I classified as "novices." These 60 minute clinical interviews involved a series of tree-thinking and related tasks. The first task of categorizing taxa from photographs is the focus of this dissertation. I asked participants if they were familiar with any of the organisms and how they would describe them if they were related. I then asked them to draw what they explained to me for a hypothetical novice listener. I then primarily used Strauss' brand of Grounded Theory to analyze the data and inductively create internally consistent categories and subcategories (Glaser & Strauss, 1967; Strauss & Corbin, 1994). Three of the primary sets of categories form the basis of this results chapter.

Overview of Results

A series of higher order categories emerged from the data some of which resembled what I anticipated (i.e., criteria for grouping organisms) and some of which I did not (i.e., *structural metaphors*) (Table 6). The first of these higher order categories forms the basis of the analysis

Table 6: Summary of the Highest-Order Categories Comprising Relatedness Understanding

Category Name	Category Description	Simple Example	Example Quote
Structural Metaphor	Metaphor describing the shape of the taxonomic groups discussed. Refers to the form of the taxonomic "slots" or "blueprints" the individual creates.	*Vertical Structure*, i.e., Groups are "higher" than others.	"And so, we can use ways, we can use things such as arrows to show that one level is, is under the level above it…"
Grouping Strategies	Strategies that determine which taxa belong in which groups. In other words, methods that determine which "slots" within the structure that taxa should inhabit.	*Isotype*, i.e., normative strategy of using synapomorphies	"And I think at the most basic level that can group a lot of life together and the classification of being alive."
Grouping Criteria	Criteria or evidence used in combination with the grouping strategies to determine which taxa should be grouped together.	*Morphology*, i.e., macroscopic physical traits like unguligrade foot posture	"I think of the number of legs, the seed cycle, I think of buoyancy, I think of whether it has gills or not. Things like that."
Group Membership	Actual groups formed as a result of these processes. Does not necessarily imply structure.	*Camelids*, i.e., group consisting of camel and llama	"We saw two camels too so I should have drawn it bifurcating as well."
Metacognitive Amendments	Modifications or understandings offered as an aside to the aforementioned components of relatedness understanding. Typically, higher order.	*Representations are Imperfect*, i.e., suggests that diagrams are insufficient to fully describe concept.	"You have to like, kind of realize that this phylogeny is a very abstract concept, right? But it's representing like, real organisms that existed billions of years ago, like this guy was there. We theorize that he was there, sometimes we have evidence that he was there."

for my first research question, "In what ways do experts and novices conceive of relatedness?" and to a lesser extent "What are our ideas of relatedness?" In the *Structural Metaphors* section, I discuss the larger category as well as the relevant subcategories, their meaning, evidence for their existence in the dataset, and the potential ramifications of each of the said categories on their own.

The second set of higher order categories formed the basis of my analysis for my second research question, "How do experts and novices group (i.e., classify) organisms?" In the *Grouping Strategies and Grouping Criteria* section I discuss the two larger categories and subsequent subcategories that constitute results pertaining to this question.

In anticipation of the passages I will describe in detail, I also provide a table of the participants in the study to orient the reader to the participants' experience and identity (Table 4). I will describe those as necessary, but this table will also provide a place for the reader's quick reference. Additionally, I remind the reader that participants were encouraged to create their own pseudonyms, so you might find some unusual. I also note that while I will refer to the undergraduate students as "novices" here, I recognize that students will have varying levels of expertise, with some even having near equivalent knowledge to socially acknowledged "experts" here like faculty members. However, given the expert-novice framework I am explicitly drawing upon (discussed in greater detail in Chapter 2 and 3) I am still assuming that on average, the student sample is representing a population, i.e., undergraduate biology students, that has comparatively less normative understanding than experts and that our goal in education is to facilitate and scaffold towards the current community's scientifically normative ideas.

It is worth noting that in this analysis few of the categories were cut and dry, especially those relating to the first research question. Several utterances showed glimpses of words or terms that fit other structural metaphors. Part of this might be explained by the complexity of cognition and how different mental representations can be drawn upon at one time as well as different normative language experience. I still however, built a definition of each metaphor as is consistent with prototype theory in a methodological sense (Lakoff, 1987), by suggesting the relative frequency that certain phenomena (such as terms and qualities) appear and why they might appear to help further delineate groups.

I additionally mention that I will discuss each of these structural metaphors and whether or not they could be fully or partially scientifically normative within evolutionary biology currently. I remind the reader that while they may or may not be currently normative, I will

discuss whether these perceptions should be revised in the context of their utility for supporting formal normative thought.

Structural Metaphors

One of the primary categories of responses observed was *structural metaphors*. These metaphors were organizational in nature and represented general patterns of taxonomic organization harbored by participants that are analogous to blueprints. They referred to the ways in which organisms were ordered or grouped, regardless of taxa included, and were typically spatial in nature.

This category of responses is distinct from other categories observed--which will be discussed in subsequent chapters--in multiple ways. *Grouping strategies* and *grouping criteria* were two of the other major categories observed in the data set and are distinct from structural metaphors (Table 6). Grouping strategies are the methods employed to group taxa while grouping criteria represented the sources of data or evidence used by participants to generate such technical taxonomic groups. Restated simply, grouping criteria are the characteristics or general ideas that participants used paired with a grouping strategy to slot taxa into their hypothetical grouping structure (structural metaphor).

To use an extended metaphor to describe and differentiate between each of these categories, structural metaphors could be analogized to blueprints, i.e., documents representing the spatial framework of the materials to be used in building construction, indicating where such materials should be placed. Materials to be used, therefore, could be analogized to group membership, referring specifically to taxonomic groups. Grouping strategy would therefore equate to the strategies used to determine which materials are used while grouping criteria would refer to those criteria by which those materials are chosen. The actual house that is created from

such processes could therefore be considered a combination of group membership and structural metaphor, i.e., the combination of blueprints/arrangement and actual materials used.

Seven primary categories of structural metaphors were observed in the dataset. These categories were: (A) vertical, (B) flat, (C) family-like, (D) tree-like, (E) anthropocentric, (F) storytelling, and (G) siloed. A summary of the definitions of each of these categories is depicted in Table 7.

Table 8 depicts the frequency at which each of these categories were observed in each group, i.e., number of respondents who were coded as such at least once. Note that subcategories represent percent of overall individuals in the group depicting that subcategory and also that some responses were coded as part of the larger category but not assigned to a subcategory because the utterance did not provide sufficient evidence. Each of these categories were typically observed in isolation within a turn but were also sometimes observed overlapping within a single turn. Many of these categories were spatial in nature, as is common in other basal conceptual metaphors observed in other domains (Lakoff, 2012; Lakoff & Johnson, 1999; Núñez, 2006), but not all of them were.

A detailed discussion of each of these categories will be discussed in subsequent sections as well as implications for the results of each category. I follow with a discussion of general patterns of category observation across many different domains as well as an examination of the patterns of category appearance across experience levels (expert versus novices). I close with a discussion of the implications of these findings.

All of these major categories varied in their scientific normativeness and can therefore considered to be "correct" by varying degrees. This result suggests that not one answer is simply "correct" and the ways that individuals think and subtly modify their understanding of

Table 7: Subcategories of Structural Metaphor

Category or Subcategory Name	Category Description	Scientifically Normative?	Qualities
Vertical	Refers to a structure that has groups stacked on one another. Groups are considered "higher" or "lower" relative to one another. Stronger sense of hierarchy in this type of response.	Yes	Strongly hierarchical, weak time alignment, strongly spatial and iconic
Top-Down	Variant of the vertical structure where the individual begins by talking about larger taxonomic groups and works their way down to the individuals given. Analogous to o*ut-in* and *traditional* tree-like structures.	Yes	
Bottom-Up	Variant of the vertical structure where the individual begins by talking about the smallest taxonomic groups, i.e., those depicted in the task, and works their way upward to larger taxonomic groups. Analogous to *bottom-up* and *twisted lineages* categories.	Yes, strongly	
Flat	Refers to a structure that does not indicate hierarchy as strongly as the vertical sense. Individuals will often speak of "dividing" or "putting together" taxonomic groups indicating nestedness to their understanding.	Yes, strongly	Weakly hierarchical, varying time alignment, somewhat spatial and iconic
Out-In	Variant of flat structure analogous to *top-down* and *traditional* tree-like structures, where individuals begin with larger taxonomic groups and work their way toward smaller taxonomic groups.		
In-Out	Variant of flat structure analogous to *bottom-up* and *twisted lineages* categories, where individuals begin with smaller taxonomic groups and work their way toward larger taxonomic groups.		
Family-Like	Structure that maps on the structure of a family tree. Often this has a stronger sense of time associated with it. Analogizes familial ancestors to organism ancestors and taxonomic groups.	Yes	Strong time alignment, possibly hierarchical, less spatial
Tree-Like	Structure that resembles a phylogenetic tree. Individuals will use tree-related terms (some not scientifically normative ones) that describe groups as the structure of a tree.	Yes, strongly	Strong time alignment is typical, somewhat hierarchical, somewhat spatial.
Twisted Lineages	Refers to a structure that may be more strongly scientifically normative in some parts of systematics. Models hypothesized ancestral lineages (and therefore analogized to groups) as lineages tied or twisted together. Expert responses often will consist of individuals describing lineages going back in time and occurring effectively "together" in the past, implying a conception that still implies some level of separation. Individuals would begin by describing tip-taxa and end with taxonomic groups closer to the root. Analogous to the *bottom-up* and *in-out* subcategories.		
Traditional	Refers to the traditional understanding of trees, more or less. Individuals would either often use tree-specific language like "branched off of" or "leaves" or "splits." Individuals would often begin by talking about groups closer to the root and end with tip taxa. Analogous to the *top-down* and *out-in* subcategories.		
Anthropocentric	Refers to a structure that relates groups spatially and with distance metaphors with a center reference of the individual or humans in general. Essentially considers groups with reference to how closely or distantly they are related to humans. This view can overlap with any of the other structures described in this, either with basic highlighting of humans or implied reorientation of said structure highlighting humans.	No (but can be)	Strongly hierarchical (humans important), weak time alignment, may be spatial.

Table 7: Subcategories of Structural Metaphor, Continued

Category or Subcategory Name	Category Description	Scientifically Normative?	Qualities
Storytelling	Refers to a presumably linear structure indicating a series of events over time. Typically, these stories do not necessarily talk about events that are occurring at similar times (such as potentially two bifurcation events).	No but some argue this is (Jenner, 2022; Matuk & Uttal, 2012)	Strong time alignment, possibly hierarchical, weakly spatial, if at all.
Siloed	Refers to another structural metaphor which can overlap with the aforementioned structures. Refers to having miniature structures instead of one large structure that can accommodate all organisms. Siloed diagrams will consist of separate groupings that may themselves employ one of the aforementioned structures. Typically, this appears to occur in cases where there is a strong indication that the individual is not recognizing that all organisms are related.	Mostly not because everything is related. However, in practice experts do focus on one small clade.	Weak time alignment, hierarchy quality depends on other structure used, weakly spatial.

Table 8: Number and Frequency of Structural Metaphor Responses

Category Name	*# Experts*	*# Novices*	*Total*	*% Experts*	*% Novices*
Structural Metaphor	*11*	*18*	*29*	*100*	*100*
Vertical	6	6	12	55	33
Top-Down	6	3	9	55	16.6
Bottom-Up	0	2	2	0	11
Flat	5	15	20	46	83
In-Out	3	5	8	27	28
Out-In	5	12	17	45	67
Tree-Like	9	7	16	82	39
Traditional – Root to Tips	7	5	12	64	28
Twisted Lineages – Tips to Root	4	2	6	36	11
Anthropocentrism	2	4	6	18	22
Family	3	2	5	27	11
Storytelling	2	6	8	18	33
Siloed	2	9	11	18	50

relatedness might also reflect arguments in the literature and in the systematics and phylogenetics community.

Vertical

Description

The *vertical* structural metaphor describes a hypothetical organizational structure, typically employing spatial terms, that places higher-order taxonomic groups (e.g., families,

phyla, etc.) that include many taxa on top of said structure and lower taxonomic groups (e.g., species, populations, etc.) on the bottom. These higher-order taxa are typically described with language such as "level" or "tier" and often individuals will describe "going up" or "going down," said levels. Individuals also described taxa as being "above" or "below" one another. The use of levels also implies hierarchical value associated with each of these taxa. A hypothesized representation (and therefore approximation) of said structure is pictured in Figure 5.

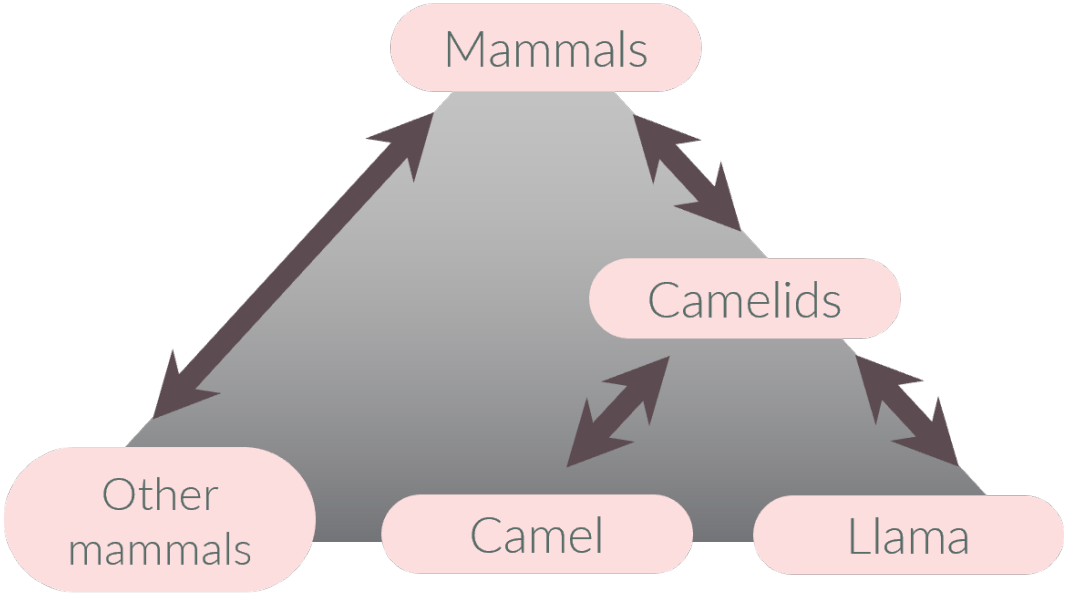


Figure 5: Author's representation of the vertical structural metaphor.

Many participants exhibited this structural metaphor: 12 out of 23 participants were coded as exhibiting this structure at least once, half of which were novices. For example, Queen, an expert, stated, "And so these are **major classifications** and scientists and naturalists have come up with ways to further organize species into **more discrete groups below** these major **levels** of organization." (Note that from this point forward, parts of quotes will be bolded to highlight key phrases and words to help the reader's comprehension.) She stated the groups she made were "below" others, indicating that she likely viewed the groups in a hypothetical vertical structure where some groups were positioned above or below others. One could, however, argue

that these references are not necessarily to a hypothetical structure, but rather are deictic and contextual references to the diagram she created. However, she made this statement before beginning to draw her diagram. In addition, the diagram was self-generated, meaning she chose to place those items in an ordered manner that reflected her understanding or, at a minimum, a manner that was negotiated (normatively) as being vertical. Later after drawing her diagram and being asked to explain it she clarified this idea:

> So these are just representing **different levels of organization**. So there's this the way that all, all organisms are **organized, um, is into different levels**, so we have, where's [UNINTELLIGIBLE] so we have kingdom, oops, **kingdom, phylum, class, order, family, genus, species**. And so these **start with really broad levels of organization and work down** to a specific species and sometimes even a specific subspecies. And so, we can use ways, we **can use things such as arrows to show that one level is, is under the level above it**… Yeah, the way I the way I thought about it was first, I scanned each of the, each of the photos and tried to identify what the species was. And then, once I knew kind of what the species was, then I tried to look at, I guess, from the **top down how I would classify, how I would try to characterize them**. And so I kind of went to the most obvious differences, plants, animals, and **then next level**, you know, like fish, mammals, and **then the next level**.

Here Queen explicitly suggested that she viewed each group as being part of a hypothetical structure with levels. She then tied this idea to normative scientific groups such as "kingdoms," etc. which not all participants tagged with this category did. However, the remark makes it clearer that she was thinking about nesting ranked taxonomic groups.

This structure was also notably sometimes associated with egocentricity, as is common in cognitive metaphors examined through the embodied perspective (Núñez et al., 2012; Núñez & Sweetser, 2006) One participant, Seline, would sometimes describe the taxonomic groups by saying "going up" or "going down" implying a hypothesized figure traveling across the structure while describing it. Novice Seline, for example, stated, "So we start with the broadest animals after we've separated plants from that. And then, to further separate these animals into two bigger, still big but more specific groups, **we can go down** to our second category, which is land

or sea animals... And then **going down on the right side**, you have a further distinction from that." In this particular turn, Seline mentioned "going down" and also placed both speaker and listener in those shoes by stating "we start" and "we've separated" and even more explicitly saying "we go down." It is as if she was calling upon a metaphor of traveling down said structure. I highlight this occurrence given that many conceptual metaphors are egocentric in nature, with, for example, people suggesting features of time, i.e., future and past, as being in relation to the body as indicated by the use of indexical forward and backward spontaneous co-speech gestures (pointing forward for the future, pointing backward for the past, pointing downward for the present) (Núñez et al., 2012; Núñez & Sweetser, 2006).

Top-Down and Bottom-Up

Two variants of the vertical *structural metaphor* were apparent in the dataset: *top-down* and *bottom-up*. Top-down referred to the subset of vertical responses that began with a description of the larger taxonomic groups and worked their way down the structure to smaller taxonomic groups. Individuals exhibiting this subtype, for example, might begin by talking about plants and animals, then further discuss the "levels" beneath that, with animals containing fish and mammals. These structures were restricted to utterances and were not necessarily always observed in every instance of vertical structures.

Gardner, a biology faculty member, provided a detailed example of such thought processes. She described in detail how she imagined the organisms given were arranged:

> **Up at the top** is the grouping that all of the **pictures fit together**, all of them are living things. Some of the characteristics of living things, they all have cells, they all grow, they all reproduce, they have homeostasis, which means they keep things about the same on the inside. They use energy. **They can be subdivided into animals on the left and plants on the right**. We'll look at the animals. Animals can't make their own food; they have to ingest their food to get energy. And they can be subdivided in lots of different ways. But one of the ways is whether they have a backbone or not. All of the ones in the picture have

> backbones. So the invertebrates is hanging out there. **And then they can be subdivided**… fish are cold blooded. They have gills, they're water dwelling, they lay eggs. Mammals are warm blooded, they have lungs, they have live young. And then they're **categorized** showing those, because there are enough pictures you **could keep subdividing beyond this.** But we're not going to.

In this instance, Gardner began her discussion of the groups "up at the top" and ended with smaller taxonomic groups. "Up at the top" she suggested that all "pictures fit together" into a large group consisting of "living things." After first implying that this group comprised a single unit or category in her mind, she said she "subdivided" them, implying that the smaller groups she conceived of were created with the organisms given through a division of the group in question, i.e., all "living things" and they are nested beneath them.

Gardner's diagram was somewhat tree-like (see *Tree-Like* structural metaphor) but also had elements of vertical structure (Figure 6). She placed larger taxonomic groups on top and smaller taxonomic groups on the bottom. Her diagram, however, was not a truly normative tree as the arrows functioned to point toward components of each category and connect taxonomic groups. Her diagram also contained descriptions of each of the categories she made.

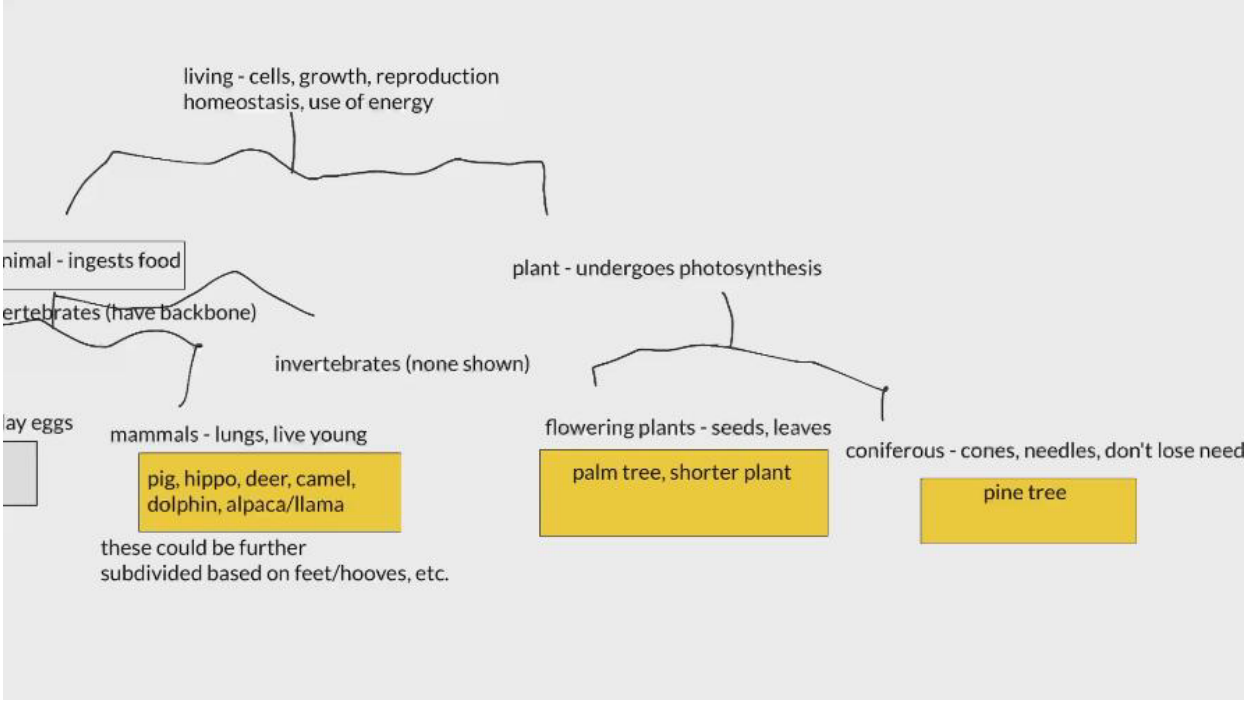


Figure 6: Gardner's relatedness diagram (expert).

Turtle, a novice, uniquely paired his vertical understanding with the use of a traditional table (Figure 7). Tables, of course, imply some verticality with the larger categories placed on the top of them with smaller categories placed below them. Turtle explained his figure as follows: “So every item on the bullet point **goes into the category that's above**. And I guess for the animals, there's two categories within the animal. So **the peak** is a land animal, and a fish and aquatic animal and poultry is just land.” Here Turtle first focused on each of the parts of the table and indicated that the categories above are higher on the hierarchy than those below. Now one could argue that they simply explained how traditional tables work, but he ended by saying that the “peak is a land animal.” The word “peak” implies a vertical iconic image of a mountain and his usage in this context implies that larger taxonomic groups sit at the top of said mountain. This utterance was therefore considered vertical; his decision to start from the top of his table is some evidence for the adoption of a vertical structural metaphor and was reinforced by this clarification that the top group was a “peak.”

Figure 7: Turtle's relatedness diagram (novice).

The opposite variant of the vertical structural metaphor was also observed, referred to here as the *bottom-up* vertical structural metaphor. In this variant, individuals would begin their explanation with the smallest taxonomic groups depicted in the task and work their way upward toward larger taxonomic groups.

Zoey, an expert, stated the following when describing her diagram: "Well, I would say, you know, probably what you would do is you would group them into some groupings, right? All of the plants are more closely related to each other. All of them, mammals are more closely related to each other **and then a bigger group** than that would be like all the animals, right? And then a pretty long time ago, all of the plants shared a common ancestor." Zoey in this example began explaining that certain organisms grouped together in smaller groups and then worked her way to larger groupings of animals and plants, referred to here as the "bigger group."

Diagrams exhibiting the vertical structural metaphor were varied. Sometimes they appeared as above where individuals would use traditional tables like Turtle. Alternatively, they used a diagram similar to a phylogenetic tree where lines would be connecting different taxonomic groups. Note that these are different than tree diagrams in that the nodes denote large taxonomic groups rather than ancestors. Louise, for example, drew a cascading set of taxonomic

groups and she in particular strongly emphasized a sense of hierarchy by adding in numbered tier system on the right side (Figure 8).

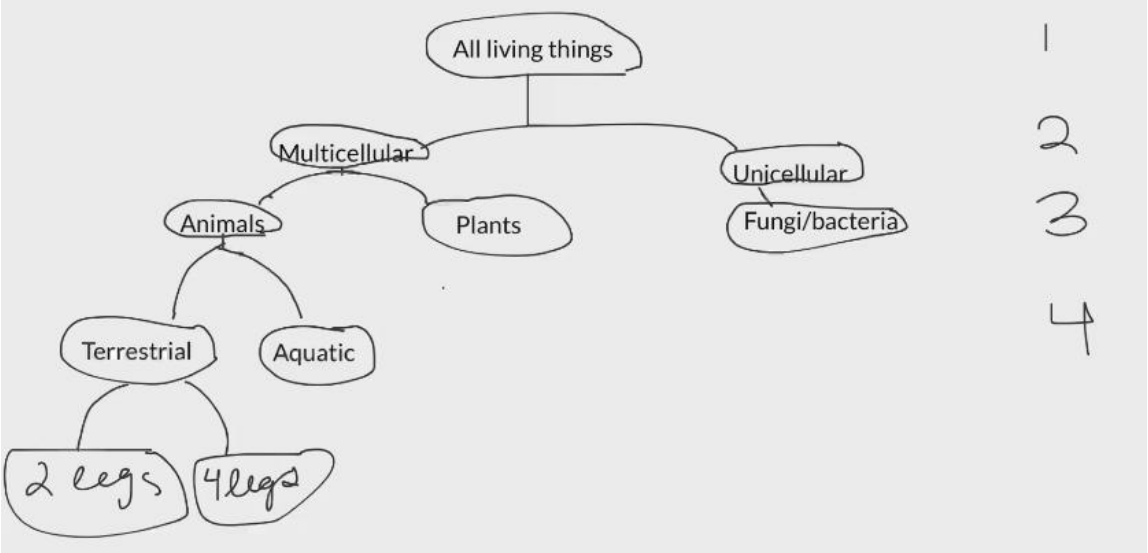


Figure 8: Louise's relatedness diagram (novice).

Scientific Normativeness

This particular structural metaphor could be considered somewhat scientifically normative. Its structure implies that splitting events may have occurred if evolutionary events are tied to the structure, though this possibility was not rigorously probed. These bifurcations leading to lineages could be considered normative, however the content that is represented is different, i.e., that nodes represent larger taxonomic groups. The vertical structural metaphor, however, assumes an extraordinarily strong level of hierarchical orderedness given its nestedness and vertical orientation. Individuals would use terms like “level” and “tier” to indicate implicit vertical structure and hierarchical value. Normative phylogenetic trees do impart some level of hierarchy, but this is potentially contentious as it implies difference of value of real taxonomic groups which is generally considered untrue. Hierarchy in terms of nesting grouping organisms (e.g., kingdom being nested within phyla) is historically scientifically normative; however, hierarchicalness in the sense of assigning value to certain taxonomic groups as being more

important than others may be problematic. The literature discusses the use of the terms “basal” or “primitive” to reference crown groups and old lineages, i.e. presumably unbroken lineages that bifurcated earlier (Yu et al., 2021). Such terminology implies these lineages are of lesser value or may have less complexity which is not necessarily the case—some researchers have argued that this language is inappropriate for describing these groups as a result (Bronzati, 2017; Krell & Cranston, 2004). Such a description also lends itself to a common misconception of teleology and misinterpretation of phylogenetic trees, i.e., that a primary branch or line of evolution exists and any bifurcations along that branch are implied to be offshoots of said “primary” branch (Ainsworth & Saffer, 2013; Halverson, 2011; Novick, Stull, & Catley, 2012; Phillips et al., 2010). This thinking suggests one lineage is favored, when in reality evolutionary mechanisms may treat them variably and it is more correct to consider the splits as being equal when using the phylogenetic tree model.

Experts versus Novices

This structural metaphor was notably prevalent within the dataset, especially among experts, with 55% of experts employing this structure. In contrast, only 33% of novices exhibited this structure. Median expression of structures among experts was 23% and 33% for novices. This suggests that experts were typically more likely to express this structural metaphor more frequently than other metaphors.

This finding is particularly intriguing given my expectation that novices would lean more heavily on this type of structure—indeed we find the opposite that experts more frequently expressing this structure. This might reflect a difference in generation. Historical perspectives on organisms, such as Aristotle's "Great Chain of Being," posited a strict hierarchical view of life (Matuk, 2007; Matuk & Uttal, 2012). Many religions and cultures also traditionally conceive of a

linear hierarchy of organisms that is typically speciesist with humans at the apex (Wilkins, 2022). The relative absence of this structure among novices might reflect a shifting cultural landscape in the United States (Pew Research Center, 2022), possibly characterized by decreasing religiosity, leading to a decline in literal interpretations of religious texts. Alternatively, it might indicate the effectiveness of K-12 education in conveying the concept that humans are part of nature which does not necessarily imply dominionism. This perception might be further associated with growing environmental consciousness in the U.S. The rising visibility of alternative diets such as veganism, plant-based diets, and vegetarianism might underscore a growing recognition of animals' rights and imply a close relation to humans (McCarthy, 2020). This would suggest a less hierarchical view of all organisms, rather than just an altered perception of humans vis-à-vis other organisms.

Flat

Description

The *flat* structural metaphor refers to an organizational structure that lacks the hierarchy of the previously mentioned vertical structure. This structure refers to the grouping of organisms by "putting things together" or "dividing" groups with no indication that these large or smaller taxonomic groups sit at any higher or lower level or that there is any indication of differential value. Individuals would most often group two taxonomic groups at each level and on occasional individuals would place either three or one taxonomic group per group (the latter appearing as "X" group being part of "Y" group such that Y group has no other members present in the given dataset for the task). A hypothetical representation of said structural metaphor is depicted in Figure 9. Also note that my reference to "flat" is itself a spatial representation of this structural

representation, where I am assuming categorized things occur on the same plane. I am essentially using a hierarchical metaphor to indicate that it is non-hierarchical.

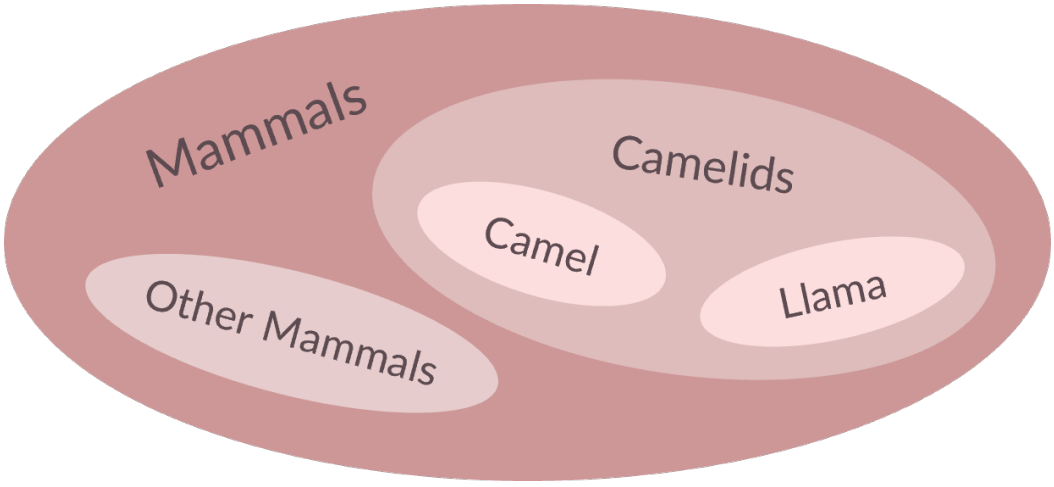


Figure 9: Author's representation of flat structural metaphor.

For example, Maira, a novice, would use the terms "separated" and "within" repeatedly suggesting containment. She stated, "And then after that, it would probably be **separated first** by fish and mammals, and then **within** the mammals, I would guess, sea mammals and land mammals **separate and within** land mammals. I think there's a **little more separation within what's there** as well." All of Maira's responses used "separated" and "within" distractingly frequently, which may be a result of the adoption of this structural metaphor in her understanding of taxonomic organization, or due to a lack of vocabulary available to adequately describe her reasoning. During four extensive turns, she only once used a word or phrase that could imply hierarchy and compared to other individuals who were marked as having vertical understandings of taxonomic organization, that usage was much more infrequent.

These categories I have created, however, are not without nuance as Maira at one point dabbles into hierarchical language, by stating all of the items she was examining should go on the "same level":

> Like fish and mammals, I'm guessing if I'm not mistaken, **separated evolutionarily** from each other before, like... The common ancestor was a lot earlier than the common ancestor, for the sea, mammals in the land mammals. And then the-- within the land, mammals, domesticated and wild aren't like, **biological separations**. It's more like **context environment related separation**. So, after, after land, they, I would **put them all on like an even level**, except for maybe the two deer looking animals and the alpaca which--and the camel, which seemed similar compared to the other like pig.

So here, though there is one instance of "level" in this utterance that coincides more with a vertical structure, the context surrounding this turn still suggested a flat hierarchy. She almost goes out of her way to avoid using any hierarchical terms though they seem to be in her lexicon by that one mention. Comparing her to other participants, she uses the terms "within" and "separate" far more frequently than others, especially those that were not marked as having a flat *structural metaphor*.

Others who exhibited this structure would also similarly use division and combination terms extensively. Professor X, an expert, said the following:

> So I would say group the mammals **together**... So I group the mammals so like, pig and hippo. I **put the camels together** with the... subtended by the other ungulates here so these five here would be... Oh, there's a deer. Sorry the deer--so the two deers. [LAUGHS] The two camels at the outgroup. Hippo, although the whales are there, so the hippo and whale I **put together**, the sperm whale. And then I **put the two sharks together** way on the other side of the vertebrate tree. So the whale shark and you know great white. Oh sorry, so the sperm whale and the bottlenose dolphin **together**.

Professor X not only repeatedly used terms like "within" and "together," but his manner of speech here suggested he was in the process of figuring out what the relationships were. He began by explicitly attempting to pair things rather than directly match the taxa to the knowledge structures he already possessed representing higher level taxonomic groups. It is noteworthy that other experts did not perform these actions (and state those corresponding actions) as often as Professor X, likely because they were not systematists like Professor X. Later in this chapter I discuss in detail subdomain experts' inclination to demonstrate the building process likely

because of the hypothesized audience (i.e., systematists teaching "students" about tree-building), but also because they are aware of the "correctness" of doing so even though they may already have definitive knowledge about the of the current state of the literature regarding normative taxonomic groups.

Responses that were coded as "flat" typically included taxonomic groups in its organizational structure (e.g., mammals contain whales and deer) but "taxa" could also be interpreted as feature-based groups (e.g., marine organisms or organisms that live in the ocean versus terrestrial organisms). In this case, individuals would group organisms into larger groups based on features, but those features that identified such "groups" were not necessarily hierarchically organized and therefore flat. In this case, this structure could be analogized to a Venn-diagram and indeed individuals who exhibited this variant of the flat structural metaphor often did draw Venn-diagrams or diagrams similar to Venn-Diagrams (Figure 10 and Figure 11). This feature-based variant using true Venn diagrams was not considered its own structure given it still lacked hierarchy and verticality. I did not consider the generating of groups based on characteristics that appear in multiple groups structurally that different from those typically coded as flat except for the lack of nestedness (in other words, true Venn diagrams were also considered flat even though they were not completely nested). The only varying element was the feature and category which were described by other higher order codes in this study (see Chapter 6). One could theoretically group a set of organisms with a feature within another group featuring another feature, as is possible with Venn diagrams but this was not observed in the dataset.

Figure 10: Sita's relatedness diagram (expert).

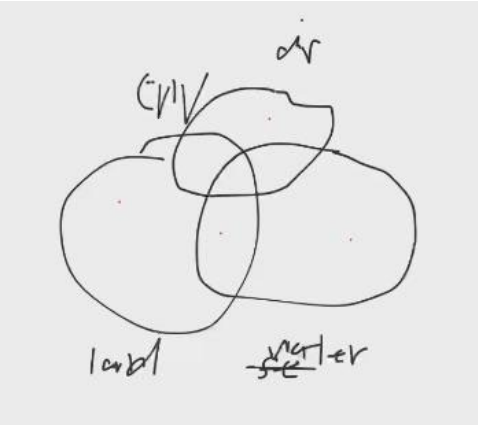


Figure 11: Elisa's relatedness diagram (novice).

Peppermint, a novice, discussed her diagram with me which made her usage of the flat metaphor much clearer.

> *Peppermint*: I'd probably do like the fish and stuff, kind of in more of a web. Like, this is a dolphin, I do like--relating the ones that seemed more related, in a web, and then separately, the trees and stuff would be in a web together, like the palm trees and the bush are all in a web together. And then, the land animals would be in another web. Yeah.
>
> *Tina*: Can you explain to me what you mean by web?
>
> *Peppermint*: Like, all connected to a central point, these would all connect to, I guess that was not that good. So, if the central point was here, this would connect here…

While her understanding of "web" could have been informed by or modulated by her understanding of a food web and/or her understanding of spiders' webs, she drew the diagram in a way that was distinct from a food web indicating that it was a different type of diagram. She

suggested that groups of organisms that are related belong to the same web and are connected to one another by one hypothetical point. There is no indication that different groups belong on different levels or any other terminology that would suggest verticality (consumer levels) in her understanding as is consistent with a conventional food web. Additionally, there was no indication that the lines referred to energy movement. She stated that "the lines are just showing the relationship, like where it belongs" where "it" refers to the taxonomic groups given.

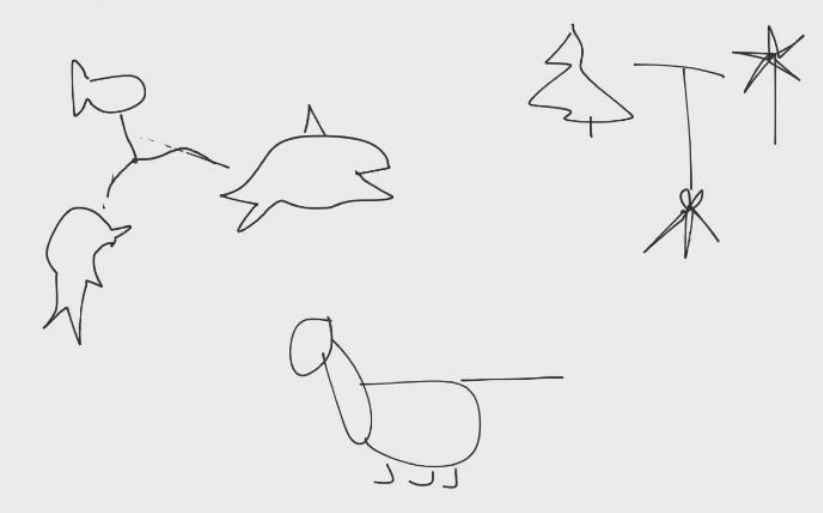


Figure 12: Peppermint's relatedness diagram (novice).

In-Out and Out-In

Much like the vertical structural metaphor, the flat structural metaphor was observed as having two variants: *in-out* and *out-in*. These variants similarly describe the ordering of taxa discussed with *in-out* responses beginning with smaller taxonomic groups and working their way outward towards the larger taxonomic groups. Similarly, the *out-in* variant involved the description of larger taxonomic groups followed by smaller taxonomic groups. "In," therefore, in these terms refers to smaller taxonomic groups and "out" refers to larger taxonomic groups.

Professor X, the expert mentioned in the previous section, exhibited the in-out variant of the flat structural metaphor. He stated he "put together" different taxa visible on the slide into

larger groups: “Hippo, although the whales are there, so the hippo and whale I **put together**, the sperm whale. And then I **put the two sharks together** way on the other side of the vertebrate tree." He began with the hippo and whale and placed them into a group together with the sperm whale then puts the sharks together into likely a shark group. He then implies that these then go into the vertebrate group by ending the clause with “way on the other side of the vertebrate tree.”

Louise, a novice, exhibited the out-in variant in one of her responses. She stated: “Oh, the lines indicate **dividing** it into categories. So all of them have living cells. And it's **divided into** either plants or animals. Plants are **divided again** into whether they're flowering or coniferous. The animals are **divided into** vertebrates, and there are n--invertebrates, and then into mammals and fish.” Louise used a lot of arguably neutral vocabulary here that does not seem to indicate hierarchy. She stated repeatedly that everything is “divided.” I similarly notice that other individuals depicting this type of response would similarly stick to a particular neutral division or categorization word, such as “within,” “separate,” and “divide” and use it over and over again.

Scientific Normativeness

The flat structural metaphor could be considered scientifically normative given that it has a small but lesser degree of hierarchicalness compared to the vertical case. Much like normative phylogenetic trees, the flat structural metaphor appropriately nests organisms into groups based on their relatedness. However, this nestedness is less hierarchical than the vertical case and therefore may be considered even more scientifically normative when compared to normative phylogenetic trees, tree-like structural metaphors, and vertical structural metaphors given the aforementioned framing about not valuing certain taxa over others (Bronzati, 2017; c.f., Diogo et al., 2015; Meisel, 2010). Indeed, certain individuals have recommended in the literature that

concentric circle diagrams be used as an alternative to tree thinking (Gibson & Cooper, 2017; Meisel, 2010; Schramm et al., 2019; Waese et al., 2017).

However, like any form of representation, certain aspects of the natural phenomenon it aims to portray might not be emphasized due to the diagram's limitations. This particular structure minimizes the focus on ancestors; a hypothetical diagram of said structure would not contain any symbol or component depicting the ancestor and instead uses the broader taxonomic group that the ancestor and its descendants belong to as the label for them (though of course this may vary though the latter was not observed in this study). This structure is also limited in the number of taxa it can depict; concentric circle diagrams are likely not used since they can only depict a handful of organisms whereas phylogenetic trees can span several vertically oriented pages (e.g., Tang et al., 2021). In addition, phylogenetic trees give a better sense of relative or absolute time depending on the type of phylogenetic tree used. A flat external representation, such as Venn and near-Venn diagrams (Figure 10 and Figure 11), can be interpreted as having time as an axis projecting through the screen or paper, or it might not be considered at all. In the case of a three-dimensional representation of that structure with nested spheres, time could be considered as radiating inward from the surface. The latter form of said representation, however, was not observed in the study, though that may be due to insufficient probing.

Experts versus Novices

Experts bafflingly were observed using the flat structural metaphor less frequently than novices despite its normative potential. Only 45% of experts exhibited this structural metaphor compared to 83% of novices (Table 8). Twenty-seven percent of individuals from both groups employed the "out-in" structure. However, experts used the "in-out" strategy 45% of the time, in contrast to novices who used it 65% of the time. This difference may speak to a variety of causal

phenomena. I had anticipated that experts would demonstrate the "in-out" structure (and related small-to-large taxonomic group substructures) more often than novices since it is sometimes considered more scientifically normative and serves as a valid alternative metaphor for describing taxonomic relatedness, apart from tree-thinking. However, these findings suggest that this assumption should be reevaluated.

One possible explanation for this discrepancy could be that experts are excessively trained with the tree metaphor, to the extent that other equally valid metaphors might not immediately come to mind. This aligns with literature suggesting that alternative conceptions might exist within a population (e.g., experts) but are only expressed during certain conditions (Smith et al., 1994). This observation additionally aligns with my claim that experts might be overly fixated on the tree metaphor and may not contemplate other methods for representing the data when given a task. Critics may argue that experts perceived me as their target audience for the task and therefore presented their thoughts tailored to a fellow expert; indeed, phylogenetic trees are the most common vehicle for communicating taxonomic relationships in the field. However, experts very often explicitly referenced their instructional experiences or used pedagogically common rhetorical moves, suggesting this was unlikely. I also attempted to safeguard against this by framing the task as though they were explaining to someone without expertise. Some experts even sought clarity on the intended audience for their explanation. Taken together, these suggest that this overrepresentation was not merely a reflection of their intent to communicate to a particular audience but reflective more of their internal understanding of phylogenetic trees and that the dominant understanding they used was the tree-metaphor.

The second factor that could explain this disparity pertains to the cultural differences between the two populations, as briefly mentioned in the previous section about vertical

structural metaphor. We know that cultural shifts can be marked by changing values within the same society. Indeed, target population data must be validated for fit to assessments to correct for changes in culture and other differences over time (American Educational Research Association et al., 2014). Student populations represented by this dataset might not be drawing upon cultural references to strict hierarchies, either in the general sense (hierarchicalness in society, for example) or in the domain-specific sense (taxonomic hierarchy) and are therefore either transferring this lack of hierarchicalness to the taxonomic structure or simply not prompted to link hierarchicalness and taxonomy. This underrepresentation of hierarchical thinking in novices could be due to a change in the societal perceptions of humans' role in nature and changes in adherence to religious perspectives that often emphasize hierarchical views of nature. Although experts are trained to adopt new knowledge structures, evidence from literature on expertise and constructivist theory suggests that novice understandings do influence expert perspectives as new knowledge frameworks are built from preexisting ones (Smith et al., 1994). Experts are older on average and were therefore exposed to different cultural references, so they might be drawing upon these older, more hierarchical understandings of the world. Even though experts may not explicitly or consciously adhere to these cultural understandings and that the normative nuanced understanding might technically not include strong hierarchicalness, these dominant hierarchical cultural references may have bled into experts' understanding of taxonomy and conversely newer perspectives about lack of hierarchy have seeped into students' understanding of taxonomy.

The third potential reason for this disparity centers on the tree diagram itself. While some critics have discussed the drawbacks of thinking about taxa in a hierarchical way such that different value is placed on different taxa others (Bronzati, 2017; c.f., Diogo et al., 2015; Meisel,

2010), the structure of the tree might inadvertently reinforce some of these hierarchical ways of thinking. Many rooted trees in the literature are oriented left-to-right with the root and relative time starting on the left, with tip taxa arranged from top to bottom. Such ladderized presentations have been shown to inadvertently bolster misconceptions about evolutionary directionality and sequentialness (Crisp & Cook, 2005; Meisel, 2010; Omland et al., 2008) where the sequence of events and relative time is assumed to either go down the tips or go down the backbone of the tree. The tree's vertical orientation of tips, which places older taxa or outgroups at the top, is consistent with misconception within the tree-thinking literature where older taxa are deemed "primitive" and "less complex" suggesting these top tip taxa hold less value by being older, value here implied by the presumed freezing of ancestral traits in the taxon discussed. While there is a grain of truth to the complexity assumption on vast macro-evolutionary scales (i.e., the range of trait sets increases over time rather than the mean), it does not hold for shorter time frames. Alternatively, trees are also sometimes depicted top-to-bottom with the root on the top. This arrangement might reinforce the concept of vertical structure shown in this study where the larger taxonomic groups are at the top when the tree is depicted with the root at the top--this is consistent with the well-documented typically Western conceptual metaphor of the top being more valuable (Gottwald et al., 2015; Meier & Robinson, 2004; Ming et al., 2022). Though these conceptions are technically at odds depending on the orientation, both at a minimum suggest that valueness and hierarchy are associated with verticality albeit in different directions. Experts might wrestle with dual misconceptions but still implicitly understand that verticality conveys a quality judgment, like "good vs. bad" or "primitive vs. modern." In short, exposure to these diagrams might reinforce this hierarchical thinking and therefore result in experts reporting the flat structure less often.

Tree-Like

Description

Tree-like structural metaphors were those that either closely or loosely resembled phylogenetic trees. In terms of utterances, tree-structures were typically described with normative jargon which itself is closely related to the tree metaphor. Many individuals would use terms like "branching off" or "splitting" to describe the lineages coming off from nodes. Some individuals would also use normative terms like "leaves" to describe terminal taxa. These descriptions would sometimes be paired with explanations or descriptions of lineages as being further back or forward in time, so this structure could be considered as having some quality of time associated with it. They also occasionally would include an egocentric description of traveling "up" and "down" a lineage. A diagram of said structure is depicted below (Figure 13).

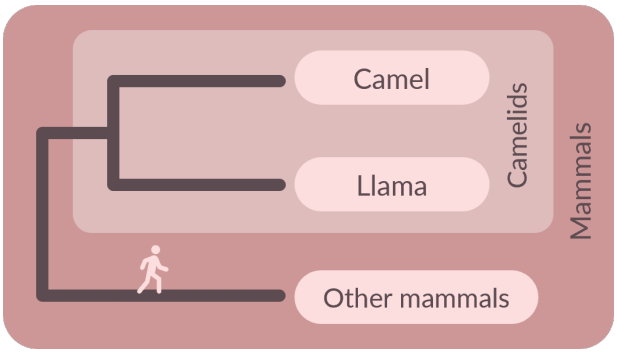


Figure 13: Author's representation of the tree-like structural metaphor.

Expert Mika, for example, repeatedly uses tree-based metaphors to describe the phylogeny; she uses "tree" and "branches" to describe the tree. She also uses terms like "leaves" and the verb "branching." She describes her diagram in great detail (Figure 16):

> The **central part of the tree**--which is this **main trunk**--you can think of that as like life itself. And life has then **branched into** plants and animals. And plants--you know I'm not super familiar with the different… So the **ends of the branches are like the leaves of the tree**, then so they can be—or they can be **leaves of the of the you know hypothetical tree** more they can be they can lead to **further branching.**

Aside from the normative metaphorical terms used, another notable term she uses is "trunk." "Trunk" is not typically used to describe the root of the tree and is non-normative. Her use of a non-normative term that corresponds to the tree-metaphor is stronger evidence for the adoption of a tree metaphor than usage of normative phylogenetic tree-based terms. This was not observed in all instances of those utterances coded as tree-like but is notable because it suggests that this conception goes beyond conventionalized discourse and implies some level of internalization of the concept.

As this structure is unique in that it resembles the most normative discourse and diagrams, special care was taken to ensure that this structure had sufficient evidence. It is possible for some of the normative discourse to be learned without internalization of said structural metaphor. As such, overly normative non-tree related speech was infrequently considered as evidence for having a tree-thinking structure. Individuals were required to state more than what was typical to be marked definitively as tree-thinking. Simply stating oft repeated and memorized scientifically normative aphorism such as "Taxa A is more closely related to B as either of them are to C" was not sufficient to be marked as tree-like. While that statement is very strongly associated with evolutionary biology and tree-thinking, it does not inherently suggest a tree metaphor and is often memorized. While the terms "branching" and "root" are normative, they were considered weaker evidence for tree-based metaphorical thinking than non-normative references to trees. Of course, some verbal uses of metaphor can arise from simple conventionalization and therefore not necessarily reflect the thinking of the individual (Kövecses, 2008) or rather that the conventionalization lacks information that may alternatively be communicated via gesture (Abner et al., 2015). Therefore, simply mentioning "branches" once or twice I did not take to mean that the participant was employing a tree-like

structural metaphor. Individuals would have to go into detail talking about branching and branching events, for example, for them to have been categorized as "tree-like."

Another way that tree-like structures were present in the dataset was simply the combination of a mostly or fully normative tree diagram with some deictic speech and gesture (in this case, pointing with the mouse) referencing their diagram. Some individuals would explicitly state their diagram was a tree and clearly labelled parts of the tree which I also took to be an indication of a tree-like structural metaphor. While those tree components did not themselves indicate tree metaphors in an informal way (i.e., talking about "nodes" rather than a "trunk" or "branches"), I took them to be highly internalized representations of the normative tree.

For example, Professor X, an expert, explained his phylogenetic tree diagram, "The lines are... so there's a **root** here. And these are called **tips**. So the **tips** are the species that we're interested in. And we **connect the tips** with these **lines**." This statement is evidence of his experience with tree instruction and that he is metacognitively aware of the representations' role in the scientific community. He explains that the diagram has certain rules and labels associated with it that he does not expect the listener to know said rules. This indicates that he has internalized some form of this representation (as he has adopted the appropriate discourse) and is aware of its role as are representation (teaching parts of it indicates recognition that it isn't inherently understandable and that it is a model with different components that symbolize different things). While he does not give other informal indications of (i.e., non-standard description of a tree) a tree-metaphor, I took this to be an example of internalized tree-metaphor, especially when taken together with the knowledge that he is a systematist and therefore an expert.

Other evidence for this category came in the form of the metaphor's repackaging. I asked the interviewees to describe how they would present their understanding of relatedness to a hypothetical lay audience. One expert, Mika, said, "All right, so you know we referred to **tree of life** right like a **phylogenetic tree** and if I were to describe it to a friend, you see an architect, or something. I **wouldn't say phylogenetic tree**, I would say, you know, this is a **tree of life**, this is a diagram that shows an idea of the relationships among these different groups of organisms." Here Mika relabels a phylogenetic tree as a "tree of life" which is a normative term in evolutionary biology–though less frequently used–referring to the idealized goal of evolutionary biology of reconstructing a comprehensive history of the organisms on earth (Hinchliff et al., 2015). However, the fact that she retains the tree metaphor, i.e., using a different version of the metaphor, to communicate that idea to the lay audience, suggests that this structural metaphor is somewhat stable--stability here meaning that her model of organismal groupings is not just transient and not highly based on context.

In addition to utterances, participant generated diagrams were also used to determine the appearance of some metaphorical thinking. Some individuals, for example, would discuss and create very scientifically normative phylogenetic trees with lots of normative notation (ex. Figure 14) rather than basic trees (Figure 15). Those that drew them very obviously were marked as such—however it is important to note that though those individuals may have used trees to describe and draw them, they may not use those structural metaphors in isolation (see Chapter 5).

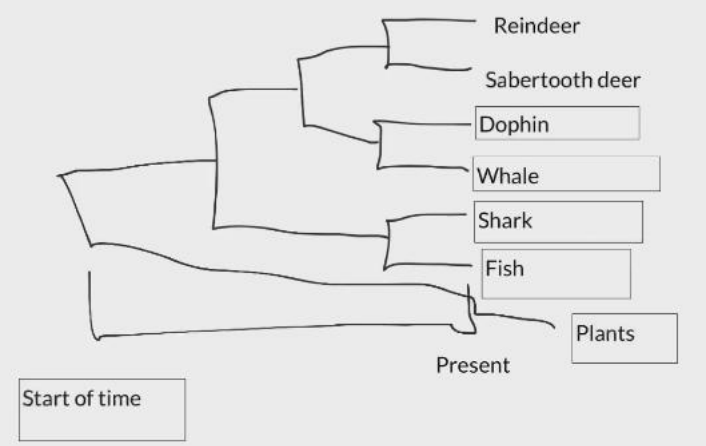


Figure 14: Ella's relatedness diagram (novice).

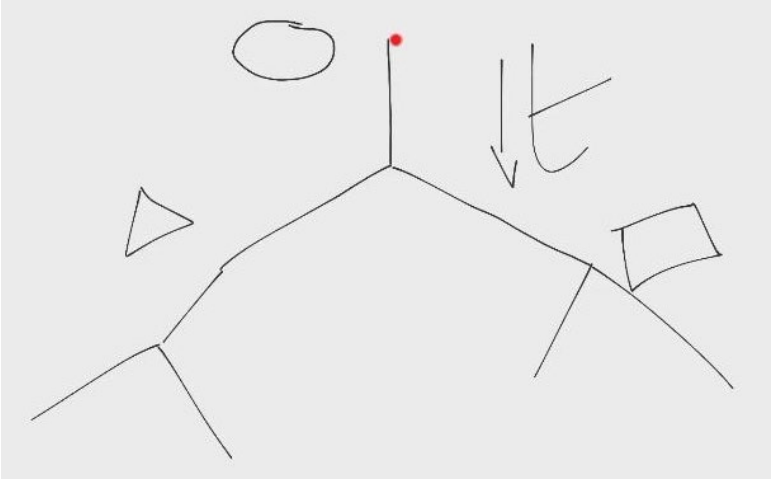


Figure 15: Witch's relatedness diagram (expert).

Also considered were utterances and pointing (i.e., with technology) gestures that were deictic in nature and highly contextual. Deictic gestures and demonstrative language indicate high correspondence between the environment (in this case the representation) and the subject the speaker is talking about (Diessel & Coventry, 2020). Student Ella states while explaining her diagram (Figure 14):

> So if here is the present, I'll draw that right **here**… So this common ancestor, for instance, I'm going to find my pointer. Because this common ancestor may have **evolved prior to this one**, or especially this one, most likely **evolved prior to this one**, given the fact that sharks were around pre-dinosaurs… So, this line **may be all the way back here**, where this line may be exactly where it is. Or even **further up**, considering the start time, it probably is right there.

Ella uses many deictic terms such as “here,” “right there,” and “further up” while using the software’s pointing ability to connect her normative ideas about what her phylogenetic tree represents. This suggests that her understanding is indeed tied to normative trees. While she did not use the terms “branching” or “splitting” in this particular turn, the associated normative concepts such as common ancestor were pointed out on the diagram suggesting some level of internalization of the normative tree-metaphor.

A notable difference was observed in the diagrams of those using vertical structural metaphors in contrast to tree-thinking metaphors. Those diagrams that were coded as exhibiting vertical structural metaphor were sometimes tree-like, in that they showed larger taxonomic groups with lines or arrows branching from them to show subgroups. In normative trees the larger taxonomic groups would be represented by the entire monophyletic clade instead and sometimes would be shown as parts of a timeline, brackets on the tree, color coding, or circles around the respective clade (Figure 16). These vertical diagrams essentially depict the larger taxonomic groups at the nodes rather than entire clades (Figure 8).

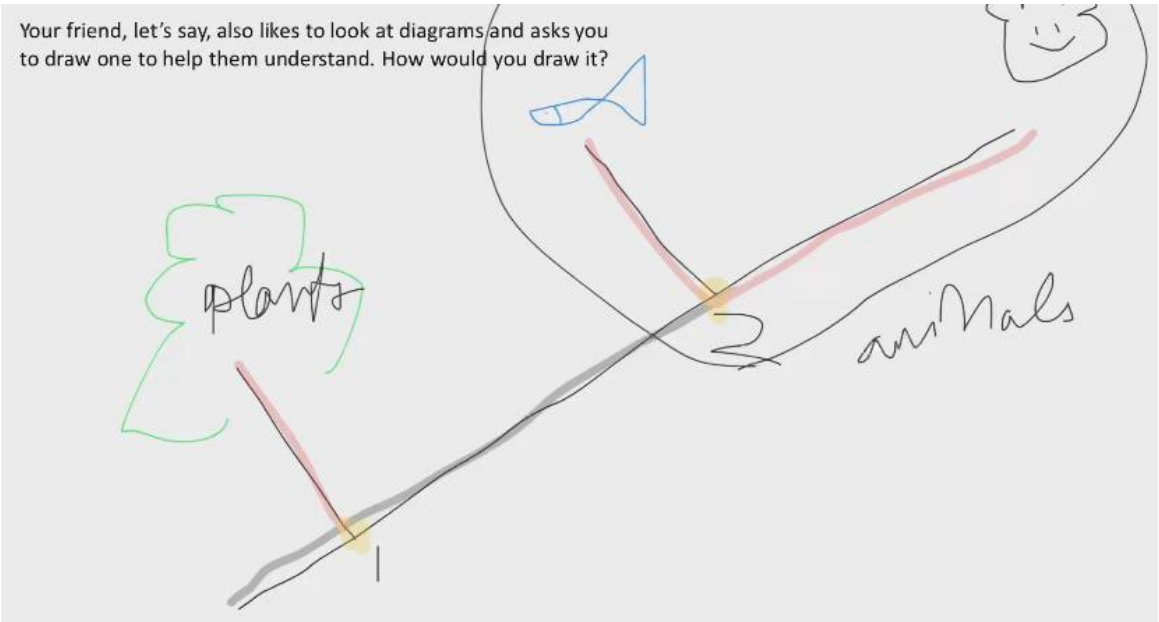


Figure 16: Mika's relatedness diagram (expert).

Twisted Lineages and Traditional

A less common structural metaphor was the *twisted lineages* structural metaphor which I differentiate from the aforementioned understanding of phylogenetic trees which I broadly consider the traditional variant. Experts that exhibited the tree-like structural metaphor were 82% compared to 39% of novices. Experts that exhibited the traditional tree-like structure (root to tip) were 64%, while 36% exhibited twisted lineages. Novices that exhibited the traditional subcategory were 28%, while only 11% exhibited the twisted lineages. *Twisted lineages* was an uncommon structure I consider a subset of the tree-like metaphor, especially given the parallelism observed in the vertical and flat structural metaphors.

Twisted lineages refers to what I would oversimplify as a sort of "backwards" tree. Trees in this context are discussed from the present to the past, rather than the traditional past to present. This structure correlates to a less frequently normative discourse used in systematics, often describing the lineages as being two distinct entities and that they "join" in the past. This normative perspective foregrounds the fact that real extant data is (typically) used to infer ancestors and relationships in order to better elucidate the drawbacks of implementing such methods. Stronger modern taxist and cladist perspectives dismiss any speculation on or reconstruction of ancestors. They cast the practice as being particularly tricky exercises, especially given that extant data likely leaves out splits or oversimplifies bifurcation events (Cunningham et al., 1998; Finarelli & Goswami, 2013), or that the practice is philosophically inconsistent with the practice of science (Fitzhugh, 2012). These perspectives also often consider lineages as being wound together rather than being turned into a single unit, again, which is likely a product of population genetics given that each population would have at some point been a portion of the initial population. An example of such thinking is illustrated pedagogically to

students in an activity created by Daniel (née Halverson, 2010) where twisted pipe cleaners depict how ancestral lineages conceptually consist of their daughter lineages.

In this particular dataset, one variant of this substructure was implied as tracing two extant (or focal) taxa backward in time until they meet. The two lineages were implied to theoretically co-occur together backward in time, but simultaneously thought of as mixing together. Experts only expressed this particular interpretation of a tree. Two of the four experts exhibiting this substructure exhibited this variant; however, their other discourse suggested that they also simultaneously called upon and likely harbored other representations of evolution and speciation. Novices, however, did exhibit backtracing but did not seem to exhibit callbacks to this discourse. This is unsurprising given that in my personal experience this model of tree-thinking has usually only been expressed at a graduate level and rarely at the undergraduate level.

Practically, if not directly explained, this structure would often be referenced as lineages "sharing," evolutionary history with the individual describing said phenomena starting from present to the past. "Shared evolutionary history" tends to be discourse often associated with the particular normative understanding I just mentioned. Witch (an expert), for example, stated the following:

> So, I am thinking about living things as they evolve, some populations, some species, some groups.... Man, this is really hard to explain. Yeah! Things that have a **shared evolutionary history could be thought of as things along a shared branch of a tree**. And so that's how in biology, we tend to think about **evolutionary history is like branching trees with things that are more like, sharing more branches,** and more history compared to things that are less like.

Here Witch is suggesting that taxa that are more alike share more evolutionary history. This particular example may be weaker evidence for said statement, especially as I do not have empirical evidence that this type of discourse in the field co-occurs with this conception.

However, it is clear that the shared evolutionary history she is referring to involves some binding of the two lineages in question.

Similarly, Zoey, an expert, mentions that two lineages were bound together:

> So, I would say this down here **represents the most recent common ancestor** of all life. So that's like some **microbial ancestor**, you know, think like microbial soup. Something like, you know, hydrothermal vents in the ocean or something like that, hot springs, whatever. So I say, the first group of organisms that diverged from that most recent common ancestor... here. So that **population separated, and it's separated** into what we know today to be the Bacteria. And then, like along this lineage, that lineage **stayed pretty much together** until this ancestor--and again that population separated into two.

Here Zoey is forgoing the formal tips to root explanation of this model but does exhibit a remnant of said model. She suggests that after the population "separated," a statement which is initially consistent with some of the flat, non-hierarchical *structural metaphors*, she then states that the lineage "stayed together." That implies a normative model like what I previously described, where lineages are thought of as being bound together in the past and separate now, almost like separate co-occurring lineages.

Students were observed tracing back as well, though without this additional attention paid to how the two lineages are modeled as co-occurring. They simply described the tree from the tips to the root. Ella, a novice, stated:

> So, fish such as these **three would be in the same group** and probably have very similar genetics. So, they're... So, different genetics would be very similar and higher percentage. Then, I probably **relate back to our whale, dolphin, and see that they are very similar for the same reason, just in a different group**, and then those **both are most likely related to our hippo friends**. And our deer and camels are very similar as well as the alpaca just in different ways. The pig and hippo are relatively similar. But in general, all of our mammals. These guys are very similar in genetics. And then, they are then closely related to our fish. **But as the lineages go on, our plants will be related to each other, more so than our other living animals.** But all of them share surprisingly amount of similar genetic composition.

Here Ella began with smaller groups and slowly worked her way up towards the larger taxonomic groups while embedded in a broader discussion about ancestry and shared genetic composition. When pointing at her diagram she stated, "Because obviously, **these lines** will not be. But if I talked about it like this in terms of genetics, I could say that these **two have a more recent common ancestor that two lines diverge from one another**." "These lines" refers to her own tree diagram (Figure 14). Note that this was a variant of novice response lacking the typical rote discourse about shared evolutionary history but still has traces of or perhaps may be a precursor mental representation of the aforementioned "shared history" understanding. Ella suggested that the two lineages were one and they have equal weight rather than saying something akin to "branching off" which is a common misconception of tree-thinking (implies one of the branches is a primary branch). Other novices were observed back-tracing but lacked this additional detail.

Scientific Normativeness

This structural metaphor, of course, is the most scientifically normative of the set of structural metaphors observed in the dataset. They could be considered similar or the same as phylogenetic trees in that they show that bifurcation assumption, i.e., that a single lineage results in two taxa. (Some individuals predictably included polytomies though the reasoning for their inclusion was not clear.) The meaning assigned to such bifurcations also implicitly appeared the same as in trees, i.e., that they refer to splitting events and/or most recent common ancestors (MRCA). This differs from the vertical structural metaphors which have the larger taxonomic groups defined at each "level" – ancestors are at each level and the ancestors and their descendants are assumed to form the taxonomic group in question (though this particular feature of tree-like structural metaphor was not rigorously probed in every instance of its appearance).

This structure, when presented in an external representation, much like normative phylogenetic trees, always appeared bottom-up, left-right, or top-down and never right-left.

I also consider this structural metaphor the most formally scientifically normative structure observed. It is "more formal" in that it is most commonly used in the field, though the level to which it "best" represents the phenomena they are trying to depict is debatable. The twisted lineages variant is, however, likely somewhat more normative given the prior discussion about the focus on back-tracing by population geneticists and systematists (Rosenberg & Nordborg, 2002).

Experts versus Novices

The tree-like structural metaphor was significantly more evident among experts than novices. Approximately 82% of experts displayed evidence of a tree-like structural metaphor, compared to 39% of novices (Table 8). This finding aligns with expectations, as this metaphor resonates with the normative views held by most evolutionary biologists and therefore should be the most prevalent in the expert group.

While many experts exhibited this structure, it is noteworthy that a considerable proportion of novices did so as well. Though most were introductory biology students, not all were first-year students. The elevated adoption of this metaphor could result from advanced schooling or the influence of instructors with advanced science pedagogical training. While this percentage is still lower than that observed in experts, it still suggests that some novices on some level internalized the understanding of tree-thinking. However, the quality of their responses should be considered; some used a lot of jargon imprecisely which suggests they lacked deep comprehension of the subject. However, some frameworks consider jargon use as a critical part of learning, i.e., the learning of the ways of speaking in a domain (Hamnell-Pamment, 2023), so

one can reasonably argue that some level of learning has occurred, though quantifying the exact extent is challenging given the study's qualitative nature.

Another area of interest is the fact that experts did not achieve 100% expression of said structure (and this may also explain some novice underexpression) even though it is likely the most normative. Several factors might explain this finding. Some experts may have felt the tree-like structure was not ideal for the given task. Some individuals were indeed observed adjusting their explanation strategy to cater to their audience, explicitly indicating that they were thinking about and selecting particular rhetorical moves to suit the intended audience. In addition, we know from constructivist perspectives that expert knowledge is constructed from existing conceptions—they can be amended, coordinated, analogized, etc., in some way (Smith et al., 1994). These experts may have constructed their understanding of tree thinking without the context of a tree and therefore solely reported non-tree related structures. Experts are required to communicate with other experts using trees, so we would expect to see them use tree-like structures. However, it may be that their understanding of organismal groupings is best served by modified versions of whatever structure they have which more or less correlates to a tree-like structure and since the task did not ask for communicating with a peer, they referred to their constituent thinking. This finding may be modulated by experience as some of the individuals who did not display any evidence of tree-like structural metaphor were not systematists—non-systematists use and generate trees less often and therefore they might be less likely to use it. Indeed, unpublished data from a pilot of another study I conducted where individuals were asked to draw evolution suggested that experts from different disciplines tend to draw diagrams that highlight aspects of evolution more pertinent to their own subdiscipline; for example, population biologists tended to draw evolution from a micro perspective showing individuals in a population

whereas macroevolutionary biologists tended to draw phylogenetic trees (Marcroft, 2019). However, it is also notable that one systematist chose to create a diagram that did not resemble a tree, indicating that experience and research interest or focus does not solely determine representation choice.

Another significant observation is the varied usage of these constructs, illustrated by the two structural subtypes: tip-to-root (twisted lineages) and root-to-tip (traditional). Both experts and novices predominantly used root-to-tip over tip-to-root tree-like structures. Given that many trees are oriented from left to right, and considering the students are US English speakers, they might be more accustomed to trees reading from left to right and translate that behavior to any other trees they might encounter. Reading order has been shown to influence tree interpretation and creation (Novick, Stull, & Catley, 2012), but no studies have been conducted to substantiate my particular hypothesis.

These variants may also differ in their usage because of what they represent cognitively, discussed in greater detail at the end of the chapter. The order of reporting might be due to the inclination to report set structures as categories. Depending on the learning theory used to frame this question, one can argue that they have already previously learned about some taxonomic group, either immediately before the reporting or much earlier, and can therefore be considered preset. In the constructivist perspective, this might be more closely aligned with assimilation, where someone determines that the new stimuli can be filed under a preexisting knowledge structure rather than change the structure in question (Piaget, as cited in Glasersfeld, 1995; Pugh, 2017). It would follow that experts would likely have more information on what taxa belong in what group and therefore exhibit the root-to-tip (and related large-to-small variants of structural metaphors) variant more often. Similarly, novices, being less experienced with this dataset

should therefore report the small-to-large variants as they presumably would be less likely to have information about these taxa. This, however, was not observed; novices exhibited root-to-tip more often than tip-to-root tree-like structures. This discrepancy could result from numerous factors, such as insufficient probing or the ambiguous nature of novice responses. Some novice answers seemed to reflect a superficial understanding, making it challenging to categorize them definitively. Further insights into this finding, couched in the global task orientation framework, are discussed later in this chapter.

Anthropocentrism

Description

In the anthropocentric category of responses, individuals referred to taxonomic structures that were often self-referential or referencing humans broadly. This could present itself as a preoccupation with what taxonomic groups were more or less closely related to humans, and very few efforts were made to compare other taxonomic groups to one another. Therefore, this conception could be considered as somewhat egocentric with a hypothesized external representation of said structure with a human or the individual at the center. It was very often also observed in conjunction with other structural metaphors, but more intrinsically so. Table 9 depicts structural metaphors that overlapped in only the transcripts, showing that anthropocentrism most often co-occurred with the tree-like, storytelling, and flat structural metaphors. A diagram of said structure is depicted below; this particular variant shows anthropocentrism paired with the tree-like structure (Figure 17).

Most examples of anthropocentric structural metaphors were typically used rhetorically. Victoria, a novice, simply liked referencing many aspects of her argument and diagram to humans, “And then we have mammals, which are usually **what we see** running around,

Table 9: Number of Individuals Displaying Overlapping Structural Metaphors (Transcripts Only)

Category	Siloed	Storytelling	Tree-Like	Family	Vertical	Flat	Anthropocentrism
Siloed	—	—	—	—	—	—	—
Storytelling	0	—	—	—	—	—	—
Tree-Like	2	0	—	—	—	—	—
Family	0	0	5	—	—	—	—
Vertical	0	0	3	0	—	—	—
Flat	2	1	3	0	0	—	—
Anthropocentrism	0	3	7	0	0	4	—
SUM	4	4	20	5	3	10	14

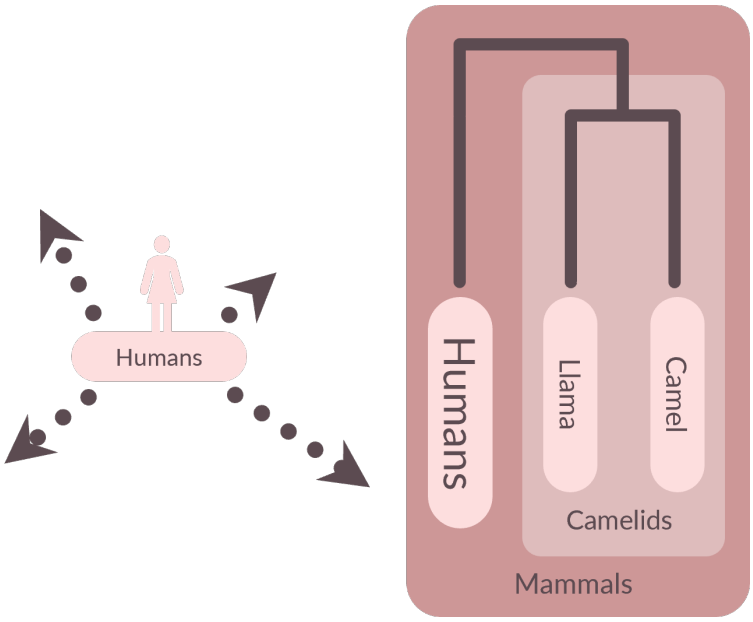


Figure 17: Author's representation of the anthropocentric structural metaphor.

and **you probably have a pet** who is a mammal." The wording "you probably have a pet" suggests she thought about a hypothetical audience she was speaking to—she did not say this to the interviewer assuming that they had a pet. She also attempts to relate her explanation to the listeners' experiences by saying that it is "what we see." Her statement implied that comparison or reference to humans might make the explanation more relatable and in doing so she centered her structural metaphor around humans.

Jean, an expert, was much like many of the other respondents when using the anthropocentric structural metaphor. She explicitly thought about her audience and implied that the usage and centering of her example with humans was a useful rhetorical tool depending on the audience she is speaking to. She stated, "Yeah, mostly **depending on who my audience** was, I'd probably stick to those really **recognizable large subgroups**. **So, as humans**, I think we tend to **recognize mammals of the things that are more closely related to us**. So, I would include, you know, the deer and the pig um, and the llama and the camel, like all of those are mammals." She then went on at length about how to create a human based diagram and how to focus her explanation on humans: "Maybe I would have a **phylogeny that is, of different populations of humans** or different individuals… Present day at the tips, and then I might say something like, oh, let's say **humans and then another closely related group**."

Novices were also notably thinking about the rhetorical use of anthropocentric explanations. Bhakthi, a novice, was very clearly drawing upon her experiences as an undergraduate teaching assistant when thinking about her response:

> As a student taking a class, this is **what I teach my tutees in an intro bio class**. You're usually looking at, like, what's the most recent common ancestor, **this guy and this guy**, and then you literally just take your highlighter, and you trace back and they trace back and wherever is the like note, like intersection where they meet first, that is the common ancestor… If you're at this period of time, then **this guy** is going to be new to you. But if you're all the way over here, then **Mr. Over-Here** is not going to look so new anymore.

Bhakthi repeatedly anthropomorphized taxa using "this guy" in an implied attempt to make the explanation more relatable. She later also referred to different taxa as "Mr." in another form of anthropomorphism. She implies that she is mimicking her typical explanation using these anthropocentric or anthropomorphic explanations for her undergraduate tutees.

Maira, a novice, reflected on when she was younger, suggesting that she thought about organisms with reference to humans, in this case, the common domesticated pets like dogs and cats.

> I think probably maybe **when I was little**, I would, I think, I probably thought, like, **dogs and cats related because they were like house pets**. I'm, I know now like, they're not as closely related as, maybe dogs and wolves, and like cats and lions and tigers. But I think while, when I was younger, I thought that **things that lived in similar contexts, like maybe everything that lived in the ocean and everything that was a pet** were more related than what they actually are now.

Maira said that she grouped dogs and cats together because they were house pets. Indeed, even though she implied that its proximity to humans is not a normative reason to group organisms (implied here as living with humans), she later spends a considerable amount of time talking about groups being domesticated or not whether or not the taxa in question were domesticated. She says:

> So specifically, I guess for like the land animals, I can tell that they are all mammals. I'm not entirely sure what two of them are, but I would guess that **neither of them are domesticated animals**. So, I was just kind of keeping them **separated**. And then also adding the hippo to that because that one is **also not domesticated**. And then the pig is kind of separate from everything else, because I think that's the only one where this, like, particular kind of pig, **I don't think is found in the wild**? But I think alpacas and camels are **both found in the wild**, **even though I guess like people have them domesticated**.

Maira mentioned domestication very frequently and even beyond the turn shown here was overly concerned with whether or not organisms are domesticated. Although she acknowledged that classifying organisms based on house pet status, a form of domestication, is not standard practice, she still used it as a grouping criterion and did not appear to make the connection to her prior statement about erroneously considering human related usage as a grouping criterion. This might indicate a particularly strong reliance on anthropocentric metaphors. This, however, is unsurprising. The embodied perspective argues that our understanding of the world is rooted in our bodily experiences and interactions, leading to some human-centered interpretations of

natural phenomena (e.g., Núñez et al., 2012; Núñez & Sweetser, 2006). Furthermore, constructivist perspectives argue our knowledge is built upon prior knowledge (Piaget, as cited in Glasersfeld, 1995; Pugh, 2017), so human-centric and/or early experiential viewpoints are therefore understandable and expected (Minstrell, 1982).

Scientific Normativeness

The anthropocentrism structural metaphor is mostly non-normative. Variants of this structure depending on the adjoining or overlapping overall structural metaphor can affect its "correctness." Those using a generally "correct" tree structure that simply highlights humans would be more correct though its highlighting of humans would not be considered so. There is discourse in evolutionary biology regarding the tree of life and how humans consist of an extremely tiny fraction of the diversity in of life and that humans are not necessarily more "evolved" than other organisms; in that sense, it can be thought of as non-normative. While we do have qualities that are unique to our species, such as enhanced cognition and tool use, other organisms outperform us in other areas, such as endurance running, respiration in water, etc. Other structures overlapping with it, such as a vertical structure, might be even more non-normative as they more greatly emphasize the relative importance of humans versus other taxa.

The perception of anthropocentrism as a valid pedagogical tool does seem to be somewhat normative, however, with several experts and novices seemingly using it as such. The true utility of this structure in terms of its pedagogical use will be discussed further at the end of this chapter.

Experts versus Novices

The anthropomorphic structural metaphor bears similarities to the family structural metaphor in its infrequent observation and its rhetorical application. However, the clarity in its

rhetorical usage is not as pronounced as with the family structure. Anthropomorphic structural metaphors were identified in roughly 18% of experts and 22% of novices. Despite these percentages appearing closely matched, it might not be startling to see a slightly higher usage among novices. Language that draws parallels between natural phenomena and human behavior is well documented and discussed in science education literature (Kallery & Psillos, 2004; Legare et al., 2013; Ochs et al., 1996; Varella, 2018) and is also discussed at length in the broader biology literature (Wynne, 2006).

Its manifestation among experts aligns with Smith et al.'s (1994) proposition that experts might resort to novice structures under certain pressures, either as a heuristic or as a foundational knowledge structure that has evolved to encapsulate the complexities of expertise in the domain. However, the exact extent and nature to which this may be occurring is unclear given the data available in this study. Alternatively, experts might be utilizing this structural metaphor as previously mentioned. Experts might believe that the anthropomorphic structural metaphor could aid in scaffolding novices' (likely they were imagining students) grasp of tree thinking and therefore anthropocentrism was used as a pedagogical tool. This could be a valuable initial approach for supporting relatedness understanding and tree-thinking, especially since anthropomorphism aligns with observations from the embodied cognition literature. Within this framework, body-based metaphors frequently play a role in forging new knowledge (Lakoff, 2012). Some cultures, for instance, employ egocentric time metaphors where individuals will map time on a horizontal axis with the future behind and the past ahead (Núñez & Sweetser, 2006). It is essential to note, though, that while some educational researchers in the embodied cognition domain propose that these body-based metaphors might bolster the assimilation of normative interpretations of phenomena, this viewpoint is not universally held among cognitive

scientists. They contend that while these concepts originate from fundamental bodily understandings, it does not inherently imply that kinesthetic learning is more efficient (R. Núñez, personal communication, 2019).

Another interesting occurrence was the extensive employment of this strategy by a novice, suggesting she could be classified as an expert in her own right. She disclosed during her interview that she was an undergraduate teaching assistant, guiding students in comprehending phylogenetic trees. This usage might indicate that some experts, through their experiences, recognize this explanatory method's potential benefits for students, whatever their experience level may be; presumably, she still has less than those formally recognized as experts in this study. However, while it might have assisted the experts and this one student or may be perceived as being helpful, it might not necessarily serve all students well, so additional investigation into this assumption is needed.

Family

Description

Another type of structural metaphor observed in the data set was that of family structure. The family structural metaphor consisted of equating patterns of evolutionary relatedness to that of family structure, and sometimes explicitly comparing patterns of taxonomic relatedness to human family trees or pedigrees. Individuals would often explicitly state, for example, that an ancestral node would approximate a specific proximal ancestor such as a grandparent. Much like the anthropocentric structural metaphor, it was most often used rhetorically. A diagram of said structure is depicted in Figure 18.

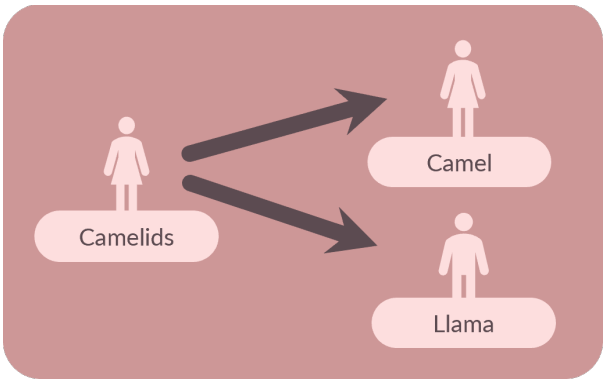


Figure 18: Author's representation of the family structural metaphor.

Zoey, an expert, used the family metaphor to describe her ideas about taxonomic relatedness:

> And so I probably would say something like, you know, if you have like what you would call a **really distant relative**. That's **somebody for whom you know, you shared an ancestor** a long time ago. Maybe like a **great, great, great, great, great, great uncle** or something like that. That would be like a **distant relative**. Whereas, you know, if you have **a sibling or whatever you share parents and that's a more recent relative**. So I would say it's like everything on here that has fur or hair, those are all mammals. And **so they shared the most recent common ancestor pretty recently**.

Here Zoey discussed how taxonomic relationships resemble familial relationships. Aside from jargon associated with tree-like structures or tree-thinking that could be considered part of this structure, i.e., ancestry related terminology of "most recent common ancestor," she also uses the non-standard examples of "sibling" and "uncle" to illustrate these relationships. She also emphasizes its perceived rhetorical value for the hypothetical novice audience by saying, "I would probably say something," which implies a particular context she is imagining when generating this response.

Novices also exhibited this structural metaphor. Bhakthi, after mentioning that she was an undergraduate teaching assistant, equated these evolutionary relationships to familial ones in this quote:

> So for example, each of these intersections are called nodes. And those refer to common ancestors when like, **for example, me and my brother would share a node with like, our parents or something like that**. But that's not exactly how evolution works. It's more like you have one organism that gives rise evolutionarily, and it evolves into different forms. So each of these lines represents an evolutionary history.

Bhakthi here exhibits some overlap with the tree-like structure and equates her tree-like thinking to family structure. She analogizes nodes as ancestors to parents and sister groups to the relatedness of siblings.

Given the embeddedness of this family metaphor within evolutionary biology and tree-thinking and the propensity for experts to be familiar with this jargon and use it casually, very obviously normative evolutionary biology terms were not considered part of this family structural metaphor. For example, biologists often refer to clades stemming from the same node as "sister groups." Work in the cognitive sciences suggests that linguistically, one reason people may use metaphorical language is simply due to conventionalism. In other words, individuals would use such a metaphorical term because that's the way they have heard everyone using it, not necessarily because they actually understand said concept using such a metaphor. Responses that went beyond the typical family terminology of phylogenetic trees (e.g., "ancestors" and "sister groups") were therefore considered as representing a family structure. What was also taken into consideration was distinct comparisons of this way of thinking about taxonomic structure to other structures. Drawing parallels between two structures suggested that individuals on some level believe the two to be distinct.

No individuals gave any indication that their external diagram belonged to this metaphor. While the diagrams may have been related to this metaphor, there was no evidence to suggest this was the case such as family explanations paired with deictic gestures towards diagram parts or labels to suggest that aspects of it equated to parentage or any other familial relationships.

Scientific Normativeness

The family structural metaphor may be considered somewhat normative in the sense that trees and family trees or pedigrees are related in that they both show variants of ancestry, but the latter can show exchange of genetic material through parentage, whereas the former would show one "parent" as an ancestor. The education literature does sometimes suggest using lay knowledge of family trees or knowledge of pedigrees from classes as a means for supporting understanding of phylogenetic trees (Gregory, 2008; Meisel, 2010; Schramm et al., 2019). Indeed, many appearances of the family structural metaphors also were rhetorical in nature suggesting that there was some recognition of the family metaphor from a pedagogical perspective. The actual pedagogical value of said structure will be discussed in the discussion portion of this chapter.

Experts versus Novices

This structural metaphor was less prevalent compared to others. Both experts and novices infrequently employed this metaphor, with 27% of experts and 11% of novices doing so. The higher adoption rate among experts may be attributed to its specific usage. Many experts leveraged the family structure rhetorically, seemingly to clarify or scaffold the concepts for the audience rather than reflecting their preferred understanding or viewpoint. Existing literature posits that family trees could serve as scaffolds for phylogenetic trees (Gregory, 2008; Meisel, 2010; Schramm et al., 2019). Given the literature on the topic and experts tendency to reference them for explanatory reasons in this dataset, one can reasonably surmise that some experts attribute some pedagogical value to this type of structure. Some novices also employed this structure, seemingly for the same reasons. This collective behavior suggests that while both groups recognize the metaphor's potential pedagogical utility, its infrequent appearance outside

of a pedagogical context as well as its decreased appearance among novices might indicate that its educational effectiveness is overestimated. If we follow the constructivist notion that new knowledge builds upon prior knowledge, and accept the educational value of this rhetorical tool, it implies that the learner possesses some familiarity with family trees. Historically, family trees held a lot of significance, especially among affluent naturalists (barring exceptions like Alfred Russell Wallace), which is arguably may be why many early trees looked like ancestry trees (for examples see O'Hara, 1997) besides their common use of the tree metaphor. Indeed, some of the sociological literature surrounding the history of biology and genetics suggests that the preoccupation with lineages and "social degradation" in the upper classes was a precursor of the geneticization and hard heredity social movements (Paul, 1995). This therefore merits inquiry into the metaphor's contemporary relevance versus its past significance, a question that could be explored in a future study using cohorts of varying ages.

On a related note, the appearance of this structure begs the question of course order as institutions tend to vary in their requirements. Institutions where students are required or permitted to take genetics before delving into phylogenetics would therefore have students with prior exposure to formal pedigrees. Furthermore, examining the correlation between familiarity with pedigrees and tree interpretation proficiency could be enlightening, assuming tree interpretation aligns with the use of a tree-like taxonomic organization. While pedigrees are a type of family tree, they are highly conventionalized and have their own sets of strict rules for formatting—such rules might impact the way individuals understand relatedness and interpret phylogenetic trees.

Storytelling

Description

The storytelling structural metaphor is an uncommon but unique structure that arose during analysis. Individuals depicting this structure would often exhibit traces of what could be filed under many other structures and were initially a cause for confusion. Other individuals who exhibited multiple *structural metaphors* were clear in their use of multiple in that there usually was an obvious delineation and switch to another metaphor type, even within the same turn. The type of language used would change, and often this was connected to the natural progression of the conversation as they, for example, switched from an explanation with no representation to incorporating their generated diagram into their explanation. At some point during the analysis, it became clear to me that the dominant feature of this individual's response was its storytelling structure in addition to a few other unique features and I therefore referred to it as the *storytelling* structural metaphor.

The storytelling structure, therefore, was characterized by highly mixed usage of language that may either indicate a mix of structures or simply that which did not tend towards one particular other structural metaphor. It was also very often characterized as a focus on the iconic aspects of ancestry and relatedness and was typically chronological in nature. This structure could therefore be described as a richly descriptive sequence of events illustrated by the speaker. Individuals would draw and/or speak in great detail about the setting and qualities of the taxa in question and if other structures were co-occurring, such as tree-like structure, aspects of those sequences of events were compared or aligned to the second structure. As in traditional stories, individuals would also imply a sequence of causal events or elucidate the causal factors of said story. I illustrate two types of said structure in Figure 19; both can be considered linear,

but one incorporates some branching when tree-like structures were involved. Both, however, highly emphasized the rich sequence of events occurring in their explanation.

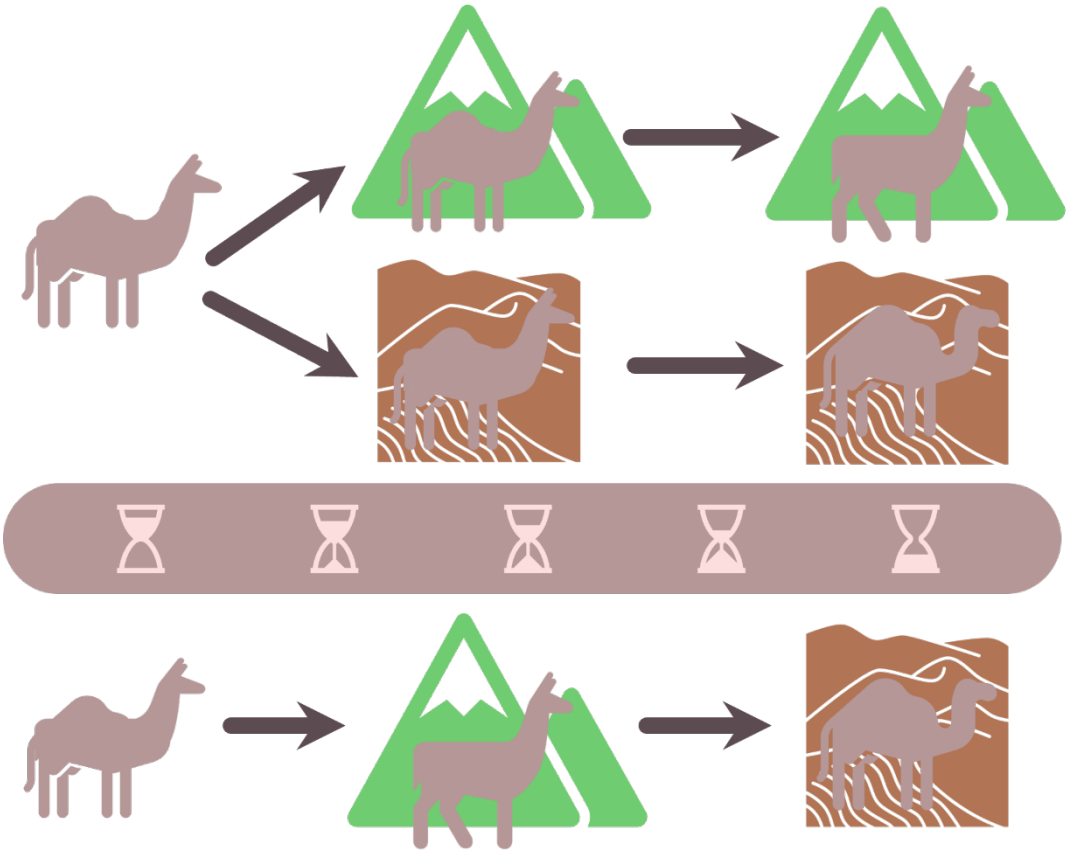


Figure 19: Author's representation of the storytelling structural metaphor.

This particular structure placed taxa in a linear orientation, likely with a horizontal quality (as in other Western conceptions of time; (L. K. Miles et al., 2010). These individuals also would speak about a series of traits and taxa appearing as a sequence of events, with the story appearing in chronological order and their depictions of it would generally go from left to right, top to bottom (Figure 20). This could therefore be considered as a telling of the history of the organisms starting from the ancestral taxa and then would proceed forward in time, indicating when the splits occurred that gave rise to the larger group containing the taxa in question. The types of words used by the individuals would vary a lot and include language found in many of the aforementioned structural metaphors, with the exception of the obvious

vertical positioning language ("above" or "level"). They would suggest that smaller taxonomic groups were contained within others which would suggest non-hierarchicalness, though this structure had a strong aspect of time associated with it given its emphasis on chronological order. They would therefore use phrases that subtly suggested a sequence of events such as "and then," "starting with," and "ended up."

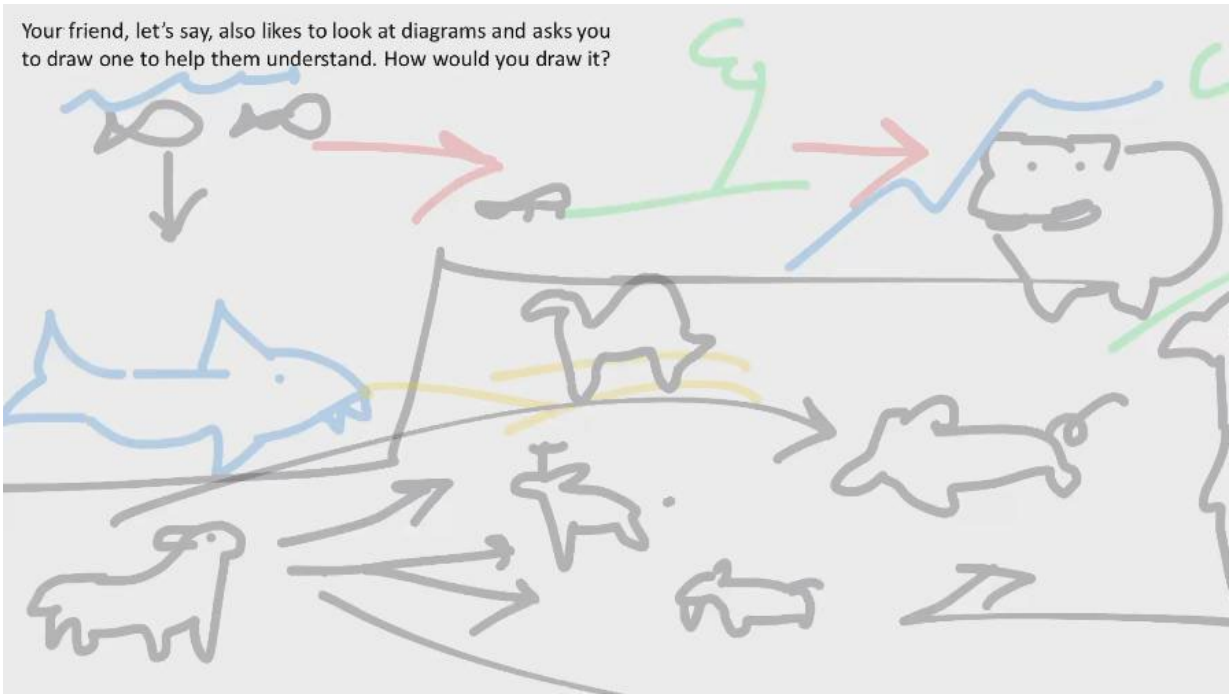


Figure 20: John's relatedness diagram (novice).

Individuals exhibiting this metaphor visually would also typically include a lot of iconic imagery, i.e., pictures, in their diagrams. These would either exhibit states of environments, or a series of events (Figure 20). Indeed, the shape of these diagrams would imply linearity and would often progress from left to right, top to bottom. They also sometimes incorporated descriptions of time. Alternatively, they would depict the current setting of the taxa in question (Figure 21) as it exists or existed in a moment in time but would be associated with multiple timeframes in their utterances.

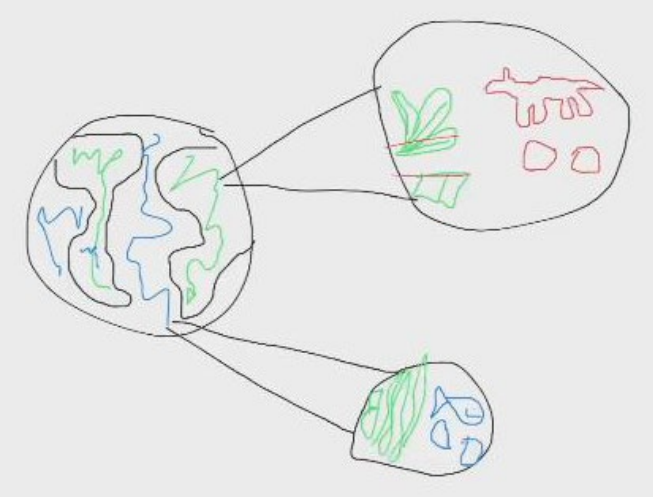


Figure 21: Victoria's relatedness diagram (novice).

John, a novice, focused more on what was happening at microscopic time scales in addition to describing what happened on macroscopic time scales. She stated (she identified as female though she chose a traditionally male pseudonym):

> Okay so **way back when** there was some type of mammal with four legs and fur. And **as time went** by um - I want to copy. Okay, um **his environment changed** to a sandy desert. And so, he **went from** a four-legged mammal **to a** camel because **he had to adapt, his body changed**. Um, while **another group just stayed in the forest**, and they became deers. The males have antlers. And then another, sub section of those deers grew fangs because they needed it for their unintelligible and another one was domesticated into pigs… And then, oh, the alpaca, that one. Um. What of the **environment they had to adapt to** they grew long necks and a lot of fur. Yeah. So, they all came from a common mammal. Um, **but due to time and change in habitats they became animals** that we now have that we now have: the camel, deer, pig, and the mammal, alpaca.

Note that she mentioned a series of events using terms like “as time went by” and described change over time. She incorporated a bit of anthropocentrism and treated the ancestral species as an individual saying that “he had to adapt” and “change” as if she were describing the protagonist of a story. While John mentioned some splitting which is consistent with the tree-like metaphor, it is still presented somewhat linearly and as a series of events. She later emphasizes linearity explicitly saying:

> So, to me it means like **along the evolution line.** They all kind of have it the same way. But either something could have happened like maybe you know, some of them died off, or maybe like some sort of you know, **change on land like maybe there was like a volcano earthquake or something separated like the pack.** Um, and so that, second group because of the separation, evolved into um a separate species because it had um - it needed **different things to survive** compared to um the one that ended up evolving into

John here seems to think about it as a linear sequence of events with her statement "along the evolution line" and, like a story, illustrates some potentially causal factors for said events. Note that while speciation and things happening at the microlevel were mentioned now and then by other participants, those were much more inconsistent. Storytelling structural metaphor was always accompanied by linearity, movement through time, and a description of the environment plus causal factors.

Two experts also exhibited this structural metaphor. Witch, for example, was more macroevolutionary in her approach and uses more precise terminology but similarly describes what occurred as a series of events. She stated:

> But, based on the animals that are here ah ancestrally so, yeah. **Hundreds of millions of years ago,** the common ancestor of all of these **animals started out** in the water. It **started out looking** like a fishy looking thing. Some of these animals, you know, think of something like, yeah so basically the fish that are on this slide, an animal that was in the water, was the **common ancestor** of all of these animals on this slide. **And then, there was their transition** hundreds of millions of years ago that then led to some of the animals staying in water and some other animals developing like limbs. So, the **fins of the fish basically over many millions of years developed** into the same appendages that are now the legs of the pig, or the legs of the hippo etc. So yeah, these animals, all of these animals have a common ancestry and we still see, you know those **same body parts are functionally different now** as in you know fins are used for swimming, limbs are used for locking but in terms of evolution, all of those instant links had the same ancestry.

Note that she described a sequence of events with more detail at smaller timescales, though broadly describing a sequence of events that occur over longer time scales, much like the aforementioned novices. She begins her story with a mention of timing "hundreds of millions of

years ago" resembling the opening of many archetypal stories "a long time ago." She indicates different time points, both absolute and relative, of the story with phrases like "started out," "transitioned," "developed," and "now." She adds detail about the various parts of the organisms and their environment gives us a sense of the setting and variables at play causing these changes to occur.

Scientific Normativeness

This structural metaphor is not considered scientifically normative. Assuming this structure is analogous to the fields' discourse on "stories," which I argue it is, some biologists likely frown upon it. Specifically, many evolutionary biologists often reject the use of "storytelling" or "just so stories" in the field to describe evolutionary phenomena or events and therefore relatedness (Andrews et al., 2002; Jenner, 2022). The argument against such "stories" implicitly is a rejection of a supposed reductivist approach to phylogenetic estimation or related methods; storytelling is implied to overemphasize the casual attribution of fewer factors as well as suggest that the researcher is attempting to overfit the data to a neat "story." It has been also alternatively construed as framing biology in an adaptationist sense in that these stories may underemphasize the importance of stochasticity, the appearance of spandrels, and other factors that are not directly causal (depending on your definition of causality) (Andrews et al., 2002). Though this is the dominant discourse in the field, there are some in the field who are also trained in the philosophy of biology who now argue that storytelling is a valid form of thinking in evolutionary biology as long as it is recognized as a heuristic (Figueredo & Berry, 2002; Jenner, 2022) and/or that trees (and their associated assumptions) themselves may be inappropriate models of evolutionary history (Bapteste et al., 2009; Doolittle & Bapteste, 2007; Podani, 2019; Velasco, 2012).

Experts versus Novices

Storytelling here does resemble different phenomena in the literature. Storytelling was so named because of its resemblance (or equivalence) to storytelling or story structure in the literature. There is extensive scholarship on the topic given that it is an innate way of organizing or thinking about the world and the storytelling practice is intrinsic to the human experience (Polletta et al., 2011; Schachtner, 2020).

Traditional story structure typically includes the existence of a protagonist who overcomes obstacles (Polletta et al., 2011; Schachtner, 2020). Similarly, story can also be considered a chronological narrative of how a series of causal factors instigate a series of events whose telling is intended to communicate some meaning. In this study, the storytelling structure refers to both formulations of the term as the stories function to show how certain factors caused certain events to occur (e.g., allopatric speciation caused by geological changes) but also that a protagonist encountered obstacles that it overcame (e.g., a lineage "had to evolve" an adaptation to overcome a challenge in the new environment). My definition of storytelling based on the dataset also includes a rich description of the context and settings of the proposed "protagonists" i.e., the taxa in question.

The storytelling structural metaphor, though generally uncommon, was more prevalent among novices. While 20% of experts displayed this structure, 33% of novices did so. The difference in pervasiveness might speak to a variety of factors relating to the innateness of storytelling and/or humanity's rich oral and written storytelling tradition. Assuming storytelling is an innate way of organizing and thinking about natural and social phenomena, it follows that novices would engage in it more often as we would expect novices to rely on lay forms of thinking when reasoning about new phenomena. As my definition of storytelling also includes

the rich description of the setting and circumstances of said "story," it may also speak to the cognitive phenomenon of novices attending to detail (non-normatively so) that are not related to the main theme or pattern at hand. Novices, for example, when asked to draw neurons will tend to include intricate details which are more illustrative of the naturalistic shape and form of the neuron. This is in contrast to experts who will often draw a simplified diagram emphasizing the parts most relevant to understanding its shape and function (i.e., a simple axon and simple dendrites rather than the irregular non-rounded appearance of cell walls) (Hay et al., 2013). Taken together, these two phenomena may explain the storytelling construct in this study; storytelling then represents a combination of innate story creation behavior as well as a potential lack of focus on the details that are most pertinent to the "story" in addition to details relevant to the causal parts of the story. They may also explain why novices were observed exhibiting this behavior more than experts. More work, however, is needed to determine the exact extent to which both of these behaviors are contributing to each.

However, given the aforementioned argument, experts should be expected to be much less likely to use this structure, yet they still do. This may be because of the previously mentioned constructivist perspective that nuanced understandings of phenomena are constructed from novice understandings and the embodied perspective suggests that they are constructed from basal embodied understandings of the world. If this is the case, traces of storytelling should be apparent in expert responses and indeed we see that here. It may be then that experts are calling upon this innate understanding as it is part of their innate understanding, or perhaps, similarly they might implicitly recognize that its innateness might be pedagogically valuable for scaffolding novices' understanding of said phenomena toward normative reasoning. Reasons for

this choice, however, were not probed in this study nor traces of said pedagogical utility were observed in the dataset.

Additionally, the usage of storytelling may confoundingly be a byproduct of expertise. Experts often have a good understanding of how phenomena are related to one another at different scales; for example, physicists can often transition between the representation types (Kohl & Finkelstein, 2008) while chemists also presumably similarly do so between representation types described by Johnstone's triangle and related models (Taber, 2013). Similarly, we would expect biologists to flexibly be able to relate the macroevolutionary scales to microevolutionary scales though I know of no study that has confirmed this. If that is the case, then we'd expect biologists to be able to elaborate what the hypothesized circumstances of the changes observed in the tree were and doing so might therefore seem story-like as the microlevel description of what is occurring would naturally be more story-like (group of protagonists, setting, etc.) as they are at an individual level; a human or group of humans in a story are individuals as are the organism or organisms in the story of the taxonomic history of some clade.

This perspective on the ability to deftly maneuver between scales also begs the question of whether we should amend our definitions of expertise in the field to also include this ability as well as the ways we might support and measure such an ability. It would be necessary to understand if, for example, constant reference to the natural context at a microevolutionary scale would help students make similar connections to macroevolutionary scales or if more direct support is required for this understanding.

The difference in usage and the potential innateness of such a structure suggests more investigation is required. The presence of this structure in novices might therefore make this structure a candidate for measurement to their proximity to normative understanding if it is

assumed as "default" in addition to a potential target for scaffolding normative understanding. However, studies would have to be conducted to see how the propensity to use such a structure changes with increasing experience. These novices, for example, did not fully exhibit this metaphor and presumably have had a lot of training and interest already in biology from high school which may mask the level to which this type of thinking may or may not be ubiquitous. If the latter is true, then this structure may prove more useful as a measurement tool in K-12 biology education and not undergraduate biology education.

In addition, given some biologists' insistence that we should not abandon storytelling as a cognitive tool for understanding natural phenomena (Jenner, 2022), additional investigation into the ways in which storytelling thinking can support normative thought is necessary. The ways in which storytelling can be buttressed or compared to other forms of thought would also have to be addressed to prevent drawbacks this model could support such as teleological thinking or sequential linear reasoning (Schramm & Schmiemann, 2019). Indeed, some similar work in evolution education suggest something similar where they suggest that teleological thinking is widespread and potentially innate but that experts constantly work on said understanding and add nuance to this thinking to address the aspects of teleological thinking that are not applicable (Hammann & Nehm, 2020).

Siloed

Description

The last category I discuss in this chapter is the *siloed* structural metaphor. The *siloed* structural metaphor was observed in combination with some of the aforementioned structural metaphors (Figure 22). It refers to the tendency for participants to consider the groups discussed in the task as being separate enough to not warrant talking about. In other words, individuals

would often talk about each group in isolation, not indicating that these groups were connected to one another, either in their diagram or in their utterances. In contrast, everyone who was not categorized as exhibiting the siloed metaphor either indicated directly or indirectly that all organisms given were related to one another. Those not exhibiting the siloed category may have indicated some level of differentiation between groups, such as suggesting that plants are vastly different from animals, however, they still either verbally or diagrammatically linked each of them together in a larger structure. Those exhibiting the siloed metaphor did not indicate that "everything is related." Their diagrams would show that taxa were not connected to one another at larger taxonomic group levels.

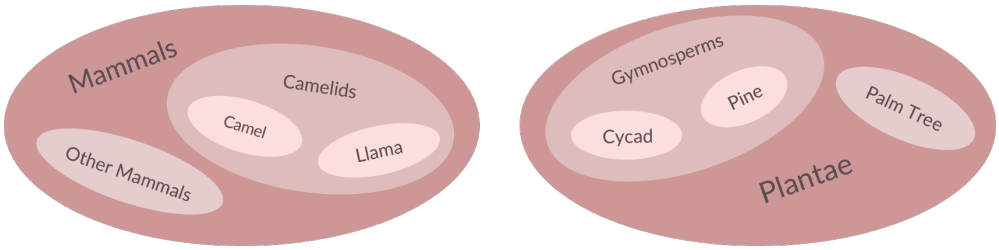


Figure 22: Author's representation of the siloed structural metaphor.

The siloed structural metaphor, therefore, would be observed in combination with other structural metaphors. Figure 22, for example, shows a case of the siloed metaphor in combination with the flat metaphor; in this case the plants and mammals are distinct from one another and there was no indication from the diagram that they are linked to one another. I will preface, however, that with pressure, most likely all of those participants exhibiting the siloed metaphor would indicate that all life is related (this was not probed) but their initial inclination to ignore or separate these groups was considered as evidence for this structure though it may be highly transient or changeable.

Sita, an expert, for example, drew a flat structural metaphor combined with the siloed structural metaphor (Figure 10). As she indicated that she is a systematist, she likely does

recognize that ontologically everything is related; however, she still indicates that they are separate on the diagram. While one could argue that the window containing the diagram could be the visual container unifying the two items, she made no indication that that was the case. Her particular diagram could be part of a larger tendency for people to consider plants and animals as vastly different such that they are presumed to operate under a different set of natural rules (Heredia et al., 2016).

Seline, a novice, similarly shows some division, again between plants and animals. However, in her case, she combined the siloed structure with a vertical structure (Figure 23). She separated plants and animals into two different diagrams and within each structure used nodes of a branching structure to represent taxonomic groups. Note that she even explicitly added a line to her diagram to indicate a high level of distinction between the two groups. When asked about her line, she states, “So each line is a further distinction between the group, I guess, at the top of the line. So we start with the broadest animals after we've separated plants from that.” While she thinks I am referring to the other lines, her statement that she’s “separated plants” after drawing a line and drawing it more distinctly (no text) in the diagram suggests that she views them very differently (Figure 24). The video makes it seem like she was sorting out the plants from the group. While this could be also due to difficulty, i.e., perhaps she thought it would be more

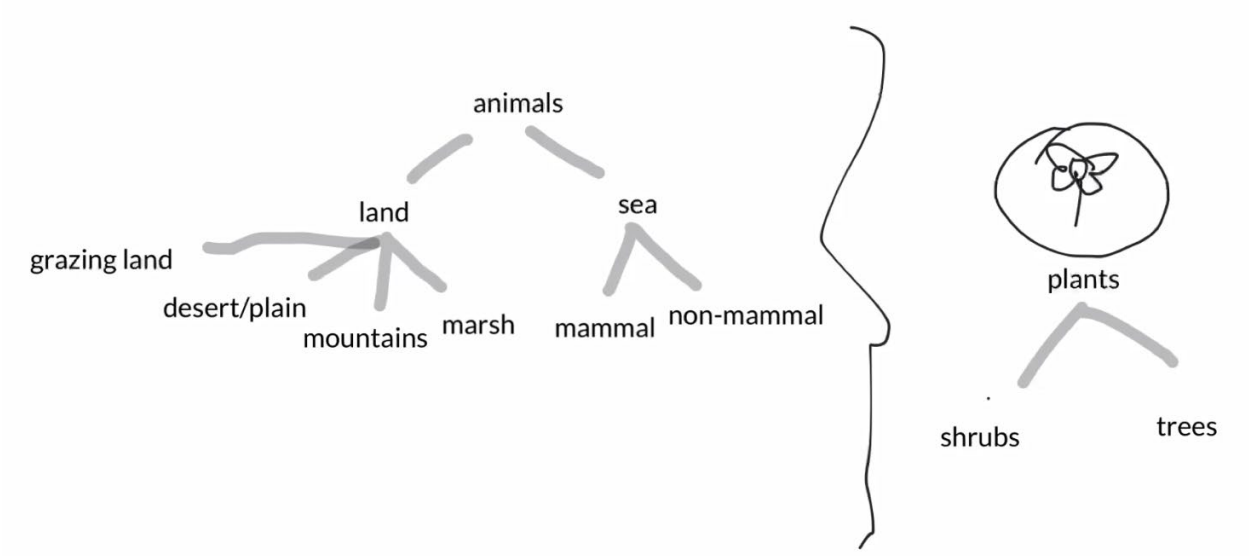


Figure 23: Seline's relatedness diagram (novice).

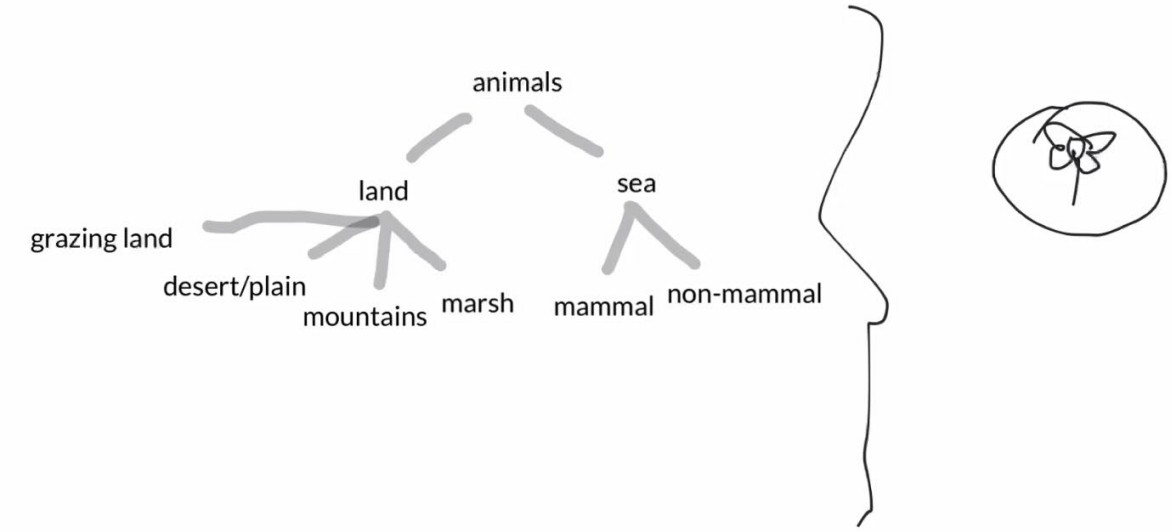


Figure 24: Seline's relatedness diagram (novice) when explaining her diagram.

difficult to do so for the plants because of the lack of familiarity, it could also be because of their differentness which is consistent with the literature on plant awareness disparity (Achurra, 2022; Wandersee & Schussler, 1999, 2001).

While the siloed structural metaphor was most apparent in the diagrams that participants created, there was also evidence in the form of utterances. King, a novice, stated, "From the photos, I would say I don't know that much about the organisms so I would **probably do the**

plants and animals separately. And sorry, I forgot about that. I would do the **plants and animals separately** and probably if this is the animal one I would start with the one that looks like kind of a plain fish and I would go this branch area would be probably the aquatic..."

Peppermint, a novice, similarly stated,

> I'd probably do like the fish and stuff, kind of in more of a web. Like, this is a dolphin, I do like, **relating the ones that seemed more related, in a web, and then separately, the trees and stuff** would be in a web together, like the **palm trees and the bush are all in a web together**. And then, the **land animals would be in another web**.

Both students indicated that they placed taxa "separately" and made diagrams that reiterated this separation. Though other students used the term "separate" they often would use it in a phrase like "separate out" indicating it was part of a larger whole or part of a larger lineages. These students did not do so and indicated that the two were wholly separate entities. It is also notable that no experts exhibited utterances that indicate siloed structures. The only experts marked as such did so in their diagrams, and experts that typically stated things that would suggest a siloed structure, e.g., suggesting that plants were unusual, would still ultimately create a diagram that suggested that they thought about the two as ultimately being connected.

Scientific Normativeness

This last structural metaphor, i.e., the siloed structural metaphor, could be considered mildly scientifically normative. It is most likely not normative in the sense that all life is technically related; therefore, when asking if and how organisms are related, one diagram should be drawn and one unifying explanation should be given. However, it can be considered normative in the sense that depending on the context of a systematics study, only a subset of taxa is focused on at one time (e.g., Buckner et al., 2015), and groups as broad as the ones given are not often investigated. However, this underrepresentation may be simply due to the way life has evolved, there are more tip taxa relationships to determine than larger relationships/taxonomic

groups. On the other hand, in a situation where a researcher has a set of taxa to determine the relationships up, all are typically placed in the tree, so in that sense, the siloed structure would still be considered non-normative.

Additionally, many of the examples of siloed structures involved plant awareness disparity (Achurra, 2022; Parsley, 2020; Wandersee & Schussler, 1999). Individuals would often use this structure in combination with the assumption that plants and animals are too distinct to be considered related (though of course individuals when probed would likely assert that everything is related—it is just that their bias towards organisms more like themselves may cause individuals to forget their relationship to us). Plant awareness disparity could be considered broadly as being non-normative though it is also observed in experts, even within this dataset. Plants are subject to the same evolutionary forces and phenomena as non-plants despite their morphological distinctness from macroscopic motile animals that causes us to assume that they are not.

Experts versus Novices

Among experts, this siloed metaphor was uncommon with only about 20% displaying such thinking compared to 50% of novices. Such prevalence among novices is not surprising. From an information processing standpoint, siloing likely arises due to a lack of interconnectedness between concepts, which in this case could be considered taxonomic categories. Experts, having undergone extensive education, are frequently exposed to the interconnectedness of all living things, aiming to grasp the comprehensive "tree of life" in biology. In contrast, novices may simplify grouping organisms inside fewer sets of categories like plants, animals, or fish without recognizing broader taxonomic connections between them. This siloed thinking typically was observed in the form of "plant awareness disparity (PAD)",

where entire organism groups, like plants, were overlooked (Achurra, 2022; Parsley, 2020; Wandersee & Schussler, 1999, 2001). PAD is recognized in the literature and in this formulation could therefore be considered a type of siloing in context. In particularly blatant cases, novices completely neglected the existence of plants or other taxonomic groups and only spoke about them when I asked about them.

While I assume most undergraduates have been exposed to the idea of universal organismal relatedness, this might be an overestimation. If accurate, it is striking that even at their educational stage, many did not instinctively recognize this interconnectedness, underscoring the persistence of biases like PAD. This underscores the importance of investigating other areas of disconnectedness that likely need to be addressed more explicitly in order to help support the generalized thinking of the relatedness of all things, in addition potentially to help the transfer of other related macroevolutionary concepts from one context to another (PAD can lead to students thinking that plant macroevolution and microevolution are not subject to the same evolutionary pressures or phenomena (e.g., allopatric speciation can occur in plants and animals). Indeed, this study showed my own biases here where two of the experts asked about the lack of reptiles and invertebrates in my example. Indeed, some of this has briefly been documented in the literature, specifically with regards to tree-thinking; Catley and Novice examined the understanding of reptiles and how that was affected by tree-reading and vice versa (Novick & Catley, 2014). However, more taxa need to be investigated. Fish, which are very distinct from most macroscopic motile land organisms which humans arguably have more experience with, are likely almost as unfamiliar as plants and might be subject to some siloing, some disconnectedness, and lack of transfer of evolutionary principles.

The siloed structural metaphor was observed alongside other structures; for instance, someone could display a disconnected array of vertical structures, which would be coded as exhibiting both "siloed" and "vertical." Of those exhibiting either overlapping or closely positioned structural metaphors such that they could be considered as describing the same composite structure, one expert and three novices exhibited the siloed structure in combination with the vertical structure. Two experts and four novices exhibited the siloed metaphor together with the flat structure. The siloed structural metaphor was also observed as co-occurring with the storytelling metaphor in one novice and with the tree-like metaphor in another novice. Most of these individuals exhibited siloing in their diagram primarily and may or may not have given indications in their utterances. This may indicate that my methods and probing questions were insufficient for uncovering more verbal evidence for such a structure or it may indicate that the content of said structure might better be revealed by diagrams. I also note that the vertical and flat structures most often coincided with them, which is plausible given the more spatial nature of the vertical and flat structures. It is unclear, however, why this may be the case. The appearance of said structure does not seem to coincide with normativity, for example, so more investigation into the subject is needed to explain this result.

Grouping Strategies and Grouping Criteria

One major aspect of understanding relatedness that soon became clear during the analysis was that the framework or structure in which individuals placed organisms was distinct both from the actual groups that were discussed as well as the procedures and characteristics associated with the generation of taxonomic groups. These two codes are referred to in this study as *grouping strategies* and *grouping criteria*.

Grouping strategies are ways in which individuals determined the group membership and (potentially) *structural metaphor*. In other words, grouping strategies are the generalized procedures used to assign taxa to taxonomic groups which are likely independent (though more investigation is needed) of task features besides taxonomic group identity. Individuals were observed sometimes using multiple procedures while others only one. These categories were derived from responses about the identification of taxonomic groups and responses to probing about how they knew that certain taxa belonged in certain taxonomic groups.

Grouping criteria on the other hand refers to the standards by which taxa were being evaluated in combination with a grouping strategy to determine taxonomic group association. Some of the grouping criteria discussed are implied by the existence of the grouping strategies but explicit evidence of such criteria was not probed further and thus not obtained. This category of responses is much more straightforward and phenomenological in nature as most instances of responses were very explicit references to the characteristics discussed; less interpretation, therefore, was necessary to group them.

In the following sections I discuss these categories of responses in greater detail and I additionally discuss the subcategories belonging to each. Grouping criteria's subcategories will be discussed in less detail given the number of subcodes and the obviousness of their appearance in the dataset; as their definitions and occurrences were straightforward, little special attention was required to differentiate between them. Within each section I also briefly discuss the implications of all the categories for research and pedagogy. I also end with a discussion of the frequency of appearance and the similarities and differences between experts and novices.

Grouping Strategy

Grouping strategy refers to a large subset of responses that employ different procedures for grouping taxa together. Individuals would often either explicitly or implicitly explain why taxa were grouped together using these strategies. Typically, these would only be apparent after explicit probing for grouping strategies, with me asking participants to explain why or how they grouped those taxa together. Four major types of grouping strategies were observed: prototype, isotype, appeal to authority, and essentialism. I discuss these strategies below. A summary of the number of individuals exhibiting each type of response is given below (Table 10). All participants exhibited at least one grouping strategy.

Table 10: Number and Frequency of Grouping Strategy Responses

Grouping Strategies Subcategories	***Experts***	***Novices***	***Total***	***%Experts***	***%Novices***
Similarity	11	18	29	100	100
Prototype	9	15	24	82	83
Isotype	11	17	28	100	94
Essentialism	1	3	4	9	17
Appeal to Authority	7	9	16	64	50

The Strategies

Similarity. Similarity refers to the first broad group of grouping strategies. These two strategies include prototype and isotype strategies. These strategies involve finding the similarity between two members of said category and using criteria for said similarity in different ways. These categories used the same types of criteria, which I discuss later in this chapter. I discuss these two categories in greater detail below.

Prototype. The prototype strategy is that which is probabilistic in nature and is likely based on the typically innate ways in which individuals identify and classify objects which in this case includes organisms. In other words, individuals harbor in their mind some sort of prototype or ideal version of said concept which represents a class of objects or being. This ideal

representation of said objects is associated with a series of defining characteristics. The representation is therefore used as reference with which the object in question is compared to. This category, therefore, appears to have a quality of centeredness with the prototype at the center and version of said class being more or less distant from said prototype and that associated with the prototype is a set of criteria which may or may not be present in each of the individual's said category.

This strategy presumes a list of characteristics or criteria is associated with each prototype, though they may not necessarily be all employed to determine whether an object is considered part of that class. In addition, lack of certain traits or different versions of traits may be present, but the object is still classed as such. So, if, for example, an individual was to examine a very small tree, like a bonsai, it is still considered a tree, even though most trees are much taller than bushes or shrubs and not that small. We know that the bonsai has a mostly bare trunk, and it may have needle leaves resembling existing trees such as pine trees but are not necessarily present in shrubs. We also know that it is forced to be that way and is related to primary trees. Therefore, we can suggest that tall height might be a trait that defines trees but in the case of bonsai this trait or criteria is violated. According to the prototype strategy, bonsais therefore would be categorized as trees given these probabilities of having certain tree like properties.

This strategy is named after an associated theory which I argue is essentially the same, i.e., *prototype theory*. Prototype theory is a theory from embodied cognition traditions that describes the hypothesized process by which humans typically classify things, whether they be objects, organisms, social categories, etc. (Lakoff, 1987; Rosch, 1973). Like previously described in Chapter 2, prototype theory suggests that individuals harbor a non-existent

prototype object or class in their mind that have a series of criteria associated with it and that these criteria may not be present in every member of the class. These criteria therefore have some probability of appearing within a member with different criteria having different weights or probabilities of being associated with said class. This ideal "prototype," however, may not exist cognitively as some iconic image, rather it represents a set of traits associated with highest probabilities of appearance.

Professor X, an expert, for example, stated, "**Plants**! LAUGHS. **Two plants**. Then let's make them [parts of the diagram] **green**!" Professor X here made a clade that included plants and indicated he made the clade green, implying that their coloration was a reason they were grouped together. Though not explicitly stated or probed, Professor X likely knows that plants are not the only organisms that exist that are colored green, especially as he is a systematist. He therefore used a trait that is not unique to the group in question and does not necessarily represent every member of the group in question, suggesting that he was employing a lay understanding of plants that uses a probabilistic understanding of traits, i.e., that plants are typically green and many other groups or organisms are not necessarily usually green.

Gardner, an expert, similarly stated, "There are **three plants**. They're **green**, they're **photosynthetic**." Similarly, Gardner uses the fact that plants are green to group them. She also uses the criterion of photosynthesis to group plants. This is yet another trait that Gardner likely knows is not unique to plants; she is a microbiologist, and some microbes are also photosynthetic but not monophyletic to plants. Again, while not probed or stated directly, I thought this was a common enough feature and known widely enough to warrant this assumption. Gardner used this same strategy again when she stated, "**There are fish that are coldblooded**. Oh, let me add one more thing because **these are warm-blooded**." She indicated in her diagram that she marked

that fish were coldblooded but when talking about her diagram (Figure 6) amended that "there **are** fish that are cold blooded" suggesting she later recognized that there are warm-blooded fish (e.g., Opah) but that the ones in the diagram were coldblooded and used the broad probabilistic assumption that fish are generally coldblooded to make the group "fish" and that she was sticking to it because the contrasting group she created consisted of warm-blooded organisms.

King, a novice, similarly used traits in a manner suggesting she was using a prototype strategy. After I asked her how she knew the two taxa she placed in "sharks" were sharks, she stated, "I just say because I've **seen sharks before**. But I see the **fins** are kind of my **big key**, I would say." Two elements of her utterance here suggest that she might be using prototype strategy. First is her use of "fins" as a "key" trait. Fins are found in bony fishes and cetaceans as well as cartilaginous fishes; the task included examples of bony fishes and cetaceans given, yet she did not include them in the group. She used a trait to define a group though it may not necessarily be unique to the group, which again leads to this idea of probabilistic understanding of traits as properties belonging to a category. The second giveaway was her vague statement "I've seen sharks before." She could simply be suggesting that she was relying on prior knowledge where she simply learned the normative use of the term "shark," recognized them as the same and believed they were such. However, more likely she recalled a series of traits, again probabilistically related to the category of "shark." Several other individuals similarly struggled to describe why they grouped organisms saying that they just "knew" they were a group because it "looked like it," meaning they were using visual cues but were not cognizant of what exact discrete properties were used. This lack of discreteness implies it is different from the classical isotype strategy where a group has a unique trait that all members of the group have and have an equal chance of having. It is much easier to recall a distinct trait than a fuzzy recollection of a set

of traits likely coming from seeing patterns in the visual field (object recognition). Indeed, I included an uncommon shark species (whale shark) for this reason and her statement "seen sharks" implies that at least one of them was not necessarily one she had seen before but fit her set of criteria for sharks.

Isotype. Isotype refers to the strategy observed that I argue is analogous to Aristotelian classical theory. Individuals exhibiting this strategy would suggest, implicitly or explicitly, that all individuals that were part of some organismal group must have all traits described and not having said trait would suggest that they are not part of the group. This is consistent with Aristotelian classical theory as described in Chapter 2 which posits that concepts or classes are so determined by members having a set of characteristics or not (Lakoff, 1987; Studtmann, 2023).

Classical strategy of course, in this context, refers to perception of the groups. Classical grouping of phenomena or objects may or may not be normative. While many human-perceived classes of natural phenomena likely more closely align to prototype theory (e.g., species has no one set definition, the edges of the species concept are quite fuzzy), this is not necessarily always true of all classes. Socially, some classes are constructed and have strict definitions – someone is either a US representative in Congress or not. This would both realistically and perception-wise align with classical theory. However, there is the case that the natural phenomena being represented has itself fuzzy boundaries and humans can either perceive said category as also having fuzzy boundaries or impose a classical understanding of categorization on said category. I mention this issue as Margolis (1994) has suggested that cognitive scientists often conflate perceived and ontologically true categories. While Margolis (1994) implies this is due to incompetence, I adopt a more charitable view that it is difficult to keep track of what is perceived

and what is real as we often assume that our categories are ontologically true and we unfortunately need to use said assumed real categories to describe and refer to said categories.

While the aforementioned framing of classical category formation for taxa might appear negative, I suggest it is not actually the case. The classical strategy is arguably used very often in science and systematics, especially in the case of synapomorphies. Science often tends to use discrete definitions and descriptions of phenomena to operationalize and analyze data. While methods like parsimony, etc., attempt to look at a series of traits and their evolution, in textbooks and in lectures, larger taxonomic groups are often defined by the appearance of a particular trait, implying that all individuals in said group have said features (e.g., mammals have hair). That therefore constitutes classical categorization.

Ridley, an expert, used classic examples of synapomorphies to describe her reasoning for grouping mammals. She stated, "Yeah, so **hair** versus I guess you could, if we come down, yeah, there's three things I guess like plants, and then **things with hair follicles and mammary glands**, and then things like scales." Hair follicles and mammary glands are classic synapomorphies instructors use to describe mammals. I therefore assume she is using in the normative fashion, i.e., is a unique property of the mammals. It is, however, worth noting that hair follicles have been lost in some species of mammals, so it is not necessarily truly present in all members of the group in question, but it is not present in other groups.

Professor X, like many other experts and students, referred to the existence of a common ancestor as the grouping category. This is technically more normative as we ideally would group based on the hypothesized ancestor which we estimate using other traits. Professor X states, "So the pig and hippo, they **share a common ancestor** very much more closely with the whales and dolphins. So they are all in just one group here. And we know that from fossils and from DNA,

it's pretty cool." Repeatedly throughout his response he refers to the ancestor as being the reason for grouping things the way they are and that the fossils and DNA are used for estimation of this grouping. Though the word "because" is missing, it is implied to be after the first clause, he suggested that the two should be grouped because they share a common ancestor and later clarifies that "fossils and DNA" could be used to determine this grouping.

Elisa, a novice, gave an advanced example of using an isotype strategy. She appeared to be familiar with phylogenetic informativeness by her statement,

> So, if a particular **pair of genes** are used to **give sharks their teeth, for instance**. And that is **very similar across** a grouping of sharks, let's say **all the sharks**, that's the **same type of gene** used. Well, that may be more of **a useful comparison** versus doing something where this gene is literally there. If we **changed it in the animal, nothing would change**.

She stated this after I asked her to clarify her statement on useful genes versus not useful genes. She gave a hypothetical example of a unique gene that functions to give all sharks a unique trait so in that sense is an example of an isotype strategy. She additionally suggests that if the gene were "changed" in the rest of the animals (confusingly stated), meaning that the rest of the animals also had the gene, she could not "change" the group or group the organisms at all. She essentially suggested that the uniqueness of the gene is how she decided to group the sharks away from the rest of the other animals.

Appeal to Authority. The next strategy that was prevalent in the dataset is what I refer to as appeal to authority. This strategy's name does indeed allude to the common "logical fallacy" of "appeal to authority." While fallacy implies negativity, I use it in a non-judgmental sense but similarly name this category given the public's familiarity with the topic.

Appeal to authority in this sense refers to individuals either directly or indirectly suggesting that certain groups belong together because other parties, typically outside authorities implied to be more knowledgeable on the topic, have determined that was the case. If individuals

suggested that, for example, instructors had told them that the organisms in question constituted a group, that was considered an appeal to authority. Novices would also often suggest that scientists had produced these groups, implying that they viewed scientists as others and did not identify with scientists, suggesting that scientists had greater knowledge or expertise on the topic. This practice, however, was also observed in experts who would reference other scientists with presumably greater expertise in the taxa being discussed. Some, for example, reminded me that the group in question was not the same taxon they focus on in their work and were uncomfortable providing definite answers. They therefore would cite other experts as a sort of reasoning for the taxa they created. Both of these cases imply that the individual suggested that someone other than themselves have established these groups and they simply reported them and used these authority figures as a warrant to explain their reasoning rather than giving their own reasoning for these groupings. Of course, one could argue that many knowledge structures ultimately come from authority, i.e., as learning science involves learning about the results of scientists' past work. Indeed, one individual dryly replied that the reason for the groupings he gave was because he was an "expert." However, authority was considered the appropriate strategy to categorize their responses given that it was the most proximal reason for the groupings that they offered.

It is important to also note that while criteria were not explicitly given for these groupings, it is arguable that the property of being given by authority is a criterion to determine groupings and therefore these responses could be considered an isotype grouping strategy as all members possess the criteria of being grouped by authority. I considered this strategy unique, however, because of the nature of said grouping. A criterion of "being grouped by authority" might be applied to any and all groups as, technically, most organisms have been given some

taxonomic designation and therefore they would be all grouped together irrespective of the specific taxonomic groups designated. In other words, if camel and llama were designated as being part of camelids by an authority and palm trees and cycads part of plantae, then using the criteria of being grouped by an authority would group camel, llama, trees, and cycads together in one group. In other words, two disparate groups could use the same criterion. Such a move would be inconsistent with classical theory. In addition, in some formulations of categorization, extrinsic properties are philosophically less likely to be used for categorization as it constitutes a property that is not "part of" the object in question (Lakoff, 1987), of which authority would be as it is not intrinsic to the taxon in question.

Alternatively, one could amend an appeal to authority criterion to be "designated by authority as part of X group" and in that sense it could be considered a property. However, this formulation is less generalizable and does not fit the utility criterion of qualitative validation standards for Grounded Theory. As I am interested in highlighting trends or themes that have strong explanatory power and are useful for practitioners (Glaser & Strauss, 1967; Kuczynski & Daly, 2003), I have isolated appeal to authority from the aforementioned strategies and adopted a broader version that encompasses all criteria. Additionally, I separated out the effect of context (i.e., in this sense, taxonomic identity) from these codes into its own code, which I do not discuss in this dissertation and is beyond the scope of this study.

Appeals to authority in general could also signal that the individual may not have internalized a conception that is well connected to other conceptions as in information processing perspectives, indicating that they have a less well-formed understanding of said category or at least they perceive themselves to have such. For example, students often give an explanation "Well my teacher said so" indicating that they may have not internalized the mechanisms by

which that fact or concept was designated as such. However, I also note that even experts have utilized this strategy. Experts have often been observed to use technically "incorrect" heuristics or language for situations with certain features (Catley et al., 2010; Hammann & Nehm, 2020). Appealing to authority may function as a heuristic filler for some problem spaces even though they likely have internalized some of the formal processes by which taxonomic groups, especially species, have formed. Alternatively, usage of this strategy might indicate an affective tendency within the community of practice to not overstate expertise in their subject. While a well-read student might confidently suggest that certain groups are related, an expert with a similar understanding of said taxa, but does not study said taxa, might decide to understate their understanding of the subject, given that they know peers might know more and/or the norms of the field suggest that they should do so in the presence of peers.

Whatever the reason, the general tendency I wish to highlight is the grouping by authority designation for which there might be multiple reasons as previously mentioned. Focusing on this group as distinct from the rest could indicate, for example, that students exhibiting this behavior require additional instruction in the reasoning behind such structures. Alternatively or additionally, efforts in the classroom may be required to increase students' perception of themselves as scientists or people who conduct science (Camacho et al., 2021; Stets et al., 2017) as acquiring that identity may alleviate the distancing observed in their discourse (i.e., they found out that it is part of X group rather than *I* know it is part of X group for Y, Z reasons) and presumably increases motivation to internalize said concepts as they are motivated to enter related professions. Adopting said position may therefore require different actions by practitioners than the aforementioned ones, hence it is separated from them, as per the practicality validation criteria (understandability) for Grounded Theory (Glaser & Strauss, 1967).

Novices, like John, would exhibit this strategy often when they struggled to provide reasoning for the grouping and instead cited prior knowledge or learning. John stated, “Um, I think I just--**I've heard that before**, and **I think I remember** they give live birth. **Um, so they're mammals**.” He implies that some outside source has determined that they are mammals and that should be true but provides none of his own evidence to suggest that they should be grouped the way they should; he simply trusts their grouping as being correct. John here is implying that some authority is the one who has determined these groups. He mentions live birth but says it in a way to suggest he did not observe that but that the authority has given such reasoning for a grouping.

Novices like Theodore, for example, would exhibit this strategy in the form of a preoccupation with classification. Much of Theodore’s responses were spent struggling to remember what the supposedly proper taxonomic groups were and their names. He stated, “All right, so I would have two separate parts basically, but they all branch out of... [UNINTELLIGIBLE]... just **in terms of classification**.” He later stated, “I know there are certain types, but this is just for the **basic ones** we have **I'm not sure** if certain plants over here are **coniferous** or if they're **Bryophytes** or what not. But for now, I'm just going to basically go with plants.” In these quotes, and many others not provided here, he was overly concerned with what the proper taxonomic classification was for each of these groups and what the proper names were though I did not ask for them. His worry about what is “correct” seems to be a preoccupation with what the outside authority thinks these groups should be. While a valid concern, especially as undergraduate biology education typically contains a unit on the current state of our knowledge of the arrangement of taxonomic groups, this preoccupation may suggest a lack of identification with the scientific process and a reticence to offer his own reasoning.

Before explaining her groupings of marine mammals, for example, Turaco stated, "Although **I'm not super good at mammal phylogeny** so **don't tell my mammologist friends** that I have here." Similarly, Sita stated, "Blue whales and... From **my past knowledge of biology**, I **have taken a lot of biology in my life**, I know. The blue whale, dolphins, and all the other terrestrial animals or mammals." Experts in this sample were coded as referring to appeal to authority when they appeared overtly preoccupied with what was "correct" and would either refer to others' expertise or use their own expertise as a sort of warrant for their argument. While common in the sciences, experts do reasonably claim that they have a very limited scope of expertise and are hesitant to claim broader knowledge of adjacent fields; it is notable that they, in a sense, are doing the same as novices. They are simply trusting what they have been told or have already found out and group things as they have memorized them. Experts might know enough to suggest that they do not have the real data to create the most realistic groups, but other experts, such as Professor X, still performed the procedures of grouping organisms. However, Professor X was a systematist who may have been used to instructing others on grouping organisms with a hypothetical small dataset. Regardless, the outcome of these statements is the same, that is that experts who exhibited this strategy seemed to use either others or their own knowledge as a justification for the groups they created rather than provide evidence.

Essentialism. Essentialism is a grouping strategy observed in the dataset that refers to individuals that provide reasonings for their generated or pre-generated taxonomic group as being inherently part of that group. Someone might say, for example, in an almost circular manner, that a camel is a camel because it is a camel. Implied in such a statement is that there is an intangible essence that makes a taxonomic group have a particular identity. This view of organisms has been documented in the literature extensively (Muradoglu et al., 2022; e.g., Sousa

et al., 2002) and has been observed in other areas of human cognition such as social categories (Ho et al., 2015). What is difficult about this particular category is the lack of reasoning given for such grouping tendencies. When probed, individuals will often simply repeat the statement as if inherent to both speakers and implied to be ontologically true. However, given enough probing, the interviewer can unintentionally imply that reasons besides essence must exist for the objects to be categorized and individuals will ultimately attempt to find a different strategy in the moment to explain said reason. In other words, though they may not have initially thought to use that strategy, they can be forced to adopt other strategies in certain situations. Indeed, two individuals showed this pattern of responses; they first would adopt an essentialist reasoning for their taxonomic groups and then would later adopt others after probing.

Turtle, a novice, exhibited an essentialist grouping strategy in his statement: "I guess I would first... one way I would relate them is **between plants and animals**. So these are **more related than like the palm tree and the shark because they're both plants**." Here Turtle indicates that his reasoning for grouping the palm trees together as plants is "because" they are plants. Another novice, Peppermint, after being asked what criteria they were using to determine what belonged to each group, stated, "I guess, the **criteria** for that was **that they are plants and not animals**, like that was the difference for them." She is, like Turtle, suggesting the reason for such groupings was that that the members were part of the group and not part of another; some intangible reason for such groupings determined group membership.

Only one expert, Sita, exhibited this strategy very briefly. She stated, "They're all related **because they are mammals**. Because the **plants are not mammals**, so they're not related." While what she stated could have been simply a case of misspeaking, she reinforces her argument after suggesting that the mammals she was pointing to were mammals "because they

are mammals," that conversely plants are not part of the mammals, "because… [they] are not mammals."

Normativeness

These grouping strategies vary in normativeness, with the essentialist strategy being the least normative. The isotypic strategy could be considered closest to the idea of the use of synapomorphies to determine group membership. However, given the nuance in conducting these studies, it can also be argued that the prototype strategy is somewhat normative in the sense that often traits that are not necessarily present in all members are sometimes used, i.e., individuals lose traits. Additionally, some traits are analogous and have arisen convergently which is difficult to discern on its own, especially if genetic. Most phylogenetic methods pick the trees that have the fewest evolutionary events that explain the pattern of traits on the tree, so it is in a sense, a type of amended understanding of similarity.

Experts versus Novices

Experts and novices in this dataset all exhibited some sort of similarity strategy, i.e., prototype or isotype (Table 10). Experts and novices exhibited prototype understanding at an equal rate: 82% of experts and 83% of novices. Isotype was similarly high but with a bit more variation: 100% of experts exhibited the isotype understanding compared to 94%. Essentialism was particularly low in both groups, with only 9% of experts, referring to one respondent, exhibiting it as compared to 17% of novices. Appeal to authority, however, was slightly higher in experts at 64% and 50% in novices.

These patterns are perhaps, not very surprising. Similarity of course being the most frequent, we would expect that individuals would likely interpret relatedness to mean similarity in the sense of series of properties that are in common between members of a group. Indeed,

normative understanding uses similarity, as I argue, in a sense when understanding taxonomic groups—though technically the field says it does not do so. Similarity is likely a very common method individuals use to categorize objects; however, it may not be innate given that some cultures have been known to group for reasons other than common properties (Lakoff, 1987).

Prototype is particularly high among both groups, which is also unsurprising. Assuming the prototype strategy observed here is consistent with the literatures' description of prototype theory, it is likely a very common or near innate manner of grouping as some imply (Lakoff, 1987). Prototype theory reflects a natural tendency to categorize things based on a probabilistic set of properties by which membership is determined—this is often reflected by vague statements of knowing that an object belongs to a category and when pressed individuals will produce traits that they have not necessarily noticed explicitly. One could imagine that this is likely a very common or potentially primary manner for categorizing objects, especially since prototype theory could be considered best suited for the categorization of objects that have a lot of visual properties given that we often lack the declarative knowledge to be able to describe why our perceptual systems have parsed visual cues to assign objects to different categories. Indeed, individuals when asked about their answers that were consistent with prototype theory seemed like (though more investigation is necessary) they were generating criteria on the fly.

The isotype strategy was surprisingly more common in the dataset. While some researchers from the embodied perspective suggest that prototype theory is the innate method or process of categorizing things (Lakoff, 1987), others have suggested that may not be the case (Margolis, 1994). Many sociological categories, for example, seem to have a set of properties that define them; a US senator can only be someone who was elected US senator, for example—there is no ambiguity about the properties that define what a US senator is. Indeed, this dataset

seems to suggest that the isotype strategy was as common, if not more common than the prototype strategy. However, this may be limited to this particular context.

The bottom line from these particular results is that similarity is the most common strategy used and that both types of similarity strategies are overwhelmingly common in both groups. This is, of course, unsurprising given that the majority of categories that people create that contain members of objects, likely are such due to their similarity. What this suggests, however, is that, assuming "similarity" is not normative or incorrect might not adequately capture the reality of what experts and novices are thinking and that, when experts discuss how "similarity" in tree-thinking and tree-building is incorrect (Baum et al., 2005; Gregory, 2008; Halverson et al., 2011; Kummer, 2017; Kummer et al., 2016), they are referring to a specific operationalization of the term. The similarity they are referring to, for example, is the comparison of the number of differences. Systematists arguably do, in a sense, think of similarity in that they assume that as an outcome of evolutionary events, organisms will tend to be more similar and that homologous traits should outnumber the analogous ones (analogous and other traits can make species look more similar than they really are, however we assume that these are homologous traits). This is what allows systematists to use these outcome values to estimate the number of evolutionary events that explain the data at hand. I discuss this issue in greater detail in Chapter 5.

Appeal to Authority was one of the less frequently observed strategies with 63% of experts and 50% of novices exhibiting it. Appeal to Authority's relatively high appearance in experts is initially surprising but not so when examined in greater detail. This category is defined as using a reference to an outside source or being excessively concerned with what is the "correct" taxonomic group. Academics tend to rely on each other's expertise in associated

subfields as their own expertise becomes increasingly niche, so it is unsurprising that some might be concerned about that. Students' responses, however, are wholly unsurprising given that it is common for students to be preoccupied with the "correct" answer rather than perform such a task on their own. This of course could be due to inadequate probing, but it is also likely a matter of context as these students knew this was an education research related interview and students have been observed to exhibit such preoccupations in a classroom setting to which this interview could be viewed as analogous to. Given this finding, it might be worth investigating if increased utility of this strategy is observed in settings where students feel more pressured by instructors and operate in an atmosphere that prioritizes performance over learning (Elliot & McGregor, 2001). In addition, motivational profiles such as performance-avoidance (Elliot & Harackiewicz, 1996; Murayama et al., 2011) might coincide with greater usage of this strategy, suggesting the need to amend pedagogical procedures around tree-thinking to account for such a strategy. Much like those offered in other circumstances, such practices would need to ensure that students feel like they are in a supportive environment for them to warrant adoption (Marcroft et al., 2022). Usage of this procedure may account for decreased attempts to engage in the relatedness process or task or tree-building and perhaps decreased adoption of appropriate relatedness methods, though of course this would also have to be investigated as well separately.

The appearance of Appeal to Authority, however, may also be a function of simply forgetting the procedures that lead to the groupings specified and an epistemological lack of recognition that these are phylogenetic relationships and are not necessarily the be-all end-all "truth" of evolutionary history. They are, in fact, hypotheses or reconstructions of evolutionary history and are subject to revision. Most experts likely know this (Morrison, 2014; Podani, 2019), but students may lack the epistemological experience of being in science to this (Peffer &

Ramezani, 2019). Not identifying with science and scientists, therefore, might lead them to think that they need to report the "capital T" Truth, so engaging in their own reasoning might not be the first impulse in such a situation, no matter how I might frame the task. This recognition, however, would suggest that experts and novices engage in this strategy for different reasons, and this would need to be explored further for investigation. In addition, the other factors mentioned, such as epistemological views and identity, would also have to be investigated as possible reasons for choosing such a strategy.

The last strategy, essentialism, was the most infrequently observed strategy in the dataset. It only appeared in 9% of experts and 16% of novices. This strategy is the most non-normative of the set. It is therefore unsurprising that it appeared comparatively infrequently. However, it is worth noting that it may be another lay manner of categorizing natural phenomena (Sousa et al., 2002; M. G. Taylor et al., 2009). Indeed, this strategy was more commonly observed in novices, who, by definition, lack the expertise of the experts and therefore should be more likely to give non-normative and more everyday reasoning.

It is worth noting, however, that an expert did indeed express this view and it was within the context of plants. This person was an animal systematist and was unfamiliar with the plant taxa given so this usage of essentialism might have been an expression of plant awareness disparity (Achurra, 2022; Parsley, 2020; Wandersee & Schussler, 1999, 2001) which experts have been observed to engage in. Plants, being beings that are less similar to us in form and biology than animals, it is understandable that people generally are unfamiliar with and therefore less likely to want to speculate about them. The individual in question grouped plants because they were "not animals," implying something about them made them plants which was intangible and difficult to describe. One avenue of future research, therefore, might be the effects of

taxonomic identity and context on relatedness understanding; this result would imply that we likely would see greater use of essentialism in unfamiliar contexts, assuming the individuals' understanding of relatedness does not transfer. I discuss these matters in greater detail in Chapter 5.

Grouping Criteria

Grouping criteria is a category that encapsulates interviewee responses that refer to the characteristics or criteria by which they grouped organisms. These characteristics range in type with some referring to extrinsic properties, i.e., properties there are more external or related to the organism in question, or intrinsic properties, i.e., properties that are more part of the being of those organisms. These traits were used in combination with aforementioned grouping strategies to generate taxonomic groups during the interview task.

The traits include scientifically normative characteristics such as morphological traits and genetic traits, as well as non-normative characteristics such as domestication status and environment. In the following paragraphs I discuss each of these criteria in greater detail. Note that each of these criteria are a bit more phenomenological in nature compared to the other categories discussed in this study in that they required less interpretation on the part of the researcher and thus will not be described in great detail here. Quotes from the dataset will be provided but not explained in detail as they are likely not useful for the reader.

A summary of the categories and the frequencies at which they were observed among experts and novices is depicted in Table 11. All participants exhibited at least one grouping criteria.

Table 11: Number and Frequency of Grouping Criteria Responses

Grouping Criteria	# Experts	# Novices	Total #	Percent Experts	Percent Novices
Emergent Traits	10	18	28	91	100
Behavioral	2	5	7	18	28
Nonphysiobehavioral	2	10	12	18	56
Physical traits	10	17	27	91	94
Metabolism	2	6	8	18	33
Living Status	3	5	8	27	28
Genetics	5	7	12	45	39
Environment	5	17	22	45	94
Domestication	2	4	6	18	22
Fossils	2	1	3	18	6
MRCA/Common Ancestor	7	7	14	64	39

The Criteria

Emergent Traits. Emergent traits were one of the most common subcategories of grouping criteria observed with 28 out of 29 individuals using them to categorize groups. Emergent traits I define as any type of physical trait where physical can refer to physiological, morphological, phenological, etc., traits that are typically visible on a macroscale. This subcategory contained 5 smaller categories: behavioral, nonphysiobehavioral, physical traits, metabolism, and living status.

General examples of these emerging traits include coat color, lactase persistence, limb length, etc.

Behavioral. Behavioral traits were a less commonly observed subcategory of grouping criteria observed with X number of individuals using them to categorize their organismal groups. Behavioral traits include any type of perceived movement or action that is commonly attributed to organisms that engage in these types of behavior. Behavior is typically considered less phylogenetically informative, meaning that normatively they are less useful for tree construction, though it is used infrequently in the literature and one could theoretically create a tree with this data with evidence of the traits' robustness and lack of plasticity.

General examples of these traits include swimming, walking, etc. Maira, a novice, for example, stated, “And they would be related to each other and then separate like the water animals that are **swimming** like they're all water animals, but some of them are mammals.” Swimming here was taken to be a behavior.

Nonphysiobehavioral. Nonphysiobehavioral traits were those that did not necessarily fall neatly under any of the other traits here. They did not refer to behavioral traits, nor to bodily traits per se. They typically referred to actions or arrangements. General examples of such traits include things like activity type (e.g., nocturnal) and life cycle.

Ella, a novice, for example, stated: “So the whales and dolphins, I know are mammals, just because they have hair and **lactate**.” Lactation here was taken to be a type of trait that is not really a physical or shape related trait, not quite a behavior, etc. Esther, a novice, conversely stated, “They have like the stem. And then the leaves, which are normally green, like in most cases, the leaves are green. And then **they can't move like, they're not relocatable**, I guess.” Esther here uses the fact that plants are sessile in order to group them.

One of the more complex traits to code was referring to spacing of organisms. Sasha, a novice, stated:

> So the ones that are like, **sociable**, I guess this one, this one, this one. I **see a lot of these together**. And then this, maybe. These can be **together**. And there's another one in the background. So there might be like, a **lot in one spot**. This **looks like it's farther away**. And this looks like it could be like in a forest. So **they would like be together**. And I don't know. That's **usually together**, I think. But usually like, these **big ones are alone**… But like, there's like, **pods of dolphins or whatever**, I guess like a **herd of deer**. And, like, I don't know, I know that these **travel a lot together**. These are, like, usually I **see as a group**.

Physical Traits. Physical traits refer to any morphological or physiological traits. They could refer to traits as vague as body shape to presence of specific structures such as antlers.

These traits range from the macroscopic (e.g., limb presence) down to the microscopic (e.g., flagellum presence).

Peppermint, a novice, for example, stated: “So the fish will be more related to the dolphin say, rather than the pig, because the fish and the dolphin are probably more alike biologically and how they survive, rather than the **pig which has fur** and lives on land, and it **doesn't have like fins or flippers or anything**.”

Metabolism. Metabolism referred to any type of trait that did not refer to body shape, structures, etc., and referred to chemical reactions occurring in organisms that help it survive. Traits, such as photosynthesis, glycolysis, and digestion of lactose were all considered macroscopic traits.

For example, a novice, Turtle, stated: “Well, we know the, these are, I guess, this is kind of prior knowledge. I know the plants are these because of their yes, because they all have like green leaves. So I'm assuming, or I guess, using prior knowledge, know that they are using those because they have like, they **get energy from photosynthesis**.”

Living Status. Living status was one of the more unusual distinct traits to appear in the dataset. When asked if organisms were related to one another, many wanted to recognize that all of the organisms were related and after that assertion almost all of the participants would then qualify the statement with the fact that all organisms are alive. While this could in a sense be considered a part of one of the aforementioned traits, it was so frequently observed and so unique that I thought it should be its own criteria.

Alice, a novice, showed an example of this reasoning, when she stated:

> So if I want to explain to somebody what the difference between **alive and not alive** is, and how the **category of life** is pretty broad. I would want to **juxtapose the idea of something that is not alive versus something that is alive** to show the categorical difference between them. So maybe, when I'm thinking of

> something that's **not alive**, I'm thinking of like a mug, weirdly enough. It's on the brain. So, there's like a mug. I hope that looks like a mug.

Domestication. Domestication could be considered less of an intrinsic trait as it is the state of having been intentionally altered by humans. Domesticated organisms normatively are those that have undergone intentional artificial selection, for traits such as lowered aggression, higher yield, etc. In this dataset individuals often did not go into detail about what types of downstream traits were a result of domestication and the state of being domesticated was itself thought to be a trait. Individuals would almost always simply refer to the organism as being domesticated without discussion of additional details of domestication, with the exception of the setting in which they lived (e.g., a pig on a farm).

Maira, a novice, stated,

> And there's like all the land mammals. I think I would maybe keep them together or separate just like the pig, camel, and alpaca **because they've been domesticated to a certain extent,** but it would depend on I guess the context on how I would... separate them.

Environment. This category refers to the usage of environmental components or general environments as a grouping criterion. Depending on the framework used, these could be considered extrinsic or intrinsic. This criterion could refer to a general type of environment, such as forest or desert, but could also include smaller biotic or abiotic factors in the environment, such as the appearance of predators, amount of rainfall, etc. This was separate from emergent traits given that it is normatively more causal in nature rather than an outcome (i.e., emergent traits) and also given that it is very commonly recognized in the misconception literature as being non-normative, i.e., grouping organisms based on their environment is considered incorrect. Organisms living in an environment may have entered that niche from an ancestral clade that did not have any individuals living in that niche.

Peppermint, a novice, stated:

> Through pictures, and also just prior knowledge. Some of the pictures the organisms are **suspended in water** because you can **either see the water**. Like you can **see under the dolphin**, which I just assumed it was **jumping out of the water**. And then, the other have **clear blue backgrounds**, I assumed was **water just from seeing a picture of fish and water before**. And then, the **ones on land**, like the camel, deer, and alpaca and that I can see that they're **walking on land**. And then, the pig, just from prior knowledge, which I **knew that it lived on land**. And then, the hippo also prior knowledge that it lives in **land and in water**.

Genetics. This trait refers to any sort of genetic information that could be used as a grouping factor. I did not probe for whether this referred to DNA or RNA or any other information containing molecules, so specifics are unknown. Normatively we use specific portions of the genome such as single nucleotide polymorphisms (SNPs) to group organisms but in this case almost all participants referred to it generically. It is notable that some even explicitly stated that it was not given in this dataset or not known but that it should be used to group organisms.

King, a novice, gave a very clear example of its usage much like many of the other participants, saying, “If I had all the information in the world, I would do it from how related they are by **comparing their entire genomes**, but from these photos, or is that the question?”

Fossils. This trait refers to the usage or presence of fossils for grouping organisms. Typically, this was not discussed in detail as in the genetics case as it refers to data that is not immediately visible in the task given. Exactly what parts of the fossils, i.e., whether they were using specific morphological traits or some other data, was not discussed. As it was not discussed as itself being grouped, but rather a grouping factor, it was categorized as a criterion. This criterion was observed the most infrequently in the dataset; only five instances were observed, especially surprising given that fossils are used normatively to construct trees or estimate divergence times.

Professor X, an expert, stated quite clearly how pigs and hippos were grouped together, "So the pig and hippo, they share a common ancestor very much more closely with the whales and dolphins. So **they're all in just one group here**. And **we know that from fossils** and from DNA, it's pretty cool." While here he gave a proximal criterion of ancestry as determining group membership, he also suggests that they are estimated by fossil data, in addition to DNA. He states that in just "one grouping" he "kn[e]w that from fossils."

Common Ancestor. This particular trait refers to the discussion of a common ancestor, ancestry, or most recent common ancestor as the reason for their grouping. They may also discuss other criteria which were also coded for, but often individuals would cite ancestry as the reasons for the groups they created.

Bhakthi, a novice, for example explained what an ancestor was and stated that it gave rise to a group, implying that it should be the reason for the groups' existence. She stated,

> And **each one of these nodes** represents a **common ancestor**. And **each one of those branching** represents like...Sorry, so let me say it like a branching represents, like an evolutionary event. **Event that gave rise to like a new group**. And that **group can be like, whatever we want it to be**. It can be a species. It can be like, insects, for example.

Similarly, expert Zoey talked at length about common ancestry explaining group membership,

> And so **they shared the most recent common ancestor** pretty recently. Way more recently than any of the mammals shared a recent, shared an ancestor. Well, I would say, you know, probably **what you would do is you would group them into some groupings, right to all of the plants are more closely related to each other**. All of them, mammals are more closely related to each other and then a bigger group than that would be like all the animals, right? And then a, pretty long time ago, **all of the plants shared a common ancestor**.

Here she suggests that the reason for the existence of the groups in question is due to ancestry by suggesting that "you would group them" and repeatedly follows with relatedness as a reasoning though she does not explicitly suggest so.

Normativeness

Generally speaking, many of these criteria can be used for normative tree construction but they vary in their phylogenetic informativeness, i.e., whether they can give valuable information on whether or not two items can group together. Certain traits that are more labile because they are easy to change (e.g., coat color) or do not change at all across many taxa are less useful and therefore less normative. Typically, genetic traits and morphological or physiological traits are most often used (Lee & Palci, 2015; e.g., Santini et al., 2013), and behavioral traits are used infrequently (e.g., Robillard et al., 2006; Stuart et al., 2002).

The use of common ancestor as an explanation for grouping poses a special case in that it does not directly describe the intrinsic or extrinsic properties of the taxa in question and can in a sense be considered the most normative. It is the most proximal to the goal of grouping, i.e., to have taxonomic groups represent evolutionary history by hypothetically grouping things according to their ancestry. Systematists use traits to estimate these ancestors to differing degrees, either estimating their presence or specifics like the age of splits. So, in this sense, one could argue that systematists are looking ultimately for "similarity" or commonality in the sense of shared ancestry by looking at the "similarity" of taxa, based on their traits. (Here I refer to a general colloquial sense of similarity and not necessarily calculations of similarity, i.e., number of trait differences rather than number of events that explain trait differences).

The trait that is likely the least normative is that of domestication. Domestication is the result of human forces. While in a sense individuals may have domesticated a single group, such as what we have done with the mustard plant (Mabry et al., 2021), the fact that organisms are domesticated is a weaker grouping criterion. The goal of reconstructing evolutionary history is typically focused on the distant past and not more recently (though of course that is not always

true)—humans were the drivers of said artificial selection and often they would drive evolution towards desired traits which can be convergent in nature (e.g., large reproductive structures).

Environment is also equally as nonnormative given that the environment is often the causal factor for evolutionary change via natural selection. As such, many organisms undergo convergent evolution and arrive at similar traits which can negatively impact the estimation of ancestry; in other words, things that look similar because they live in similar circumstances can mislead us into thinking that they are more closely related than they really are.

Experts versus Novices

In this dataset, experts and novices were both observed using emergent traits at exceedingly high rates. Experts that used emergent traits to explain their grouping categories were 91%, while 100% of novices did so. This is unsurprising given that many of these traits are typically visible macroscopic physical traits. Indeed 91% of experts cited physical traits (these include morphology, color, etc.) while 94% of novices did so. The task included visual "data" as it contained photos, so this was unsurprising. However, it is additionally unsurprising given that we likely use visible data to categorize these outside the context of this task. I omitted the usage of common names in the task or any writing as I did not want the often misleading common names to sway individuals' perception of a group. Additionally, I did not want interviewees to lean towards a classroom context and what was considered "more correct." As the "natural" setting of categorization likely comes from visual data primarily, this presentation of the task is likely closest to what we individuals would be doing outside the context of an interview or classroom. Individuals are unlikely to be learning about new organisms without visual information. It is likely reasonable therefore to suggest that experts and novices both still use emergent traits, typically visible, to group organisms, regardless of their expertise.

It is also notable that novices were more likely to report behavioral, nonphysiobehavioral, and metabolic traits more often (Table 11). These generally tend to be used less frequently by systematists for determining taxonomic groups and evolutionary relationships either because not phylogenetically informative (like they are liable to change too often or do not differ at all between taxa), difficult to characterize, or only provide a small source of data for what it is (e.g., being diurnal is a large part of their existence but that one trait would only provide a single datapoint for estimation). As novices lack access to, both in the context of this task and in the classroom, other less obvious data like genetic information, it would make sense that they use this other (mostly) visible or easily apparent data to characterize organisms. Indeed, fewer novices reported using genetics; however, it is notable that the number is still close to that of experts. Experts typically use both genetic and morphological data to build trees nowadays, so it is surprising that less than half of experts cited it. This might suggest either a stronger reliance or tendency for people to stick to visual cues for grouping despite likely knowing the normative ways of grouping organisms. Alternatively, it could speak to the tendency for some of the experts to use rhetorical moves that suggested they had students in mind or were performing what they would have or have done in a prior lecture.

Another significant finding of this study is the pervasiveness of the environment and domestication criteria in novices' responses. Novices used the environment as criterion for grouping taxa 95%, while 50% of experts did so. Similarly, 22% of novices compared to 18% of experts exhibited the domestication criterion. Both of these criteria are notable in that they are much less scientifically normative compared to the other criteria in this dataset. It is perhaps therefore unsurprising that novices used these criteria much more often than experts, especially the environment criterion. This speaks to several potential issues. In the classroom, instructors

often observe that evolution and natural selection are conflated. Evolution refers to a change in traits whether they are genotypes or phenotypes in a population over time while natural selection could be considered a force or mechanism that causes evolution to occur. To my knowledge, the reasons for this conflation are not clear but taken together with these results they could indicate a confusion or conflation of what factors are causal or outcomes. In the case of evolution, nature and natural selection would be the causal factors and evolution would be the outcome. In this case of determining relatedness, the outcome is the reconstruction of evolutionary history and the "causal" factor would be the appropriate intrinsic properties including genetic traits, etc. Environment would be an outcome of this estimation process. This is not however a criticism of lay methods of grouping organisms, but rather a recognition that intrinsic properties, while useful in other contexts of grouping objects (e.g., pencils are often used in classrooms), they may not be useful here and that clarification might be necessary during instruction on causality.

This of course relates to one of the primary concerns of science as characterized by Next Generation Science Standards (NGSS; NGSS Lead States, 2013). NGSS identifies "cause and effect" as a cross cutting concept that underlies much of science. These findings may suggest that additional attention to these types of larger scientific conceptual practices may be necessary in this particular context of understanding evolutionary relatedness and tree thinking. To my knowledge, no such study investigates the pedagogical utility of connecting epistemological views of science regarding causality to this domain (relatedness and tree-thinking) and therefore could be a future avenue of research.

Another striking finding is the usage of living status as a grouping factor. Individuals who often recognized that all the organisms were related to one another did think to offer a criterion for such a grouping and that was always the fact that organisms are alive. One particular

novice even went so far as to make a distinction between what was alive and what was in her description of what constitutes each taxonomic group she proposed. Experts and novices were observed equally reporting this criterion (of course, I again remind the reader that no statistical tests were performed given the study's small sample size). This lack of difference seems reasonable as there would likely be no difference in the rate of noticing that the single feature that groups all living things and relates all of them is the fact that they are alive. I can think of no reason experts would be more or less attuned to such a feature nor the reverse. Additionally, such a recognition is a fundamental understanding of the world and also likely taught early suggesting that its appearance in this dataset may constitute the maximum rate of noticing by a population that might occur for such a criterion. In other words, this may be fully "learned" by a population such that the proportion exhibited here is what we would "normally" expect and that the "full" expression would thus be present in both groups as there is nothing more to learn. In other words, the novices in this study would also be the experts.

Fossils were the least frequently observed criteria in the dataset with only 18% of experts mentioning it and only 6% of novices. That 6% consisted of a single individual mentioning fossils as a grouping factor. Both of these findings are perhaps somewhat surprising as I expect that fossils are not an uncommon or unknown criterion. Were I to survey novices about their knowledge, I would anticipate that most people would know what they were; dinosaurs and their fossils are a frequent topic of interest among small children so I would assume they are not that unheard of. Therefore, assuming that individuals have heard of fossils, it is surprising to note that no one mentioned them, especially as a large proportion of the individuals in this study did indeed cite genetic traits as being a potential grouping criterion. Genetic traits are not given parts of the dataset but they were still cited. Fossils were also not given and they are widely known yet

they were not given. This could be for twofold reasons, (1) individuals might not consider non-extant organisms as being important for groupings; and (2) the mention of fossils here might be more of a nod toward more normative methods that involve using fossils for time calibration. These fossils are implicitly grouped with organisms but are not necessarily the target of said reconstructions or groups. Individuals in this task might therefore be attuned more toward extant taxa and not consider extinct taxa. More inquiry is required to investigate the reasons for said omission, though I offer that said omission might be warranted given the type of relatedness understanding we would like to instruct novices about.

Finally, the last and one of the more unusual criteria is the discussion of an ancestor or the normative term of "most recent common ancestor." This criterion was provided more frequently by experts; 64% of experts exhibited this criterion compared to 39% of novices. While its appearance as a criterion is surprising (as participants would word their responses such that ancestry was the reason for the groups provided), it is also unsurprising. Ancestry is essentially the ultimate grouping factor as the goal of systematics is to reconstruct the history of organisms and classify them accordingly. However, we have no time machine to investigate what actually occurred so we can only estimate what may have occurred using other criteria. The criteria used in estimation of ancestors one could argue as being more distal reasoning for generating taxonomic groups compared to the more proximal criterion of ancestry.

Authority and Essentialism: Special Cases of Criteria and Strategies

Before closing this chapter, it is also worth mentioning that appeal to authority and essentialism were not associated with criteria but could be argued as containing their own criteria. They could also be considered special cases of the isotype trait in that every member of the group is assumed to share the same criteria. Appeal to authority represents individuals using

outside resources as their reasoning or explanation for why taxonomic groups were grouped the way they were. In that sense, one could argue that the outside authority's decision on what is "true" is itself the grouping criteria. Similarly, essentialism argues that members of the same category all share some sort of intangible essence that characterizes them. In this sense, the criteria would then be the presence of said vague essence. I mention this here before I discuss these categories and their relationships in greater detail in Chapter 5.

Chapter Summary

In this chapter I have discussed the results of my analyses of the data for two research questions: (1) What are the ways in which experts and novices understand relatedness and what is relatedness? and (2) How do experts and novices group organisms? I have discussed my three primary categories that came out of a Grounded Theory analysis of transcripts and artifacts from 11 experts and 18 novices. The three primary observed components of relatedness understanding are structural metaphors, grouping strategies, and grouping metaphors. Structural metaphors refer to the structure or blueprints to which taxa are assigned and are typically spatial in nature. Grouping strategies are the ways in which individuals decided to assign taxa to these structures and they refer to the different ways in which different criteria were used. Grouping criteria were those properties or criteria by which individuals decided, in combination with grouping strategies, how taxa should fit inside the structural metaphors discussed. Experts and novices were observed using each of these subcategories at varying degrees and in varying ways. Individuals often displayed more than one of each subcategory.

In the next chapter, I discuss some of the broader takeaways from these data. I discuss what these data mean as well as implications for these findings in terms of both research and pedagogy.

Chapter 4 contains unpublished material coauthored with Komperda, Regis and Lo, Stanley. The dissertation author was the primary researcher and author of this chapter.

CHAPTER 5 – RELATEDNESS: WHAT DO THESE FINDINGS MEAN

Summary

In this chapter, I discuss the major implications of the findings of this study. I begin by discussing the results pertaining to research question one, where I discuss the ways in which this study impacts our understanding of the structure of relatedness understanding. I follow with the discussion of research question two focusing on the ways in which participants grouped organisms. I then discuss some broad implications for the entirety of relatedness understanding. I then discuss the implications for research and teaching. I end with the discussion of future directions, limitations, and the summary of this chapter.

Research Question 1: The Nature of Taxonomic Reasoning Structure

The first objective of this study was to explore the differences in how experts and novices understand the concept of "relatedness" in taxa. I inquired about participants' interpretations of relatedness and how they would group organisms based on this understanding. A significant finding of this study was that a primary factor in their understanding of taxonomic relatedness was the structure of taxonomic groups. In essence, I uncovered patterns of taxonomic organization irrespective of taxon identity or other factors such as categorization strategy.

I found that such structuring aligns with the embodied perspective's conception of conceptual metaphors. Often, participants employed spatial language, among other linguistic tools, indicating they were mapping these taxa onto some kind of metaphorical structural framework that was largely spatial in nature. I identified seven distinct structural metaphors in the participants' responses: vertical, flat, tree-like, anthropocentrism, family, storytelling, and siloed. Several of these structural metaphors exhibited subtypes relating to the directionality of

said structure. Participants often used a variety of structural metaphors and the type of structure chosen by individuals appeared to vary with their level of expertise.

In the following paragraphs, I summarize these findings as they pertain to the first research question (*In what ways do experts and novices conceive of organism relatedness?*) as well as their implications for research and teaching. I discuss the nature of taxonomic reasoning that these results imply, specifically the ubiquity of structural metaphors across experience groups as well as common features of the structures that arose in the results. I also describe the role of the tree as a metaphor in normative speech about relatedness and what this study might suggest about its pervasiveness in the intervention literature when compared to reality.

Commonalities Among Multiple Structures: Group Size, Hierarchicalness, and Time

One critical finding of this study is that the subcategories of structural metaphors appeared to vary along three dimensions or qualities. These qualities were present at varying levels in each of the structural metaphors observed in this study: group size (directionality), time, and hierarchicalness. In the following sections, I discuss the form these qualities take shape in, how they vary, and what this variation might mean for relatedness understanding and tree-thinking research.

Hierarchy

One primary theme associated with the majority, if not all, of the structural metaphors described is hierarchy. Depending on the definition of hierarchy employed, it is arguable that each variant adheres to hierarchy to some degree. Hierarchy at a minimum could refer to nestedness and therefore be argued as being a part of all of the structural metaphors found here with, for example, the tree-like structure suggesting some lineages are part of or nested as part of some larger taxonomic group which includes the ancestor and those tips. Alternatively, in the

case of the vertical structure, some larger taxonomic groups are named by the participant and members of said category are enumerated as belonging to said category.

However, as all of the individuals in this study are from institutions in the United States, they may also have been calling upon common Western conceptions of hierarchy which also include elements of verticality (which is intrinsic here to the vertical structural metaphor described) and value-ladenness. Indeed, the linguistics literature has very often discussed the Western conception of verticality being associated with value, where "high" things are "good" and "low" things are "bad" (Gottwald et al., 2015; Meier & Robinson, 2004). It follows of course that this understanding bleeds into other understandings of hierarchicalness in an everyday sense with, for example, "higher" positions in a company, of course, being considered more important or prized by some in Western society.

This latter definition of hierarchy, therefore, alternatively suggests that some of the structures are more hierarchical in nature than others. The vertical structure in particular would be the most egregious of the bunch, but others would vary in their hierarchicalness based on these definitions taken together. This variation in hierarchicalness could suggest a potential reason for the expression of certain misconceptions in tree-thinking described in the literature aside from those referring to parsing of the visual field. The "backbone" misconception where individuals perceive in diagonal trees the backbone to be a single unit signifying a linear sequence of events (Novick & Catley, 2007) as well as the misconceptions suggesting that certain sister lineages or older lineages are more "primitive" or "basal" implying differences in value (Bronzati, 2017; Krell & Cranston, 2004) might be indirectly supported by this understanding. It may be that in addition to the aforementioned interpretation due to the features of the diagram that the underlying understanding of relatedness may be inadvertently supporting

these ideas. Indeed, there is some support for alternative understandings supporting some of these misconceptions with Catley et al. (2010) finding that careful wording of prompts and interventions does impact student responses.

That being said, it is noteworthy that a larger proportion of novices exhibited the flat case, even more so than the experts, which out of all the structural metaphors is essentially the least hierarchical and could be considered one of the most normative outside of the tree-like structure. This may reflect the process of learning, whereby the use of trees, which could perhaps be considered more hierarchical than the flat case, imposes a somewhat more hierarchical understanding onto expert thinking.

Alternatively, this difference in expression between the two groups might be indicative of cultural differences between the two groups. Indeed, experts and novices, aside from their experience, are different populations, with, for example, faculty notably being older and less diverse (Kaplan et al., 2018; O'Brien et al., 2020) than their undergraduate counterparts. Differences in generation, for example, as mentioned in Chapter 4, could explain differential expression of hierarchal views, first with more exposure to strictly historical perspectives such as Aristotle's "Great Chain of Being" (Matuk, 2007; Matuk & Uttal, 2012). Alternatively, the United States has been decreasing in religiosity and affiliation of which the former could coincide with traditional understandings of nature and the roles of humans in nature where organisms are linearly hierarchical with humans present at the apex of such a sequence (Wilkins, 2022). (Note that here I am intentionally conflating anthropocentrism with hierarchical perspectives of nature as they are typically closely related as in the "Great Chain of Being" as humans usually are depicted at the apex of such hierarchical understandings.) These findings may also be explained by improvements in education that attempt to minimize dominionism

(human dominance over nature) in students' understanding of nature which may be more successful than anticipated (Almeida & Vasconcelos, 2013; Quinn et al., 2016; Ross, 2020). Finally, the increased popularity of animal rights and associated movements such as veganism (McCarthy, 2020) could also underscore this altered view of hierarchicalness and orderedness of nature in American society.

Directionality (Group Size)

Another area that these findings recommend additional attention is the idea of directionality or group size. Many of the structures elucidated in this study, specifically the vertical, tree-like, and flat structural metaphors, seemed to vary in terms of the way they were discussed by participants, with participants choosing to either start their reasoning explanation with smaller taxonomic groups and working their way to larger taxonomic groups or starting from larger taxonomic groups and working their way to smaller taxonomic groups.

One way to interpret these differences is that they may simply reflect a difference in the ways in which these structures are presented. In other words, these may simply be just subvariants of major categories that exist simply due to differences in other prior knowledge. For example, a simple preference for looking at smaller groups before explaining larger taxonomic groups may have explained this expression or perhaps some tendency to examine smaller components before larger components of a problem. Were that the case, then we could actually argue that students are engaging in this behavior, i.e. specifically smaller to larger taxonomic groups, much more than anticipated and, from a formal perspective, which could be considered more normative. (Typically, we describe tree construction as an estimation of ancestors or evolutionary relationships backwards in time using evidence from extant organisms.) In other words, students are engaging in more normative understandings than we may have anticipated. If

that is the case, that suggests their everyday understanding is not translating into their tree-interpretation, assuming that students in the novice group were experiencing difficulty with tree-interpretation. This would therefore highlight the need for examination into the ways in which these ideas are or are not recruited into tree-thinking understanding. If students are indeed using more normative structures, then ideally they should be integrated and scaffolded into their normative tree-interpretation which appears to not be happening.

Another way to interpret these data is that the subcategories represent differences in cognitive processing. I noticed during the analysis that the ways in which individuals often discussed their reasoning seemed to coincide with the stage of reasoning they were engaged in. For example, individuals might slowly work their way through the smaller taxonomic groups and work their way upward. The pace and pauses could indicate that they were in the process of using these new stimuli to generate new knowledge based on their prior structures. From a constructivist perspective, that behavior could be considered a type of accommodation where they attempted to reconcile the new stimuli, presumably unfamiliar like the deer with large canines, with their prior understandings of what these groups were, such as deer. Given the nestedness of all of the structures, they would need to start at the bottom and rework their understandings of the higher-level groups to report their understandings. Alternatively, those that engaged in top-down types of thinking may have been undergoing assimilation. They have already decided what taxonomic groups were and were just taking the time to recognize and fit each taxon into the preexisting concept of each group. Perhaps some amendment might occur here or there where individuals would largely suggest that the musk deer would be related to the other hooved animals but appears somewhat different because of that particular feature.

It is notable, however, that experts and novices engaged in both behaviors and that the reasons for doing so in experts might be different. It follows that experts likely have some idea about the ways in which taxa are grouped and therefore should exhibit more of the top-down understanding but many also exhibited the bottom-up variant. That variant is considered a normative, operationalized, and formalized method of tree construction. Two experts in particular (Professor X and Zoey) who clearly had more systematics instruction experience exhibited cues that they were going through the motions of "performing" these steps for an imaginary student audience. It is therefore worth noting that experts and novices engage in both of these behaviors for different reasons.

Time

The final theme apparent in the structural metaphors present in this study is that of time. Time is an intrinsic part of evolution education and, much like chemistry, may be a cause for confusion given its inconceivable scale. Where chemistry is often concerned with the microscopic, i.e., particle behavior (Taber, 2013), organismal biology is often concerned with overly macroscopic, i.e., evolution (Stenlund & Tibell, 2019), which occurs at massively large timescales.

Each of the structures, at least from an external representation perspective, seemed to prioritize or deprioritize the appearance of time with time not being explicitly modeled or referenced in that structure. The storytelling case, for example, would be a structure that very explicitly models or contains time in that a series of events occur—indeed many individuals who discussed large timescales or incorporated discussion of time often did so in the context of storytelling which makes sense. In contrast, the flat structural metaphor offered little to no explanation or distinct representing of time. If we take the flat structural metaphor to be

represented by a concentric circle diagram, we could argue that relative time is pointing at the viewer and, while partially discernable, less so than something like a tree where branch length could scale with time.

The variation in said categories begs the question whether they have any impact on tree-interpretation and normative relatedness understanding. Those structures that possess greater qualities of time, for example, might be responsible for (or coincide with) those nonnormative conceptions that incorporate senses of time. Indeed, human conceptions of time are a universal aspect of our cognition—we are organisms that can conceive of relatively large timescales that often incorporate egocentric (Núñez & Sweetser, 2006) or world-based (Núñez et al., 2012) spatial metaphors that are typically linear. The storytelling structure, for example, might predict more teleological views of evolution (Jenner, 2022; Kampourakis, 2020; Trommler & Hammann, 2020; Werth & Allchin, 2020) and certain interpretations of phylogenetic trees (Matuk & Uttal, 2012). Assuming this is true and that the visibility of time is linearly correlated with the appearance of such understandings, then those understandings which have some moderate aspects of time (e.g., tree-like, or phylogenetic trees) might correspond to the appearance of those misconceptions. Therefore, not only would the visual attributes of the phylogenetic tree as theorized by Novick and Cately (Catley & Novick, 2008; MacDonald, 2014; Novick & Catley, 2007; Novick & Fuselier, 2019; Novick et al., 2010, 2011) correspond to misinterpretations of trees but the conceptions underlying said diagrams (whether or not these underlying conceptions were normative) are also impacting their understanding.

Research Question 2: Taxonomic Reasoning Strategies and Criteria

The second objective of this study was to investigate the ways in which experts and novices engaged in taxonomic reasoning. I found that two major components constituted these "ways" of reasoning about organismal grouping, i.e., grouping strategies and grouping criteria.

Grouping strategies referred to the ways in which individuals determined whether or not taxa belonged to a proposed taxonomic group. This strategy was distinct from grouping criteria which referred to the types of evidence or data individuals would use to determine these taxonomic groups. Therefore, the "ways" of reasoning about organismal grouping consisted of a strategy, often combined with a criterion, to determine taxonomic groups.

The three primary grouping strategies were found—similarity, essentialism, and appeal to authority—of which similarity referred to two major subcategories of strategies: prototype and isotype. A range of criteria were found which I will not enumerate at length here and are available for your perusal in Chapter 4. Essentially, they referred to data that included morphological, genetic, environmental, behavioral, and ancestry traits.

Experts were predictably more likely to use normative criteria, but some of which were unexpected like the use of ancestry more often than genetic data, though the two are essentially both normative but differ in their proximity to the taxonomic group in question (proximity here I am referring to criteria that are dependent on one another and those in closer proximity are more highly dependent). Novices, on the other hand, tended to use less normative outcome or causal type criteria such as the environment.

Experts and novices exhibited for the most part all of the strategies at similar rates, but with the two less normative ones (essentialism and appeal to authority) slightly more in novices. However, it is notable that experts did indeed exhibit some of these with appeal to authority in

particular being expressed differently in experts. This difference was more of a reflection of lack of confidence about different domain or reflective of preformed organismal knowledge structures for which the details have been forgotten as is common in experts. What is particularly notable, however, is that both experts and novices exhibited very similar rates of isotype and prototype strategies, though novices were expected to rely more on prototype similarity understandings.

In the following sections I discuss some of the broader takeaways of these results, specifically the sources and effects of essentialism, a commonly discussed misconception in the child psychology organismal categorization literature (Shtulman & Schulz, 2008; Sousa et al., 2002) and in the evolution education literature (Betz et al., 2019; Coley & Tanner, 2012). I also discuss the similarity-based strategies and how they compare to the existing tree-thinking literature.

Similarity: Prototype Theory, and Isotype Strategies

One primary finding of this study is the appearance and relative prevalence of the two similarity strategies: prototype and isotype. Given the definition of said similarity and given the vagueness of the "similarity equals relatedness" misconception in the tree-thinking literature, these results suggest that this, and potentially other misconceptions, require additional attention and review.

The tree-thinking literature at length discusses how "similarity equals relatedness" is a very common misconception with somewhat varying references (Baum et al., 2005; Dees et al., 2014; Gregory, 2008; Halverson, 2011; Halverson et al., 2011; Kummer et al., 2016). Essentially it can refer (1) broadly to simple similarity determining relatedness, (2) to the tendency for analogous traits (due to convergent evolution) and phylogenetically uninformative traits to impede or not impact estimation of relationships, or (3) to specific methodologies and

operationalizations of the term "similarity" (referring to distance) between organisms which are now less favored. In like manner, similarity referred to the state of having traits, properties, qualities, etc. which may be intrinsic or extrinsic to two individuals or an individual and a group that determines group membership. In other words, these strategies assume that properties must be in common in some respect for them to be called similar and therefore grouped together.

Taken together, depending on the definition of similarity employed, it is reasonable to assert that experts exhibited this "misconception" very frequently and at similar levels to novices. This could suggest that our persistence in trying to eliminate this misconception warrants some revision. Indeed, experts do have some elements of similarity that are different; definition (3) above refers specifically to methodology which could be considered analogous to strategy and criteria. Rather than repeating this phrase to students (as I have often experienced myself) without nuance, it may behoove students to experience an intervention that makes more explicit the intricacies of relatedness understanding as this study suggests. In other words, these results highlight the fact that the strategies, criteria, etc., can be considered distinct components of relatedness understanding and such a highlight perhaps helps students achieve the understanding that experts want, i.e., that some "relatedness" is essentially being estimated by similarity of taxa's properties that are outcomes of the evolutionary process (morphology, etc.).

These results also suggest that normative understanding is somewhat more complex perhaps than the two subcategories of similarity observed. Indeed, normative understanding, depending on the level of operationalization and methods used, could feasibly span both prototype and isotype strategies. The isotype strategy, for example, could be analogous to a purely synapomorphic strategy, especially those cases that are used in biology classes with mammals as being defined as being able to produce milk. In that sense, the trait used to group

mammals is indeed present in all daughter taxa and there is no uncertainty about that. However, that perspective is turned on its head when you realize that taxa often lose or regain traits. Mammals are often described as having hair; however, some mammals like cetaceans do not always have hair. Hair in that formulation, would not necessarily be present in all members and therefore “mammal membership” would not necessarily be defined by the strict presence or absence of traits. Rather different traits have different weights or probabilities of occurring in members, which is more analogous to a prototype understanding. The difference between the normative understanding and prototype understanding, if there is one at all, could potentially be greater formalization and metacognitive recognition of those steps; indeed, some of those marked as exhibiting prototype theory struggled to describe why they grouped things the way they did, simply expressing that the organisms “looked like” they belonged to a group.

Essentialism and Its Possible Effects

A significant concern of evolutionary biology education literature, as well as the child psychology literature, is the usage of essentialist reasoning to characterize taxa or describe evolution. Essentialism is a cognitive phenomenon recognized across various domains ranging from biology to simplified perceptions of sociological concepts like gender. Essentialism is often described as a belief, often noncognizant, that members of categories possess some intrinsic "essence" defining their membership. In sociological contexts, this might appear as the belief that a particular essence defines gender. When pressed to explain why a particular entity belongs to a category, individuals with this essentialist belief or perspective might circularly argue, "A woman is a woman because she’s a woman." Essentialist reasoning is typically considered insufficient to describe natural phenomena, and researchers, education or otherwise, suggest it is a misconception and should be avoided.

The findings of this study suggest that undergraduate students, and to a limited extent experts, engage in some essentialist reasoning, though it appears to be at a lower rate than young children. Young children have been shown to exhibit essentialist reasoning in reference to taxonomic groups, and these general ideas about large taxonomic groups typically change over time, presumably with education and life experience (Gelman, 2003; Gelman & Coley, 1990; Muradoglu et al., 2022). Therefore, it was not unexpected to find traces of essentialism within this study's data set, reflecting taxonomic groupings among undergraduates. While this study does not directly generalizably quantify the overall prevalence of such beliefs across the population, there is sufficient evidence in this dataset to suggest it continues to be present, albeit diminished, in undergraduates and therefore a potential target for amelioration.

Essentialist understandings in this dataset, however, appear dynamic and malleable in adult learners. Whenever there was any indication, whether intentional or not, that their reasoning might be incorrect, participants would provide an alternative strategy and use alternative evidence or essentially repeat their essentialist statements. This fluidity might stem from the fact that adults recognize the insufficiency of essentialist explanations and feel pressured by the environment (an education study) to provide a more detailed response. Alternatively, this fluidity may be a result of the inherent stability of the connections to these ideas (Gelman, 2003). While some essentialist beliefs, like those regarding gender, might tie more deeply into core convictions and are therefore more stable, taxonomic groupings—anchored to evolutionary principles—may be less controversial and therefore more changeable (Pugh, 2017). Still, considering the potential conflict of evolutionary principles with certain religious views, particularly in the United States (J. D. Miller et al., 2022), it is worth investigating students' beliefs about evolution in various evolutionary thinking aspects. Probing

deeper into these issues might influence participants' responses on taxonomic grouping methodologies. Other potential areas of exploration include humans' perceived relation to nature, which might influence hierarchical views of organisms and the required educational approaches to establish a normative view.

An intriguing aspect of essentialism is its sporadic adoption by experts as a heuristic. One expert in my dataset, for example, occasionally employed an essentialist strategy. Under certain pressures, experts that use essentialism may recognize its inappropriateness but still use it for brevity as has been found in other situations where abbreviated heuristics were employed based on the features of the task (National Research Council, 2000; Smith et al., 1994). Some may have been prompted by the potential simplicity of the task (photo grouping) and therefore employed a simpler heuristic for grouping them. Alternatively, experts are not immune to misconceptions, themselves often displaying biases such as "plant awareness disparity," a phenomenon where individuals overlook plants' importance (Achurra, 2022; Jose et al., 2019; Wandersee & Schussler, 1999, 2001). For instance, the systematist, who did not specialize in plants, essentially remarked (I am paraphrasing) that "plants are plants because they are plants." This is not a critique of the expert but rather an observation of how deeply ingrained biases, shaped by our experiences or lack thereof, influence our perceptions. Essentialism, therefore, might simply in some cases be an expression of gaps in knowledge which individuals may or may not be cognizant of.

The Ubiquity of Relatedness Understanding

Essential Components of Relatedness Understanding

A major takeaway of all of the findings of this study is the broad ubiquity of the components of relatedness understanding. While efforts were indeed made to describe all

individuals as is commonly part of Grounded Theory (Glaser & Strauss, 1967), it is notable that each set of understandings do share similar properties. In other words, the fact that all participants exhibited structures, strategies, and criteria suggests that these hypothetical constructs might be essential components of relatedness understanding heretofore not described in the tree-thinking literature.

While categorization itself and broad categorization of taxa has been explored in other domains (Gelman, 2003; Muradoglu et al., 2022; M. G. Taylor et al., 2009), this has not been discussed with reference to biology education and tree-thinking education. As discussed in Chapter 2, the only individuals who have begun to examine these potentially underlying understandings of relatedness outside the context of a tree are Matuk and Uttal (2020) who asked non-majors to draw and describe the ways in which they grouped hypothetical organisms.

Dynamic Taxonomic Reasoning

These findings also suggest that these relatedness understandings are particularly labile, with many individuals exhibiting many different strategies, criteria, and structures over the course of the interview. From a knowledge-in-pieces perspective (diSessa, 2018), this is perhaps unsurprising. Different contexts, here recognized as perhaps the evidence or taxa given to the participants, might spark use of different structures or meanings associated with structures based on whatever experiences they might have had. This was particularly true of the case with plant awareness disparity, with many of the novices exhibiting siloed structures, likely because they lacked information about the constituent smaller taxonomic groups or because plants in particular were particularly unfamiliar. Individuals, for example, might separate plants out of their understanding of relatedness while giving greater detail, and therefore a different structure, to more familiar, charismatic taxa such as land vertebrates (Trimble & Van Aarde, 2010).

Alternatively, this dynamism might also suggest that individuals' understanding of grouping might be, for whatever reason, inherently more labile compared to other types of understandings described in the education literature. Of course, traditional constructivist perspectives suggest that some understandings, especially those that are more fundamental, are particularly resistant to perturbation and are therefore more stable (Piaget, 2001; Pugh, 2017). As these understandings likely do not interact closely with core understandings of the world (e.g., ideas about the self and the world) they are more easily changeable. To my knowledge this high level of changeability has not been previously shown in the literature and is cause for additional investigation. This may suggest that, while tree-thinking itself is difficult, the ideas underlying tree-thinking, i.e., relatedness, are more easily changeable. If it is more easily changeable and we can create interventions that could better support scaffolding these relatedness understandings without the context of a tree to those that use trees, then this finding might be great use to practitioners and researchers.

Implications for Research and Teaching

In the following section, I discuss the implications of all of these findings broadly for both research and teaching. I discuss the ways in which these results suggest that we may be too closely married to our use of the tree metaphor. I also discuss the ways in which structural metaphor can be further investigated using gestural studies and potentially be used for determining student conceptions. I also discuss how these results suggest how we might need to amend our learning goals to accommodate some of the findings of this study. I also broadly discuss how process differences in taxonomic groupings could correspond to particular tree-interpretation behaviors and that observation of the latter might also give insight into students'

behaviors. I finally discuss what types of phylogenetics interventions might need to be tried in lieu of these results.

Prevalence of Metaphors in the Field

During the process of constructing and implementing this study, it became exceedingly apparent the extent to which the tree and related family metaphors pervade the language of the field. While obviously the tree is an explicit hypothesis or model of evolutionary relationships (University of California Museum of Paleontology, 2021b), it did not initially occur to me how intrinsic this understanding has become as other metaphors or ways of knowing are (Trommler & Hammann, 2020; Werth & Allchin, 2020). We often employ terms like "crown group," "branches," "leaves," and "branching," all of which likely invoke tree imagery and therefore elements of the tree metaphor. Additionally, systematics' normative terminology borrows from the related concept of family trees, such as "sister group," "daughter," and "ancestor." These linguistic choices highlight the extent to which this metaphor pervades the field and, in my opinion, may have unwittingly impacted the way we have conducted interventions to support tree-thinking.

The prevailing focus in tree-thinking literature centers on interpreting trees as diagrams, often overlooking the foundational concepts they represent, such as taxonomic relatedness, bifurcation speciation, and other core ideas. While much of the literature emphasizes identifying the most easily interpretable tree type, these approaches, though valuable for cases where novices need to interpret diagrams quickly (MacDonald & Wiley, 2012; MacFadden et al., 2007; i.e., lay people in museums) or with a lower cognitive load (Ballen & Greene, 2017; i.e., lessons about taxonomic groups themselves), they do not necessarily scaffold's students understanding of trees' underlying principles or assumptions. Indeed, it is quite telling that out of all of the

interventions I have come across in the literature, none let students create their own unique representations of said phenomena (Bokor et al., 2014; Firneno et al., 2023; Huang & Whittall, 2018; Kuzoff et al., 2009; Lents et al., 2010). Almost all, if not all, assume that the tree-structure should be given, or at least highly recommend it.

This is in contrast to the K-12 science education literature which supports engagement in and learning of authentic research practices of modeling, representation generation, and negotiation of norms among students (NGSS Lead States, 2013; Windschitl et al., 2018). These approaches posit that students might be better supported in acquiring normative understanding if they are encouraged to formulate their own hypotheses, models, and representations of phenomena, and then refine with peers. Additionally, there is some evidence to suggest, that beyond given multiple representations which have a spotty record at best, learner generated representations (Ainsworth & Scheiter, 2021; Matuk & Uttal, 2020; Van Meter & Garner, 2005) may better support learning, at least when scaffolded appropriately. The mechanism by which it may do so may involve a number of factors such as exposure to multiple representations, identification as a doer of science, engagement in "authentic" science practices, among others. Regardless of its efficacy, the dearth of these methods in the literature suggests that at a minimum it should be tried. In addition, many of the interventions that exist in the literature have very little evidence for their efficacy beyond broad metrics such as unvalidated test scores or grades if any are given at all (Davenport et al., 2015; Firneno et al., 2023; Gendron, 2000; Gibson & Cooper, 2017; Huang & Whittall, 2018; Kumala, 2010b; Perry et al., 2008). They also do not attempt to understand the mechanisms by which these work, which may lead to improvements in subsequent iterations of said interventions.

In summary, these findings suggest that the pervasiveness of the tree metaphor may have inadvertently negatively impacted our attempt to support tree-thinking in that it ignores the underlying understanding of phylogenetic trees, i.e., relatedness, as evidenced by the tree-thinking interventions that only prioritize tree-building or inherently assume that trees are the only and best way to represent these relationships. It also ignores other metaphors or other ways of thinking (represented by the results of this study) that could be therefore recruited into tree-thinking. These findings therefore suggest that interventions investigate the scaffolding toward normative understanding but leverage these underappreciated understandings of taxonomic relatedness. To my knowledge, only one such study has examined the explicit epistemological role of tree as metaphor in instruction (Marcelos & Nagem, 2012) but it primarily focused on teacher's knowledge of its epistemological role rather than its explicit use in pointed intentionally designed intervention.

Impact of Metaphorical Speech on Learning

A potential area of concern that this research may bring up is the use of metaphor in speech in both the positive and negative sense. Metaphors are used in all areas of communication and cognition and are likely a fundamental aspect of cognition (Hart & Long, 2011; Lakoff & Johnson, 2011). Their use in discourse may therefore have an effect on the adoption of ideas by students, or at a minimum, the language that they use. Teleological language use by instructors may lead to teleological language use by students and possibly a teleological understanding of evolution, for example, though they most likely do not necessarily adopt that view (Betz et al., 2019). A similar effect has been found regarding certain aspects of tree-thinking (Catley et al., 2010). While these particular examples of metaphorical utterances are not always used as metaphors by the speaker (they may use terms for reasons of conventionalization), it follows that

metaphorical ones might similarly be adopted, either conceptually or in discourse, by learners. It is in this sense that the use of metaphorical speech, especially ones that are more likely to encourage language or constructs, lead students to unproductive models of the phenomenon being described.

These findings therefore suggest that usage of these tree metaphors be investigated for their impact on everyday understandings of relatedness and tree-interpretation. And it follows that the ways in which novices' understanding of relatedness might be amended would be of interest. Indeed, using no tree language, for example, in the context of a learner generated representation intervention initially may help scaffold their existing understanding into normative understanding. Alternatively, not using such metaphors might impede their understanding, or at a minimum negatively impact their understanding and usage of such jargon, which I remind the reader is still a part of enculturation into science, i.e. the learning of social languages (Hamnell-Pamment, 2023; Marchenkova, 2005).

Another area in which the usage of utterances relating to metaphors may impact learning is in their metacognitive and epistemological identification. Indeed, some cognitive scientists suggest that one general outcome of the cognitive metaphor literature is that making students epistemologically aware of the fact that metaphor invades every aspect of their cognition might make them in turn more aware of the affordances and limitations of each metaphor (R. Núñez, personal communication, 2018). The implication is that metacognitive awareness might make learning about said concepts less difficult. While this has been true of the author's firsthand experiences, to my knowledge empirical evidence is limited though its usage in the classroom has been recommended in other domains (González Galli et al., 2020; Hammann & Nehm, 2020).

Structure, Strategy, and Criteria as Diagnostics

One consequence of these results is the potential identification and usage of the constructs illustrated in this dissertation to potentially scaffold students' understanding of evolutionary relatedness and therefore tree-thinking, assuming that they are related. Once the effect of these constructs on tree-thinking are understood, practitioners could use them to determine exactly what types of additional work might be needed to guide students toward those understandings of relatedness that are more productive for tree-thinking.

It may be that certain structures may coincide with certain types of interpretations of phylogenetic trees. For example, individuals who employ structures that closely resemble (i.e., share more features that more directly match onto their understanding—easiest example are iconic diagrams) formal phylogenetic trees may find it easier to interpret them. Alternatively, those trees might require less processing to understand given the similarities between them and mental representations. Those individuals who do not harbor such mental structures may require exposure to, and manipulation of, external representation similar to both.

Of course, interventions that measure these constructs would have to be developed and the prevalence of these constructs would need to be ascertained. But once that is conducted, the variants observed here may be of use by practitioners to determine the ways in which students currently understand relatedness such that they can be scaffolded toward normative understanding.

Backtracing as Indicators of Progress

As mentioned in the directionality portion of this dissertation, one major outcome of this study is the recognition that individuals would engage in the description or construction of taxonomic groups in different directions across taxonomic structures. This is of potential interest

in that engaging in this behavior may give insight into the cognitive structures at which individuals have constructed taxonomic groups. In other words, learners who trace from the tips to the top may not have yet settled on specific taxonomic groups. If this is the case, then it may suggest that individuals could be more receptive to changing their ideas about what constitutes taxonomic groups and that more appropriate or alternative methods for generating or structuring taxonomic groups at this point could be provided. Alternatively, it may suggest that since they are in the process of sense-making they might be more sensitive to situations requiring higher cognitive load as they are already in the process of engaging in cognitive processes that require many cognitive resources.

These practices might also coincide with other tree-thinking behaviors assuming these relatedness understandings are recruited. Students, for example, anecdotally have been observed "tracing" trees from tips to tips to determine normative taxonomic groups whereas experts typically do not. While this is likely at a minimum due to their lack of experience of looking at representations and parsing out major patterns (e.g., recognizing at a glance what a monophyletic group is) (Talanquer, 2022), it may also be the case that their unformed understandings of these taxa may also be impacting this behavior. If they are observed to co-occur, then it can suggest to instructors that students are still in the preliminary stages of tree-interpretation and can amend their instruction accordingly. Future eye tracking studies could, for example, investigate the prevalence of this behavior and aspects of relatedness understanding that could coincide or interact with tree-interpretation, in combination with interviews about said tree-interpretation.

Targets of Learning

These results also prompt a reevaluation of the tree-thinking misconception literature and accordingly tree-thinking representational competence. While some tree-thinking researchers

have indicated both directly (Halverson & Friedrichsen, 2013; Jenkins et al., 2022; Kong et al., 2022; Kummer et al., 2019; Schramm et al., 2021a; through assessment and competency guidelines) and indirectly the goals for learning in tree-thinking (Baum et al., 2005; Gregory, 2008; Meir et al., 2007; Meisel, 2010; through identification of misconceptions), they typically refer very strictly to tree-interpretation; rotation about nodes, for example, is a property that is intrinsic to trees but not to other potential representations of relatedness. We must therefore reconsider our understanding of what constitutes expert knowledge and determine how it can be nurtured through instruction, especially if we are to consider that these underlying understandings of relatedness should be considered in tandem with normative tree-interpretation. These findings suggest on some level that everyday understandings of relatedness understanding are likely used in tree-interpretation; experts were observed to use novice structures. As we assume that experts do have some different understanding of the phenomena of interest, it follows that they therefore are adding nuance in areas that are not accounted for by this study. That means, however, that our learning goals must be amended to account for them.

So, in the case of similarity or "similarity equals relatedness" we may need to amend the ways that we support normative relatedness understanding by explicitly indicating to students that they already have some ideas about how to group organisms. Alternatively, we can support epistemological recognition of those practices through, for example, brief introductions to theory of knowledge (i.e., asking students, "how do we know a plant is a plant," and having students share their thoughts with their groups). Through both of these practices we might be able to support a more nuanced understanding that suggests that we *do* in some sense examine "similarity," but that we amend our understanding to incorporate the existence of phenomena that can impact our estimation (e.g., analogous traits) as well as the problems we encounter

during operationalization of our ideas about relatedness (e.g., using one individual of a species to represent the traits of an entire species). Reconsidering these ideas about similarity, therefore, might have downstream effects on the ways in which we conduct interventions targeted toward supporting tree-thinking.

In addition, recognizing and reconsidering our misconceptions and our learning goals would also impact assessment. Given the aforementioned example once again, a reformulation or reconsideration of "similarity" would therefore change our assessments to instead of looking for any possible indication of using "similarity" to looking for those nuanced amendments and how they are employed by students to ascertain whether or not they achieved the learning goals desired.

These findings are consistent with others in evolutionary biology and evolution education. There has been recent discourse surrounding the role of teleology in expert understandings of evolution (González Galli et al., 2020; Hammann & Nehm, 2020; Kampourakis, 2020; Werth & Allchin, 2020). A variety of perspectives—philosophical, educational, etc.—suggest that teleological construals are likely intrinsic to human cognition (Kelemen et al., 2013) as scientists have been observed to fall back on teleological heuristics when pressured (Kelemen & Rosset, 2009) and Alzheimer's patients revert back to teleological explanations more often than healthy populations (Lombrozo et al., 2007). As experts likely construct their understanding of evolution from these nascent teleological schemas and amend, repress, or regulate them (depending on the model of cognition you employ), they argue it is imperative that we amend our approach to evolution education in that we support the learning trajectory by which experts have presumably arrived to this amendment, repression, or regulation

(González Galli et al., 2020; Hammann & Nehm, 2020; Kampourakis, 2020; Werth & Allchin, 2020).

Evolutionary biology also appears to be contending with such an outlook from a non-pedagogical perspective as well. Some evolutionary biologists (Jenner, 2022) have, for example, pushed back on the discourse in evolutionary biology surrounding the inappropriateness of storytelling or, as often referred to, "just-so stories." Evolutionary biologists have in the past suggested that framing reconstructions of ancestry or relatedness as stories can be reductive and obscure other factors that may explain the evolutionary history of the taxa being examined. Jenner (2022) suggests that storytelling is a part of the way people naturally think about these natural systems and that, while they do have their drawbacks as any representation of reality, they can still be potentially useful tools for formulating and motivating new questions.

These parallel developments in these domains related to tree-thinking lends some credence to the idea that our everyday understandings of the natural world might be somewhat more fundamental to human cognition and might therefore require a different approach to facilitating student learning of relatedness and, by extension, tree-thinking.

Grouping Criteria Proximity

Another way in which these findings may prove useful to the tree-thinking literature is the concept of proximity when it comes to grouping criteria. What is clear from the analyses is that individuals appear to use a wide variety of criteria to determine taxonomic groups, with novices leaning more heavily on those non-normative extrinsic properties such as the environment in which a taxon exists. It is notable that these criteria appear to exist on a continuum when viewed with a lens of argumentation. Assuming Toulmin's (2008) model of argumentation, individuals could be characterized as using varying criteria as warrants for their

assertions. While this was not explicitly examined as part of this analysis, I can make a cursory observation that both experts and novices did so. However, taken in a realist sense, these criteria vary in the proximity of their use as evidence or reasons for categorization. In other words, some of the individuals in this analysis would likely suggest, beyond those who explicitly cited ancestry, that these groups are ultimately determined by their evolutionary history. However, what is assumed and not explicitly stated is that other traits are downstream effects of evolution even though we use them to estimate evolutionary history. Stated differently, we might say these taxa group together because they share some morphological trait, when really they are related due to their history and those traits are an outcome of that history. Additional research is required in this area to determine the extent to which individuals recognize this nuanced take on evidence for their groupings and their proximity. Indeed, many experts cited these other criteria as warrants but they may still recognize that the ultimate grouping criteria is that of ancestry or evolutionary history.

Being Explicit About Methods and Cognition

The aforementioned point about criteria also coincides with a larger point that the rest of the analyses suggest, i.e., that interventions likely require more explicit attention to the more nuanced aspects of tree-construction and tree-interpretation. While this is not necessarily a new idea, these findings can suggest a few areas where some additional attention can be paid.

Criteria, for example, is one such area. As stated previously, tree-thinking interventions often do address the phylogenetic informativeness of traits and phenomena that might negatively impact tree-estimation (Huang & Whittall, 2018). However, the criteria found in this study highlight the fact that a subset of these criteria are actually outcome variables rather than intrinsic variables that define the group. In other words, ancestry is normatively the real criteria

we all should be using to determine groups; however, we also use properties such as traits which are outcomes of this evolutionary process used for estimation of these evolutionary relationships. Proximity therefore refers to the level to which a criterion could be considered as determining a group, with ancestry being the direct and technically correct reason for the proposed existence of taxonomic groups; however, we would say traits are also reasonable explanations for taxonomic groups despite being less proximal. That suggests that interventions be more careful in the ways that they discuss and approach criteria pyramid. This is already done to some extent with many interventions discussing the impact of analogous traits on tree building; however, students might benefit from this discussion about what constitutes evidence and what constitutes properties that determined membership.

What might aid these discussions surrounding the process of determining taxonomic groups, either in an everyday sense or normatively, i.e., phylogenetic trees, may potentially be discussions about theory of knowledge with students, i.e., why it is that we categorize plants the way that we do. This may lead to fruitful epistemological discussions about criteria, methods, etc., that could potentially lead to an increased awareness of these processes and perhaps better tree-interpretation. This epistemological recognition could be thought of as a metacognitive awareness which has shown to either improve or coincide with better understanding of concepts in other domains, presumably because it occurs in experts. This may also lead to a discussion about the different ways in which a phenomenon could be externally represented. The effect of these epistemological discussions about the role of representations on the interpretations of formal representations, to my knowledge, has not been examined.

Therefore, in summary, researchers should investigate the ways in which metaphorical jargon may positively or negatively impact both relatedness understanding and tree-thinking.

From these findings, further interventions can be generated that appropriately address this language, where instructors can, for example, make more explicit certain aspects of relatedness or epistemological understandings of models. There is some evidence that "being more explicit" about certain aspects of a phenomenon may have some positive impacts on learning (Ashman et al., 2020; Kruit et al., 2018) and such an approach therefore could be recommended in the meantime as these interventions are developed and assessed.

Future Directions

This work may lead to multiple productive avenues of future research. One such avenue is the generalization and quantification of some of the ideas presented. This study is qualitative and as such can describe the breadth and types of conceptions or phenomena observed in learners; however, it cannot provide information on the relative abundance and extent to which each contribute to relatedness understanding, and by extension, tree-thinking.

Indeed, I plan to use the results from this study to construct an evidence-based assessment of such conceptions to give instructors and researchers insight into students' nascent ideas about classification and how they might relate to tree-thinking. In addition, to my knowledge, few if any representation studies have documented how varied factors come together to influence representational competence (Halverson & Friedrichsen, 2013; Schramm et al., 2021a). I intend to use such an assessment in combination with preexisting assessments (e.g., Jenkins et al., 2022) to model tree thinking quantitatively using methods such as structural equation modeling. Such an examination is necessary since factors that are thought to influence tree-thinking, for example, may be modulated by other factors; evolution acceptance, for example, may influence whether spatial reasoning is deployed. Just because a student is capable

of mental rotation of models, does not necessarily mean they will be deployed in this context, especially if they do not understand evolution.

Toward a Synthetic Model of Taxonomic Thinking

A primary takeaway of this study is the potential investigation of taxonomic thinking with or without the involvement of trees. Indeed, two other major categories were described by this analysis but were not pertinent to this study, i.e., group membership (actual groups determined) and metacognitive amendments (epistemological recognition of methodologies used, etc.).

While investigating the links between each of these hypothesized constructs was beyond the scope of this study, I do offer a tentative model upon which other researchers can build. The links between said constructs were either derived from direct observations in the dataset or via deductive reasoning based on the form of said constructs. I provide a tentative representation of said constructs and their hypothesized connections below (Figure 25).

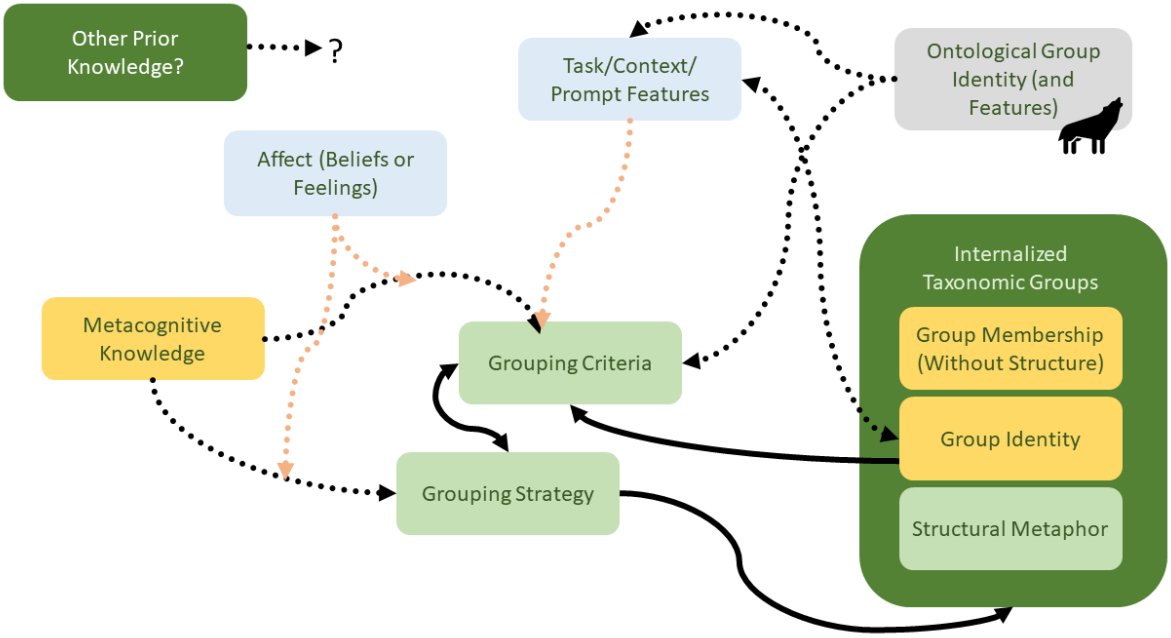


Figure 25: Tentative proposed model of relatedness understanding.

In this model, the light green boxes represent the constructs discussed in this study, i.e., grouping criteria, grouping strategy, and structural metaphor. The yellow boxes represent the two constructs that were part of the analysis but not part of this study as they did not address the first two research questions, i.e., group membership (here split into group membership without structure and group identity) and metacognitive knowledge. Group membership refers to the finalized groups determined such as "canids containing dogs and wolves," while metacognitive knowledge referring to the broad swath of behaviors observed which included epistemological knowledge about, for example, representation limitations or operationalization of particular methods. The blue boxes were those constructs which were completely beyond the purview of this study and were not investigated at all—some affect codes were initially obtained in the first coding round but that was done to make the separation between the codes of interest clearer. These refer to affect which we know is also a critical and until recently underappreciated aspect of learning (e.g., J. Y. K. Chan & Bauer, 2016; Fink et al., 2020; Ni et al., 2018; Wilton et al., 2019) as well as task or context features which refer to, for example, the use of photos, the use of visuals, or the particular taxa of interest. The grey box is a philosophical realist statement that ontologically true groups exist (in some capacity, though of course our estimations of said groups are insufficient; species determination of course being a classic example of the paradoxes that these real phenomena afford). The large green box represents what could be considered a complete construct representing a taxonomic group as understood by an individual, and note that it consists of (and is therefore affected by) structural metaphor, group membership, and group identity. The final box on the top left-hand corner refers to other ideas which may impact these constructs in ways that I have not yet accounted for, which include things like beliefs about nature which could impact the structural metaphors chosen (linear ordered relationships of

organisms rather than nested, branching relationships). The lines and arrows refer to the hypothesized effect of each construct on one another with a few interacting in both directions as shown by double headed arrows. The dotted lines indicate those relationships for which I have less evidence, given the scope of the study, while the solid lines are those that are more strongly implied by this study. Those lines that end on other lines indicate mediation in the form of controlling deployment (i.e., the construct in question can restrict the deployment of particular metacognitive understandings of methods, for example, used by individuals).

I will not go into exhaustive detail at this moment for the sake of brevity and given that it is tentative, but suffice it to say, many of these, outside of the evidence suggested by the study, can be argued deductively. For example, structure, identity, etc., could be considered as components of the entire understanding of a taxon given that perhaps using a strict tree metaphor where lineages are restricted to having only two daughter lineages, the smallest taxonomic groups containing members resulting from such a formulation would therefore only have two members unlike a broader structure such as the vertical case where each smallest group could conceivably contain more than two members.

This model also suggests that grouping criteria and grouping strategy may have a reciprocal relationship. Certain criteria might be "available" to individuals given the task context. In other words, the task may only include visual data, therefore, they may only employ visual evidence to construct or support their understanding of that particular taxon. Those criteria, therefore, might impact the choice of strategy. Broad photos of organisms might therefore encourage the use of prototype strategy wherein the criteria are often inaccessible (they are not cognizant of all of the aspects of said criteria) while giving a table of appearance of morphology might encourage a more isotype strategy. Alternatively, experts, for example, having more

metacognitive understanding of the limitations of their models and representations, might therefore restrict their understanding to the use of particular strategies, such as the isotype strategy, because they are attempting to estimate relationships based on the number of hypothesized evolutionary events (parsimony). That therefore restricts the types of criteria that can be used to construct these taxonomic groups.

In sum, these findings suggest there is a potentially very productive novel area of research which has the ability to potentially support formal tree-interpretation (assuming that leveraging everyday underlying understandings can support formal tree-interpretation). Obviously, these constructs would need to be further validated, examined for pervasiveness among target populations as desired, and the connections between them further theorized. Additionally, innovations that can either assess or support students' understanding based on these findings could also be potentially investigated.

Limitations

In this section, I discuss the limitations of this study. They include the lack of attention paid to gesture, especially given the partial embodied framework adopted, inauthentic settings in which these tasks were conducted, insufficient sampling of certain groups, and inappropriate task coverage. I discuss these limitations in detail below.

Lack of Gesture Documentation

As previously discussed in prior chapters, one issue of the usage of utterances to mine human thought is the conventionalization of language. While mutual negotiation between individuals can help researchers find shared meaning and thus communicate an understanding of the subjects' mind through that shared meaning, utterances may also be misleading because of

conventionalization which is why some cognitive scientists prefer to use spontaneous co-speech gesture to investigate cognition (Núñez et al., 2012; Núñez & Sweetser, 2006).

Participants might use words that are common in their culture without reference to their original meanings, obscuring the mental structures I wish to probe. A student, for example, might state that they interpret a tree from the roots "down" to the tips but point from left to right. One could argue that they may have a spatial understanding of time. Though they point left to right, the use of the word "down" might suggest, for example, that their conception of time is direction and oriented downward, where "downness" is more recent and thus more focused on the tips of the tree.

However, this word could have been used because of conventionalization—they learned the word socially without adopting the metaphorical aspects of it. In other words, they say "down" because most people in their culture say "down" in similar situations. Perhaps the metaphor of "downness" was invoked in the past when the term first appeared, but it is no longer used as such by the interviewee. In other words, participants may say something like "down" because they have simply heard it repeatedly. They do not say it because they have a cognitive structure that maps conceptions of time onto conceptions of position—they may not actually be relating time and space.

In addition, as this study was conducted on Zoom, I was unable to see clearly any gestures made by the participants given the highly variable video capturing abilities of participants. One interview had no visual data as the participants encountered problems with their internet and a few others had poor video quality. The camera was placed such that participants' hands were often not visible. In addition, the usage of the Explain Everything app required that the participants use a mouse and the apps tools, meaning that their hands were

unavailable for gesturing--doing things with your hands can prevent the appearance of gestures or themselves can be considered a subset of gestures that are primarily deictic and employ tools (Genest & Gutwin, 2011; Gutwin & Penner, 2002; Ou et al., 2003). Future studies can investigate this in more detail by using in-person interviews.

Inauthentic Settings

Using video conferencing tools to conduct interviews is potentially problematic. Not only are there issues with how people relate to one another using video conferencing—interactions and norms are quite different on video than in person (O'Malley et al., 1996). As I conducted a study that makes claims about the types of ideas that might appear in classrooms, differences in task context may have limited the claims I was able to make. I may have, for example, been unable to read certain body cues from a flat screen that might clue me into a student's discomfort with a problem. Not perceiving that in the moment would prevent me from inquiring about it, resulting in a loss of data.

Additionally, clinical interviews themselves have been criticized for being inauthentic settings in which to probe student thought. Students may act differently in the classroom than in interview settings (diSessa, 2007). Environmental cues have been shown to affect student behavior. Students in different classrooms will often use different strategies to solve the same problem (Saljo & Wyndhamn, 1993). Alternatively, students may be motivated differently in a classroom. Work in the classroom, for example, is likely higher stakes than in an interview; students will receive a grade for their work. Earning a grade may differently affect students' affect and behavior in classrooms—some might be more relaxed in a non-assessment setting while others less so (Chamberlin et al., 2023). Alternatively, the power differential in a clinical interview setting might differ from that of the classroom (Scheurich, 2014). Regardless, this

could result in the documentation of inauthentic behavior, i.e., behavior that may not be in reference to the contexts we hope to understand: learning in a classroom environment. If we are to document these student understandings and we know that human behavior is cued by the environment, having interviewees not in the environments where we want them to exhibit the behavior may give us an understanding of thoughts or cognitive processes that are not generalizable to a classroom and are thus less useful.

Expert-Novice Comparison

Another potential issue with this study is the comparison between experts and novices. While it is inadvisable to suggest my sample is representative of biology faculty, the faculty and post-doctoral researchers interviewed skew heavily white (Table 4). This coincides with the literature that documents that faculty and graduate students in evolutionary biology and a few other disciplines are not diverse (Kaplan et al., 2018; O'Brien et al., 2020), especially in comparison to undergraduate populations. This may, therefore, suggest that some of the comparisons may not necessarily be of use (e.g., that experts derived from these novices in the future might not express the same patterns of expert understanding).

Task Coverage

The tasks I designed were, of course, incapable of accounting for every organismal context that may arise. Many organisms exist and I could not feasibly account for every major group of organisms in a short interview. I did restrict the tasks to vertebrates and plants; however, within vertebrates I omitted a few major groups as an oversight. I did not include amphibians and reptiles which themselves may elicit some unusual responses. I did not recognize this oversight until after the interviews were completed. This oversight, however, may be

forgiven given that some studies exist that probe for this specific context in the tree-thinking education literature (Novick & Catley, 2014).

An oversight that is more concerning, however, is the exclusion of macroscopic invertebrates. These would have been particularly interesting to probe given that many have body plans that are similar to vertebrates (cephalization, presence of limbs and other appendages) but are not necessarily homologous (e.g., insect wings are derived from patches of tissue on the abdomen instead of converted limbs). Alternatively, homologous structures may not necessarily be easily recognizable. Invertebrates too, like plants, may additionally be subject to similar biases; indeed, initial my exclusion of them might be the result of my own vertebrate-centered biases.

Final Takeaways

In this study, I showed for research question one that relatedness understanding consisted of structural metaphors that were typically spatial in nature. These understandings often employed different spatial metaphors and essentially described the ways in which taxa were positioned with respect to one another. Several different variants of structural metaphors were observed, some of which were observed at similar rates within experts and novices and some that were quite disparate. Unsurprisingly, experts did exhibit more normative structures (trees); however, it is notable that they also used a wide variety of other understandings which coincided with novice structures. These suggest that students and instructors could leverage these structures to support normative relatedness understanding and, potentially, formal tree-interpretation.

I showed for research question two that relatedness understanding also consisted of the strategies employed to create taxonomic groups as well as the criteria used to determine membership. These were similarly ubiquitous as in the structure case; all participants exhibited at

least one of each of the major components of relatedness understanding, suggesting they are fundamental components of taxonomic understanding. These coincide with the embodied cognition literature on categorization; however, they also consist of other forms of categorization such as classical (isotype) categorization. Experts and novices were observed using both of these major types of strategies at similar rates, indicating that novices do have the potential building blocks for normative understanding of these components of relatedness but also that our relationship to categorization might be more complex than anticipated—normative understandings, depending on the formulation can fall under either of those major subcategories (prototype and isotype) of strategies.

Research question two's results also suggested, unsurprisingly, that experts favored more normative criteria and that novices favored nonnormative criteria such as causal or downstream outcome variables such as the environment. These results, however, also highlight that experts more often used ancestry as a warrant for explaining the composition of taxonomic groups which is distinct from the usage of traits as warrants. This highlights a difference in the ways in which we argue for the existence of taxonomic groups and that greater care is likely necessary in clarifying to students the level to which criteria and evidence suggests taxonomic group membership.

These results taken together suggest that the tree-thinking literature further investigate the underpinnings of tree-thinking, i.e., relatedness understanding, and its potential impact on formal tree-interpretation. They suggest that the misconception literature as well as our formulations of learning goals be reassessed to appropriately account for, assess for, and target the nuanced aspects of normative understanding which are modified versions of everyday novice understandings. They also suggest that further investigations into interventions that scaffold

toward normative understanding and representational competence using modeling-based intervention methods that are in use in K-12 education might be good candidates for supporting student understanding. Finally, they also suggest that instructors and researchers may need to consider the impact of the metaphorical jargon on student understanding as well as the formulation of new research questions in tree-thinking.

My contribution to the literature is that it centers and separates out the underlying understanding of tree-thinking, i.e., relatedness understanding, for investigation and targeting by researchers and educators. It provides a tentative hypothetical model for the components of relatedness understandings and seeks to direct the field's attention to other models or representations that could feasibly represent evolutionary relatedness and be leveraged to support students' ultimate formal understanding of relatedness, vis a vis, phylogenetic trees. I also provide qualitative evidence for these constructs in the form of artifacts and utterances, which, until recently (Kummer et al., 2016) outside of a handful of studies (Halverson et al., 2011), was uncommon in the tree-thinking literature. Furthermore, I compare expert and novice understandings, the former of which has been very rarely examined (Halverson et al., 2011; Novick & Catley, 2007) and, as I argue, is necessary for appropriately understanding what our goals are as instructors for instruction as well as identifying everyday understandings that can be leveraged for supporting normative understanding.

Chapter 5 contains unpublished material coauthored with Komperda, Regis and Lo, Stanley. The dissertation author was the primary researcher and author of this chapter.

APPENDIX A. PRIMER ON PHYLOGENETIC TREES

Phylogenetic Trees

As phylogenetic trees are uncommon outside of biology, I would first like to briefly discuss what phylogenetic trees are, how they are made, and how they are currently used. I am doing so to not only to provide the reader with a systematics primer, but also to orient the reader to the complexity of tree usage in the literature. Experts wade through a large variety of trees with slight differences in meaning that are communicated implicitly, yet our research on teaching tree-thinking unfortunately only focuses on a subset of these (e.g., Dees et al., 2017, 2018; Novick & Fuselier, 2019).

What Are They?

Phylogenetic trees are biology diagrams that depict evolutionary relatedness. Evolutionary relatedness refers to the relationships between organisms (or genetic sequences) that arose from diversification processes such as speciation and extinction. In other words, phylogenetic trees are reconstructions of evolutionary history.

Phylogenetic trees have a variety of names including *cladogram, phylogeny, evolutionary tree,* and *tree*. Often these terms are used interchangeably to refer to any type of tree, though some argue that cladograms should only show evolutionary relatedness (no other data) and the term phylogeny only be reserved for cladograms exhibiting the currently accepted hypothesis of evolutionary history. However, this perspective is uncommon (University of California Museum of Paleontology, 2021a). Like many modern biologists, I will use these terms interchangeably to refer to any type of tree.

All phylogenies have the following components: branches and nodes. Branches—also known as lineages, leaves, or edges—refer to the lines in the diagram (Figure 1). The meaning

assigned to branch length varies according to tree types. Ratograms depict branch length as substitution rate, i.e., rate of change of morphological traits or nucleotide bases. Phylograms depict the number of substitutions or the "distance" observed between lineages. Phylograms are also referred to as trees with scaled branches—other trees are unscaled. Chronograms depict absolute time, usually with an axis indicating the timescale (Santini et al., 2013). In empirical systematics and phylogenetics, chronograms are the most common as they intend to reconstruct not only the evolution relationships but timing of splitting events. All these trees are not considered ultrametric; ultrametric trees, such as cladograms (Figure 1), have all tips reach the end of the diagram (equidistant from the root).

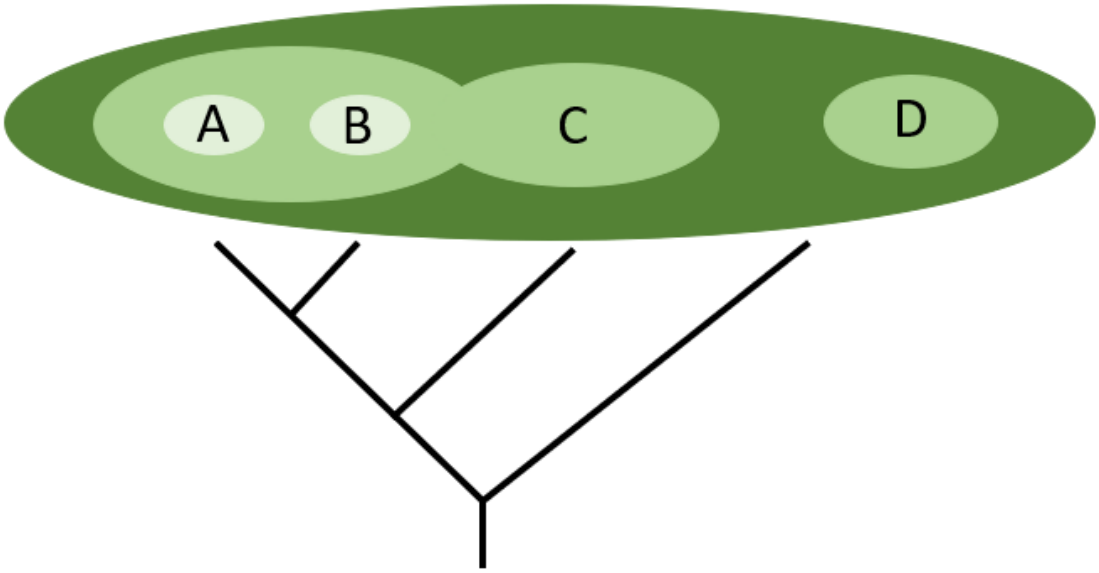


Figure 26: An example of a topophylogeny.

Interior nodes (aka *internal nodes*) are the points at which branches divide. One important assumption of phylogenetic trees (and evolutionary relatedness) is that lineages bifurcate, *i.e.*, one branch always gives rise to two branches. These bifurcations are also sometimes referred to as *branching events, splitting events,* or *speciation events*. The latter-most term is used when the lineages represent species, though often lineages can represent any number

of taxa (Figure 27). In the unusual cases where these splits refer to convergences instead of divergences, they are referred to as *coalescences*.

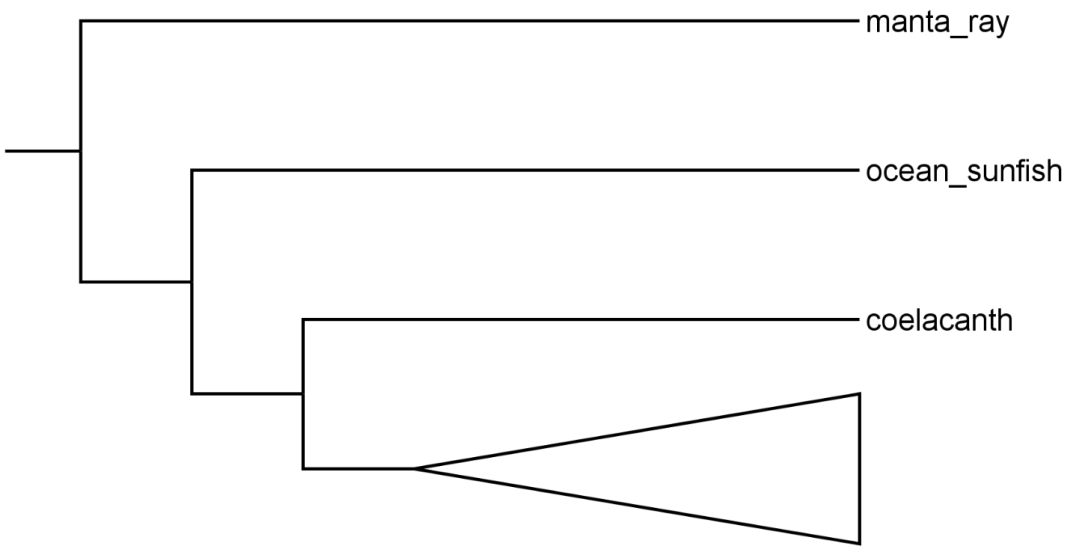


Figure 27: Tree featuring a collapsed clade.

Interior nodes are often simply referred to as *nodes* by many biologists. These nodes are so named to differentiate between the very end of the tips of branches and the branches leading up to them. They refer to these taxa as *tips, exterior nodes, terminal nodes, operational taxonomic units* (*OTUs*), or *terminal taxa* (Figure 1). These tips represent a wide range of biological phenomena including individuals, species, genera, phyla, genes, and populations. Tips can represent extinct and/or extant (living) organisms, though the latter is more common. The node at the beginning of a *rooted tree* (see below) is referred to as a *root*. In this document, however, I will simply refer to all interior nodes as nodes and tip taxa as tips or extant taxa.

Phylogenies also sometimes depict nodes that give rise to more than two branches, referred to as polytomies (Figure 28). However, phylogeneticists typically assume bifurcation because of probability theory or parsimony—we assume that one species (gene or other taxonomic group) will likely split into only two species in an infinitesimal unit of time. Splitting into more than two species is likely rare in nature, if it occurs at all (Braun & Kimball, 2001).

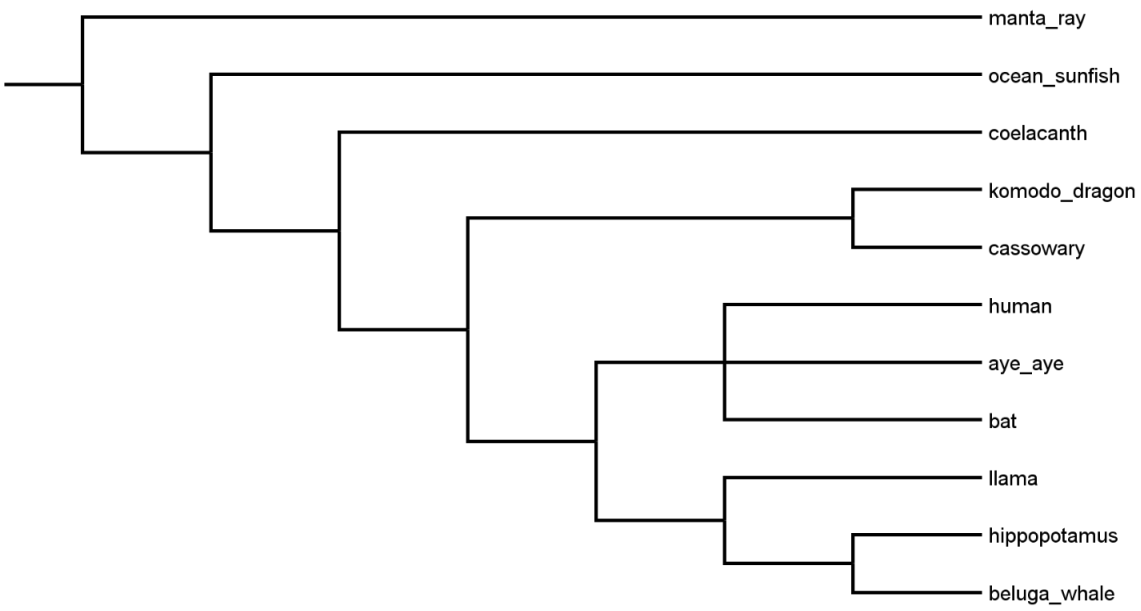


Figure 28: Tree containing a polytomy.

Interestingly, polytomies are frequently observed in constructed trees. However, they are often assumed to result from either (1) having too little data to resolve rapidly bifurcating nodes that thus appear as one polytomy or (2) constructing a tree with a timescale so large that the branching events are so close together it appears as a single event (Braun & Kimball, 2001).

Generally speaking, the appearance of trees can be variable. Most often they are bracketed/squared or diagonal/triangular (Figure 29). Sometimes they can appear rounded (Figure 30) and especially large phylogenies are fanned out into a circle (Figure 31). The thickness of the branches and the branch lengths may also differ. Paleontologists often use thicker trees (e.g., MacFadden, 2005) and romerograms (also known as spindle trees) that show phenotypic or species diversity (e.g., Keesey, 2007). Infrequently used dahlgrenograms are cross-sections of trees that show species diversity via branch thickness but obscure some evolutionary relationships (e.g., Verstraete et al., 2014). Empirical phylogeneticists very often use bracket trees (e.g., Buckner et al., 2015; Santini et al., 2013) and when such trees are too large will often arrange them in a circle instead of depicting them on several pages (e.g.,

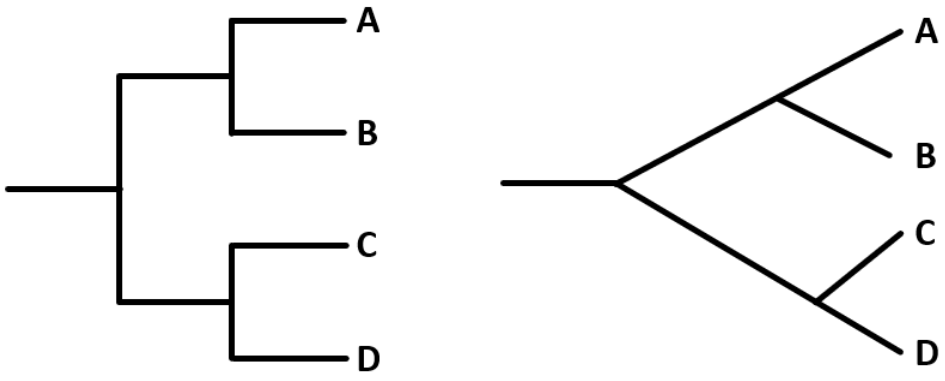


Figure 29: Bracketed/squared tree (left) and diagonal/triangular tree (left) with identical tree topologies.

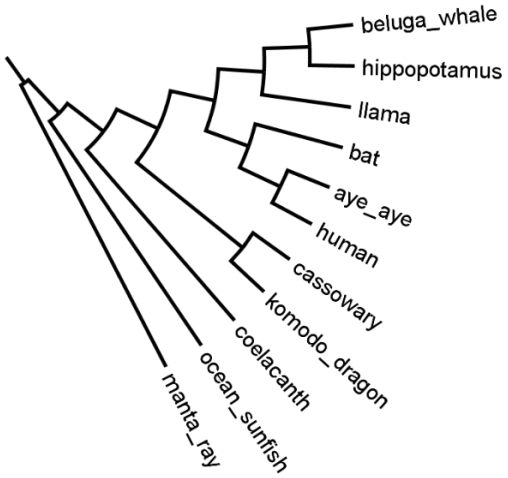


Figure 30: Rounded bracket tree.

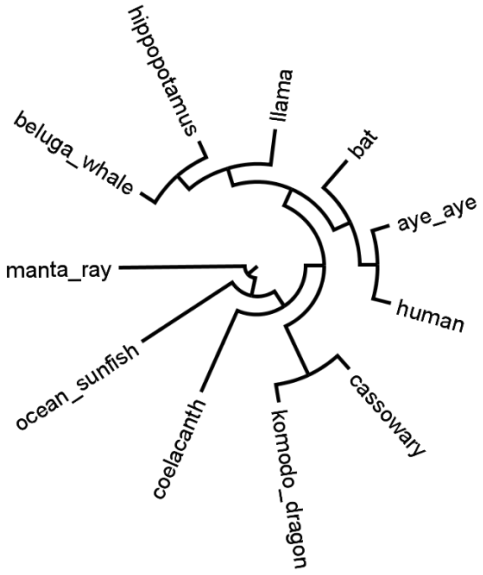


Figure 31: Bracket tree fanned out into a circle.

Espeland et al., 2018). All aforementioned trees are community recognized conventional trees and as a result they were or do appear very frequently in the literature.

Biologists also use a variety of unique diagrams or trees. Theoretical evolutionary biologists and phylogeneticists use a wide variety of trees that do not have as strict conventions. They can, for example, resemble more traditional trees but vary in material (e.g., Vaux et al., 2016), or they can include concentric circle diagrams (e.g., Waese et al., 2017), or they can be wavy (e.g., Baum & Smith, 2013). They also can be used in combination with other diagrams (e.g., Waese et al., 2017). Most famously they can be used to show nuanced relationships between phylogenies at different levels as seen in coalescence, aka gene tree-species tree, diagrams where a gene tree is embedded within a species tree (e.g., Rosenberg & Nordborg, 2002).

Phylogenetic trees may be rooted or unrooted (Figure 32). Rooted trees are those that depict either absolute time or relative time where the root can be considered the first bifurcation in the tree, or the most recent common ancestor of the organisms considered (Figure 1). The inclusion of an outgroup—an organism that is markedly different and thus equidistant from the study group—helps determine the first branching event and thus the placement of the root (Figure 33). Ideally, the outgroup is chosen because it is still similar enough to the taxa being examined such that it has characters that are comparable, but distinct enough to safely assume that the most recent common ancestor of the ingroup (study group) is not shared with the outgroup. Rooted trees that show absolute time have branch lengths that are proportional to time. Rooted trees that show relative time either do not have meaningful branch lengths (Figure 1), or they have branch lengths that represent something else such as number of substitutions. Relative time refers to the general ordering of branching events, *e.g.,* speciation event C happened after

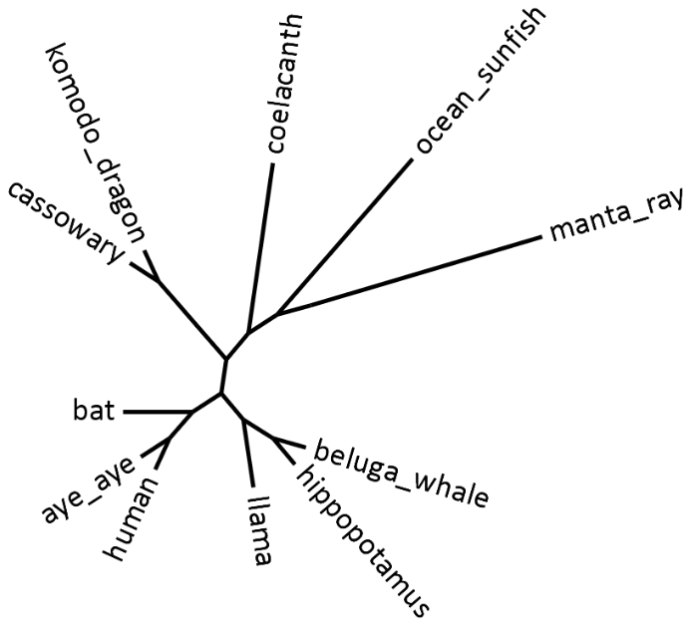


Figure 32: An unrooted tree.

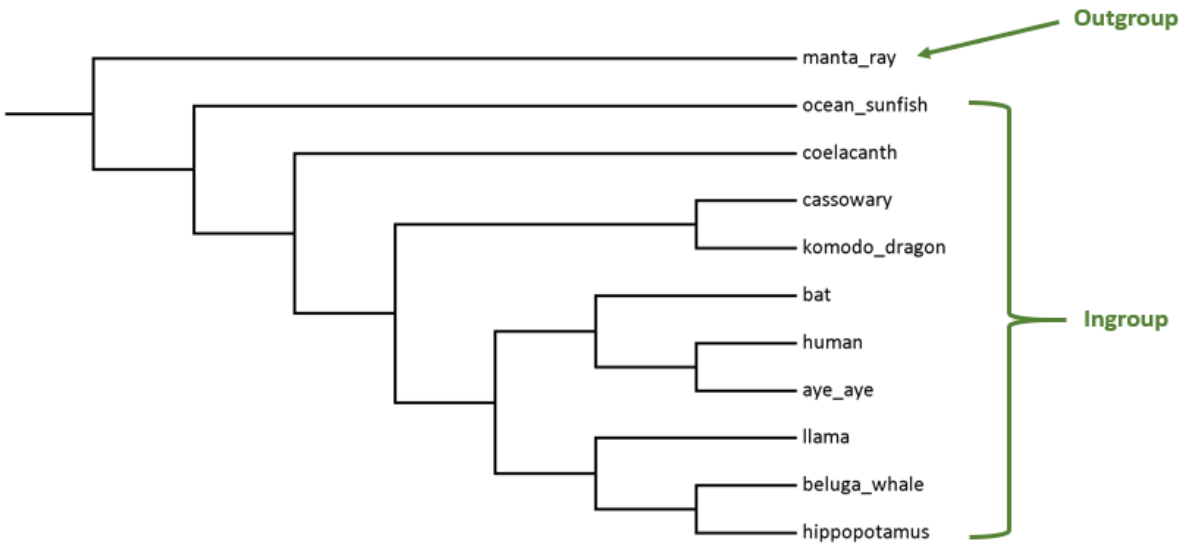


Figure 33: Tree with ingroup and outgroup labeled.

speciation event B. Note, though, that relative time does not provide the ordering of branching events of equivalent node depth (Figure 34).

The main key feature of trees that makes them interpretable is their ability to be rotatable about nodes. Were, for example, two sister taxa swapped (sister taxa are two lineages stemming

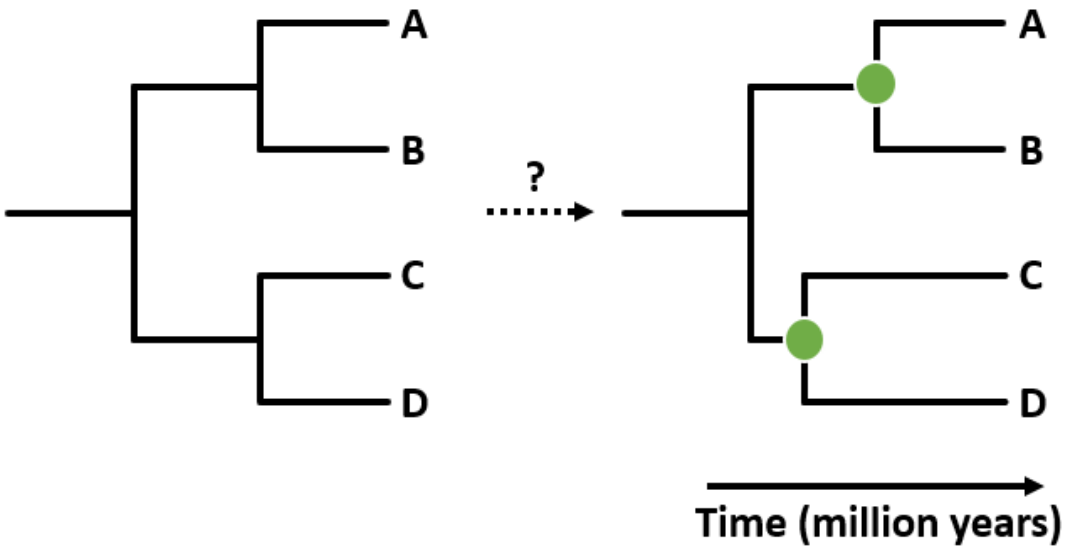


Figure 34: Cladogram showing relative time where branch lengths are meaningless (Left). Chronogram showing absolute time where branch lengths refer to time (Right).

from the same ancestor), the tree would still communicate the same information (Figure 2). These two trees would be considered as having the same tree topology, meaning they show the same patterns of evolutionary history. This property of trees is sometimes referred to as the three-taxon statement and posits that sister taxa are more closely related to each other than to any other branches in the tree (Novick & Catley, 2007; Novick, Catley, & Schreiber, 2012).

How Are They Made?

Morphology

Historically, evolutionary relationships were determined by morphological characteristics such as wing shape, bone shape, bone presence, and muscle presence. A morphologist would perform dissections or closely examine specimens to accumulate a broad variety of traits and determine their *polarity*, *i.e.*, which ones were more recent (*derived*) and old (*ancestral*). Systematists would then use these traits to reconstruct an evolutionary tree using parsimony.

Parsimony is a fundamental assumption that underlies both phylogenetics and evolution. Changes in traits and changes in genes are assumed to be relatively rare, so when two individuals share the same trait, they are likely related.

The application of parsimony, however, is nuanced. Biological phenomena such as convergent evolution can cause two very distantly related species to have the same trait. Marine organisms serve as a good example. Tuna, seals, penguins, and even ichthyosaurs, for example, share the same fusiform torpedo-like body shape. However, tuna are fish, seals are mammals, penguins are birds, and ichthyosaurs are reptiles. Despite originating from vastly different groups, they all share the same shape. Most organisms in these groups have recent ancestors who walked on land, meaning that they evolved traits suited for aquatic environments sometime in their evolutionary history. In other words, the environments in which these disparate organisms lived selected for the same solution to the same problem.

Constructing an evolutionary tree using this one trait, however, is problematic. A systematist might conclude that all these organisms are very closely related when they are not. Additionally, such a conclusion would imply that the source of their body shape was their ancestor and not derived later. Morphological datasets thus needed to be chosen carefully such that natural phenomena such as convergent evolution did not result in the incorrect determination of evolutionary relatedness. The inclusion of many stable traits can overcome the problems contributed by convergent traits during tree construction by being comparatively more common. This can be accomplished by excluding known rapidly changing traits (such as fur color) or if convergent evolution is infrequent in the target population. Convergent evolution is, however, often common.

Earth contains a finite number of habitats, so convergent evolution is highly likely to occur. Thus, morphological traits alone are often insufficient for reconstructing evolutionary histories.

Molecular Phylogenetics

A paradigm shift thus occurred with the advent of genetic sequencing technologies. Stretches of DNA that code for functional genes like enzymes are much easier to group as being *homologous*. Homology refers to the idea that traits appear due to common ancestry—for example, both humans and cats have femurs because their common ancestor had femurs. Homologous traits are critical to the determination of evolutionary relatedness and homology of morphological traits is much more difficult to determine. Systematists began using molecular data to infer evolutionary relatedness because of this property and its abundance; genetic changes became the primary source of trait data and not morphological traits.

A variety of methods were developed to create trees using either morphological traits or molecular sequence data. Methods such as neighbor-joining and unweighted pair group method with arithmetic mean (UPGMA) construct trees using simple similarity metrics. The number of differences between each pair of sequences determines relatedness. Other methods use parsimony to construct trees; often the entire *tree space* (the set of all possible trees for a number of taxa) is searched to find the tree with the lowest parsimony score, *i.e.*, the tree exhibiting the fewest evolutionary events.

However, these algorithms produce a single hypothesis (tree) for a given group's evolutionary history. Consequently, some have argued that these methods are insufficient for finding well-supported trees. First, pairwise similarity matrices may oversimplify the evolutionary history of these groups and lead to incorrect tree inference. Secondly, by producing

a single tree and not being probabilistic, they do not indicate the likelihood of these trees are accurate given data used to infer them. In response, other algorithms were developed that employ Bayesian statistics—these methods can report probabilistic information using evaluation methods like maximum likelihood and maximum parsimony. In addition, Bayesian methods can estimate other parameters of interest, such as speciation rate.

As sequencing became more common and cheap, more sequence data became available. Systematists grew bolder and attempted to reconstruct giant trees. As mentioned, some methods required searching entire tree space, which is quite computationally taxing. Significantly increasing the number of taxa makes examining the entire tree space nearly impossible. To accommodate these huge trees, some systematists began using heuristics to sample tree space, such as random walk methods like Markov Chain Monte Carlo. These methods were combined with a wide range of tree evaluation methods such as maximum parsimony and maximum likelihood.

Molecular or genetic approaches, however, are not radically simpler than morphological approaches. As genes often undergo indel mutations, *i.e.*, mutations that result in the insertion or deletion of bases in a gene, the length of genes varies. This variation obscures which positions, or "slots", are comparable, making it difficult to determine what changes have occurred (Figure 35). An adenosine base in the gene of one organism may match to an adenosine base in the same gene of another, suggesting that no mutation occurred at that site in either sequence since their most recent common ancestor. However, the adenosine of the first organism may instead match a cytosine in the second organism, meaning that either the ancestor of both organisms had either an adenosine or cytosine and one of the two organisms experienced a permanent substitution. Alternatively, the adenosine in the first organism might match another adenosine in a different

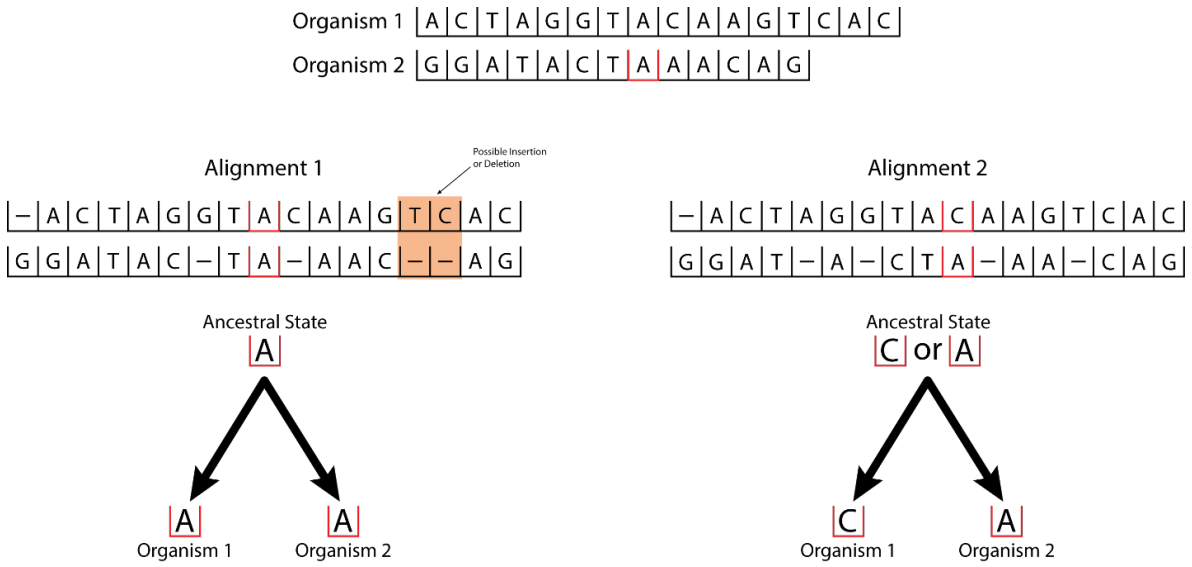


Figure 35: Example of issues with sequence alignment.

location—perhaps a deletion mutation occurred, making one of the sequences shorter and obscuring the bases' ordering. Even more confusingly, perhaps the gene in question experienced many mutations and ultimately mutated back to the same ancestral base, adenosine. If we are using these substitutions to determine evolutionary relatedness, this substitution history would be hidden. Often systematists use only sequences from living organisms, a snapshot of each lineage's history. These are a few examples of the complexities involved in using genetic information to determine relatedness.

Alignments are intended to determine which slots on sequences are homologous and thus comparable. The alignments on the bottom are two possible scenarios (out of many) that an alignment algorithm might evaluate when trying to align the two sequences shown at the top. Below each alignment are hypothesized reconstructions of the evolutionary history of that slot. Note that both alignments show two different slots that might be matched to the red slot on organism 2's sequence. The A might be matched with another A, implying that the ancestor passed an A to both species. Alternatively, the C might be matched with another A, implying that one of the lineages experienced a mutation. Also note that because sequences are never the same

length, some insertion or deletion events are assumed to have occurred as indicated by the dash symbol.

Sequence alignment algorithms were developed in response. These algorithms used a modified conception of similarity that used empirical data to determine which types of substitutions were more or less likely. Sequence alignment became an essential part of evolutionary relatedness determination. It was a first step in determining what traits—in this case genetic—could be comparable (*i.e.*, homologous) before trees could be constructed from the aligned sequences.

An affordance of using genetic sequence data to determine evolutionary relatedness was the determination of *divergence time*. Divergence time refers to the time when the most recent common ancestor of two lineages may have arisen. Determination of this time allowed for the creation of chronograms that show absolute time (e.g., Figure 34). Divergence time could be determined using the idea of a *molecular clock*. The molecular clock is a model that assumes, given the rarity and randomness of mutations that occur in genes that code for functional products like enzymes, which appear on average regularly like the ticking of a clock. Systematists use this *clock rate* to estimate the ages of ancestral nodes. By knowing the general rate of substitution in the gene(s) used in sequence alignment and tree construction, systematists can determine what lineages are related and when splits occurred. Morphological traits are not used comparably as their changes may be less regular due to the tremendous selective pressure the environment puts on them (see *punctuated evolution*). In addition, the quantification of morphological traits into rates of morphological evolution can be difficult.

Biologists once hoped that the molecular clock would be universal, but that is unfortunately not the case. Clock rates change depending on the gene used, the organisms

examined, and sometimes merely as a function of time. Systematists, therefore, cannot apply one universal clock rate to a tree to determine split times. A newer approach was thus required.

Modern Phylogenetic Inference

In modern systematics, slightly different data are used to reconstruct evolutionary history. Systematists use genetic data in combination with phenotypic traits such as bone presence and muscle attachment orientation. They primarily use extant organisms' genetic data to determine tree topology and use morphological data in the form of fossils to calibrate the tree. These trees are referred to as time-calibrated trees, or fossil calibrated trees. While systematists will not commonly use only morphology to determine relatedness of extant organisms, they will use it to place fossils within the tree. Doing so constrains the nodes to possible time intervals and helps more effectively estimate split times. In addition, they help calibrate the usage of changing substitution rates in genes.

Despite these new developments, the work is not yet, and will likely never be, complete. Systematists continue to grapple with new limitations of these methods, such as long-branch attraction, sampling bias, and gene-tree species-tree issues (Bergsten, 2005; Hillis, 1998; Rosenberg & Nordborg, 2002). While not widespread, some have begun rejecting the base tree metaphor for understanding evolutionary relatedness (Bapteste et al., 2009; Doolittle & Bapteste, 2007; Podani, 2019; Velasco, 2012). Some believe that coral better approximates the relatedness because its shorter and buried branches can represent extinct organisms better (Podani, 2019). Some microbiologists believe we ought to adopt a network model in place of the traditional tree framework; microbes and plants exchange genetic material or hybridize very frequently, making tree building incredibly complex. Despite these controversies, we will address tree-thinking as is currently common among biologists.

APPENDIX B. FINAL INTERVIEW PROTOCOL

Interviewer Interview Protocol

Legend

Q = Official Question Or Task Not on Participant Sheet That I Am Asking The Participant.

PQ = Participant Question. Question or Task That is Listed on the Participant Sheet.

APQ = Additional Probing Questions. Questions not on Participant Sheet that Are Optional Given the Interview Circumstances.

Notes for Interviewer

- Be sure to restart recording with each new slide.
- Viewer will see your cursor, participant will not. Have participant use the laser pointer and/or pen.
    - Unfortunately, there are no hotkeys.
- Video created may not reflect actual timing of video b/c pauses between slides and switching back and forth. Announce slide #.

Introduction

Hello, my name is Tina. I'm a grad student studying science education. It's nice to meet you! Just to let you know a little bit about me. I'm from California, been here most of my life. Originally studied biology but I found I liked thinking about how students think so I switched to education. I'm really happy you've decided to participate! Thank you very much!

Anyway, as you know, I'm doing these interviews so I can understand how people think about biology diagrams. It's going to be about 60 minutes or so long. Like I mentioned in (*[undergrad]* class | *[grad/faculty/post-doc]* my email), I just wanted to let you know that this interview is being recorded. I'm going to ask you a series of questions and activities—this isn't a test or anything and there are no right or wrong answers. Just follow my instructions and then explain your reasoning. I just want to know what you think and giving me extra explanations about why you chose or decided certain things is going to be very helpful to me. If you want, you can talk while doing some of the activities, but don't feel like it is necessary or if it is really hard to think about the questions while talking to me. Also you are encouraged to use this Explain Everything software I have set up—it's basically an online scratchpad. One more thing is to let me know if the question is confusing or is worded weirdly, I can reexplain. Finally, if you need a moment to think about the question, feel free to take as long as you like.

So let us just first set up this software we are going to use… [Help the participant open up Explain Everything and instruct them how to use it.]

Okay, so that's done. So, before we get started, tell me a little about yourself.

[PRESS RECORD ON BOTH EXPLAIN EVERYTHING AND ZOOM]

Q1. Background

[Student] What classes are you taking? Are you also volunteering in any labs or doing any extracurriculars, like clubs and stuff?

[Faculty/Post-Doc] What projects are you working on at the moment? What other kinds of things do you like doing on your spare time?

Transition.

Ah, okay. Awesome! Let's get started then. My first question is… [say first question]. Again just to remind you, feel free to draw all over the activity if you want.

Q2. PQ1. Evolutionary Relatedness – Conception and Drawing (RQ1 and RQ3)

2a. Here are some different organisms. Have you seen any of these before?

PICTURES OF ORGANISMS HERE

Stock photos of the following organisms will be shown but mixed up:

- White Tailed Deer
- Musk Deer (Male)
- Llama
- Camel
- Pig
- Hippo
- Bottlenose dolphin
- Sperm whale
- Whale shark
- Great white
- Tuna
- Pine tree
- Cycad
- Coconut tree

2b. Are these organisms related or similar to one another at all? How so and how do you know?

2c. Let's say you were trying to explain to your friend how those organisms you mentioned were related. How would you explain how they are related? Your friend,

let's say, really likes looking at diagrams when trying to understand things and asks you to draw a diagram to help you explain. How would you draw it? Go ahead and use the drawing board if you'd like.

APQ2a. Have you ever thought about how organisms are related to one another? Like, "Oh gee that animal must be related to that animal" for whatever reason?

APQ2b. How do you know that some organisms are related?

APQ2c. What kinds of organisms have you thought about as being closely related or not closely related?

APQ2d. What does this [*cursor points to part of the diagram*] mean?

APQ2e. [*If they as the interviewer if their friend knows about biology or not*] Let's assume they don't have a lot of biology experience. Also, how would it look if they did have a lot of experience?

APQ2f. What kind of information would you need to figure out how they are related?

APQ2g. Are they all equally related to one another?

APQ2h. What does it mean for them to be "related" – what does that mean to you?

APQ2i. What does it mean for them to be "similar" – what does that mean to you?

Transition

Okay so I will be showing you a series of biology diagrams. We're going to be using them for all of the next series of questions. Remember if you need a moment to think about these, take as long as you'd like.

[Portion of the Protocol Originally Included In This Study Has Been Omitted From This Dissertation]

Conclusion. <u>Clarifications</u>

Thank you so much for your time! That was wonderful and I learned a lot!

7. I was wondering if you had any comments or thoughts about what we just did?

Thank you so much again! I will be compensating you via Venmo. Thanks again!

APPENDIX C. FINAL TASK SLIDES

Note that organismal images have been blurred for publication reasons. These are ordered left to right, top to bottom.

Interpreting Biology Diagrams

The consent form is located here:
https://sdsu.co1.qualtrics.com/jfe/form/SV_d5qCNzDNlnuInCC

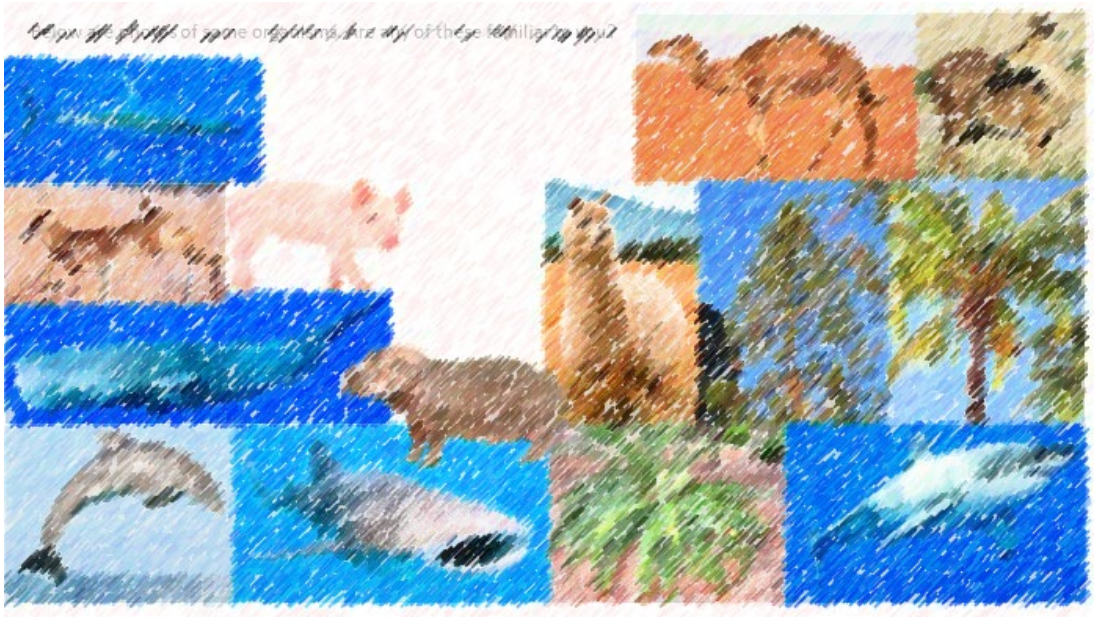
Thank you so much for participating!

(Interviewer: Please **Record** Now)

Let us test out the Explain Everything tools.

You can draw anywhere on this slide using the tools on the left.

Try it out!



Your friend, let's say, also likes to look at diagrams and asks you to draw one to help them understand. How would you draw it?

Note: Images of organisms above have been used in accordance with Fair Use. Permissions have for their usage have not been sought. These images were used for research purposes only.

Website Sources:

- Tuna: https://www.facebook.com/zubatacomisalj/photos/a.105862068622834/105861895289518/?comment_id=121721777036863&paipv=0&eav=AfbyCfcPhq2ADzuxICV9U78KY20VPCnrzmnGGTNlXiF-Z8YchbcNw8kGAFwhlfLH6a8&_rdr
- Deer: https://www.paducahsun.com/news/next-week-hunting-s-biggest-deal-gun-season/article_9625c912-c36f-5b59-b6fc-53e5979ce91d.html
- Pig: https://kids.nationalgeographic.com/animals/mammals/facts/pig
- Humpback Whale: https://www.sciencefocus.com/news/smart-sperm-whales-can-teach-each-other-to-avoid-hunters
- Hippo: https://twitter.com/thatnakedvegan/status/1131274818677018624
- Dolphin: https://www.thoughtco.com/learning-about-dolphins-1834133
- Whale Shark: https://travel.padi.com/snorkeling-trip/yucatan-peninsula/whale-shark-snorkeling-13216/
- Cycad:

https://www.indefenseofplants.com/blog/2017/10/2/the-nitrogen-fixing-abilities-of-cycads

- Alpaca: https://www.farmbrite.com/post/7-differences-between-llamas-and-alpacas
- Camel: https://foodtank.com/news/2021/01/new-research-finds-method-to-create-safe-camel-milk-products/
- Pine: https://en.wikipedia.org/wiki/Knobcone_pine
- Water Deer: https://www.nationalgeographic.com/animals/article/141224-deer-fanged-vampire-animals-science-krampus-christmas
- Palm Tree: https://photos.com/featured/coconut-tree-holger-leue.html
- Great White Shark: https://www.wired.com/story/dun-dun-duun-duuun-the-great-white-shark-genome-is-here/

APPENDIX D. CONTACT SUMMARY SHEET

Note that the following was intended for a prior variant of the study before it was pared down. The parts of this sheet that are pertinent to this study are questions 1, 2a, and 3.

Contact Note and Summary Sheet

Expert and Novice Understanding of Phylogenetic Trees

Participant Code:

Pseudonym:

Date:

NOTES DURING THE INTERVIEW:
QUESTIONS TO BE COMPLETED UPON COMPLETION OF THE INTERVIEW:

1. **What were the main issues or themes that struck you during this interaction?**

2. **Summarize information your obtained and/or failed to obtain for each task:**

   a. *Understanding of Evolutionary Relatedness*

   b. *Interpretation of Simple Tree*

   c. *Interpretation of Biogeography Tree*

   d. *Interpretation of Inward Circle Tree*

   e. *Comparison of Trees*

   f. *Location of Evolution*

3. **What struck you as salient, interesting, illuminating, or important in this contact?**

APPENDIX E. CODEBOOK

Code System

1 Grouping/Structural Metaphors	185
1.1 SILOED	16
1.2 STORYTELLING	19
1.3 TREE-LIKE	69
1.3.1 GENERAL TREE-LIKE	29
1.3.2 TRADITIONAL - Root to Tips	32
1.3.3 PARALLEL LINEAGES - Tips to Root	10
1.4 FAMILY	10
1.5 VERTICAL	25
1.5.1 GENERAL VERTICAL	8
1.5.2 TOP DOWN	15
1.5.3 BOTTOM UP	2
1.6 NO HIERARCHY - FLAT	62
1.6.1 GENERAL FLAT	21
1.6.2 IN-OUT	8
1.6.3 OUT-IN	35
1.7 ANTHROPOCENTRISM	18
2 Relatedness Criteria	251
2.1 EMERGENT TRAITS	157
2.1.1 BEHAVIORAL	11
2.1.2 NONPHYSIOBEHAVIORAL	27
2.1.3 PHYSICAL TRAITS	123
2.1.4 METABOLISM	18
2.1.5 LIVING STATUS	21
2.2 GENETICS	27
2.3 ENVIRONMENT	91
2.4 DOMESTICATION	10
2.5 FOSSILS	5
2.6 MRCA/COMMON ANCESTOR	51
3 Grouping Strategy	263
3.1 SIMILARITY	239
3.1.1 SIMILARITY UNCLEAR	50
3.1.2 PROTOTYPE (Non-existent ideal + weighted traits)	66
3.1.3 ISOTYPE/EQUITYPE (All members have same traits)	141
3.2 ESSENTIALISM	5
3.3 APPEAL TO AUTHORITY	34

1 Grouping/Structural Metaphors

Metaphors employed describe the structure or arrangement of taxa. Some might be normative and some might not be. Typically uses spatial language but not always. May or may not be scientifically normative. Does not describe specific taxonomic groups (so coding birds would not work but noticing that taxon A vertically includes taxon B and C does). Can be overlapping given the definitions of each of them. Codes in this section are prototype in nature in that not every instance will have 100% of the items or properties described in them.

1.1 Grouping/Structural Metaphors >> SILOED

Refers to structures that have distinct and separate taxonomic groups. These may or may not be hierarchical. They can overlap with any of the other structures. So, for example, completely ignoring plants as existing would count as siloed because they aren't even considering plants as part of the structure that they are describing. Similarly drawing two separate taxonomic arrangements would also be considered siloed. Cases where individuals use a line to divide two or more sections should be taken on a case-by-case basis. Textual evidence can suggest whether they meant simply to separate the two and prevent confusion or if they have wording to suggest they are separating two parts of a whole which would not be considered "siloed."

1.2 Grouping/Structural Metaphors >> STORYTELLING

Refers to a structure that suggests taxa are arranged linearly and separated by a sequence of events. Response to storytelling is often met with initial confusion as to how to characterize it. It may use many features in a very mixed fashion from all or many of the structures without any clear transitions or shifts between them. Typically storytelling suggests linearity of structure due to the story having a chronological element. In other words, it is linear likely because of its

strong sense of time. Usually is exhibited as a person comparatively emphasizing the story's iconic aspects, i.e. thinking a lot about the habitat the organism was in, maybe the selection pressures that were at play (or not), or simply thinking about the order of changes over time. This category is distinct from the tree-like structure in that it downplays the discourse about "branching" and "ancestors." Often these responses will imply or are at least worded in a way to suggest teleological thinking about taxa.

Diagrams depicting storytelling will typically be iconic (pictures) in nature and often include a sequence of events linked together in some way.

1.3 Grouping/Structural Metaphors >> TREE-LIKE

Tree-like metaphorical structures are those that are essentially scientifically normative. Individuals exhibiting this metaphorical structure may use a lot of normative language such as branching, splitting, root, nodes, taxa, sister group, crown group, etc. These structures consist of connections between taxa listed as well as hypothesized ancestors of said taxa and usually involve bifurcation, i.e. a single lineage splitting into two daughter lineages. Individuals may also use non-standard, non-normative language such as "trunk" or (sometimes) "base." Individuals exhibiting this structure will often overemphasize splitting events or that splits are occurring along lineages. They also might describe ancestors. While typically these structures do not necessarily include the identity of items within the groups, mentioning ancestors or parental lineages is important as it differentiates the uses of terms like "splitting" to refer to "splits" in a tree rather than "splitting" a non-hierarchical group into smaller ones. Also, this is one of the rare cases, especially since this is scientifically normative, where the individual explicitly stating that they are making a phylogenetic tree or cladogram was accepted as a use of this case (this was accepted given the assumption that normative diagrams might be one type of representation or

model of a phenomenon they have internalized, though they may also have other internal models that are less clear). Individuals exhibiting this structure also often used the term "related" more often than others, though that is not a foolproof indicator of what other participants used "related" in conjunction with other structures, especially since my prompt included the term "related." Individuals exhibiting this structure tended to use spatial language in the sense of taxa being more or less "closely" related. Individuals might also talk about evolutionary "pathways" and journeying down these "pathways." This type of thinking is hierarchical and may include terms like "level" and "tier" like the vertical structure but they occur less often.

Tree-like diagrams are essentially normative. They will use and sometimes talk about their diagram in normative ways, describing the parts suggested. The inclusion of lines in a diagram is not necessarily indicative of tree-like thinking. Vertical structures can also use lines but the "nodes" of those diagrams usually consist of the name of the category they belong to rather than the ancestor--taxonomic groups in trees are either typically thought of as a monophyletic clade or at a minimum (non-normatively) as the group of extant taxa at the tips. A larger taxonomic group can be circled, written next to a node, or on top of a lineage. Some individuals made simplified tree diagrams that were situated as tree diagrams by the accompanying utterances. Lines would more commonly be referred to as "lineages" or "ancestors" rather than "connections" or showing "belonging" as in the vertical metaphorical structures.

1.3.1 Grouping/Structural Metaphors >> TREE-LIKE >> GENERAL TREE-LIKE

See tree-like definition. Fits the definition with no indication of directionality.

1.3.2 Grouping/Structural Metaphors >> TREE-LIKE >> TRADITIONAL - Root to Tips

Fits the definition of tree-like but also indicates that they are thinking about it from the root (larger taxonomic groups or older ancestors) to the tips (smaller taxonomic groups and more recent ancestors and extant taxa). Individuals would often order their discussion of taxa in response to a question about how they are related by first starting at the largest taxonomic group and ending up with smaller ones.

1.3.3 Grouping/Structural Metaphors >> TREE-LIKE >> PARALLEL LINEAGES - Tips to Root

In this metaphorical structure, individuals exhibited the typical properties of the tree-like structure described above, except that they had an added element of reverse directionality. Individuals would start with the tips and work towards the root when asked how organisms were related. In novices, this tended to appear as described but in experts, this had an added element of likely referring to the lesser-known normative way of thinking of using extant taxa to "trace back" in time to the ancestor. Typically people will describe the two lineages as "occurring together" or cooccurring when existing as the ancestor. This likely comes from the coalescent literature which has a stronger emphasis on thinking about time in reverse to better understand how extant data might correctly or erroneously estimate ancestral state and evolutionary relationships. Kristy's pipe cleaner activity is an example of making that into a pedagogical intervention to support that type of thinking (and tree-thinking in general). Essentially if they talk about trees starting at the tips, they are exhibiting this structure.

Drawing the tree starting from the tips would also count but no one did this so this likely will not be used as a characteristic of it (likely because it is easier to draw trees starting from the root.

1.4 Grouping/Structural Metaphors >> FAMILY

Individuals illustrating this metaphorical structure will often refer to the arrangement of taxa like a family tree. Individuals could have discussed this without the context of a tree-like structure but they often did anyway. Notably, this seemed to often occur in a rhetorical sense with individuals trying to draw parallels from family trees (and by extension, pedigrees) to phylogenetic trees for the sake of the hypothesized listener. This is considered distinct from tree-like thinking, however, given that traditional normative family trees and pedigrees are not strictly tree-like. As they require drawing connections between parents to generate offspring, they do not have the property of trees that single lineages bifurcate (or result in more than 2 lineages in low-resolution trees). It is notable to indicate that tree-thinking often uses family-based normative terms to describe trees such as "sister taxa" and "daughters." These mentions were not considered family structure unless they made some other non-normative mention of family metaphors or made some other mention. Explicit attempts to describe them as family trees were coded as such but other discourses such as equating, for example, ancestors to "grandparents" suggested a family structure. This structure may or may not be related to themselves and overlap with anthropocentric structures in that they might talk about their relatives.

Diagrams that suggested family trees would have been categorized as such but no such artifact was created.

1.5 Grouping/Structural Metaphors >> VERTICAL

In this particular metaphorical structure, individuals would assign a very hierarchical understanding of taxonomic groups with taxa arranged "above" or "below" others. Sometimes these would be described as "levels" or "tiers." Individuals would use terms that overlap with other metaphorical structures like "divide", "separate", and sometimes "within" but these were almost always linked with the above strictly hierarchical vertical ordered ness terms. This

particular metaphorical structure has a component of linearity in that the taxa were almost always discussed in an ordered fashion from bottom to top or top to bottom.

Diagrams depicting vertical structures would sometimes use implicit labeling of "tiers" of groups of taxa with things like numbers, for example. Diagrams depicting larger taxonomic groups on top and smaller taxonomic groups on bottom were considered. Those that had lines linking taxonomic groups were included when they included larger taxonomic groups at "nodes" of structures that superficially resembled trees. Lines in the diagrams were often described as "connections" or as shown as "belonging to groups" rather than lineages.

1.5.1 Grouping/Structural Metaphors >> VERTICAL >> GENERAL VERTICAL

See the broad vertical definition above. No indication of the direction of traveling along the structure.

1.5.2 Grouping/Structural Metaphors >> VERTICAL >> TOP DOWN

See the broad definition of vertical structural metaphor above. This variant of the vertical structural metaphor involves discussing taxonomic groups in order of largest to smallest. The individuals might work their way "going down" said hypothetical structure.

1.5.3 Grouping/Structural Metaphors >> VERTICAL >> BOTTOM UP

This particular hypothetical structure has all of the properties of the aforementioned vertical metaphorical structure but the participant indicates that they are first thinking about individual organisms and how to put them together variably to determine the higher-level groups. They will begin often by talking about putting certain organisms together first and then creating larger taxonomic groups. This particular code may have elements that suggest the individual is reasoning aloud how they are forming taxonomic groups.

1.6 Grouping/Structural Metaphors >> NO HIERARCHY - FLAT

Individual groups organisms and talks about relatedness in a way that suggests no clear orderedness or hierarchy. There is very little sense of linearity but there might be some sequential properties given the order of taxonomic groups they chose to discuss. Individuals showing said category will often use words like "within" and "divide" suggesting some level of subsetting is occurring within the taxonomic groups that they are suggesting. They will tend to use "within" far more frequently than some of the other nesting/division languages compared to the vertical structure. This particular structure, while typically implying a nested structure to the taxa, can also include any other cases where they organize taxa without verticality or hierarchy.

Given the aforementioned description, one could describe the structure in two different ways, i.e. one that is "flat" and resembles a Venn diagram and another that is essentially a sphere or a network with connections between items to be described.

Diagrams depicting this code will draw either normative or non-normative diagrams. Individuals sometimes drew nested circle diagrams or Venn diagrams. Individuals would also sometimes use lines to show connections between taxonomic groups (not depicting ancestry or evolutionary history) but it would be clear that the way they spread the taxa on the diagram was either horizontal or not ordered.

1.6.1 Grouping/Structural Metaphors >> NO HIERARCHY - FLAT >> GENERAL FLAT

See the description of the general no-hierarchy structure.

1.6.2 Grouping/Structural Metaphors >> NO HIERARCHY - FLAT >> IN-OUT

Much like the vertical and tree-like cases, this structure represents a tendency by the individuals to focus on the smaller taxa first and work their way towards larger taxa. This is

considered non-hierarchical and can be approximated by a sphere where the individual successively works towards larger spheres.

1.6.3 Grouping/Structural Metaphors >> NO HIERARCHY - FLAT >> OUT-IN

Refers to the opposite of in-out. Individuals will begin by talking about larger taxonomic groups and work their way toward smaller taxonomic groups.

Traditional nested circle diagrams will show this.

1.7 Grouping/Structural Metaphors >> ANTHROPOCENTRISM

Says things that suggest anthropocentrism in conceptions of self in comparison to nature. Suggests that people might be more separate from everything else. Might save humans for last, for example, or technically we are one species but may talk about all of the other groups a family or phyla level but then suddenly talk about humans which are one species. Can also refer to the use of anthropocentric rhetorical turns to explain taxa. The structure still might resemble the aforementioned structures but have variants that are anthropocentric (includes humans or centers humans, for example).

2 Relatedness Criteria

Refers to any reasoning they may have for grouping the way that they do. This can refer to what they use in the task, what they say they use, or what was implied (implied ones like for appeal to authority and essentialism were not documented separately). Note that if individuals explicitly put on the rhetorical role of a novice and say "Oh I might notice X but I don't", X is not counted as criteria used for relatedness. Note also that simply stating "traits" will not be coded as anything as all of these groups could be argued as being a "trait." Also note that if an individual talks about certain properties as outcomes of another property those aren't necessarily coded unless the individual suggests that those outcomes are used to group organisms.

Note that these might appear on diagrams as criteria used to sort things, such as in headers. They might also be used in conjunction with trees, for example, to show that certain traits were used to establish taxonomic groups.

2.1 Relatedness Criteria >> EMERGENT TRAITS

This category encompasses a subset of traits used by participants to determine taxonomic groupings. I consider these traits "emergent" because they result from intrinsic factors like genetics (genotype) and external factors such as their environment. These traits include morphological, physiological (metabolic), and reproductive traits. Morphological traits are those that are more likely to be traditionally associated with normative tree-building and innate categorization of organisms; they refer to mostly visible macroscopic traits relating to body shape, composition, or color. Physiological traits include those less visible microscopic traits such as metabolic processes. I also include other traits such as life cycle type and organisms undergoing as being emergent traits that are a phenotypic outcome of genotype but do not necessarily neatly fall under the categories of morphology or physiology. One special type of trait included in this category is living status which refers to the state of being alive. Being alive is a trait of certain objects in the universe and is a function of those physiological processes so I consider that a part of this category though they hold special status as they are always used to group all living things into one group. As we are interested in all of the types of data used to determine taxa, this was also included.

2.1.1 Relatedness Criteria >> EMERGENT TRAITS >> BEHAVIORAL

Uses behavioral traits such as interactions with other organisms or the environment, or how they learn. These are things animals do in response to the environment that are not necessarily on the molecular level (so a bacteria secreting some substance in response to

something in the environment would not be considered a behavior). While behavior is often ultimately dictated (or at least modestly influenced) by genetics, it is very often more responsive to the environment and in that sense could be more "external" and less intrinsic than something like, say, body shape. This set of responses was considered distinct from physical traits given that difference and given these traits are not often used when building trees and are mostly non-normative given they can be so plastic (users of these traits argue that homology needs to be shown before use). It is also notable that I am using a more normative definition of behavior here than one that is emergent from the data given the complexity of parsing out meaningful non-metabolic, non-morphological traits such as life cycle, etc. To achieve some clarity and also to help determine the edge cases of behavior, my definition is more normative. Additionally, it is notable that this "behavior" is unique in that it is self-limiting, i.e. it will only occur in animals given that only animals are considered as exhibiting "behaviors."

Examples of behaviors include diel activity patterns, courtship, social behaviors, stalking, defending territory, making vocalizations, etc.

2.1.2 Relatedness Criteria >> EMERGENT TRAITS >> NONPHYSIOBEHAVIORAL

This refers to non-physiological traits that don't neatly fall under morphology or metabolism. They can also refer to things that are not purely behavioral. Examples include things like diet, types of food creation/ingestion, life cycle, metamorphosis, etc. Note that all of these aren't necessarily a part of their body shape or metabolism nor are they a characteristic of their environment, they are often an interaction between the two (e.g., diet can be considered a function of its environment, its morphology, and its behavior). Individuals might also explicitly talk about the interaction between the environment and the individual as a criterion for creating taxonomic groups.

2.1.3 Relatedness Criteria >> EMERGENT TRAITS >> PHYSICAL TRAITS

Physical traits refer to macroscopic or microscopic morphological or physiological features. Typically refers to macroscopic features that are what is normatively considered morphology. Can be internal or external and can also refer to microscopic features in microbes such as the presence of flagella. Can include physiological traits such as the presence of certain organs, etc. This can be coded if an individual uses a morphological trait to name a mostly non-standard group. So "spiky plant" rather than just "plant" when referring to a cycad, for example. May also refer to groups such as "multicellular organisms." Note that multicellular organisms are a group that is often taught about but it isn't monophyletic since multicellularity evolved several times. The most vague version of physiology coding will refer to individuals saying that certain organisms "look" like they should be in a certain group, implying they are using macroscopic morphological features to determine taxonomic groups. This group can also refer to any physiological features that do not change shape and are also typically not normative for tree-building such as organism color.

Examples include things like body shape, hoof type, presence of fur, color, etc. Individuals can also explicitly refer to morphology, assuming they provide some examples of said morphology to ensure that their definition is the same.

2.1.4 Relatedness Criteria >> EMERGENT TRAITS >> METABOLISM

This category encompasses any traits that refer to the metabolic functioning of an organism or refer to the generation of molecular products where the small molecular product is the endpoint (e.g., the creation of functional lactase in lactase-persistent individuals). This category will often include things like photosynthesis, glycolysis, etc. If talking broadly about eating a specific food, that is instead filed under nonphysiobehavioral.

2.1.5 Relatedness Criteria >> EMERGENT TRAITS >> LIVING STATUS

Consistently and frequently refers to whether an organism is alive or not, indicating that living status may be a major trait by which they categorize objects. Seems to be preoccupied with this feature and will refer to it repeatedly as an organizing trait. Typically refers to individuals grouping all organisms in a category of "living." Occasionally will be contrasted with a category of objects that are not alive. While metabolism is responsible for the state of being alive, this was considered distinct from metabolism since it was so pervasive an explanation and differs somewhat in quality from the typical metabolism-related responses (such as photosynthesis). This category is sometimes couched as being capable of interaction with the world, rather than the physiological state of being alive and undergoing metabolic reactions.

2.2 Relatedness Criteria >> GENETICS

Organisms are grouped because of genetic similarity (similarity here used in a broad sense, rather than the typical normative sense of genetics being used to infer evolutionary events). As this information was not provided, they typically would bring this up on their own, and would, for obvious reasons, not refer to specifics given that they need a lot of data that is not uniquely remembered i.e., they will not say sites 328 and 291 of gene X are synapomorphic for Y clade--no one can remember that. They also often do not give broad specifics such as general genetic sequences that were used to construct trees as those might also be difficult to remember.

2.3 Relatedness Criteria >> ENVIRONMENT

Any mention of using the environment to group organisms; can occur via naming ("marine animals") or explicit grouping ("I thought these were related because they both live in the water"). This can also include a subset of responses that refer to a specific location, i.e. "I know X and Y are related because they are both in the forests near my home." This refers to

typically natural or near natural environments. I suggest "near-natural" as there are many "natural" locations that are considered to be "natural" that exist that are not "natural" -- for example, the current distribution of pinnipeds around the world some may consider "natural" but is the result of centuries of human poaching which limited their global distribution. Typically individuals would not make such a distinction but I outline this for the sake of the reader who might be familiar with what is considered "natural" or not.

2.4 Relatedness Criteria >> DOMESTICATION

This criterion is when individuals use domestication status to determine taxonomic groups. This category may suggest a preoccupation with human-related activities. This can described as focusing on anthropomorphic life history, i.e. distinguishing between those organisms that have and have not been domesticated and who are often raised in farming environments. While this can refer to an "environment" in that sense, this was considered its group given its proximity to humans (which is a known cognitive bias) The results of that domestication are often not discussed (i.e. morphological, behavioral, etc. traits that arose as a function of domestication).

2.5 Relatedness Criteria >> FOSSILS

Individual messages using fossil evidence to group organisms. This can be used in the morphological sense, where fossils can be used to identify groups or distinguish between groups. However, it can alternatively refer to the more common use recently as a resource for time calibration on phylogenetic trees. This was also accepted because time calibration can move the nodes relative to one another. While it may not necessarily change the content of the node, it can change the relative appearance of nodes which, non-normatively, does change in some sense the properties of the group, i.e. if it is an older or younger lineage.

2.6 Relatedness Criteria >> MRCA/COMMON ANCESTOR

Individual indicates that groups are determined by the presence of a most recent common ancestor or simply a common ancestor. In other words, they might use the ancestor as a reason for groups being the way they are. Note that this category is complex and participants can implicitly suggest this is a grouping factor. This can also refer to within-species variation if they talk about grouping individuals of a population into the same species.

Ancestry might be explicitly referred to as a grouping factor but also acceptable are terms such as "evolutionary relatedness" or "related biologically." While the latter might be more vague of the two, typically those who stated that biological relatedness was a grouping factor implicitly distinguished between that explanation for grouping versus their description of organisms prior. Suggests that things are related "biologically" and have something to do with evolutionary history.

Simply mentioning common ancestors, however, should not necessarily mean that they are coded for. They should give some indication in the way that they've worded that they consider the ancestor as determining the group identity.

3 Grouping Strategy

Grouping strategy broadly refers to how individuals choose (consciously or not) how to include organisms in a category. This can therefore refer to the ways traits are used to determine category membership. Common ways to do so include ensuring all members of a group share some trait or that members of a group share a subset of a pool of traits with each trait having a certain probability of appearing. Note that they usually need a little more than the naming of a group (e.g., furry animals) to suggest a strategy. They need to state something to suggest more about how they are grouping it, otherwise almost all groups mentioned would be marked as

similar. This may have been stated earlier in the transcript so subsequent namings can be coded as such.

3.1 Grouping Strategy >> SIMILARITY

This broad category refers to any argument about how organisms may or may not be similar. This can be thought of very broadly. The similarity in this sense can be thought of as a downstream or upstream consequence of being in a taxonomic group, i.e., that individuals may vary in their cognizance and acceptance of traits being a function of genetics, for example, and their selection history. Normatively we now use similarity as a downstream consequence of various properties to estimate the splits in the tree because we assume that generally, with certain exceptions, individuals that are more closely related tend to have more traits in common than those that do not but we also examine the chain of evolutionary events that lead to the pattern of similarity and dissimilarity to reduce the number of evolutionary events. I bring this up given the discourse surrounding "similarity" being a common misconception among students which is a topic I will discuss in my results.

Note that for this group in particular, while most individuals referred to similarity as a deciding factor, a few also referred to dissimilarity as a reason to split groups. I consider this the same construct given they are simply suggesting the same thing, i.e., that similar things are likely to be more closely related.

I also mention that variants of similarity (i.e., prototype versus isotype) are often dependent on the type of criteria given, given that sometimes certain reasonings were not adequately probed. Be careful, however, with the phrasing. Sometimes individuals will refer to various traits or criteria but some of them will explain as outcomes of the other criteria, implying that the source criteria was used for creating the groups and not the outcome criteria. Make sure

to base your selection of strategy based on the source criteria. (Individuals might say something like "And I use the environment to figure out how these guys were grouped. These organisms though were shaped by natural selection so their body shapes are also different." Here the individual is indicating they used the environment to group and imply they did not use body shape because they conceive it to be an outcome of the environment.)

3.1.1 Grouping Strategy >> SIMILARITY >> SIMILARITY UNCLEAR

Suggests similarity is what determines how things are related but talks about it in such a way that it is difficult to distinguish between isotype or prototype variants of the definitions.

If individuals say that there are "common features" or "common traits" with no other evidence to suggest one of the other strategies, then this would be coded as similarity unclear. Common features could reasonably refer to sets of synapomorphies or sets of traits that may or may not be present. As it is difficult to ascertain with no additional evidence, this should be coded as similarity unclear.

However, you might encounter a "common ancestor" as a reason for generating groups. This was coded as isotype given that it is normative jargon and likely not referring to probabilistic sets of traits; rather an ancestor is present or not. This was coded as isotype typically and not similarity unclear.

Also note, sometimes you might see an idea embedded within another and not know how to code it. Overlap among strategies is okay but be careful about similarity unclear. It may be tempting to mark every tiny phrase that isn't specific as unclear. Only mark sections as unclear if it seems like there is a considerable chunk of text within a turn that is unclear. If there are multiple within a turn, it'll be pretty obvious, for example, that the first half is unclear and the second half is a specific type. That's okay to code separately. However, if a whole section is

using an isotype strategy and there is one phrase that is vague where you can't tell what type of similarity they are referring to, do not mark it as similarity unclear. Likely that vague phrasing is referring to the rest of the strategy used in that passage.

3.1.2 Grouping Strategy >> SIMILARITY >> PROTOTYPE (Non-existent ideal + weighted traits)

Prototype strategy refers to the categorization of members of a group based on a series of characteristics that are typically not present in all members of the group or may be present in many members but its not necessarily unique to that group. As probing for some of these traits was not extensive, we took any very obvious traits used to make such groups as being evidence for such strategies, i.e. suggesting that plants are a group because they are green. Though it was not explicitly asked, most people would likely agree that other living things exist that are green, and, less likely, they might also know that not all plants are green. Other more specific cases where edges are less commonly known were not assumed to be referencing prototype theory. In other words, traits that are not just synapomorphies are used to define a group, symplesiomorphies and other traits might be used to determine the identity and membership of a given taxon.

Prototype responses also can be very fuzzy and extremely vague. Note areas where after additional probing or explanation, the participant struggles to describe similarity and does so in vague holistic terms such as taxa "looking like" one another and they "generally look kind of similar, like the general body shape." The traits they may be using to probabilistically categorize taxa they likely do not have conscious knowledge of or they might invent traits in response to a prompt about the criteria they use.

People referring to this strategy will sometimes include language that suggests some "average" or prototype hypothetical taxa has certain traits but does not suggest that every

member of said taxa exhibits all of these traits. Additionally, individuals will also use language that implies a sense of centrality of the group, i.e., they use metaphorical language to suggest that a member of the group resembles a hypothetical taxon.

Participants referring to organisms as "type of X" might suggest a hypothetical "prototype but it also might not, it may be referring to the entire group name." Therefore "type of X" was not sufficient for being referred to as an example of a prototype strategy. However, saying something "looks like X" was coded as prototype as the phrasing suggests some fuzziness in the way they determined similarity which sounds more prototype than isotype so this was accepted.

One other term that might clue the coder into this strategy is the use of "typically." “Typically” implies there are exceptions which implies a probabilistic understanding of the set of probable traits. However, be careful, as some synonyms of "typically" are often used simply because the participant might be not confident. As those cases are more difficult to ascertain whether or not they are referring to prototype strategy, simply do not mark them and use other evidence to determine what the strategy might be.

Also, note they might use a mix of traits in the same sentence or set of sentences within a turn. If some are unique and some aren't, mark them as “prototype.” But if they are all unique, then mark the quote as “isotype.”

3.1.3 Grouping Strategy >> SIMILARITY >> ISOTYPE/EQUITYPE (All members have same traits)

Refers to the usage of common features to determine grouping. These features may need to be present in all daughter taxa. They may be referred to as "synapomorphies" or "defining characteristics." Features they refer to may be extrinsic or intrinsic so they can include things like morphological traits but can also consist of things like the environment. Note that simply using a common name that includes a feature such as "land animals" will not be automatically

coded, they either need to repeat that type of usage several times in the turn or they need to have some other evidence to suggest that they are basing the grouping based on the ubiquitous presence of said trait.

Also, note they might use a mix of traits in the same sentence or set of sentences within a turn. If some are unique and some aren't, mark them as “prototype” but if they are all unique, then mark them as “isotype.”

3.2 Grouping Strategy >> ESSENTIALISM

Suggests that "criteria" for deciding them is the thing that they are describing; participant implies that something is fundamentally different between the groups. Individuals will say something like "These are split into two because they are plants and animals" or "This and this go together because they are plants" or that the "criterion I am using is the fact that they are plants."

3.3 Grouping Strategy >> APPEAL TO AUTHORITY

Seems to focus on the "correct" taxonomic groups. Suggests that they are preoccupied with what is scientifically normative while potentially suppressing their impulses. Might be familiar with the term "classification" and uses it frequently. Suggests that things need to be classified appropriately. May or may not give information suggesting that the features they have assigned to organisms colloquially may be used to determine such "correct" or scientifically normative groups. Can also be used by experts who have essentially "placed" organisms in their knowledge structures already and struggle to "remember" it and refer to the evidence as being their "knowledge of systematics" for example. Also uses words like "proper." Other ways of expressing it suggest that they "happen to know." Essentially this wording suggests an external agent has identified the correct grouping. If the individual simply cites "prior knowledge"

without indicating the source of said knowledge or using rhetorical moves that suggest an external agent, their utterance is not considered an "appeal to authority" except in the case where someone where the individual explicitly states their role as being the reason for their grouping. If individuals state that they simply forgot and they do not strongly invoke or suggest they are very preoccupied with formal taxonomic groups like family, etc., that does not count as an appeal to authority. I also did not count when people responded to my question about data sources and the participant mentioned being taught things. As it was prompted, I thought this wasn't an accurate representation of the tendency to be overly concerned with what an outside force thinks about taxonomic groupings.

REFERENCES

A2Z Colleges. (n.d.). *Directory and list of community colleges, technical colleges and county colleges in USA*. Retrieved February 10, 2021, from http://www.a2zcolleges.com/Comm/commusa.html

Abner, N., Cooperrider, K., & Goldin-Meadow, S. (2015). Gesture for linguists: A handy primer. *Language and Linguistics Compass*, *9*(11), 437–451. https://doi.org/10.1111/lnc3.12168

Achurra, A. (2022). Plant blindness: A focus on its biological basis. *Frontiers in Education*, *7*, Article 963448. https://doi.org/10.3389/feduc.2022.963448

Adams, F. (2010). Embodied cognition. *Phenomenology and the Cognitive Sciences*, *9*(4), 619–628. https://doi.org/10.1007/s11097-010-9175-x

Ainsworth, S. E., & Scheiter, K. (2021). Learning by drawing visual representations: Potential, purposes, and practical implications. *Current Directions in Psychological Science*, *30*(1), 61–67. https://doi.org/10.1177/0963721420979582

Ainsworth, S. E., & Saffer, J. (2013). Can children read evolutionary trees? *Merrill-Palmer Quarterly*, *59*(2), 221-247.

Almeida, A., & Vasconcelos, C. (2013). Teachers' perspectives on the human-nature relationship: Implications for environmental education. *Research in Science Education*, *43*(1), 299–316. https://doi.org/10.1007/s11165-011-9272-z

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist*, *51*(4), 355–365. https://doi.org/10.1037/0003-066X.51.4.355

Anderson, J. R., Reder, L. M., & Simon, H. A. (1999). *Applications and misapplications of cognitive psychology to mathematics education*. Carnegie Mellon University, Department of Psychology.

Andrews, P. W., Gangestad, S. W., & Matthews, D. (2002). Adaptationism – How to carry out an exaptationist program. *Behavioral and Brain Sciences*, *25*(4), 489-504. https://doi.org/10.1017/S0140525X02000092

Arjoon, J. A., Xu, X., & Lewis, J. E. (2013). Understanding the state of the art for measurement in chemistry education research: Examining the psychometric evidence. *Journal of Chemical Education*, *90*(5), 536–545. https://doi.org/10.1021/ed3002013

Ashman, G., Kalyuga, S., & Sweller, J. (2020). Problem-solving or explicit instruction: Which should go first when element interactivity is high? *Educational Psychology Review*, *32*(1), 229–247. https://doi.org/10.1007/s10648-019-09500-5

Ballen, C. J., & Greene, H. W. (2017). Walking and talking the tree of life: Why and how to teach about biodiversity. *PLOS Biology*, *15*(3), Article e2001630. https://doi.org/10.1371/journal.pbio.2001630

Bapteste, E., O'Malley, M. A., Beiko, R. G., Ereshefsky, M., Gogarten, J. P., Franklin-Hall, L., Lapointe, F.-J., Dupré, J., Dagan, T., Boucher, Y., & Martin, W. (2009). Prokaryotic evolution and the tree of life are two different things. *Biology Direct*, *4*(1), Article 34. https://doi.org/10.1186/1745-6150-4-34

Barbera, J., & VandenPlas, J. R. (2011). All assessment materials are not created equal: The myths about instrument development, validity, and reliability. In D. M. Bunce (Ed.), *Investigating classroom myths through research on teaching and learning* (pp. 177–193). American Chemical Society.

Barnes, M. E., & Brownell, S. E. (2017). A call to use cultural competence when teaching evolution to religious college students: Introducing religious cultural competence in evolution education (ReCCEE). *CBE—Life Sciences Education*, *16*(4), Article es4. https://doi.org/10.1187/cbe.17-04-0062

Barnes, M. E., Supriya, K., Dunlop, H. M., Hendrix, T. M., Sinatra, G. M., & Brownell, S. E. (2020). Relationships between the religious backgrounds and evolution acceptance of Black and Hispanic biology students. *CBE—Life Sciences Education*, *19*(4), Article ar59. https://doi.org/10.1187/cbe.19-10-0197

Bauerle, C., DePass, A., Lynn, D., O'Connor, C., Singer, S., & Withers, M. (2009). *Vision and change in undergraduate biology education*. American Association for the Advancement of Science.

Baum, D. A., DeWitt Smith, S., & Donovan, S. S. S. (2005). The tree-thinking challenge. *Science*, *310*(5750), 979–980. https://doi.org/10.1126/science.1117727

Baum, D. A., & Offner, S. (2008). Phylogenetics & tree-thinking. *The American Biology Teacher*, *70*(4), 222–229. https://doi.org/10.1662/0002-7685(2008)70[222:PT]2.0.CO;2

Baum, D. A., & Smith, S. D. (2013). *Tree thinking: An introduction to phylogenetic biology*. Roberts and Company Publishers.

Bergsten, J. (2005). A review of long-branch attraction. *Cladistics*, *21*(2), 163–193. https://doi.org/10.1111/j.1096-0031.2005.00059.x

Bernard, H. R. (1988). *Research methods in cultural anthropology*. SAGE Publications.

Bernard, H. R. (2017). *Research methods in anthropology: Qualitative and quantitative approaches* (4th ed.). AltaMira Press.

Betz, N., Leffers, J. S., Thor, E. E. D., Fux, M., de Nesnera, K., Tanner, K. D., & Coley, J. D. (2019). Cognitive Construal-Consistent Instructor Language in the Undergraduate

Biology Classroom. *CBE—Life Sciences Education*, *18*(4), Article ar63. https://doi.org/10.1187/cbe.19-04-0076

Blacquiere, L. D., Fawaz, A., & Hoese, W. J. (2020). Who's related to whom? Use published phylogenies and make customized tree-thinking assessments. *Evolution: Education and Outreach*, *13*(1), Article 20. https://doi.org/10.1186/s12052-020-00134-8

Blair, C., & Murphy, R. W. (2011). Recent trends in molecular phylogenetic analysis: Where to next? *Journal of Heredity*, *102*(1), 130–138. https://doi.org/10.1093/jhered/esq092

Blake, K. (2018). *On the origins of "the march of progress."* Washington University ProSPER. https://sites.wustl.edu/prosper/on-the-origins-of-the-march-of-progress/

Bloomberg, L. D., & Volpe, M. (2019). *Completing your qualitative dissertation* (4th ed.). SAGE Publications.

Bodner, G. M., & Orgill, M. (2007). *Theoretical frameworks for research in chemistry/science education*. Pearson/Prentice Hall.

Bokor, J. R., Landis, J. B., & Crippen, K. J. (2014). High school students' learning and perceptions of phylogenetics of flowering plants. *CBE—Life Sciences Education*, *13*(4), 653–665. https://doi.org/10.1187/cbe.14-04-0074

Bong, C. L., Fraser, K., & Oriot, D. (2016). Cognitive load and stress in simulation. In V. J. Grant & A. Cheng (Eds.), *Comprehensive healthcare simulation: Pediatrics* (pp. 3–17). Springer International Publishing. https://doi.org/10.1007/978-3-319-24187-6_1

Boni, M. F., Lemey, P., Jiang, X., Lam, T. T.-Y., Perry, B. W., Castoe, T. A., Rambaut, A., & Robertson, D. L. (2020). Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nature Microbiology*, *5*(11), 1408–1417. https://doi.org/10.1038/s41564-020-0771-4

Boyce, C. J. (2015). *Investigating how students communicate tree-thinking*. University of Southern Mississippi.

Braun, E. L., & Kimball, R. T. (2001). Polytomies, The power of phylogenetic inference, and the stochastic nature of molecular evolution: A comment on Walsh et al. (1999). *Evolution*, *55*(6), 1261–1263.

Brear, M. (2018). Swazi co-researcher participants' dynamic preferences and motivations for, representation with real names and (English-language) pseudonyms – An ethnography. *Qualitative Research*, *18*(6), 722–740. https://doi.org/10.1177/1468794117743467

Bronzati, M. (2017). Should the terms "basal taxon" and "transitional taxon" be extinguished from cladistic studies with extinct organisms? *Palaeontologia Electronica*, *20.2.3E*, 1-12. https://doi.org/10.26879/173E

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *18*(1), 37–42.

Brownell, S. E., Kloser, M. J., Fukami, T., & Shavelson, R. (2012). Undergraduate biology lab courses: Comparing the impact of traditionally based "cookbook" and authentic research-based courses on student lab experiences. *Journal of College Science Teaching*, *41*(4), 36-45.

Bruce, C. (2007). Questions arising about emergence, data collection, and its interaction with analysis in a grounded theory study. *International Journal of Qualitative Methods*, *6*(1), 51–68. https://doi.org/10.1177/160940690700600105

Brunger, F., & Russell, T. (2015). Risk and representation in research ethics: The NunatuKavut experience. *Journal of Empirical Research on Human Research Ethics*, *10*(4), 368–379. https://doi.org/10.1177/1556264615599687

Bucklin, C. J., & Daniel, K. L. (2017). Using word associations as a formative assessment for understanding phylogenetics. *The American Biology Teacher*, *79*(8), 668–670. https://doi.org/10.1525/abt.2017.79.8.668

Buckner, J. C., Lynch Alfaro, J. W., Rylands, A. B., & Alfaro, M. E. (2015). Biogeography of the marmosets and tamarins (Callitrichidae). *Molecular Phylogenetics and Evolution*, *82*, 413–425. https://doi.org/10.1016/j.ympev.2014.04.031

Bussey, T. J., Lo, S. M., & Rasmussen, C. (2020). Theoretical frameworks for STEM education research. In C. C. Johnson, M. J. Mohr-Schroeder, T. J. Moore, & L. D. English (Eds.), *Handbook of research on STEM education* (1st ed., pp. 51–62). Routledge. https://doi.org/10.4324/9780429021381-6

Camacho, T. C., Vasquez-Salgado, Y., Chavira, G., Boyns, D., Appelrouth, S., Saetermoe, C., & Khachikian, C. (2021). Science identity among Latinx students in the biomedical sciences: The role of a critical race theory–Informed undergraduate research experience. *CBE—Life Sciences Education*, *20*(2), Article ar23. https://doi.org/10.1187/cbe.19-06-0124

Canning, E. A., Muenks, K., Green, D. J., & Murphy, M. C. (2019). STEM faculty who believe ability is fixed have larger racial achievement gaps and inspire less student motivation in their classes. *Science Advances*, *5*(2), Article eaau4734. https://doi.org/10.1126/sciadv.aau4734

Carnegie Classification of Institutions of Higher Education. (n.d.). *About Carnegie classification*. Retrieved February 5, 2021, from https://carnegieclassifications.iu.edu/

Catley, K. M. (2006). Darwin's missing link—A novel paradigm for evolution education. *Science Education*, *90*(5), 767–783. https://doi.org/10.1002/sce.20152

Catley, K. M., & Novick, L. R. (2008). Seeing the wood for the trees: An analysis of evolutionary diagrams in biology textbooks. *BioScience*, *58*(10), 976–987. https://doi.org/10.1641/B581011

Catley, K. M., Novick, L. R., & Funk, D. J. (2012). The promise and challenges of introducing tree thinking into evolution education. In K. S. Rosengren, S. K. Brem, E. M. Evans, & G. M. Sinatra (Eds.), *Evolution challenges* (1st ed., pp. 93–118). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199730421.003.0005

Catley, K. M., Novick, L. R., & Shade, C. K. (2010). Interpreting evolutionary diagrams: When topology and process conflict. *Journal of Research in Science Teaching*, *47*(7), 861–882. https://doi.org/10.1002/tea.20384

Catley, K. M., Phillips, B. C., & Novick, L. R. (2013). Snakes and eels and dogs! Oh, my! Evaluating high school students' tree-thinking skills: An entry point to understanding evolution. *Research in Science Education*, *43*(6), 2327–2348. https://doi.org/10.1007/s11165-013-9359-9

Chamberlin, K., Yasué, M., & Chiang, I.-C. A. (2023). The impact of grades on student motivation. *Active Learning in Higher Education*, *24*(2), 109–124. https://doi.org/10.1177/1469787418819728

Chan, J. Y. K., & Bauer, C. F. (2016). Learning and studying strategies used by general chemistry students with different affective characteristics. *Chemistry Education Research and Practice*, *17*, 675–684. https://doi.org/10.1039/C5RP00205B

Chan, K. K. H., & Hume, A. (2019). Towards a consensus model: Literature review of how science teachers' pedagogical content knowledge is investigated in empirical studies. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 3–76). Springer Nature Singapore. https://doi.org/10.1007/978-981-13-5898-2_1

Chandler, D. (2019, July 31). *Signs*. Semiotics for Beginners. https://www.cs.princeton.edu/~chazelle/courses/BIB/semio2.htm

Charmaz, K. (2008). Constructionism and the grounded theory method. In J. A. Holstein & J. F. Gubrium (Eds.), *Handbook of constructionist research* (pp. 397–412). Guilford Press.

Cisler, J. M., & Koster, E. H. W. (2010). Mechanisms of attentional biases towards threat in anxiety disorders: An integrative review. *Clinical Psychology Review*, *30*(2), 203–216. https://doi.org/10.1016/j.cpr.2009.11.003

Clement, J. (2000). Analysis of clinical interviews: Foundations and model viability. In R. Lesh & A. Kelly (Eds.), *Handbook of research design in mathematics and science education* (pp. 547–589). Lawrence Erlbaum.

Clement, J., Zietsman, A., & Monaghan, J. (2005). Imagery in science learning in students and experts. In J. K. Gilbert (Ed.), *Visualization in science education* (pp. 169–184). Springer Netherlands. https://doi.org/10.1007/1-4020-3613-2_10

Cochran, K. F., DeRuiter, J. A., & King, R. A. (1993). Pedagogical content knowing: An integrative model for teacher preparation. *Journal of Teacher Education*, *44*(4), 263–272. https://doi.org/10.1177/0022487193044004004

Coley, J. D., & Tanner, K. D. (2012). Common origins of diverse misconceptions: Cognitive principles and the development of biology thinking. *CBE—Life Sciences Education*, *11*(3), 209–215. https://doi.org/10.1187/cbe.12-06-0074

Collins, C. S., & Stockton, C. M. (2018). The central role of theory in qualitative research. *International Journal of Qualitative Methods*, *17*(1), 1-10. https://doi.org/10.1177/1609406918797475

Cook, M. P. (2006). Visual representations in science education: The influence of prior knowledge and cognitive load theory on instructional design principles. *Science Education*, *90*(6), 1073–1091. https://doi.org/10.1002/sce.20164

Corbin, J., & Strauss, A. (1990). Grounded theory research: Procedures, canons and evaluative criteria. *Zeitschrift Fur Soziologie*, *19*(6), 418–427.

Creswell, J. W. (2013). Philosophical assumptions and interpretive frameworks. In *Qualitative inquiry and research design: Choosing among five approaches* (pp. 15–41). SAGE Publications.

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry & research design: Choosing among five approaches* (4th ed.). SAGE.

Crisci, J. V., McInerney, J. D., & McWethy, P. J. (1993). *Order & diversity in the living world: Teaching taxonomy & systematics in schools*. National Association of Biology Teachers.

Crisp, M. D., & Cook, L. G. (2005). Do early branching lineages signify ancestral traits? *Trends in Ecology & Evolution*, *20*(3), 122–128. https://doi.org/10.1016/j.tree.2004.11.010

Cunningham, C. W., Omland, K. E., & Oakley, T. H. (1998). Reconstructing ancestral character states: A critical reappraisal. *Trends in Ecology & Evolution*, *13*(9), 361–366. https://doi.org/10.1016/S0169-5347(98)01382-2

Damianova, M. K., & Sullivan, G. B. (2011). Rereading Vygotsky's theses on types of internalization and verbal mediation. *Review of General Psychology*, *15*(4), 344–350. https://doi.org/10.1037/a0025627

Daniel, K. L., Bucklin, C. J., Austin Leone, E., & Idema, J. (2018). Towards a definition of representational competence. In K. L. Daniel (Ed.), *Towards a framework for representational competence in science education* (pp. 3–11). Springer International Publishing. https://doi.org/10.1007/978-3-319-89945-9_1

Davenport, K., Milks, K. J., & Tassell, R. V. (2015). Investigating tree thinking & ancestry with cladograms. *The American Biology Teacher*, *77*(3), 198–204. https://doi.org/10.1525/abt.2015.77.3.8

Davis, L. P., & Museus, S. D. (2019). What is deficit thinking? An analysis of conceptualizations of deficit thinking and implications for scholarly research. *NCID Currents*, *1*(1), 117-130. https://doi.org/10.3998/currents.17387731.0001.110

Dees, J., Bussard, C., & Momsen, J. L. (2018). Further effects of phylogenetic tree style on student comprehension in an introductory biology course. *CBE—Life Sciences Education*, *17*(2), Article ar17. https://doi.org/10.1187/cbe.17-03-0058

Dees, J., Freiermuth, D., & Momsen, J. L. (2017). Effects of phylogenetic tree style on student comprehension in an introductory biology course. *The American Biology Teacher*, *79*(9), 729–737. https://doi.org/10.1525/abt.2017.79.9.729

Dees, J., & Momsen, J. L. (2016). Student construction of phylogenetic trees in an introductory biology course. *Evolution: Education and Outreach*, *9*(1), Article 3. https://doi.org/10.1186/s12052-016-0054-y

Dees, J., Momsen, J. L., Niemi, J., & Montplaisir, L. (2014). Student interpretations of phylogenetic trees in an introductory biology course. *CBE—Life Sciences Education*, *13*(4), 666–676. https://doi.org/10.1187/cbe.14-01-0003

De Lenos Meira, L. R. (2002). Mathematical representations as systems of notations-in-use. In K. Gravemeijer, R. Lehrer, B. Van Oers, & L. Verschaffel (Eds.), *Symbolizing, modeling and tool use in mathematics education* (pp. 87–103). Springer Netherlands. https://doi.org/10.1007/978-94-017-3194-2_6

Diessel, H., & Coventry, K. R. (2020). Demonstratives in spatial language and social interaction: An interdisciplinary review. *Frontiers in Psychology*, *11*, Article 555265. https://doi.org/10.3389/fpsyg.2020.555265

Diogo, R., Ziermann, J. M., & Linde-Medina, M. (2015). Is evolutionary biology becoming too politically correct? A reflection on the scala naturae, phylogenetically basal clades, anatomically plesiomorphic taxa, and 'lower' animals. *Biological Reviews*, *90*(2), 502–521. https://doi.org/10.1111/brv.12121

diSessa, A. A. (2007). An interactional analysis of clinical interviewing. *Cognition and Instruction*, *25*(4), 523–565. https://doi.org/10.1080/07370000701632413

diSessa, A. A. (2018). A friendly introduction to "knowledge in pieces": Modeling types of knowledge and their roles in learning. In G. Kaiser, H. Forgasz, M. Graven, A. Kuzniak, E. Simmt, & B. Xu (Eds.), *Invited lectures from the 13th international congress on mathematical education* (pp. 65-84). Springer. http://link.springer.com/10.1007/978-3-319-72170-5

Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *The American Biology Teacher*, *35*(3), 125–129. https://doi.org/10.2307/4444260

Dodick, J., & Aharonson, M. (2009, August). *Phylogeny exhibits and understanding geological time*. Understanding the Tree of Life Harvard Conference.

Doolittle, W. F., & Bapteste, E. (2007). Pattern pluralism and the tree of life hypothesis. *Proceedings of the National Academy of Sciences*, *104*(7), 2043–2049. https://doi.org/10.1073/pnas.0610699104

Eddy, S. L., Crowe, A. J., Wenderoth, M. P., & Freeman, S. (2013). How should we teach tree-thinking? An experimental test of two hypotheses. *Evolution: Education and Outreach*, *6*(1), Article 13. https://doi.org/10.1186/1936-6434-6-13

Elliot, A. J., & Harackiewicz, J. M. (1996). Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis. *Journal of Personality and Social Psychology*, *70*(3), 461–475. https://doi.org/10.1037/0022-3514.70.3.461

Elliot, A. J., & McGregor, H. A. (2001). A 2 × 2 achievement goal framework. *Journal of Personality and Social Psychology*, *80*(3), 501–519. https://doi.org/10.1037/0022-3514.80.3.501

Ericsson, K. A., & Simon, H., A. (1993). Introduction and summary. In *Protocol analysis: Verbal reports as data* (2nd ed.). The MIT Press.

Espeland, M., Breinholt, J., Willmott, K. R., Warren, A. D., Vila, R., Toussaint, E. F. A., Maunsell, S. C., Aduse-Poku, K., Talavera, G., Eastwood, R., Jarzyna, M. A., Guralnick, R., Lohman, D. J., Pierce, N. E., & Kawahara, A. Y. (2018). A comprehensive and dated phylogenomic analysis of butterflies. *Current Biology*, *28*(5), 770-778.e5. https://doi.org/10.1016/j.cub.2018.01.061

Figueredo, A. J., & Berry, S. C. (2002). "Just not so stories": Exaptations, spandrels, and constraints. *Behavioral and Brain Sciences*, *25*(4), 534-547. https://doi.org/10.1017/S0140525X02530090

Finarelli, J. A., & Goswami, A. (2013). Potential pitfalls of reconstructing deep time evolutionary history with only extant data, a case study using the Canidae (Mammalia, Carnivora): Brief communication. *Evolution*, *67*(12), 3678–3685. https://doi.org/10.1111/evo.12222

Fink, A., Frey, R. F., & Solomon, E. D. (2020). Belonging in general chemistry predicts first-year undergraduates' performance and attrition. *Chemistry Education Research and Practice*, *21*(4), 1042–1062. https://doi.org/10.1039/D0RP00053A

Firneno, T. J., Welsh, G. T., Nimlos, P. J., & Larson, E. L. (2023). "I think": Integrating project-based learning and case study to teach fundamentals of evolutionary tree-thinking. *Evolution: Education and Outreach*, *16*(1), Article 13. https://doi.org/10.1186/s12052-023-00192-8

Fisher, M., & Keil, F. C. (2016). The curse of expertise: When more knowledge leads to miscalibrated explanatory insight. *Cognitive Science*, *40*(5), 1251–1269. https://doi.org/10.1111/cogs.12280

Fitzhugh, K. (2012). The limits of understanding in biological systematics. *Zootaxa*, *3435*(1), 40-67. https://doi.org/10.11646/zootaxa.3435.1.2

Florence, M. K., & Yore, L. D. (2004). Learning to write like a scientist: Coauthoring as an enculturation task. *Journal of Research in Science Teaching*, *41*(6), 637–668. https://doi.org/10.1002/tea.20015

Foster, A. (2008). Games and motivation to learn science: Personal identity, applicability, relevance and meaningfulness. *Journal of Interactive Learning Research*, *19*(4), 594-614.

Franklin, C. S., Cody, P. A., & Ballan, M. (2010). Reliability and validity in qualitative research. In B. A. Thyer (Ed.), *The handbook of social work research methods* (2nd ed.). SAGE.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford University Press.

Gelman, S. A., & Coley, J. D. (1990). The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children. *Developmental Psychology*, *26*(5), 796-804.

Gelman, S. A., & Markman, E. M. (1987). Young children's inductions from natural kinds: The role of categories and appearances. *Child Development, 58*(6), 1532-1541.

Gendron, R. P. (2000). The classification & evolution of caminalcules. *The American Biology Teacher*, *62*(8), 570-576.

Genest, A., & Gutwin, C. (2011). Characterizing deixis over surfaces to improve remote embodiments. In S. Bødker, N. O. Bouvin, V. Wulf, L. Ciolfi, & W. Lutters (Eds.), *ECSCW 2011: Proceedings of the 12th European conference on computer supported cooperative work* (pp. 253–272). Springer London. https://doi.org/10.1007/978-0-85729-913-0_14

Gibson, J. P., & Cooper, J. T. (2017). Botanical phylo-cards: A tree-thinking game to teach plant evolution. *The American Biology Teacher*, *79*(3), 241–244. https://doi.org/10.1525/abt.2017.79.3.241

Gilbert, J. K. (Ed.). (2007). *Visualization in science education*. Springer.

Ginsburg, H. P. (1997). *Entering the child's mind: The clinical interview psychological research and practice*. Cambridge University Press.

Giusti, P. E., & Scott, M. (n.d.). *Yale Peabody museum of natural history: Tree of life visitor study*. National Science Foundation.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory*. Aldine Publishing.

Glasersfeld, E. von. (1995). *Radical constructivism: A way of knowing and learning*. Falmer Press.

Goldsmith, D. W. (2003). The great clade race. *The American Biology Teacher*, *65*(9), 679–682.

González Galli, L., Peréz, G., & Gómez Galindo, A. A. (2020). The self-regulation of teleological thinking in natural selection learning. *Evolution: Education and Outreach*, *13*(1), Article 6. https://doi.org/10.1186/s12052-020-00120-0

Goodwin, C. (1995). Seeing in depth. S*ocial Studies of Science*, *25*(2), pp. 237–274.

Gottwald, J. M., Elsner, B., & Pollatos, O. (2015). Good is up—Spatial metaphors in action observation. *Frontiers in Psychology*, *6,* Article 1605. https://doi.org/10.3389/fpsyg.2015.01605

Gould, S. J., & Eldredge, N. (1977). Punctuated equilibria: The tempo and mode of evolution reconsidered. *Paleobiology*, *3*(2), 115-151.

Gradoville, M. (2023). The future of exemplar theory. In M. Díaz-Campos & S. Balasch (Eds.), *The handbook of usage-based linguistics* (1st ed., pp. 527–544). Wiley. https://doi.org/10.1002/9781119839859.ch29

Gregory, T. R. (2008). Understanding evolutionary trees. *Evolution: Education and Outreach*, *1*(2), 121–137. https://doi.org/10.1007/s12052-008-0035-x

Grinyer, A. (2002). The anonymity of research participants: Assumptions, ethics and practicalities. *Social Research Update*, *36,* 1-4.

Gutwin, C., & Penner, R. (2002). *Improving interpretation of remote gestures with telepointer traces* [Paper presentation]. Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work, New Orleans, LA, United States.

Halverson, K. L. (2009). *Investigating the development and use of phylogenetic representations by college undergraduates in a plant systematics course* [Doctoral dissertation, University of Missouri-Columbia]. MoSpace. https://doi.org/10.32469/10355/9678

Halverson, K. L. (2010). Using pipe cleaners to bring the tree of life to life. *The American Biology Teacher*, *72*(4), 223–224. https://doi.org/10.1525/abt.2010.72.4.4

Halverson, K. L. (2011). Improving tree-thinking one learnable skill at a time. *Evolution: Education and Outreach*, *4*(1), 95–106. https://doi.org/10.1007/s12052-010-0307-0

Halverson, K. L., Abell, S. K., Friedrichsen, P. M., & Pires, J. C. (2009). *Testing a model of representational competence applied to phylogenetic tree thinking* [Paper presentation]. National Association of Research in Science Teaching (NARST) Annual Conference, Garden Grove, CA, United States.

Halverson, K. L., Boyce, C. J., & Maroo, J. D. (2013). Order matters: Pre-assessments and student generated representations. *Evolution: Education and Outreach*, *6*(1), Article 24. https://doi.org/10.1186/1936-6434-6-24

Halverson, K. L., & Friedrichsen, P. (2013). Learning tree thinking: Developing a new framework of representational competence. In D. F. Treagust & C.-Y. Tsui (Eds.), *Multiple representations in biological education* (Vol. 7, pp. 185–201). Springer Netherlands. https://doi.org/10.1007/978-94-007-4192-8_11

Halverson, K. L., Pires, C. J., & Abell, S. K. (2011). Exploring the complexity of tree thinking expertise in an undergraduate systematics course. *Science Education*, *95*(5), 794–823. https://doi.org/10.1002/sce.20436

Hammann, M., & Nehm, R. H. (2020). Teleology and evolution education: Introduction to the special issue. *Evolution: Education and Outreach*, *13*(1), Article 16. https://doi.org/10.1186/s12052-020-00130-y

Hamnell-Pamment, Y. (2023). The role of scientific language use and achievement level in student sensemaking. *International Journal of Science and Mathematics Education*. https://doi.org/10.1007/s10763-023-10405-7

Hampton, J. A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *18*(4), 441–461. https://doi.org/10.1016/S0022-5371(79)90246-9

Hampton, J. A. (1981). An investigation of the nature of abstract concepts. *Memory & Cognition*, *9*(2), 149–156. https://doi.org/10.3758/BF03202329

Harrison, A. G., & Treagust, D. F. (2000). A typology of school science models. *International Journal of Science Education*, *22*(9), 1011–1026. https://doi.org/10.1080/095006900416884

Hart, K. R., & Long, J. H. (2011). Animal metaphors and metaphorizing animals: An integrated literary, cognitive, and evolutionary analysis of making and partaking of stories. *Evolution: Education and Outreach*, *4*(1), 52–63. https://doi.org/10.1007/s12052-010-0301-6

Hay, D. B., Williams, D., Stahl, D., & Wingate, R. J. (2013). Using drawings of the brain cell to exhibit expertise in neuroscience: Exploring the boundaries of experimental culture. *Science Education*, *97*(3), 468–491. https://doi.org/10.1002/sce.21055

Heredia, S. C., Furtak, E. M., & Morrison, D. (2016). Exploring the influence of plant and animal item contexts on student response patterns to natural selection multiple choice items. *Evolution: Education and Outreach*, *9*(1), Article 10. https://doi.org/10.1186/s12052-016-0061-z

Hewson, P. W., & Hewson, M. G. A. (1989). Analysis and use of a task for identifying conceptions of teaching science. *Journal of Education for Teaching*, *15*(3), 191–209. https://doi.org/10.1080/0260747890150302

Hillis, D. M. (1998). Taxonomic sampling, phylogenetic accuracy, and investigator bias. *Systematic Biology*, *47*(1), 3–8.

Hinchliff, C. E., Smith, S. A., Allman, J. F., Burleigh, J. G., Chaudhary, R., Coghill, L. M., Crandall, K. A., Deng, J., Drew, B. T., Gazis, R., Gude, K., Hibbett, D. S., Katz, L. A., Laughinghouse, H. D., McTavish, E. J., Midford, P. E., Owen, C. L., Ree, R. H., Rees, J. A., … Cranston, K. A. (2015). Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proceedings of the National Academy of Sciences*, *112*(41), 12764–12769. https://doi.org/10.1073/pnas.1423041112

Ho, A. K., Roberts, S. O., & Gelman, S. A. (2015). Essentialism and racial bias jointly contribute to the categorization of multiracial individuals. *Psychological Science*, *26*(10), 1639–1645. https://doi.org/10.1177/0956797615596436

Horsfall, M., Eikelenboom, M., Draisma, S., & Smit, J. H. (2021). The effect of rapport on data quality in face-to-face interviews: Beneficial or detrimental? *International Journal of Environmental Research and Public Health*, *18*(20), Article 10858. https://doi.org/10.3390/ijerph182010858

Huang, S., & Whittall, J. B. (2018). A tree of trees: Using campus tree diversity to integrate molecular, organismal, and evolutionary biology. *The American Biology Teacher*, *80*(2), 144–151. https://doi.org/10.1525/abt.2018.80.2.144

Imenda, S. (2014). Is there a conceptual difference between theoretical and conceptual frameworks? *Journal of Social Sciences*, *38*(2), 185–195. https://doi.org/10.1080/09718923.2014.11893249

Jenkins, K., Mead, L., Baum, D. A., Daniel, K. L., Bucklin, C. J., Leone, E. A., Gibson, J. P., & Naegle, E. (2022). Developing the BETTSI: A tree-thinking diagnostic tool to assess individual elements of representational competence. *Evolution, 76*(4), 708-721.

Jenner, R. A. (2022). *Ancestors in evolutionary biology: Linear thinking about branching trees*. Cambridge University Press.

Jiang, L., & Elen, J. (2011). Why do learning goals (not) work: A reexamination of the hypothesized effectiveness of learning goals based on students' behaviour and cognitive processes. *Educational Technology Research and Development*, *59*(4), 553–573. https://doi.org/10.1007/s11423-011-9200-y

Johnson, S. C., & Solomon, G. E. A. (1997). Why dogs have puppies and cats have kittens: The role of birth in young children's understanding of biological origins. *Child Development*, *68*(3), 404-419. https://doi.org/10.2307/1131668

Jose, S. B., Wu, C., & Kamoun, S. (2019). Overcoming plant blindness in science, education, and society. *Plants, People, Planet*, *1*(3), 169–172. https://doi.org/10.1002/ppp3.51

Julius, M. L., & Schoenfuss, H. L. (2006). Phylogenetic reconstruction as a broadly applicable teaching tool in the biology classroom. *Journal of College Science Teaching*, *35*(7), 40–45.

Kallery, M., & Psillos, D. (2004). Anthropomorphism and animism in early years science: Why teachers use them, how they conceptualise them and what are their views on their use. *Research in Science Education*, *34*(3), 291–311. https://doi.org/10.1023/B:RISE.0000044613.64634.03

Kampourakis, K. (2020). Students' "teleological misconceptions" in evolution education: Why the underlying design stance, not teleology per se, is the problem. *Evolution: Education and Outreach*, *13*(1), Article 1. https://doi.org/10.1186/s12052-019-0116-z

Kaplan, S. E., Gunn, C. M., Kulukulualani, A. K., Raj, A., Freund, K. M., & Carr, P. L. (2018). Challenges in recruiting, retaining and promoting racially and ethnically diverse faculty. *Journal of the National Medical Association*, *110*(1), 58–64. https://doi.org/10.1016/j.jnma.2017.02.001

Karimi, N., Parks, B. M., Rouse, D., Martin, K., Dong, X., Rajangam, P. C., Baum, D. A., & Heitz, J. G. (2017). *Building trees: Introducing evolutionary concepts by exploring Crassulaceae phylogeny and biogeography*. CourseSource. https://doi.org/10.24918/cs.2017.16

Keesey, M. (2007, November 21). Trees, bubbles, and hooves. *A Three-Pound Monkey Brain*. http://3lbmonkeybrain.blogspot.com/2007/11/trees-bubbles-and-hooves.html

Kelemen, D., & Rosset, E. (2009). The human function compunction: Teleological explanation in adults. *Cognition*, *111*(1), 138–143. https://doi.org/10.1016/j.cognition.2009.01.001

Kelemen, D., Rottman, J., & Seston, R. (2013). Professional physical scientists display tenacious teleological tendencies: Purpose-based reasoning as a cognitive default. *Journal of Experimental Psychology: General*, *142*(4), 1074–1083. https://doi.org/10.1037/a0030399

Kelly, R. M., Akaygun, S., Hansen, S. J. R., & Villalta-Cerdas, A. (2017). The effect that comparing molecular animations of varying accuracy has on students' submicroscopic explanations. *Chemistry Education Research and Practice*, *18*(4), 582–600. https://doi.org/10.1039/C6RP00240D

Kember, D., Ho, A., & Hong, C. (2008). The importance of establishing relevance in motivating student learning. *Active Learning in Higher Education*, *9*(3), 249–263. https://doi.org/10.1177/1469787408095849

Keyton, J. (2018). Interaction analysis: An introduction. In E. Brauner, M. Boos, & M. Kolbe (Eds.), *The Cambridge handbook of group interaction analysis* (1st ed., pp. 3–19). Cambridge University Press. https://doi.org/10.1017/9781316286302.002

Kimble, M., Boxwala, M., Bean, W., Maletsky, K., Halper, J., Spollen, K., & Fleming, K. (2014). The impact of hypervigilance: Evidence for a forward feedback loop. *Journal of Anxiety Disorders*, *28*(2), 241–245. https://doi.org/10.1016/j.janxdis.2013.12.006

Kohl, P. B., & Finkelstein, N. D. (2008). Patterns of multiple representation use by experts and novices during physics problem solving. *Physical Review Special Topics - Physics Education Research*, *4*(1), Article 010111. https://doi.org/10.1103/PhysRevSTPER.4.010111

Kong, Y., Apodaca, J., & Olimpo, J. T. (2022). Implementation and evaluation of the Model of the Use of Evolutionary Trees (MUET) curricular module in an introductory organismal biology course. *International Journal of Science Education*, *44*(15), 2381–2396. https://doi.org/10.1080/09500693.2022.2124132

Kortz, K. M., Swaddle, J. P., & Fastovsky, D. E. (2012). How undergraduate students misread cladograms. *The Paleontological Society Special Publications*, *12*, 123–134.

Kosinski, R. J. (2006). An introduction to phylogenetic analysis. *Tested Studies for Laboratory Teaching*, *27*, 57-106. https://www.ableweb.org/biologylabs/wp-content/uploads/volumes/vol-27/05_Kosinski.pdf

Kövecses, Z. (2008). Conceptual metaphor theory: Some criticisms and alternative proposals. *Annual Review of Cognitive Linguistics*, *6*, 168–184. https://doi.org/10.1075/arcl.6.08kov

Kozma, R. B. (2020). Use of multiple representations by experts and novices. In P. Van Meter, A. List, D. Lombardi, & P. Kendeou (Eds.), *Handbook of learning from multiple representations and perspectives* (1st ed., pp. 33–47). Routledge. https://doi.org/10.4324/9780429443961-4

Kozma, R. B., Chin, E., Russell, J., & Marx, N. (2000). The roles of representations and tools in the chemistry laboratory and their implications for chemistry learning. *Journal of the Learning Sciences*, *9*(2), 105–143. https://doi.org/10.1207/s15327809jls0902_1

Kozma, R. B., & Russell, J. (1997). Multimedia and understanding: Expert and novice responses to different representations of chemical phenomena. *Journal of Research in Science Teaching, 34*(9), 949-968.

Kozma, R. B., & Russell, J. (2005). Students becoming chemists: Developing representational competence. In J. K. Gilbert (Ed.), *Visualization in science education* (pp. 121–145). Springer Netherlands. https://doi.org/10.1007/1-4020-3613-2_8

Krell, F.-T., & Cranston, P. S. (2004). Which side of the tree is more basal? *Systematic Entomology*, *29*(3), 279–281. https://doi.org/10.1111/j.0307-6970.2004.00262.x

Kruit, P. M., Oostdam, R. J., Van Den Berg, E., & Schuitema, J. A. (2018). Effects of explicit instruction on the acquisition of students' science inquiry skills in grades 5 and 6 of primary education. *International Journal of Science Education*, *40*(4), 421–441. https://doi.org/10.1080/09500693.2018.1428777

Kuczynski, L., & Daly, K. (2003). Qualitative methods for inductive (theory-generating) research: Psychological and sociological approaches. In L. Kuczynski (Ed.), *Handbook of dynamics in parent-child relations* (pp. 373–392). SAGE Publications. https://doi.org/10.4135/9781452229645.n18

Kumala, M. (2010a). A natural history of you. *Evolution: Education and Outreach*, *3*(4), 532–538. https://doi.org/10.1007/s12052-010-0276-3

Kumala, M. (2010b). The gummy tree challenge—Building connections one treat at a time. *Evolution: Education and Outreach*, *3*(4), 520–525. https://doi.org/10.1007/s12052-010-0275-4

Kummer, T. A. (2017). *Assessing and improving student understanding of tree-thinking*. Brigham Young University.

Kummer, T. A., Whipple, C. J., Bybee, S. M., Adams, B. J., & Jensen, J. L. (2019). Development of an Evolutionary Tree Concept Inventory. *Journal of Microbiology & Biology Education*, *20*(2). https://doi.org/10.1128/jmbe.v20i2.1700

Kummer, T. A., Whipple, C. J., & Jensen, J. L. (2016). Prevalence and persistence of misconceptions in tree thinking. *Journal of Microbiology & Biology Education*, *17*(3), 389–398. https://doi.org/10.1128/jmbe.v17i3.1156

Kunin, V. (2005). The net of life: Reconstructing the microbial phylogenetic network. *Genome Research*, *15*(7), 954–959. https://doi.org/10.1101/gr.3666505

Kuzoff, R. K., Kemmeter, S. B., McKinnon, J. S., & Thompson, C. P. (2009). Phylogenetic analysis: How old are the parts of your body? *Evolution: Education and Outreach*, *2*(3), 405–414. https://doi.org/10.1007/s12052-009-0157-9

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. University of Chicago Press.

Lakoff, G. (2012). Explaining embodied cognition results. *Topics in Cognitive Science*, *4*(4), 773–785. https://doi.org/10.1111/j.1756-8765.2012.01222.x

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh*. Basic Books.

Lakoff, G., & Johnson, M. (2011). *Metaphors we live by: With a new afterword*. University of Chicago Press.

Latour, B. (1990). Visualisation and cognition: Drawing things together. In M. E. Lynch and S. Woolgar (Eds.), *Representation in scientific practices* (pp. 19–68). MIT Press.

Lee, M. S. Y., & Palci, A. (2015). Morphological phylogenetics in the genomic age. *Current Biology*, *25*(19), R922–R929. https://doi.org/10.1016/j.cub.2015.07.009

Legare, C. H., Lane, J. D., & Evans, E. M. (2013). Anthropomorphizing science: How does it affect the development of evolutionary concepts? *Merrill-Palmer Quarterly*, *59*(2), 168-197. https://doi.org/10.13110/merrpalmquar1982.59.2.0168

Lents, N. H., Cifuentes, O. E., & Carpi, A. (2010). Teaching the process of molecular phylogeny and systematics: A multi-part inquiry-based exercise. *CBE—Life Sciences Education*, *9*(4), 513–523. https://doi.org/10.1187/cbe.09-10-0076

Liu, J., & Novick, L. R. (2022). Perceptual grouping affects students' propensity to make inferences consistent with their misconceptions. *Journal of Experimental Psychology: Applied, 29*(2), 322-340. https://doi.org/10.1037/xap0000443

Lobato, J., Orrill, C. H., Druken, B., & Jacobson, E. (2011). *Middle school teachers' knowledge of proportional reasoning for teaching* [Paper presentation]. Annual Meeting of the American Educational Research Association, New Orleans, LA, United States.

Lobato, J., & Walters, C. D. (2017). A taxonomy of approaches to learning trajectories and progressions. In J. Cai (Ed.), *The compendium for research in mathematics education* (pp. 74-101). National Council of Teachers of Mathematics.

Lomborg, K., & Kirkevold, M. (2003). Truth and validity in grounded theory - A reconsidered realist interpretation of the criteria: Fit, work, relevance and modifiability. *Nursing Philosophy*, *4*(3), 189–200. https://doi.org/10.1046/j.1466-769X.2003.00139.x

Lombrozo, T., Kelemen, D., & Zaitchik, D. (2007). Inferring design: Evidence of a preference for teleological explanations in patients with Alzheimer's disease. *Psychological Science*, *18*(11), 999–1006. https://doi.org/10.1111/j.1467-9280.2007.02015.x

Luft, J. A., Jeong, S., Idsardi, R., & Gardner, G. (2022). Literature reviews, theoretical frameworks, and conceptual frameworks: An introduction for new biology education researchers. *CBE—Life Sciences Education*, *21*(3), Article rm33. https://doi.org/10.1187/cbe.21-05-0134

Mabry, M. E., Turner-Hissong, S. D., Gallagher, E. Y., McAlvay, A. C., An, H., Edger, P. P., Moore, J. D., Pink, D. A. C., Teakle, G. R., Stevens, C. J., Barker, G., Labate, J., Fuller, D. Q., Allaby, R. G., Beissinger, T., Decker, J. E., Gore, M. A., & Pires, J. C. (2021). The evolutionary history of wild, domesticated, and feral *Brassica oleracea* (Brassicaceae). *Molecular Biology and Evolution*, *38*(10), 4419–4434. https://doi.org/10.1093/molbev/msab183

MacDonald, T. E. (2014). *Evolutionary tree design: An exploratory study of the influence of linear versus branching format on visitors' interpretation and understanding across age groups*. University of Kansas.

MacDonald, T. E., & Wiley, E. O. (2012). Communicating phylogeny: Evolutionary tree diagrams in museums. *Evolution: Education and Outreach*, *5*(1), 14–28. https://doi.org/10.1007/s12052-012-0387-0

MacFadden, B. J. (2005). Fossil horses—Evidence for evolution. *Science*, *307*(5716), 1728–1730. https://doi.org/10.1126/science.1105458

MacFadden, B. J., Dunckel, B. A., Ellis, S., Dierking, L. D., Abraham-Silver, L., Kisiel, J., & Koke, J. (2007). Natural history museum visitors' understanding of evolution. *BioScience*, *57*(10), 875–882. https://doi.org/10.1641/B571010

Mann, K., & MacLeod, A. (2015). Constructivism: Learning theories and approaches to research. In J. Cleland & S. J. Durning (Eds.), *Researching medical education* (1st ed., pp. 49–66). Wiley. https://doi.org/10.1002/9781118838983.ch6

Marcelos, M. F., & Nagem, R. L. (2012). Use of the "tree" analogy in evolution teaching by biology teachers. *Science & Education*, *21*(4), 507–541. https://doi.org/10.1007/s11191-011-9370-3

Marchenkova, L. (2005). Language, culture, and self: The Bakhtin-Vygotsky encounter. In J. K. Hall & G. Vitanova (Eds.), *Dialogue with Bakhtin on second and foreign language learning: New perspectives*. L. Erlbaum.

Marcroft, T. A. (2019). Comparing drawings of evolution by experts and novices [Unpublished manuscript]. University of California, San Diego, and San Diego State University.

Marcroft, T. A., Rasmussen, C., & Kelley, S. T. (2022). Computing in bioinformatics and engaged student learning: Student perspectives on anticipatory activities and innovative apps. *Journal of College Science Teaching*, *52*(2), 80-87.

Margolin, J., Pinerua, I., & Gerdeman, D. (2022). *Modeling in science education: a synthesis of recent discovery research prek-12 projects*. American Institutes for Research.

Margolis, E. (1994). A reassessment of the shift from the classical theory of concepts to prototype theory. *Cognition*, *51*(1), 73–89. https://doi.org/10.1016/0010-0277(94)90009-4

Martin, A. J. (2016). *Can educators reduce students' cognitive load and boost motivation and engagement?: Integrating explicit instruction and discovery learning through load reduction instruction (LRI)*. British Psychological Society.

Martinovic, D., & Manizade, A. (2017). Using grounded theory to extend existing PCK framework at the secondary level. *Education Sciences*, *7*(2), Article 60. https://doi.org/10.3390/educsci7020060

Matuk, C. (2007). Images of evolution. *Journal of Biocommunication, 33*(3), E54-E61.

Matuk, C. (2010). *Narratives in mind and media: A cognitive semiotic account of novices interpreting science media*. Northwestern University.

Matuk, C., & Uttal, D. H. (2012). Narrative spaces in the representation and understanding of evolution. In K. S. Rosengren, S. K. Brem, E. M. Evans, & G. M. Sinatra (Eds.), *Evolution challenges* (pp. 119–144). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199730421.003.0006

Matuk, C., & Uttal, D. H. (2020). The effects of invention and recontextualization on representing and reasoning with trees of life. *Research in Science Education*, *50*(5), 1991–2033. https://doi.org/10.1007/s11165-018-9761-4

McCarthy, J. (2020). *Nearly one in four in U.S. have cut back on eating meat*. Gallup. https://news.gallup.com/poll/282779/nearly-one-four-cut-back-eating-meat.aspx

McCullough, E. L., Verdeflor, L., Weinsztok, A., Wiles, J. R., & Dorus, S. (2020). Exploratory activities for understanding evolutionary relationships depicted by phylogenetic trees: United but diverse. *The American Biology Teacher*, *82*(5), 333–337. https://doi.org/10.1525/abt.2020.82.5.333

McPeek, M. A., Shen, L., Torrey, J. Z., & Farid, H. (2008). The tempo and mode of three-dimensional morphological evolution in male reproductive structures. *The American Naturalist*, *171*(5), E158–E178. https://doi.org/10.1086/587076

Meier, B. P., & Robinson, M. D. (2004). Why the sunny side is up: Associations between affect and vertical position. *Psychological Science*, *15*(4), 243–247. https://doi.org/10.1111/j.0956-7976.2004.00659.x

Meir, E., Perry, J., Herron, J. C., & Kingsolver, J. (2007). College students' misconceptions about evolutionary trees. *The American Biology Teacher*, *69*(7). https://doi.org/10.1662/0002-7685(2007)69[71:CSMAET]2.0.CO;2

Meisel, R. P. (2010). Teaching tree-thinking to undergraduate biology students. *Evolution: Education and Outreach*, *3*(4), 621–628. https://doi.org/10.1007/s12052-010-0254-9

Mervis, C. B., Catlin, J., & Rosch, E. (1976). Relationships among goodness-of-example, category norms, and word frequency. *Bulletin of the Psychonomic Society*, *7*(3), 283–284. https://doi.org/10.3758/BF03337190

Miles, L. K., Stuart, S. B., & Macrae, C. N. (2010). Moving through time. *Psychological Science*, *21*(2), 222-223.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). SAGE Publications.

Miller, C. R., & Hartzog, M. (2020). "Tree thinking": The rhetoric of tree diagrams in biological thought. *Poroi*, *15*(2). https://doi.org/10.13008/2151-2957.1290

Miller, J. D., Scott, E. C., Ackerman, M. S., Laspra, B., Branch, G., Polino, C., & Huffaker, J. S. (2022). Public acceptance of evolution in the United States, 1985–2020. *Public Understanding of Science*, *31*(2), 223–238. https://doi.org/10.1177/09636625211035919

Ming, X., Fu, J., & Yang, J. (2022). Obtaining help by standing higher: The mediating role of moral reputation. *Frontiers in Psychology*, *13*, Article 1013656. https://doi.org/10.3389/fpsyg.2022.1013656

Minstrell, J. (1982). Explaining the "at rest" condition of an object. *The Physics Teacher*, *20*(1), 10–14. https://doi.org/10.1119/1.2340924

Mørk, B. E., Hoholm, T., Ellingsen, G., Edwin, B., & Aanestad, M. (2010). Challenging expertise: On power relations within and across communities of practice in medical innovation. *Management Learning*, *41*(5), 575–592. https://doi.org/10.1177/1350507610374552

Morrison, D. A. (2014). Is the tree of life the best metaphor, model, or heuristic for phylogenetics? *Systematic Biology*, *63*(4), 628–638. https://doi.org/10.1093/sysbio/syu026

Munfaridah, N., Avraamidou, L., & Goedhart, M. (2021). The use of multiple representations in undergraduate physics education: What do we know and where do we go from here? *Eurasia Journal of Mathematics, Science and Technology Education*, *17*(1), Article em1934. https://doi.org/10.29333/ejmste/9577

Mungan, E. (2023). Gestalt theory: A revolution put on pause? Prospects for a paradigm shift in the psychological sciences. *New Ideas in Psychology*, *71*, Article 101036. https://doi.org/10.1016/j.newideapsych.2023.101036

Muradoglu, M., Marchak, K. A., Gelman, S. A., & Cimpian, A. (2022). Formal explanations shape children's representations of animal kinds and social groups. *Developmental Psychology*, *58*(12), 2322–2335. https://doi.org/10.1037/dev0001447

Murayama, K., Elliot, A. J., & Yamagata, S. (2011). Separation of performance-approach and performance-avoidance achievement goals: A broader analysis. *Journal of Educational Psychology*, *103*(1), 238–256. https://doi.org/10.1037/a0021948

Murphy, G. L. (2016). Is there an exemplar theory of concepts? *Psychonomic Bulletin & Review*, *23*(4), 1035–1042. https://doi.org/10.3758/s13423-015-0834-3

National Research Council. (2000). *How people learn: Brain, mind, experience, and school: Expanded edition*. National Academy Press.

National Research Council. (2012). *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*. National Academies Press. https://doi.org/10.17226/13362

Nee, S. (2005). The great chain of being. *Nature*, *435*, Article 429.

NGSS Lead States. (2013). *Next generation science standards: For states, by states.* https://www.nextgenscience.org/

Ni, Y., Zhou, D.-H. R., Cai, J., Li, X., Li, Q., & Sun, I. X. (2018). Improving cognitive and affective learning outcomes of students through mathematics instructional tasks of high cognitive demand. *The Journal of Educational Research*, *111*(6), 704–719. https://doi.org/10.1080/00220671.2017.1402748

Novick, L. R., & Catley, K. M. (2007). Understanding phylogenies in biology: The influence of a Gestalt perceptual principle. *Journal of Experimental Psychology: Applied*, *13*(4), 197–223. https://doi.org/10.1037/1076-898X.13.4.197

Novick, L. R., & Catley, K. M. (2013). Reasoning about evolution's grand patterns: College students' understanding of the tree of life. *American Educational Research Journal*, *50*(1), 138–177.

Novick, L. R., & Catley, K. M. (2014). When relationships depicted diagrammatically conflict with prior knowledge: An investigation of students' interpretations of evolutionary trees: Interpreting evolutionary trees. *Science Education*, *98*(2), 269–304. https://doi.org/10.1002/sce.21097

Novick, L. R., & Catley, K. M. (2016). Fostering 21st-century evolutionary reasoning: Teaching tree thinking to introductory biology students. *CBE—Life Sciences Education*, *15*(4), Article ar66. https://doi.org/10.1187/cbe.15-06-0127

Novick, L. R., & Catley, K. M. (2018). Teaching tree thinking in an upper level organismal biology course: Testing the effectiveness of a multifaceted curriculum. *Journal of Biological Education*, *52*(1), 66–78. https://doi.org/10.1080/00219266.2017.1285804

Novick, L. R., Catley, K. M., & Funk, D. J. (2010). Characters are key: The effect of synapomorphies on cladogram comprehension. *Evolution: Education and Outreach*, *3*(4), 539–547. https://doi.org/10.1007/s12052-010-0243-z

Novick, L. R., Catley, K. M., & Schreiber, E. G. (2012). *Understanding evolutionary history: An introduction to tree thinking*. https://cdn.vanderbilt.edu/vu-my/wp-content/uploads/sites/2396/2017/02/14141743/TT_instruct_Novick_Catley_8-12.pdf

Novick, L. R., & Fuselier, L. C. (2019). Perception and conception in understanding evolutionary trees. *Cognition*, *192*, Article 104001. https://doi.org/10.1016/j.cognition.2019.06.013

Novick, L. R., Pickering, J., MacDonald, T., Diamond, J., Ainsworth, S., Aquino, A. E., Catley, K. M., Dodick, J., Evans, E. M., Matuk, C., Sacco, J., & Scott, M. (2014). Depicting the tree of life in museums: Guiding principles from psychological research. *Evolution: Education and Outreach*, *7*(1), Article 25. https://doi.org/10.1186/s12052-014-0025-0

Novick, L. R., Schreiber, E. G., & Catley, K. M. (2014). Deconstructing evolution education: The relationship between micro- and macroevolution. *Journal of Research in Science Teaching*, *51*(6), 759–788. https://doi.org/10.1002/tea.21161

Novick, L. R., Shade, C. K., & Catley, K. M. (2011). Linear versus branching depictions of evolutionary history: Implications for diagram design: Topics in cognitive science. *Topics in Cognitive Science*, *3*(3), 536–559. https://doi.org/10.1111/j.1756-8765.2009.01077.x

Novick, L. R., Stull, A. T., & Catley, K. M. (2012). Reading phylogenetic trees: The effects of tree orientation and text processing on comprehension. *BioScience*, *62*(8), 757–764. https://doi.org/10.1525/bio.2012.62.8.8

Núñez, R. E. (2006). Do real numbers really move? Language, thought, and gesture: The embodied cognitive foundations of mathematics. In R. Hersh (Ed.), *18 unconventional essays on the nature of mathematics* (pp. 160–181). Springer New York. https://doi.org/10.1007/0-387-29831-2_9

Núñez, R. E., Cooperrider, K., Doan, D., & Wassmann, J. (2012). Contours of time: Topographic construals of past, present, and future in the Yupno valley of Papua New Guinea. *Cognition*, *124*(1), 25–35. https://doi.org/10.1016/j.cognition.2012.03.007

Núñez, R. E., & Sweetser, E. (2006). With the future behind them: Convergent evidence from Aymara language and gesture in the crosslinguistic comparison of spatial construals of time. *Cognitive Science*, *30*(3), 401–450. https://doi.org/10.1207/s15516709cog0000_62

O'Brien, L. T., Bart, H. L., & Garcia, D. M. (2020). Why are there so few ethnic minorities in ecology and evolutionary biology? Challenges to inclusion and the role of sense of belonging. *Social Psychology of Education*, *23*(2), 449–477. https://doi.org/10.1007/s11218-019-09538-x

Ochs, E., Gonzales, P., & Jacoby, S. (1996). "When I come down I'm in the domain state": Grammar and graphic representation in the interpretive activity of physicists. In E. Ochs, E. Schegloff, & S. Thompson (Eds.), *Interaction and grammar* (pp. 328-369). Cambridge University Press.

Offner, S. (2016). Reconsidering animal phylogeny: Tree-thinking. *The American Biology Teacher*, *78*(5), 376–379. https://doi.org/10.1525/abt.2016.78.5.376

O'Hara, R. J. (1992). Telling the tree: Narrative representation and the study of evolutionary history. *Biology & Philosophy*, *7*(2), 135–160. https://doi.org/10.1007/BF00129880

O'Hara, R. J. (1997). Population thinking and tree thinking in systematics. *Zoologica Scripta*, *26*(4), 323–329. https://doi.org/10.1111/j.1463-6409.1997.tb00422.x

Oliver, R., Herrington, J., & Reeves, T. (2002). *Authentic activities and online learning* [Paper presentation]. Quality Conversations, Perth, Western Australia.

O'Malley, C., Langton, S., Anderson, A., Doherty-Sneddon, G., & Bruce, V. (1996). Comparison of face-to-face and video-mediated interaction. *Interacting with Computers*, *8*(2), 177–192. https://doi.org/10.1016/0953-5438(96)01027-2

Omland, K. E., Cook, L. G., & Crisp, M. D. (2008). Tree thinking for all biology: The problem with reading phylogenies as ladders of progress. *BioEssays*, *30*(9), 854–867. https://doi.org/10.1002/bies.20794

Opfermann, M., Schmeck, A., & Fischer, H. E. (2017). Multiple representations in physics and science education – Why should we use them? In D. F. Treagust, R. Duit, & H. E. Fischer (Eds.), *Multiple representations in physics education* (Vol. 10, pp. 1–22). Springer International Publishing. https://doi.org/10.1007/978-3-319-58914-5_1

Osborne, J. (2014). Teaching scientific practices: Meeting the challenge of change. *Journal of Science Teacher Education*, *25*(2), 177–196. https://doi.org/10.1007/s10972-014-9384-1

Ou, J., Fussell, S. R., Chen, X., Setlock, L. D., & Yang, J. (2003). *Gestural communication over video stream: Supporting multimodal interaction for remote collaborative physical tasks*. Proceedings of the 5th International Conference on Multimodal Interfaces, New York, NY, United States.

Painchaud-Nash, R. (2020). Untangling the web of deficit discourses in mathematics education. *For the Learning of Mathematics*, *40*(1), 42–46.

Pande, P., & Chandrasekharan, S. (2021). Expertise as sensorimotor tuning: Perceptual navigation patterns mark representational competence in science. *Research in Science Education, 52*, 725-747. https://doi.org/10.1007/s11165-020-09981-3

Park, B.-Y., Rodriguez, L., & Campbell, T. (2019). Commentary: Using models to teach science. *The Science Teacher*, *87*(4), 8–11. https://doi.org/10.2505/4/tst19_087_04_8

Parsley, K. M. (2020). Plant awareness disparity: A case for renaming plant blindness. *Plants, People, Planet*, *2*(6), 598–601. https://doi.org/10.1002/ppp3.10153

Patton, M. (1990). Purposeful sampling. In *Qualitative evaluation and research methods* (pp. 169-186). SAGE Publications.

Paul, D. B. (1995). *Controlling human heredity, 1865 to the present*. Humanities Press.

Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *Journal of the Learning Sciences*, *13*(3), 423–451. https://doi.org/10.1207/s15327809jls1303_6

Pearce, J., Chiavaroli, N., & Tavares, W. (2023). On the use and abuse of metaphors in assessment. *Advances in Health Sciences Education*, *28*(4), 1333–1345. https://doi.org/10.1007/s10459-022-10203-w

Peffer, M. E., & Ramezani, N. (2019). Assessing epistemological beliefs of experts and novices via practices in authentic science inquiry. *International Journal of STEM Education*, *6*(1), Article 3. https://doi.org/10.1186/s40594-018-0157-9

Pennington, C. R., Heim, D., Levy, A. R., & Larkin, D. T. (2016). Twenty years of stereotype threat research: A review of psychological mediators. *PLoS One*, *11*(1), Article e0146487. https://doi.org/10.1371/journal.pone.0146487

Perry, J., Meir, E., Herron, J. C., Maruca, S., & Stal, D. (2008). Evaluating two approaches to helping college students understand evolutionary trees through diagramming tasks. *CBE—Life Sciences Education*, *7*(2), 193–201. https://doi.org/10.1187/cbe.07-01-0007

Pew Research Center. (2022). *Modeling the future of religion in America*. https://www.pewresearch.org/religion/2022/09/13/modeling-the-future-of-religion-in-america/

Philippe, H., Brinkmann, H., Lavrov, D. V., Littlewood, D. T. J., Manuel, M., Wörheide, G., & Baurain, D. (2011). Resolving difficult phylogenetic questions: Why more sequences are not enough. *PLoS Biology*, *9*(3), Article e1000602. https://doi.org/10.1371/journal.pbio.1000602

Phillips, B. C., Novick, L. R., Catley, K. M., & Funk, D. J. (2010). Interactive effects of diagrammatic format and teleological beliefs on tree thinking. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *32*, 2656-2661.

Phillips, B. C., Novick, L. R., Catley, K. M., & Funk, D. J. (2012). Teaching tree thinking to college students: It's not as easy as you think. *Evolution: Education and Outreach*, *5*(4), 595–602. https://doi.org/10.1007/s12052-012-0455-5

Piaget, J. (2001). *The psychology of intelligence*. Routledge.

Podani, J. (2019). The coral of life. *Evolutionary Biology*, *46*(2), 123–144. https://doi.org/10.1007/s11692-019-09474-w

Polletta, F., Chen, P. C. B., Gardner, B. G., & Motes, A. (2011). The sociology of storytelling. *Annual Review of Sociology*, *37*(1), 109–130. https://doi.org/10.1146/annurev-soc-081309-150106

Posner, G. J., & Gertzog, W. A. (1982). The clinical interview and the measurement of conceptual change. *Science Education*, *66*(2), 195–209. https://doi.org/10.1002/sce.3730660206

Pugh, K. J. (2017). *Computers, cockroaches, and ecosystems: Understanding learning through metaphor*. Information Age Publishing.

Quinn, F., Castéra, J., & Clément, P. (2016). Teachers' conceptions of the environment: Anthropocentrism, non-anthropocentrism, anthropomorphism and the place of nature.

Environmental Education Research, *22*(6), 893–917. https://doi.org/10.1080/13504622.2015.1076767

Rapp, D. N., & Kurby, C. A. (2008). The 'ins' and 'outs' of learning: Internal representations and external visualizations. In J. K. Gilbert, M. Reiner, & M. Nakhleh (Eds.), *Visualization: Theory and practice in science education* (pp. 29–52). Springer Netherlands. https://doi.org/10.1007/978-1-4020-5267-5_2

Rhodes, M., & Gelman, S. A. (2009). Five-year-olds' beliefs about the discreteness of category boundaries for animals and artifacts. *Psychonomic Bulletin & Review*, *16*(5), 920–924. https://doi.org/10.3758/PBR.16.5.920

Robbins, J. (2005). 'Brown paper packages'? A sociocultural perspective on young children's ideas in science. *Research in Science Education*, *35*(2–3), 151–172. https://doi.org/10.1007/s11165-005-0092-x

Robillard, T., Legendre, F., Desutter-Grandcolas, L., & Grandcolas, P. (2006). Phylogenetic analysis and alignment of behavioral sequences by direct optimization. *Cladistics*, *22*(6), 602–633. https://doi.org/10.1111/j.1096-0031.2006.00125.x

Rodriguez, J.-M. G., Nardo, J. E., Finkenstaedt-Quinn, S. A., & Watts, F. M. (2023). The use of frameworks in chemistry education research. *Chemistry Education Research and Practice*, *24*(4), 1109–1126. https://doi.org/10.1039/D3RP00149K

Roller, M., & Lavrakas, P. J. (2015). *Applied qualitative research design: A total quality framework approach*. Guilford Press.

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, *4*(3), 328–350. https://doi.org/10.1016/0010-0285(73)90017-0

Rosenberg, N. A., & Nordborg, M. (2002). Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genetics*, *3*(5), 380–390. https://doi.org/10.1038/nrg795

Ross, N. (2020). Anthropocentric tendencies in environmental education: A critical discourse analysis of nature-based learning. *Ethics and Education*, *15*(3), 355–370. https://doi.org/10.1080/17449642.2020.1780550

Saljo, R., & Wyndhamn, J. (1993). Solving everyday problems in the formal setting: An empirical study of the school as context for thought. In S. Chaiklin & J. Lave (Eds.), *Understanding practice: Perspectives on activity and context* (pp. 327-342). Cambridge University Press.

Sandvik, H. (2008). Tree thinking cannot taken for granted: Challenges for teaching phylogenetics. *Theory in Biosciences*, *127*(1), 45–51. https://doi.org/10.1007/s12064-008-0022-3

Santini, F., Sorenson, L., Marcroft, T., Dornburg, A., & Alfaro, M. E. (2013). A multilocus molecular phylogeny of boxfishes (Aracanidae, Ostraciidae; Tetraodontiformes). *Molecular Phylogenetics and Evolution*, *66*(1), 153–160. https://doi.org/10.1016/j.ympev.2012.09.022

Scaife, M., & Rogers, Y. (1996). External cognition: How do graphical representations work? *International Journal of Human-Computer Studies*, *45*(2), 185–213. https://doi.org/10.1006/ijhc.1996.0048

Schachtner, C. (2020). Storytelling as a cultural practice and life form. In *The narrative subject* (pp. 29–75). Springer International Publishing. https://doi.org/10.1007/978-3-030-51189-0_2

Scheurich, J. (2014). *Research method in the postmodern*. Routledge. https://doi.org/10.4324/9781315043258

Schneider, B., Strait, M., Muller, L., Elfenbein, S., Shaer, O., & Shen, C. (2012). *Phylo-genie: Engaging students in collaborative "tree-thinking" through tabletop techniques* [Paper presentation]. Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems, Austin, TX, United States. https://doi.org/10.1145/2207676.2208720

Schooler, J. W., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General*, *122*(2), 166–183.

Schramm, T., Jose, A., & Schmiemann, P. (2021a). Modeling and measuring tree-reading skills in undergraduate and graduate students. *CBE—Life Sciences Education*, *20*(3), Article ar32. https://doi.org/10.1187/cbe.20-06-0131

Schramm, T., Jose, A., & Schmiemann, P. (2021b). Seeing the woods for the trees again: Analyzing evolutionary diagrams in German and US university-level textbooks. *Education Sciences*, *11*(8), Article 367. https://doi.org/10.3390/educsci11080367

Schramm, T., Schachtschneider, Y., & Schmiemann, P. (2019). Understanding the tree of life: An overview of tree-reading skill frameworks. *Evolution: Education and Outreach*, *12*(1), Article 11. https://doi.org/10.1186/s12052-019-0104-3

Schramm, T., & Schmiemann, P. (2019). Teleological pitfalls in reading evolutionary trees and ways to avoid them. *Evolution: Education and Outreach*, *12*(1), Article 20. https://doi.org/10.1186/s12052-019-0112-3

Schwartz, R., & Schäffer, A. A. (2017). The evolution of tumour phylogenetics: Principles and practice. *Nature Reviews Genetics*, *18*(4), 213–229. https://doi.org/10.1038/nrg.2016.170

Sherin, B. L. (2000). How students invent representations of motion: A genetic account. *The Journal of Mathematical Behavior, 19*(4), 399-441.

Sherin, B. L., Reiser, B. J., & Edelson, D. (2004). Scaffolding analysis: Extending the scaffolding metaphor to learning artifacts. *The Journal of the Learning Sciences*, *13*(3), 387–421.

Shtulman, A., & Schulz, L. (2008). The relation between essentialist beliefs and evolutionary reasoning. *Cognitive Science: A Multidisciplinary Journal*, *32*(6), 1049–1062. https://doi.org/10.1080/03640210801897864

Singer, F., Hagen, J. B., & Sheehy, R. R. (2001). The comparative method, hypothesis testing & phylogenetic analysis—An introductory laboratory. *The American Biology Teacher*, *63*, 518-523.

Smith, J. P., diSessa, A. A., & Roschelle, J. (1994). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *Journal of the Learning Sciences*, *3*(2), 115–163. https://doi.org/10.1207/s15327809jls0302_1

Sousa, P., Atran, S., & Medin, D. (2002). Essentialism and folkbiology: Evidence from Brazil. *Journal of Cognition and Culture*, *2*(3), 195–223. https://doi.org/10.1163/15685370260225099

Springer, K. (1996). Young children's understanding of a biological basis for parent-offspring relations. *Child Development*, *67*(6), 2841-2856. https://doi.org/10.2307/1131755

Stebner, F., Schuster, C., Weber, X.-L., Greiff, S., Leutner, D., & Wirth, J. (2022). Transfer of metacognitive skills in self-regulated learning: Effects on strategy application and content knowledge acquisition. *Metacognition and Learning*, *17*(3), 715–744. https://doi.org/10.1007/s11409-022-09322-x

Stenlund, J. I., Schönborn, K. J., & Tibell, L. A. E. (2021). Zooming in time—Exploring students' interpretations of a dynamic tree of life. *Journal of Science Education and Technology*, *30*(1), 125–138. https://doi.org/10.1007/s10956-020-09893-x

Stenlund, J. I., & Tibell, L. A. E. (2019). Visualizing macroevolutionary timescales: Students' comprehension of different temporal representations in an animation. *Evolution: Education and Outreach*, *12*(1), Article 8. https://doi.org/10.1186/s12052-019-0099-9

Stets, J. E., Brenner, P. S., Burke, P. J., & Serpe, R. T. (2017). The science identity and entering a science occupation. *Social Science Research*, *64*, 1–14. https://doi.org/10.1016/j.ssresearch.2016.10.016

Stieff, M., Ryu, M., Dixon, B., & Hegarty, M. (2012). The role of spatial ability and strategy preference for spatial problem solving in organic chemistry. *Journal of Chemical Education*, *89*(7), 854–859. https://doi.org/10.1021/ed200071d

Storms, G., Boeck, P. D., & Ruts, W. (2000). Prototype and exemplar-based information in natural language categories. *Journal of Memory and Language*, *42*, 51–73.

Strauss, A. L., & Corbin, J. (1994). Grounded theory methodology: An overview. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (1st ed.). SAGE Publications.

Stuart, A. E., Hunter, F. F., & Currie, D. C. (2002). Using behavioural characters in phylogeny reconstruction. *Ethology Ecology & Evolution*, *14*(2), 129–139. https://doi.org/10.1080/08927014.2002.9522751

Studtmann, P. (2023). Aristotle's categories. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy*. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2023/entries/aristotle-categories/

Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (1st ed., pp. 19–30). Cambridge University Press. https://doi.org/10.1017/CBO9780511816819.003

Taber, K. S. (2013). Revisiting the chemistry triplet: Drawing upon the nature of chemical knowledge and the psychology of learning to inform chemistry education. *Chemistry Education Research and Practice*, *14*(2), 156–168. https://doi.org/10.1039/C3RP00012E

Talanquer, V. (2022). The complexity of reasoning about and with chemical representations. *JACS Au*, *2*(12), 2658–2669. https://doi.org/10.1021/jacsau.2c00498

Tang, K. L., Stiassny, M. L. J., Mayden, R. L., & DeSalle, R. (2021). Systematics of damselfishes. *Ichthyology & Herpetology*, *109*(1), 258-318. https://doi.org/10.1643/i2020105

Tarłowski, A. (2019). Parent-child conversations about biological kinds as a potential contributor to the variability in biological knowledge. *Psychology of Language and Communication*, *23*(1), 238–276. https://doi.org/10.2478/plc-2019-0011

Taylor, J. R. (2015). Prototype theory in linguistics. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (pp. 286–289). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.57034-2

Taylor, M. G., Rhodes, M., & Gelman, S. A. (2009). Boys will be boys; Cows will be cows: Children's essentialist reasoning about gender categories and animal species. *Child Development*, *80*(2), 461–481. https://doi.org/10.1111/j.1467-8624.2009.01272.x

Terwel, J., van Oers, B., van Dijk, I., & van den Eeden, P. (2009). Are representations to be provided or generated in primary mathematics education? Effects on transfer. *Educational Research and Evaluation*, *15*(1), 25–44.

Themudo, G. E., Rey-Iglesia, A., Tascón, L. R., Jensen, A. B., da Fonesca, R. R., & Campos, P. F. (2020). Declining genetic diversity of European honeybees along the twentieth century. *Scientific Reports*, *10*, Article 10520.

Toulmin, S. E. (2008). *The uses of argument*. Cambridge University Press.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, *133*(5), 859–883. https://doi.org/10.1037/0033-2909.133.5.859

Trimble, M. J., & Van Aarde, R. J. (2010). Species inequality in scientific study. *Conservation Biology*, *24*(3), 886–890. https://doi.org/10.1111/j.1523-1739.2010.01453.x

Trommler, F., & Hammann, M. (2020). The relationship between biological function and teleology: Implications for biology education. *Evolution: Education and Outreach*, *13*(1), Article 11. https://doi.org/10.1186/s12052-020-00122-y

Trowbridge, J. E., & Mintzes, J. J. (1988). Alternative conceptions in animal classification: A cross-age study. *Journal of Research in Science Teaching*, *25*(7), 547–571. https://doi.org/10.1002/tea.3660250704

Tsou, J. Y. (2022). Biological essentialism, projectable human kinds, and psychiatric classification. *Philosophy of Science*, *89*(5), 1155-1165. https://doi.org/10.1017/psa.2022.64

University of California Museum of Paleontology. (2021a). *Identify your learning goals*. Understanding Evolution. https://evolution.berkeley.edu/evolibrary/teach/912goals.php

University of California Museum of Paleontology. (2021b, June 21). *Trees as hypotheses*. Understanding Evolution. https://evolution.berkeley.edu/the-tree-room/how-to-build-a-tree/trees-as-hypotheses/

Van Bruggen, J. M., Kirschner, P. A., & Jochems, W. (2002). External representation of argumentation in CSCL and the management of cognitive load. *Learning and Instruction*, *12*(1), 121–138. https://doi.org/10.1016/S0959-4752(01)00019-6

Van Meter, P., & Garner, J. (2005). The promise and practice of learner-generated drawing: Literature review and synthesis. *Educational Psychology Review*, *17*(4), 285–325. https://doi.org/10.1007/s10648-005-8136-3

Varella, M. A. C. (2018). The biology and evolution of the three psychological tendencies to anthropomorphize biology and evolution. *Frontiers in Psychology*, *9*, Article 1839. https://doi.org/10.3389/fpsyg.2018.01839

Vaux, F., Trewick, S. A., & Morgan-Richards, M. (2016). Lineages, splits and divergence challenge whether the terms anagenesis and cladogenesis are necessary. *Biological Journal of the Linnean Society*, *117*(2), 165–176. https://doi.org/10.1111/bij.12665

Velasco, J. D. (2012). The future of systematics: Tree thinking without the tree. *Philosophy of Science*, *79*(5), 624–636. https://doi.org/10.1086/667878

Verstraete, B., Moon, H.-K., Smets, E., & Huysmans, S. (2014). Orbicules in flowering plants: A phylogenetic perspective on their form and function. *The Botanical Review*, *80*(2), 107–134. https://doi.org/10.1007/s12229-014-9135-1

Waese, J., Provart, N. J., & Guttman, D. S. (2017). Topo-phylogeny: Visualizing evolutionary relationships on a topographic landscape. *PLoS One*, *12*(5), Article e0175895. https://doi.org/10.1371/journal.pone.0175895

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & Von Der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin*, *138*(6), 1172–1217. https://doi.org/10.1037/a0029333

Wandersee, J. H., & Schussler, E. E. (1999). Preventing plant blindness. *The American Biology Teacher*, *61*(2), 82–86. https://doi.org/10.2307/4450624

Wandersee, J. H., & Schussler, E. E. (2001). Toward a theory of plant blindness. *Plant Science Bulletin*, *47,* 2-9.

Wassenbergh, S. V., van Manen, K., Marcroft, T. A., Alfaro, M. E., & Stamhuis, E. J. (2015). Boxfish swimming paradox resolved: Forces by the flow of water around the body promote manoeuvrability. *Journal of The Royal Society Interface*, *12*(103), Article 20141146. https://doi.org/10.1098/rsif.2014.1146

Werth, A., & Allchin, D. (2020). Teleology's long shadow. *Evolution: Education and Outreach*, *13*(1), Article 4. https://doi.org/10.1186/s12052-020-00118-8

West, B. T., & Blom, A. G. (2016). Explaining interviewer effects: A research synthesis. *Journal of Survey Statistics and Methodology*, *5*(2), 175-211. https://doi.org/10.1093/jssam/smw024

Wilkins, J. S. (2022). Species, God, and dominion. In B. Swartz & B. D. Mishler (Eds.), *Speciesism in biology and culture: How human exceptionalism is pushing planetary boundaries* (pp. 95-110). Springer International Publishing. https://doi.org/10.1007/978-3-030-99031-2

Williams, B., Entwistle, V., Haddow, G., & Wells, M. (2008). Promoting research participation: Why not advertise altruism? *Social Science & Medicine*, *66*(7), 1451–1456. https://doi.org/10.1016/j.socscimed.2007.12.013

Wilton, M., Gonzalez-Niño, E., McPartlan, P., Terner, Z., Christoffersen, R. E., & Rothman, J. H. (2019). Improving academic performance, belonging, and retention through increasing structure of an introductory biology course. *CBE—Life Sciences Education*, *18*(4), Article ar53. https://doi.org/10.1187/cbe.18-08-0155

Windschitl, M., Thompson, J., & Braaten, M. (2018). *Ambitious science teaching*. Harvard Education Press.

Wittgenstein, L. (1986). *Philosophical investigations* (G. E. M. Anscombe, Trans.). Basil Blackwell.

Wright, L. K., Cardenas, J. J., Liang, P., & Newman, D. L. (2018). Arrows in biology: Lack of clarity and consistency points to confusion for learners. *CBE—Life Sciences Education*, *17*(1), Article ar6. https://doi.org/10.1187/cbe.17-04-0069

Wynne, C. D. L. (2006). What are animals? Why anthropomorphism is still not a scientific approach to behavior. *Comparative Cognition & Behavior Reviews*, *2,* 125-135. https://doi.org/10.3819/ccbr.2008.20008

Yilmaz, K. (2013). Comparison of quantitative and qualitative research traditions: Epistemological, theoretical, and methodological differences. *European Journal of Education*, *48*(2), 311–325. https://doi.org/10.1111/ejed.12014

Young, A. K., White, B. T., & Skurtu, T. (2013). Teaching undergraduate students to draw phylogenetic trees: Performance measures and partial successes. *Evolution: Education and Outreach*, *6*(1), Article 16. https://doi.org/10.1186/1936-6434-6-16

Yu, G., xu, s., Dai, Z., Guo, P., Fu, X., Liu, S., Zhou, L., Tang, W., Feng, T., Chen, M., Zhan, L., Wu, T., & Hu, E. (2021). ggtreeExtra: Compact visualization of richly annotated phylogenetic data. *Molecular Biology and Evolution.* Advance online publication. https://doi.org/10.21203/rs.3.rs-155672/v1

Zsidó, A. N., Stecina, D. T., Cseh, R., & Hout, M. C. (2022). The effects of task-irrelevant threatening stimuli on orienting- and executive attentional processes under cognitive load. *British Journal of Psychology*, *113*(2), 412–433. https://doi.org/10.1111/bjop.12540