12-2000

An Account of Novice Phylogenetic Tree Construction from the Problem-Solving Research Tradition

Terrance E. Brisbin
Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations
Part of the Evolution Commons, and the Science and Mathematics Education Commons

Recommended Citation
https://scholarworks.wmich.edu/dissertations/1439

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact wmu-scholarworks@wmich.edu.
AN ACCOUNT OF NOVICE PHYLOGENETIC TREE CONSTRUCTION
FROM THE PROBLEM-SOLVING RESEARCH TRADITION

by

Terrance E. Brisbin

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Science Studies

Western Michigan University
Kalamazoo, Michigan
December 2000
AN ACCOUNT OF NOVICE PHYLOGENETIC TREE CONSTRUCTION
FROM THE PROBLEM-SOLVING RESEARCH TRADITION

Terrance E. Brisbin, Ph.D.
Western Michigan University, 2000

Both the content and strategic knowledge used in phylogenetic tree construction are described to account for novice performance. A review of the literature is used to place this research with respect to studies in: problem solving studies (both expert, and novice), the history and practice of evolutionary systematics, and the teaching of evolution. The computer-based problem-solving environment is discussed along with model problems used for the research.

Each of the five research problems consists of an encoded matrix containing phylogenetic data organized by taxa and characters. Twelve participants (high school students) volunteered to solve the research problems, on their own time, before and after school. Each participant generated three sets of data: (1) an audio tape dialogue of activities while solving each problem, (2) the completed tree as a computer print-out, and (3) any paper and pencil drawings or notes that the participant may have made.

A procedural model of novice performance for phylogenetic tree construction is described with the associated strategies needed to do so. There are four activities that make up the procedural model of novice performance. Those activities are: matrix analysis, building a tree topology, character assignment, and checking. The
chief variation between novices and experts is in matrix manipulation and character assignment. Novices did not create any alternative trees nor did they rearrange the data matrix. Once a solution was reached, it appeared to be written in stone. All the novices, however, believed there were alternate trees as evidenced by post research interviews.
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
ACKNOWLEDGMENTS

I wish to express my thanks and deepest appreciation to my advisor, Dr. Robert Hafner, who always had ideas and suggestions whenever I was at an impasse. He carefully read the many revisions of the chapters that compose the final manuscript and willingly offered his time and expertise.

I dedicate my dissertation to my dad, Dr. Lawrence Brisbin, whose inspiration caused me to begin my doctoral studies. Although he did not live to see the dissertation completed, his memory and my prayers have given me the determination to complete the manuscript.

Terrance E. Brisbin
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ......................................................................................................... ii

LIST OF TABLES .................................................................................................................. ix

LIST OF FIGURES ................................................................................................................ x

CHAPTER

I. INTRODUCTION ............................................................................................................. 1

   Evolution and Phylogenetics ......................................................................................... 1

   Problem Solving in Education ..................................................................................... 3

   Problem Solving in Evolution ...................................................................................... 5

   Significance .................................................................................................................. 6

II. REVIEW OF THE LITERATURE .................................................................................. 9

   Introduction ................................................................................................................. 9

   Biology and Evolution ................................................................................................. 11

   Introduction ................................................................................................................. 11

   Scientific Revolution and Paradigm Change ............................................................. 12

   Essentialism and Popular Thinking ............................................................................. 14

   Evolutionary Thinking Has Two Lenses .................................................................... 16

   Macroevolution .......................................................................................................... 16

   Microevolution .......................................................................................................... 17

   Systematics and Evolution ......................................................................................... 17
<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Three Philosophies of Systematics</td>
<td>18</td>
</tr>
<tr>
<td>Phenetic Taxonomy</td>
<td>19</td>
</tr>
<tr>
<td>Evolutionary Taxonomy</td>
<td>20</td>
</tr>
<tr>
<td>Cladistic Taxonomy</td>
<td>21</td>
</tr>
<tr>
<td>Ancestral and Derived Character States</td>
<td>22</td>
</tr>
<tr>
<td>Phylogenetic Trees</td>
<td>24</td>
</tr>
<tr>
<td>Importance of Phylogenetic Trees</td>
<td>26</td>
</tr>
<tr>
<td>Summary of the Biology and Evolution Section</td>
<td>27</td>
</tr>
<tr>
<td>The Problem Solving Tradition</td>
<td>29</td>
</tr>
<tr>
<td>Introduction</td>
<td>29</td>
</tr>
<tr>
<td>What Does Problem Solving Involve?</td>
<td>30</td>
</tr>
<tr>
<td>Early Conceptions of the Problem Solving Process</td>
<td>32</td>
</tr>
<tr>
<td>Experts and Chunking</td>
<td>34</td>
</tr>
<tr>
<td>Problem Solving in Physics</td>
<td>36</td>
</tr>
<tr>
<td>Problem Solving in Genetics</td>
<td>40</td>
</tr>
<tr>
<td>Computer Simulation and Genetics Problem Solving</td>
<td>42</td>
</tr>
<tr>
<td>Problem Solving in Evolution</td>
<td>45</td>
</tr>
<tr>
<td>Recapitulation on Problem-Solving</td>
<td>51</td>
</tr>
<tr>
<td>Teaching and Learning Evolution</td>
<td>52</td>
</tr>
</tbody>
</table>
Table of Contents—Continued

CHAPTER

Introduction ..........................................................................................52
Conceptual Change Studies .................................................................53
Evolutionary Misconceptions ..............................................................55
Learning Evolution Other Than by Conceptual Change ....................56

III. METHODOLOGY ................................................................................59

Introduction ...........................................................................................59
Description of the Research .................................................................59
Selection of Participants ......................................................................61
Description of the Research Site ...........................................................61
Nature of Instruction ............................................................................62
Nature of Research Problems .................................................................64
Research Problem Number 1 .................................................................65
Research Problem Number 2 .................................................................66
Research Problem Number 3 .................................................................69
Research Problem Number 4 .................................................................70
Research Problem Number 5 .................................................................73
Data Analysis .......................................................................................75

IV. RESULTS ............................................................................................81

Introduction ..........................................................................................81
<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Components of a Good Solution</td>
<td>83</td>
</tr>
<tr>
<td>Descriptive Statistics of the Results</td>
<td>84</td>
</tr>
<tr>
<td>A Procedural Model for Novice Phylogenetic Tree Construction</td>
<td>85</td>
</tr>
<tr>
<td>Success of the Model Pathways/Strategies Chosen</td>
<td>88</td>
</tr>
<tr>
<td>Matrix Analysis</td>
<td>88</td>
</tr>
<tr>
<td>Character Assignment</td>
<td>89</td>
</tr>
<tr>
<td>Explanation of the Major Sections of the Novice Model</td>
<td>90</td>
</tr>
<tr>
<td>Matrix Analysis</td>
<td>91</td>
</tr>
<tr>
<td>Building a Tree Topology</td>
<td>94</td>
</tr>
<tr>
<td>Character Assignment</td>
<td>97</td>
</tr>
<tr>
<td>Checking</td>
<td>101</td>
</tr>
<tr>
<td>Advantages and Disadvantages of the Three Methods of Character Assignment</td>
<td>102</td>
</tr>
<tr>
<td>Student 8 Solving Research Problem 3</td>
<td>104</td>
</tr>
<tr>
<td>Correct Solutions Organized by Problem Type</td>
<td>114</td>
</tr>
<tr>
<td>Introduction</td>
<td>114</td>
</tr>
<tr>
<td>Correct Solutions of Problem One</td>
<td>115</td>
</tr>
<tr>
<td>Correct Solutions of Problem Two</td>
<td>116</td>
</tr>
<tr>
<td>Correct Solutions of Problem Three</td>
<td>118</td>
</tr>
<tr>
<td>Correct Solutions of Problem Four</td>
<td>119</td>
</tr>
</tbody>
</table>
CHAPTER

The Prescriptive Model of Desired Performance and High School Students .......................................................... 153

Instruction in Phylogenetic Tree Construction .......................................................... 154

Suggested Methods for Use in a High School Biology Course .............. 170

Conclusion ................................................................................................... 173

APPENDICES

A. A Primer of Phylogenetic Terms Taken From
   Steven Brewer's (1996) Dissertation ................................................................. 175

B. Protocol Clearance From the Human Subjects
   Institutional Review Board .................................................................................. 179

C. Parental Consent Form ....................................................................................... 181

D. Participant Consent Form ................................................................................... 183

BIBLIOGRAPHY ............................................................................................................ 185
LIST OF TABLES

1. Results of Problem Solution by Student ...............................................................85
2. Results of Problem Solution by Problem ..............................................................86
3. Data Matrix for Problem #3 ....................................................................................92
4. A Key to Program Operations...............................................................................105
5. Problem Three Solved by S8 Using the Overall Model of Novice Phylogenetic Tree Construction ............................................................................106
6. Common Mistakes by Problem and Participant...................................................122
7. Comparison of the Expert and Novice Models of Performance.........................150
LIST OF FIGURES

1. A Cladogram Showing B and C Sister Groups and A as an Outgroup ...........22
2. A Typical Phylogenetic Tree ........................................................................25
3. Problem 1, Blood Flukes, Matrix .................................................................66
4. Problem 1, Blood Flukes, Single Solution ....................................................67
5. Problem 2, Gulls Behavior, Matrix ...............................................................68
6. Problem 2, Gulls Behavior, Convergence Option ........................................69
7. Problem 2, Gulls Behavior, Reversal Option ...............................................70
8. Problem 3, Quail, Matrix ..............................................................................71
9. Problem 3, Quail, Solution 1 .........................................................................72
10. Problem 3, Quail, Solution 2 .......................................................................73
11. Problem 4, Mustards, Matrix ......................................................................74
12. Problem 4, Mustards, Solution 1 .................................................................75
13. Problem 4, Mustards, Solution 2 .................................................................76
14. Problem 4, Mustards, Solution 3 .................................................................77
15. Problem 5, Caminacules, Matrix .................................................................78
16. Problem 5, Caminacules, Solution 1 ...........................................................79
17. Problem 5, Caminacules, Solution 2 .............................................................80
18. A Procedural Model of Novice Phylogenetic Tree Construction ...............87
19. Matrix Analysis for Problem #3 .................................................................93

x

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
List of Figures—Continued

20. Tree Building Topology .............................................................................................95
21. Showing Connection, With and Without a Node to the Common Ancestor ..........96
22. Character Assignment ...............................................................................................97
23. Character Assignment for Problem 3......................................................................98
24. Complete Tree Problem 3 Character 4 Convergent .................................................99
25. Checking .....................................................................................................................102
27. A Prescriptive Model of Desired Performance in Phylogenetic Tree Construction ..........................................................152
28. Example Tree and Matrix Showing Autapomorphy ..............................................156
29. Example of Matrix and Tree Showing Synapomorphy 1 ....................................157
30. Example of Matrix and Tree Showing Synapomorphy 2 ....................................159
31. Example of Matrix and Tree Showing Synapomorphy 3 ....................................160
32. Example of Tree and Matrix Showing Homoplasy 1, Convergence Option ......162
33. Example of Tree and Matrix Showing Homoplasy 2, Reversal Option ...............164
34. Example of Tree and Matrix Showing Homoplasy 3, Convergence in Two Clades ...........................................................................165
35. Example of Tree and Matrix Showing Homoplasy 4, Reversal Option in One Clade ...........................................................................167
36. Example of a Tree and Matrix Showing Convergence and Reversal ...............169
37. Example of a Tree and Matrix Showing Two Fossil Taxa ..................................171
CHAPTER I

INTRODUCTION

Evolution and Phylogenetics

Evolution is one of the over-arching concepts that binds together the domain of biology. Evolution is useful in explaining the diversity of both living and fossil organisms. Using Darwin's definition of "descent with modification", evolution can be thought of as a "change in form and behavior of organisms between generations" (Ridley, 1993). The change in living things is the result of external environmental change, which produces external forces, and of random genetic events which produces the variation upon which the selective forces act. In addition, this change has proceeded in a branching tree-like pattern with present day species having been generated by speciation from a single common ancestor. Change and splitting provide two of the main theses of evolutionary theory (Ridley, 1993).

The task of phylogenetic analysis is to determine relationships between species and to describe patterns of evolutionary change which gave rise to these relationships (Avers, 1989). Phylogenetics involves the construction of lines of descent traced back to a common ancestor. The product is an evolutionary tree consisting of individual branches which represent species with a shared evolutionary history. In order to construct such a tree, character states of a species are used. These character states may be structural phenotypes, behaviors, geographical distributions, or
molecular analyses but must always be distinctly different from those that are found in ancestral organisms. Ridley (1993) calls these changed character states “evolutionary novel” and goes on to say that only shared, derived homologies indicate phylogenetic groups.

Every living species is a mixture of ancestral character states, conserved with no significant change from the ancestors, and derived character states which have changed from the ancestral state. Those character states that remain unchanged are termed plesiomorphic and those character states that do change are termed apomorphic. In order to determine which character states are plesiomorphic and which character states are apomorphic, the polarity of the characters must be established. There are three general methods for determining polarity: outgroup comparison, early embryology, and the fossil record. In outgroup comparison the method is to choose, sometimes quite arbitrarily, a closely related species to the species being studied (e.g., cartilaginous fish as closely related to bony fish). Whatever character state is present in the outgroup is considered plesiomorphic to the group being studied or the in-group. Similarly, character states found in early embryology and found in fossils are considered to be plesiomorphic character states. Any character state found in later embryology or in present day living species is considered to be an apomorphic character state.

Once character polarity has been determined, a tree is constructed using the shared, apomorphic character states to denote an evolutionary pathway that requires the least number of transitions. A tree is a branching diagram that is used to explain
the distribution of apomorphic character states between the taxa (species) under consideration (Brewer, 1996). The goal of tree construction is to postulate common ancestors of living (extant) and nonliving (extinct) species. The process is based on the assumption that there is only one true phylogeny (Brooks & McLennan, 1991). Trees must also be monophyletic meaning that they must contain a complete set of all descendants. If this is not the case, the true evolutionary relationship between species cannot be determined. There are two ways that this can happen. First, species may be added to the evolutionary pathway, a practice called polyphyletic grouping. Second, a species may be removed from an evolutionary pathway, a process called paraphyletic grouping.

Problem Solving in Education

During the last twenty years there has been an increasing number of studies on learning and problem solving in science education. The reason for this has been the desire to develop theories of learning and problem solving that will improve science teaching. A problem-based approach to the teaching and learning of evolution may have four learning outcomes. The student may: (1) gain a better understanding of the conceptual nature of evolution, (2) learn general problem-solving heuristics and techniques, (3) learn evolutionary content specific heuristics, and (4) gain increased knowledge of science as an intellectual activity (Stewart, 1988). The problems used in a problem-based teaching approach are of two types: cause to effect and effect to cause. Cause to effect problems are those typically found in textbooks and
are usually solved by invoking algorithms. One simply applies the formula (algorithm) or procedure that guarantees an answer if correctly followed. Effect to cause problems, however, broaden the horizon and students must branch out from the strict application of algorithms to the application of general and domain specific heuristics. The solver is faced with a general, open-ended problem and must not only form an hypothesis to be tested, but must also decide on the best course to achieve a solution. The solution is not guaranteed and the solver may have to change his or her strategies, several times, in order to achieve a solution. This is how a scientist actually works.

Most problem-solving research has been conducted from an expert-novice framework (Smith & Good, 1984). The usual procedure is to compare and contrast the performance of both experts and novices as they solve problems. An expert is usually defined in terms of the knowledge he or she possesses because of experience in the domain. However, experts and novices also differ in how they structure their knowledge. When faced with a problem, experts seek a solution path rather than an answer. They do this by working in a forward manner, by checking one variable at a time, and by looking for evidence that would invalidate previous assumptions. Further, experts are able to justify their solutions in terms of underlying concepts. As a result, experts identify critical problem cues, generate and test alternate hypotheses, and recognize the need to change strategies based on problem conditions. Novices, having little knowledge organization and a different perception of problem solving, tend to work backward and in a piece meal fashion. They try for an answer and, if
the first trial does not work, they try again using an unplanned approach and with little thought to removing variables. They are interested only in completing the task and arriving at an answer. Many researchers report that while novices will not become experts unless given the proper amount of time and experience, they can be taught to become more like experts in solving problems. This should be the goal of any problem-based educational approach.

Problem Solving in Evolution

In a 1996 study entitled *An Account of Expert Phylogenetic Tree Construction From the Problem Solving Research Tradition in Science Education*, Steven Brewer (1996) produced a model of expert performance in phylogenetic tree construction. Working with college professors and practitioners of phylogenetic systematics, he collected think aloud protocols as they solved problems in phylogenetic tree construction. Subjects used a computer program called *Phylogenetic Investigator* which was developed for the purpose of creating phylogenetic trees in an open-ended, problem-solving environment. From the data generated, Brewer produced a model of eight specific actions that experts undertake when solving problems in phylogenetic tree construction. Brewer indicated that his research project extends the problem-solving research tradition into a new domain of biology. As such, it changes the complexion of the evolutionary world by examining phylogenetic relationships from a new perspective. The model of expert performance in phylogenetic tree construction serves as a model of how problem-solving can be employed in the study
of evolution. The model also serves as guide for the construction of a problem-based curriculum in the teaching of evolution.

Brewer also indicated that a model of novice performance in phylogenetic tree construction could be compared with a model of expert performance in order to illustrate specific points that could be addressed in instruction to help students improve their problem solving skills. The purpose of this study is to develop such a model of novice performance. Since it is assumed that such a model would involve the strategic knowledge used by novices in creating phylogenetic trees, the research question is: What procedural models do novices use when solving effect to cause problems involving the construction of phylogenetic trees from encoded data?

Significance

If learning is a transformation from an initial state to a final state as Reif (1983) indicated, it is necessary to measure both the learner's preinstructional concepts and strategies as well as the postinstructional concepts and strategies needed for desired performance. The postinstructional concepts and strategies are components of a model of desired performance that reflects all of the adequate knowledge about the learner's final state to be attained as a result of instruction. With the existence of an expert model of phylogenetic tree construction, the final learning state needed for novices to successfully complete the same task, is clearly outlined. That is not to say that the final model of desired novice performance must be the expert model but that the expert model definitely delineates desired elements of the final task, that is,
Explaining novice performance in phylogenetic tree construction has theoretical significance within the problem-solving research tradition in science education and in the practice of teaching evolution. Understanding how novices create phylogenetic trees breaks new ground in problem-solving research. Problem-solving research began with content free problems but has, over time, shifted to domain specific areas such as those in physics and genetics. This study applies methods of the problem-solving research tradition to an area not previously viewed as problem-based, that is, the study of evolution.

This study also has the potential to inform the practice of teaching evolution. Evolution, so often controversial among the American public, is perceived in a light far removed from scientific methodologies such as the dating of fossils. Instead, evolution is viewed as a line of organisms delineating a path from apes to humans. This view, more than any other, has entered into popular thinking as the default appearance of evolution. The teaching of phylogeny has the ability to alter this view by having students construct phylogenies using the same techniques as scientists. Images gained from this activity could allow students to recognize that popular views of linear evolutionary progress do not agree with current scientific theory and that evolution, properly conceived, is a branching tree-like pattern from a common ancestor.

Most students are taught the existence of scientific schemes of classification. They recognize that lions, tigers, and panthers are all members of the same class that
does not include wolves, dogs, and coyotes. Further, students recognize that the larger class, Mammalia, includes all of these. However, few students can provide any theoretical basis as to why these organisms are classified together. Evolution provides such a basis. Without understanding the mechanisms that have produced the diversity of life on earth, the study of classification becomes nothing more than vocabulary memorization.

This study has practical implications for those who construct curricula in the teaching of evolution. With a model of novice phylogenetic tree construction in place, curriculum developers are provided with a tool with which to access novice performance and compare it with the expert model produced by Steven Brewer (1996). After such a comparison, curriculum developers can plan specific activities designed to transform the novice performance into one that more closely parallels the performance of experts. Such activities have the potential of completely changing the way the subject of evolution is taught.
CHAPTER II

REVIEW OF THE LITERATURE

Introduction

Since this research is concerned with the creation of phylogenetic trees by novices and is conducted from a problem-solving perspective, this chapter provides the knowledge base for undertaking such a research project. This chapter is divided into three major sections: biology and evolution, the problem-solving research tradition, and the teaching and learning of evolution.

The biology and evolution section provides the theoretical basis for the construction of phylogenetic trees. Evolution, as a concept, cannot be entertained until there is a clear recognition of two important tenets: (1) that populations of species can and do change their phenotypic characteristics over time; and (2) closely related species, that is, lions, tigers, and panthers are related to each other through a shared common ancestor. Once these two tenets are accepted, it is legitimate to ask the question: how can these phenotypic changes over time and common ancestral lineages be known? The answer to that question has been answered by three different philosophical approaches: (1) evolutionary taxonomy, (2) phenetic taxonomy, and (3) cladistic taxonomy. This research is based on the third philosophy: cladistic taxonomy. Cladistic taxonomy began in the year 1950 with the classic work of Willi Henig (Ridley, 1993). Cladistics is based on the principle that evolution proceeds in a
branching pattern from a common ancestor. What determines where the branching will take place are the phenotypic character changes that are shared between populations of species. Cladists depict this branching pattern in what they initially called a cladogram but what is now known as a phylogenetic tree.

Because this research is being conducted in the problem-solving research tradition, it is necessary to elaborate what that tradition is carried on. For many researchers it has come to mean the study of the performance of some problem-solving task by both experts and novices. The idea is to compare the expert or desired state of performance and the novice state of performance with an eye to making novices perform more like experts. In 1996, Steven Brewer conducted a study in which he constructed a model of expert performance in the construction of phylogenetic trees. In this research, Brewer (1996) forms the basis for the expert approach to the process of phylogenetic tree construction. Since no model of novice performance presently exists, this research project seeks to provide such a novice model of phylogenetic tree construction. The ultimate goal is to determine what the parallels and differences are between expert and novice performance in order to plan a course of instruction to make novices perform more like experts.

The teaching and learning of evolution section is included because, in order to make novices perform more like experts, some instruction must take place. While most of this research has been based on instruction promoting conceptual change theory (Bishop & Anderson, 1986), it is recognized that there are alternative methods to present the subject matter. One such alternative method is the creation of
phylogenetic trees from encoded data. This method is desirable because it adds a new dimension of "open endedness" to a field of instruction that has not been traditionally thought of in such a manner. Students are allowed to create their own solutions to phylogenetic problems and, in so doing, work more like an "on the job" scientist. Often there is no single answer to a particular set of data but two, three, or even more equally correct answers are possible. Students need to understand that there is more than one route to the solution of a problem. The creation of phylogenetic trees allows them to visualize those alternative routes.

Biology and Evolution

Introduction

The six basic components of this section are: (1) scientific revolution and paradigm change, (2) essentialism and population thinking, (3) macroevolution and microevolution, (4) three philosophies of systematics, (5) ancestral and derived characters, and (6) phylogenetic trees. The first three components, scientific revolution and paradigm change, essentialism and population thinking, and macroevolution and microevolution, have been structured to impart the background changes that had to occur in order for evolution, as we know it today, to be recognized as a scientific concept. Once evolution reached conceptual status, there have been three philosophical schools or ways of studying it. The fourth component introduces those three philosophical schools with particular emphasis on cladistics which forms the basis for this study. The fifth and sixth sections, ancestral and derived characters, and
phylogenetic trees, follow next in order to explain how the study of cladistics is carried out. It points out the importance of derived characters and how they become the important indicators of where the branches will occur on a phylogenetic tree. Finally, the structure and meaning of phylogenetic trees is presented since such trees are the actual objects that research participants are expected to create.

**Scientific Revolution and Paradigm Change**

The publication of the Copernican treatise, *De Revolutionibus Orbium* in the year 1541, marked the beginning of a great scientific revolution in the domain of astronomy. With the determined stroke of a pen and a firm resolve against the skepticism of doubting clerics, Copernicus declared that the sun was the center of the universe. That was not all. In addition, the earth was in constant motion around the sun and so were all of the planets. The paradigm that had served so well in the observations of Ptolemy and Brahe had broken down because improved instruments and measuring processes had revealed motions not previously visible. These observations were not easily explained with a geocentric (earth-centered) universe. However, if the earth were put into rotating motion about two poles and the sun remained motionless at the center of the system (heliocentric), the motions of all heavenly bodies could be explained from the reference point on a moving earth. Complex armillary spheres could be replaced and the earth could be viewed, quite acceptably, as the center of the moon’s orbit. Astronomy would never be the same. The stage was set for the discovery of Kepler’s laws of planetary motion in the year 1610. The
Copernican revolution, having changed the paradigm, set the stage for a number of discoveries ranging from planets, to comets, to double stars and asteroids. Because of the new way of thinking, the field of astronomy became fertile ground for the germination of the knowledge taken as so common place today.

In the year 1859, with the publication of Charles Darwin's *Origin of Species*, biology was to undergo a similar scientific revolution. The existing essentialist paradigm of the immutability of species was no longer able to explain the wide array of observed species. Darwin had observed, in South America as well as the Galapagos Islands, a wide variety of species. Those on the island chain were very different from those on the mainland. After cataloging these differences into a notebook and reflecting on possible causes, Darwin put forth two new ideas. These ideas were to eventually do away with the essentialist paradigm that had so long ruled both biology and philosophy. The first idea was that species and populations can and do vary randomly. The second idea is that nature selects which of those variations will be passed on to offspring. The net result of all of this is that species can and do change and, further, that nature determines which changes will be preserved through reproduction and subsequent genealogy. Biology would never be the same. The veneration of essence and archetype of species would eventually give way to a concept of variable species that could change and even, over time, appear in new forms or disappear altogether. As was true of astronomy, this shift in paradigm set the stage for future discoveries in the fields of genetics and molecular biology. Biology was on its way to becoming a science that could explain the vast diversity of life forms that
inhabit the planet. There was a focus on the idea that life forms, no matter how different in structure and geography, are related and proceed from a common origin into the myriad of life forms that have existed both in the past as well as today.

**Essentialism and Popular Thinking**

The essentialist species concept, which dominated evolutionary thinking until the middle of the nineteenth century, was that each species are characterized by an unchanging essence and each is separated from all other species by a sharp discontinuity (Mayr, 1988). Variety, when it does exist, is an aberration from an archetype or essence. Variety is to be explained away and, when it does occur, has to be the product of a sudden drastic mutation or as the result of hybridization. Species cannot evolve or bud from other species because the tendency of nature is to revert back to the essence or type that each species possesses. In other words, there is some characteristic unique to and shared by all members of a species which explains why they are the way they are (Sober, 1994). With such a concept, evolutionary thinking is neither allowed nor is it possible.

With Darwin's *Origin of Species* a new conception of species was born. This conception is what some have termed population thinking (Eldredge & Cracraft, 1980). Populations evolve because some of the individual members of the population possess a reproductive advantage in their randomly changed genetic traits. These discrete changes in individual organisms are selected for or against by nature itself (Simpson, 1994). If selected for, they are allowed to reproduce and, in so doing, pass
on their altered genetic traits. If they are selected against, they do not reproduce and, eventually, the organism and its altered genetic material disappear. After repeated selection for or against a genetic trait, the population changes in such a way that new organisms and varieties of organisms are present in large numbers in the population (Bishop & Anderson, 1990). There are two very important ideas expressed here. The first idea is that genetic change is not need driven in the sense that Lamarkian adherents believe. The organism does not develop a longer neck, shorter legs, or a darker color because the environment dictates a need for it. The second idea is that the changes needed to create new species do not take place within the lifetime of a single organism. Evolution is a gradual process that requires several generations of offspring in order to observe variations in a population. One of Darwin's great contributions was to foster the idea of populations as the basis of evolutionary change. There are two distinct processes that influence traits exhibited by a population over time: (1) new traits originate by random change in the genetic material, and (2) these changes are retained or disappear due to selection by environmental factors (natural selection). Thus, the evolutionary process is driven by reproduction and death of individuals and not by changes occurring in their lifetime. Fitness, in the evolutionary sense, does not mean physically strong and robust but rather refers to the ability of individuals and their genes to produce offspring. Thus, evolution must be inferred in the light of the conceptual framework of populations and the ability of their individual members to pass on their genetic variety over several generations (Mayr, 1988).
Macroevolution

Most evolutionary biologists view their work through one of two lenses: macroevolution or microevolution. Macroevolution is the more panoramic of the two and has been termed “evolution on a grand scale”. Macroevolution takes place by the processes of natural selection and adaptive improvement over a long period of time (Ridley, 1993). Macroevolution is often focused on morphology because the taxonomic and fossil evidence for morphology is greater than for other kinds of characters (Ridley, 1993). The classical view of macroevolution, from Darwin to the present day, is that species, in the course of gradual evolution through time, change to such a degree that they will become different genera or higher taxa and will have all of the adaptations and specializations needed to explain the diversity that our world shows us. However, as Eldredge and Cracraft (1980) point out, evolutionary phenomena such as origins, histories, patterns and modes of extinctions must also be considered by those who practice macroevolutionary thinking.

Macroevolution is often based on the study of phenotype rather than genotype (Mayr, 1988). The macroevolutionary processes of adaptation, convergence, rate of evolution, and shift in adaptive zones all relate to phenotype (structural appearance) and can be studied without reference to their genetic basis (Mayr, 1988). The focus of this research will be through the lens of macroevolution. The focal point will be how the changes in phenotype have produced different branches of an evolutionary
tree that is rooted in one or more common ancestors.

**Microevolution**

While the focus of this research will not be through the lens of microevolution, a brief description of this type of thinking is made for the purposes of contrasting this study with the current work that occurs in many of the molecular biology laboratories of today. Mayr (1982) believes that microevolution is the study of populational phenomena often studied by geneticists. Scientists with this method of thinking seek to reduce structure and function to gene frequencies. Microevolutionary research focuses on short-time scale as well as a small spacial scale. It incorporates population ecology, population genetics, and statistical analyses of phylogenetic processes (Brooks & McLennan, 1991). Microevolution is where much of the current research is being done, often with great dependence on technology and little concern with the classical ideas of macroevolution.

**Systematics and Evolution**

Systematics is the science that is concerned with the classification and naming of organisms. The science of systematics reflects phylogenetic relationships or genealogy, genetic similarity and other factors that reflect an assemblage of evolutionary information (Eldredge & Cracraft 1980). Systematic biologists also believe that a classification system is the reflection of a branching diagram that is generally considered to be one of two types: a phenogram or a phylogenetic tree. The
phenogram is a construction born of scientists who study large numbers of characteristics and use statistical methods for clustering the characters into a diagram. Phylogenetic trees, however, reflect common ancestry and are made up of links that connect various species and their ancestors together. The creation of phylogenetic trees as evidence of what has occurred over time is the very basis of this research. Participants will be asked to create phylogenetic trees.

Systematics, unlike taxonomy, undertakes the study of organic diversity as a product of evolution (Mayr, 1976). Every classification is recognized as a scientific theory that is both explanatory and predictive. It is explanatory in the sense that it delineates the existence of natural groups as products of common descent. It is predictive in the sense that accurate predictions can be made about the pattern of variation and the placing of newly discovered species. Further, systematics tries to explain the differences between organisms in the light of natural selection. It's concerned not only with variation but with an overall history of how organisms adapted themselves, through phenotypic change, to the environmental challenges of nature. Phylogenetic trees, as products of a systematic way of thinking, can delineate paths of common descent from ancestors to extant (presently living) organisms. They can and do demonstrate, in a visual way, what has been as well as how present day organisms are related.

Three Philosophies of Systematics

Three of the current philosophical approaches to the study of systematics are
phenetic (numerical) taxonomy, evolutionary taxonomy and cladistics. This section will explain all three philosophies but will give special attention to cladistics because that is the basis of the phylogenetic tree construction used in this study.

**Phenetic Taxonomy**

Phenetic taxonomy was pioneered by Skokal and Sneath (1963), Skokal and Camin (1965), and Skokal, Rohlf, and Sneath (1965). This philosophy was a reaction to the evolutionary taxonomists who treated phenotypic characters in an unequal manner because of a belief that some characters are more indicative of phylogeny than others. Evolutionary taxonomists refer to this practice as character weighting and phenetic taxonomists are opposed to it. The aim of phenetic taxonomy is to bring about repeatability and objectivity. What the system represents is phenotypic similarity and not evolution (Ridley, 1985). Phylogenetic divergence took place in the past but phylogenetic relations cannot be known. Therefore, organisms are classified on the basis of overall similarity. Similarity is determined by the presence of numerous, equally weighted characteristics. These characteristics, once determined, are placed into a computer and a cluster statistic is used to group the species. From this clustering technique, a diagram of species similarity is produced by grouping species with like numbers of similar characteristics in the same group or category. The result is a branching diagram known as a phenogram. The diagram shows the record of character change without any regard for one character as being more truly indicative of past history than any of the others. The whole procedure is repeatable.
and, to some extent, mathematical although the exact hierarchy might differ if the practitioner chose to use a different number of characters or a different cluster statistic. The obvious criticism of this philosophy lies in the fact that it is extremely subjective. Subjectivity is reflected in the choice of characteristics, how they will be described, and how the character distance is to be measured as well as what cluster statistic will be used. There are no criteria to guide one preferred choice above another (Ridley, 1985). The philosophy is based on a system of similarity without regard to genealogy and ancestral relationships.

**Evolutionary Taxonomy**

Evolutionary taxonomy predates phenetic taxonomy and was pioneered by Dobzhansky (1937), Mayr (1942, 1969), and Simpson (1944). Evolutionary taxonomy is built on the philosophy that evolution has produced all of the natural groups of life. Therefore, evolution has to be known and a classification system must reflect what happened in time. The first step in the process of phylogenetic classification is the analysis of taxonomic characters in order to determine which ones are homologous and which ones are derived from a common ancestor. Homologous characters are termed “good” characters and are weighted highly. Analogous characters, those that are similar in function but not in structure, are termed “bad” characters and are usually not considered. The good characters are used to draw evolutionary tree diagrams that represent monophyletic groups defined by homologies (Ridley, 1985). A monophyletic group is one that includes a common ancestor and all of its
descendants. Ridley (1985) says that evolutionary taxonomists sometimes choose to leave certain descendants out of the tree. This practice changes the group from being monophyletic to being paraphyletic. Parallel with the branching tree that indicates phylogeny, there is a story of how the tree came to be. There is an historical narrative that accounts for evolution. Evolution has two main aspects: the order of splitting and the rate of divergence. Evolutionary taxonomists are concerned with both. Evolutionary processes have produced a tree like hierarchical pattern of phylogeny. They are interested in similarity of characters but only those characters that seem to explain an evolutionary history or narrative. Ridley (1985) says their ability to do this is somewhat ambiguous because much of their tree narrative is not verifiable and their practice of character weighting is subjective and follows few guidelines other than intuition.

**Cladistic Taxonomy**

The cladistic philosophy of taxonomy, the one that is used as a basis of this research, was pioneered by the entomologist Willi Hennig who published his first English-translated work *Phylogenetic Systematics* in 1966. The main idea of this school is that the results of the evolutionary process can be expressed in terms of a branching diagram called a cladogram. These cladograms are used to create sister groups or clades. Two species are alike if they share a recent common ancestor. The notion of recent common ancestor is extremely important. If two taxa, B and C, are sister groups, they are more closely related to each other than a third taxon A (see
Figure 1). B and C will have shared derived characters that A lacks. B and C must share a more recent common ancestor than either does with A (Hull, 1988). For Hennig species are the basic units of the evolutionary process. Species are the things that evolve as a result of mutation and selection. Species are also the basic units of classification. The characters used to infer species relationships must be evolutionary homologies (Hull, 1988). Homologies are characters shared from a common ancestor and only these characters indicate phylogenetic relationship.

![Figure 1. A Cladogram Showing B and C Sister Groups and A as an Outgroup.](image)

Ancestral and Derived Character States

There are two types of character states that are important in the construction of phylogenetic trees. The characters may be termed “ancestral” or “derived” depending on how the character is passed on through evolutionary time. An ancestral character is any character that remains the same or unchanged through time. These characters are found in the common ancestor and, since they do not change, are poor
indicators of evolutionary change in taxa groups. Because there is no change between extinct and extant organisms, ancestral characters are not used in the construction of phylogenetic trees except when they are used to handle homoplasy. (In this case homoplasy refers to a situation where a character that has changed has reverted back to the ancestral state and must be placed on the phylogenetic tree as a reversal). Derived characters, however, are indicators of evolutionary change between a common ancestor and its descendants. Derived characters do indicate a phylogenetic relationship and cladists use them in the construction of phylogenetic trees.

There are three methods for distinguishing an ancestral character from a derived character. They are the paleontological criterion, the embryological criterion, and outgroup comparison. The paleontological criterion is the least reliable because of the incompleteness of the fossil record. Many characters leave no record at all since many of the characters do not fossilize in hardened tissues such as bones or shells. Cladists consider most of the characters found in fossils to be ancestral. The embryological criterion makes use of VonBaer’s law which divides the embryology of an animal into two parts, early development and late development. Early development is a general state and is found in all animals. It includes such phenomena as cleavage, gastrulation, and differentiation of tissue layers. Cladists consider the general state characters found in early embryology to be ancestral. Late development includes the formation of specialized organs such as mammary glands, hair, or a large cerebrum. Cladists consider the characters found in later embryology to be derived.
characters. Outgroup comparison is a method whereby a closely related species to the species being studied is carefully examined (i.e., cartilaginous fish such as a shark with a bony fish such as a perch). The state of the character in the outgroup is considered ancestral while the state of the character in the ingroup is considered derived. Since a cartilaginous skeleton is found in the shark (the outgroup), a cartilaginous skeleton is considered to be ancestral. While a bony skeleton is found in a perch (the ingroup), a bony skeleton is considered to be a derived character. The bony character, because it is derived, becomes a character that would be mapped onto a phylogenetic tree in order to denote an evolutionary difference between a shark and a perch.

Phylogenetic Trees

A typical phylogenetic tree depicts patterns of ancestry and descent among a series of taxa (Eldredge & Cracraft, 1980). All phylogenetic trees are rooted in a common ancestor which is placed at the bottom of the tree. The extant (living) organisms or taxa are placed at the top of the tree with branches or links connecting each of the taxa with the common ancestor. Derived characters (those that have changed and are not present in the common ancestor) are placed along the links (branches) of the tree. In Figure 2, which follows, the extant taxa are L, M, N, and O. N and O are a sister group that share derived character states numbered 3 and 4. M, N, and O are a monophyletic group that share derived characters numbered 1 and 2. L is the outgroup with no derived character states at all and therefore has no character.
state numbers placed along its branch. (It would be no different than its common ancestor). When characters are mapped on a phylogenetic tree they must follow the rule of parsimony. That rule states that character states must be placed on as few branches as is possible to explain the evolutionary pathway. Once the tree is constructed it represent a true phylogeny or pattern of descent from a common ancestor to the species that are currently alive today. The entire tree is based on the assumption that there is only one true phylogeny (Brooks & McLennan, 1991). The primary rule for the construction of a phylogenetic tree is Hennig’s Auxiliary Principle which states that homology (similarity in structure) is the truest gauge of descent in the absence of any other evidence. Therefore, homology is assumed as often as possible and characters shared for other reasons are not considered.

![Common Ancestor](image_url)  

**Figure 2. A Typical Phylogenetic Tree.**
Importance of Phylogenetic Trees

Comparison is an indispensable tool for evolutionary biologists. The hylo-
genetic approach is the newest component of comparative biology. Such an approach involves the using of genealogical relationships among species, along with ecological and behavioral information for those species, to investigate questions about character origin, character divergence patterns, and the processes underlying those patterns. The phylogenetic approach covers a wide range of topics that include the evolution of ecological association between host and parasite to the interaction of complex communities and systems.

Since systematic biology is involved in the discovery, description, and naming of species and higher taxa, phylogenetic trees can provide a tool that is useful in cataloging diverse species and their relationships to each other. In addition, agriculture, which is increasingly faced with problems of pest control and crop damage, can turn to phylogenetic trees to look for organisms that may serve as biological controls against pests, thus avoiding the use of insecticides. Medicine and public health may also benefit from the use of phylogenetic trees. It is well known that there are geographically isolated organisms that produce substances that fight against diseases like cancer. Often these creatures are the only source of this treatment substance and their death or extinction would mean a loss of their medical value. If phylogenetic trees are constructed for these and closely related species, medical research can turn to other, tree identifiable, closely related organisms to isolate additional amounts of the identified treatment material. In the area of public health, phylogenetic trees could be
used to help discover new vectors of dreaded diseases. This would occur by construct­ ing trees of known vectors and then analyzing the trees carefully for similar species that might have developed a parasitic transmission pathway. The known vec­ tor and its close relative could then be removed through biological control or other medical means.

Summary of the Biology and Evolution Section

The evolution and biology section may be summarized in the paragraphs that follow. Evolution and the idea of mutability (ability to change) of species was a new way of looking at biology. Prior to 1859 it was commonly believed that each species was unchanging (immutable) and even archetypal in both form and nature. With this paradigm in place, evolutionary could neither exist nor even be entertained. With Darwin's paradigm change in 1859, the scene changed and people of science began to entertain the bold idea that species not only could change but were, through a common ancestor, actually related to each other. This did not mean that individual members of a species were related but rather that populations or groups of species were related. This was and is a very important idea since evolutionary change occurs in a population of species rather than through individual members of a species. At any rate, with this shift in thinking, evolutionary concepts were on the way to explaining how the diversity in the biological world may have come to be.

There were two ways of looking at this diversity, macroevolution and microevolution. Macroevolution was and is the more sweeping of the two and seeks to
know histories of populations and what their common ancestor(s) may have been. Macroevolution is concerned with phenotypical homologies between populations and seeks to tell the story of evolution on a grand scale. Microevolution, however, seeks to explain populations in terms of gene frequencies, population genetics, and statistics. Microevolution is concerned with putting evolution into a more quantitative perspective and one that, somehow, make evolution a more quantifiable commodity and less of a philosophical exercise.

With the rise of the science of systematics and the desire to know both evolutionary history and the phylogenetic relationships of species, three separate philosophies were developed: phenetics; evolutionary taxonomy; and cladistics. Of those three philosophies, cladistics, with its emphasis on the branching of taxa (species) from a common ancestor, is the philosophy on which this research is based. The cladogram of the cladists, known as the phylogenetic tree today, is a way of visually tracing an evolutionary pathway from a common ancestor. As the tree branches from its root, special phenotypic, derived character states (those character states not present in the common ancestor) are placed along the branches of the tree. The order of branching determines which taxa are closely related and which taxa are not closely related. The diagram even shows what pivotal character states separated one group of taxa from another. In so doing, a phylogenetic tree is a new way of both thinking about and looking at evolution. The tree not only demonstrates what has been and how it may have come to be but it may also be used to ascertain where evolution may be going through the careful examination of species relationships and the character
states that separate them from each other.

The Problem Solving Tradition

Introduction

Because this study is conducted in the problem-solving research tradition, it is necessary to know not only what a problem is but also how the solving of problems has been studied. Accordingly, this section has the following sequence: definition of a problem, early conceptions of problem solving (psychology and computer programming), experts and chunking, problem solving in physics, problem solving in genetics, and problem solving in evolution. The first two sections, definition of a problem and early conceptions of problem solving, provide the background for determining what a problem is and how the study of problem solving began in non subject matter content domains. The third section, experts and chunking, provides the information needed to understand how experts solve problems in the ways that they do. The fourth and fifth sections, problem solving in physics and genetics respectively, provide information on problem-solving carried on in two specific science domains, physics and genetics. The sixth section is an account of Steven Brewer’s problem-solving study in the field of evolution. This section is important for this research since it provides the model of expert phylogenetic tree construction that this study uses to compare the performance of novices in the same task. This study provides a counterpart to Brewer’s (1996) study and uses the model of expert phylogenetic tree construction as a model which novices should emulate.
What Does Problem Solving Involve?

A problem is any task that requires both reasoning and analysis in order to bring about a solution (Smith, 1991). The problem may be trivial, difficult, or impossible to solve, depending on the complexity of the solution. The number of paths leading to a solution determines the complexity of a problem. The idea is to choose one path that leads to successful attainment of the desired goal. Reif (1983) states that there are three general procedures for solving any problem:

1. Initial description and analysis of the problem in order to reduce the problem into a form that is more easily solved.

2. Construction of a solution by methods involving planning and choices in order to search for a single path leading to a solution.

3. Testing the solution produced in order to determine correctness and to bring about any revisions that strengthen it.

How these general procedures for solution are followed is dependent on the status of the solver. Experts, having a broader knowledge base and a wealth of experience, are more successful in problem solving because they use a patterned and planned approach. They tend to solve a problem by working in an inductive (forward) manner while drawing from information stored in memory. Novices, lacking both an informational base and past experience in problem solving, tend to use an unplanned, piecemeal approach in solving a problem. They tend to work deductively (backward) in a process known as "means end" analysis. In this analysis they determine where they are, where they have to go, and then try various trial and error methods to
achieve their goal (Ericsson & Smith, 1991). Novices often lack a store of problem solving skills and usually try to gain an immediate solution without devoting much time to problem redescription and analysis (Reif, 1983).

Through the years there have been various attempts to list the desired characteristics of a successful problem solver (Bransford & Stein, 1984; Hayes, 1981; Polya, 1945). Smith (1991) gives the following characteristics of successful problem solvers:

1. He or she forms an hypothesis as to what is needed for the problem to be solved.

2. He or she creates an internal problem space or a personal understanding of the problem.

3. He or she extracts the most important parts of the problem in such a way that it contributes to a solution (creates sub problems).

4. He or she may redescribe the problem in a different manner.

5. He or she plans a basic outline of solution strategies.

6. He or she uses general and domain specific heuristics (rules of thumb) to attack the problem and bring about a solution.

All of the above are done in order to prove or disprove an hypothesis which the solver has formed early on in the solution plan. The data, generated in the solution process, are used to validate or falsify this hypothesis.
Early Conceptions of the Problem Solving Process

Early conceptions of problem solving have their roots in the domain of psychology and consist of four distinct approaches (Voss, 1989). The first approach, held by William James, views problem solving as a description of the steps needed to solve a problem. The second approach, held by Piaget (Voss, 1989), is one of knowledge development evolving simultaneously with the mental growth of a child (i.e., problem solving ability increases as mental growth proceeds through various stages to maturity). The third approach, held by Gestalt (Voss, 1989), views problem solving as a perceptual process in which the solution to a problem is the restructuring of the solver's organization of the problem. Insight and experience often interact to produce a perceptual change that has come to be known as the "Gestalt Shift". This change in looking at the problem often produces a solution. The fourth approach, held by Thorndike, focuses on stimulus-response activity where the acquisition of certain responses, under particular stimuli, permit the goal to be attained. Since the solution of problems is intimately connected to such thought processes as thinking, learning, memory, perception, transfer, and motivation, it cannot be considered apart from these psychological processes.

In the era of the 1950's, 60's, and 70's, problem solving became the chief focus of computer programmers who sought to solve problems by means of computer simulation and artificial intelligence. Their chief goal was to process information. Information processing research concentrates on how an information processor, be it human or computer, takes an external problem, translates it internally into operational
form, carries out appropriate operations, and then outputs the result. Such research seeks to determine what knowledge individuals bring to the task, what information in the problem they key on, and how they use these items of information. In order to do so individuals must be examined as they solve problems and a frequently used technique is to gather think-aloud protocols where the solvers describe their thoughts as they go about solving problems. It is useful to think of problem solving as involving an initial state (i.e., problem statement), a goal state (i.e., problem solution), and any constraints or limiting conditions in-between. The task is to determine all possible solution states and the actions or operators needed to go from one state to another. “Viewed in terms of states and operators, the solution involves going from an initial state through a series of intervening states, and reaching a goal state, with each state to state transition occurring by the application of an operator” (Voss, 1989, p. 256). Solving a problem becomes a search process in which the solver looks for a path from the initial state to the final goal. To get there, however, one must choose operators. How does operator selection occur? Usually it involves a strategy or systematic method that has the potential to solve a problem. A frequently used strategy is “means end” analysis in which the solver considers the current state, the goal state, and, by working backwards, selects the moves that will accomplish the goal.

Problem solving has also been studied in connection with proficiency at games such as poker, bridge, or chess. The strategy here is to categorize individuals at various skill levels with respect to playing a particular game and compare the performance of each one. Thus, experts and novices may be given a particular
memory task such as recalling the position of chess pieces on a board or a particular configuration of cards in order to determine whether or not experts have better recall.

The basic assumption being that

we can approach the understanding of skill development in a particular domain (whether game, academic subject matter or some other field) by determining the processes that characterize high level performance and how these processes are similar or different from processes characteristic of lower level performance. (Voss, 1989, p. 259)

Experts and Chunking

One of the major results of studying expert-novice problem solving, is that experts have a greater ability to reproduce an array or pattern of chess pieces on a chess board when the pieces are displayed in game position. If the pieces were randomly placed, experts did not have the advantage over novices (Chase & Simon, 1973). Chase and Simon (1973) explained this by saying that the experts "chunked" information. That is to say the experts remembered more information and that one informational chunk pointed to another informational chunk whereas, in less skillful players, the chunks remained independent of one another (Chase & Simon, 1973). Thus, individuals with higher chess playing skills are able to maintain a larger amount of information in memory, possibly by chunking, that gives direct access to long-term memory (Voss, 1989). Using the information processing model of Ericsson and Simon (1980), information may be stored in two locations: short-term memory (STM) and long-term memory (LTM). Short-term memory, as the name implies, has limited capacity and consists only of the more recently heeded
information. In contrast, long-term memory has a large capacity with rather permanent storage but with slow fixation and access time. Information in long-term memory must be transferred to short-term memory before it can be reported. A chunk is a familiar pattern that is produced by previous exposure to certain stimuli. The amount of information that can reside in short-term memory is limited to between four and six chunks. New information is retained in short-term memory during the time that the sense organs are attending to it. To create a representation in long-term memory, associations must be built up by coding and imaging and the forming of branches in the cognition network. A time period of 8 to 10 seconds is required to assemble each new chunk from short-term memory and into long-term memory as a new chunk (Ericsson & Simon, 1980). As particular processes become highly practiced, they become more and more automated with the steps being carried out without interpretation and without using inputs and outputs but using short-term memory alone. The whole affect of practice is to increase this automation process and this is why experts in game strategies do so much better that novices do. If novices can learn to chunk in the same manner that the experts do, their chances of being successful problem solvers are greatly increased. Any process that will extend the memory, serves as a valuable tool in problem solving. The chief difficulty, of course, is the limited experience that novices bring to problem situations. Novice success in problem solving is often curtailed because of the lack of a trained eye to observe repeatable patterns in game situations and other areas where a choice of outcomes is possible.
Much of the problem-solving work in physics has involved the contrasting of experts and novices and the collection of "think-aloud" protocols where students are asked to think-out-loud while solving problems. In an early study of Reif, Larkin, and Brackett (1976) using think-aloud protocols, it was found that students rush in and immediately start calculating something in any unorganized manner and have no strategy, whatsoever, for directing their activities. When some of the students were taught a simple four-step problem strategy of: (1) description, (2) planning, (3) implementation, and (4) checking, they showed dramatic problem-solving improvement.

In a study involving one proficient subject and problems in thermodynamics, Bhaskar and Simon (1977), using a think-aloud protocol, found that the subject used an approach similar to means-end analysis tempered by a knowledge of the role that the conservation of energy plays in such problems. Means-end analysis is a general heuristic process where the goal is identified, the current state is compared to the goal, and operations are performed to reduce the difference between the current state and the goal state. General properties are usually used by experts but means-end analysis, often focusing on certain equations, is a common approach of novices.

Simon and Simon (1978) conducted a study in which two students, one experienced in working dimensional kinematics problems and the other not, were asked to solve a series of problems. The more experienced solver used a working forward approach in which the given information was plugged into an equation or
series of equations that were previously identified, until the unknown value was found. The less experienced solver used a working backward approach in which there was no series of planned equations. The solver started with one equation, solved for the unknown, and searched for another equation to satisfy the next part. The process took much more time and involved means-end analysis. Simon and Simon (1973) further stated that the more experienced solver has a physical intuition that guided him/her toward effective and efficient problem solving.

Larkin and Reif (1979) reported on how two subjects, a novice and an experienced problem solver, solved five standard problems in mechanics. These researchers developed two models to describe the main features of how the expert and novice solved the problems. Both of the models were similar in the fact that the original description involved only general knowledge with no specific physics concepts. The novice went from the original description to a mathematical description and then combined equations to eliminate unneeded quantities. The expert, however, moved from the original description to a physical description and then to a mathematical description. The obvious difference between the two models is the physical description of the problem. This is a domain-specific heuristic that requires physics knowledge that novices lack or do not possess in a usable form. Larkin and Reif (1979) say that the instructional implication is to teach students to organize the physical properties relating to a problem and then approach problems by going from a global description to more detailed aspects of the problem.

In Larkin’s 1980 study, experts and novices were given physics problems to
solve and protocols were collected during their solution. Analysis of the protocols revealed that experts sought to explain the problem by first drawing a diagram and then carefully studying the relationship between problem variables. The experts would categorize the problem and then proceed to write down appropriate equations for solution. In other words, experts planned their strategy by working in a forward manner in order to achieve a solution. Experts also tended to apply equations in bursts because they were stored in memory as chunks while novices tended to solve the problem in a piecemeal fashion (Larkin, 1980). The ability of the novice to plan a strategy of solution is much more limited, often focusing on a single variable at a time and then directly applying an equation to solve a particular problem rather than considering the problem as a whole.

Heller and Reif (1984) carried out experiments where they designed a detailed program of step-by-step methods and knowledge needed for a human subject to act in accordance with a specified model of performance. In the experimental procedure an individual was asked to carry out certain tasks such as the description and solution of various physics problems. The problems could be solved using Newton's second law of motion (F=ma) but two of them were more perplexing because of the employment of several operating forces. The study involved twenty-four undergraduate physics students who were given a pretest involving the problem just mentioned. The subjects had considerable uncertainty in solving these problems and did not know how to begin or proceed. After completing these pretest problems, a week of subsequent practice sessions was held during which time students were supplied with
printed sample solutions. The students could refer to these sample solutions at any
time during their practice sessions. About one week later the test subjects were again
asked to solve three problems in mechanics. The subjects were read standard direc-
tions from a script with each direction having to be performed before the next one
was attempted. Subjects were also asked whether or not any object, other then those
already named, touched the system of interest. The directions were easily followed
and some of the subjects, after the first problem, took over and began working on
their own. However, the subjects were always asked to state the next step and to do
so with the instructor intervening only when a step was missed. A control group was
also employed that worked with no model of problem solution whatsoever. The
results showed that those subjects working with a model of guidance improved their
performance over the pretest score dramatically. Those in the control group experi-
enced no gains but actually showed a retrogression from their original pretest score.
Heller and Reif (1984) conclude that their work is highly relevant to teaching stu-
dents improved scientific problem-solving skills. Such instruction requires the
employment of a prescriptive model specifying how good problem solving is to be
achieved. Such an instructional unit must do more than mimic the actions of expert
problem solvers; it must also explicitly agree with the processes whereby such expert
performance can be learned.

Several studies have been conducted to explore the relationship between the
knowledge base of an individual and his or her problem-solving behavior. Chi,
Feltovich and Glaser (1981) studied how novices and experts categorize standard
physics problems involving mechanics. Using eight experts (Ph.D. students) and eight novices (students who had just completed a physics course), the researchers asked subjects to categorize twenty-four problems according to similarity of solutions. After analysis, the researchers examined four pairs of problems that both the experts and novices considered the most similar. The novices categorized primarily on the basis of what the researchers called "surface structure"; that is, they used literal objects, explicit physics terms, or specific physical arrangement as the basis for categories. The experts, however, categorized on the basis of physical principles or what the researchers called "deep structure" or the principles governing the domain.

Velhuis (1986) replicated and broadened the research of Chi et al. (1981). He used 94 novices, 5 intermediate subjects, and 5 experts to categorize sets of problems that contained either surface or deep structure characteristics. The results were similar to those of Chi et al. (1981) except that the novices in this study used both surface and deep structure in their problem solving.

Problem Solving in Genetics

Genetics problems typically fall into two categories: those that involve cause to effect reasoning and those that involve effect to cause reasoning. The problems encountered in high school and introductory biology textbooks are in the cause to effect category. In solving cause to effect problems one need only apply an algorithm (formula) in order to achieve a solution. Such problems often contain enough
clues in the problem statement that the solver may easily choose which algorithm to employ and quickly arrive at a solution. Effect to cause problems present a different scenario. As mentioned above, there are often few solution cues in the problem statement, the solver must form an hypothesis and then evaluate the hypothesis as the solution proceeds. Of course, in order to formulate and test a hypothesis, a great deal of conceptual knowledge is needed. The solver may have to consider several hypotheses before finding the correct one. Therefore he or she must be open to alternative hypotheses as well as the one originally proposed. Such problems not only reinforce the need for a conceptual knowledge base but also may create situations where new knowledge can be created. In an effect to cause problem there is no one correct path to a solution but many, depending on how the conceptual knowledge of the solver is ordered.

Hackling (1988, 1990) was interested in the differences between experts and novices in solving genetics problems. His work mirrored the results of both Larkin (1980) and Heller and Reif (1984) in the domain of physics. Having produced a list of strategies used by experts in solving pedigree problems in 1984, he was aware that experts used four strategies in solving such problems. These strategies involve searching the problem for possible cues to an inheritance pathway, generating an hypothesis by using genotypes for each individual in question, testing alternative hypotheses to accept or reject the original hypothesis, and searching for evidence to determine which of the hypotheses (original or alternative) is most likely operating. In his 1990 study he compared novices and experts in their ability to solve five
pedigree problems. Protocols (verbal records of activities while in problem solution) were collected and analyzed for three items: cues critical to problem solution, best interpretation of each cue, and inheritance pattern of the trait in each problem. The successful solvers were better able to utilize cues and to make use of multiple inheritance pattern hypotheses in order to reject alternate patterns of inheritance. The best solutions were characterized by two features: extensive use of cues and a minimum use of genotypes for testing various hypotheses (Hackling, 1990). Novices, while far less successful at correct solutions, can be instructed to use the cues in pedigree problems as well as other expert strategies to achieve correct solutions. Hackling (1990) suggests that instruction of novices in pedigree problem analysis should focus on the following knowledge and activities:

1. Assign autosomal and X-linkage genotypes to individuals.
2. Use multiple hypotheses testing to reject alternative hypotheses.
3. Recognize and interpret critical cues in order to choose between dominant and recessive patterns as well as autosomal and X-linked patterns of inheritance.
4. Develop an awareness of the need to search the pedigree for cues to resolve conflicting hypotheses.

Computer Simulation and Genetics Problem Solving

A commonly used computer program for solving genetics problems is Genetics Construction Kit (GCK) developed by Jungck and Calley (1985). With the GCK program it is possible to simulate a problem much as a practicing scientist.
might look at it. Each problem begins with a field population of organisms to which a sex and phenotype have been assigned. Students then select individuals from this population field to be the parents for crosses. Several generations of offspring can be produced until the student is ready to explain an inheritance pattern of phenotypes. Because the program requires the solver to both generate and interpret data, it requires more knowledge to solve a typical problem than the mechanical manipulation of a few algorithms. Using the GCK program, Slack and Stewart (1990) analyzed their data and produced a profile of novice problem solving in genetics. The profile includes the following characteristics of novice solvers:

1. Novices have neither a well-organized plan of solution nor a well-thought out plan of solution strategy.

2. Novices are less systematic in data description.

3. Novices read all of the initial information given by the program and do not attempt to identify the number of traits or the kinds of variation.

4. Novices look at the initial population and immediately begin to do crosses.

5. The knowledge of novices is neither extensive nor well structured.

6. Novices make hypotheses about individual crosses but do not use inheritance patterns in their solutions.

7. Novices could not identify cues to inheritance patterns from the overall data.

8. Novices provide little evidence of hypothesis testing and usually did not confirm their solutions.
9. Many misconceptions block the ability of novices to solve inheritance problems.

Specific misconceptions were:

1. Dominant trait was confused with the phenotype that is most often present.
2. Did not think in terms of generations but viewed each problem as a new cross.
3. Did not think genotypically but only phenotypically.
4. Held an alternative concept of meiosis that blocked their ability to form the proper gametes.
5. Viewed ratios as the determining cross limits.
6. Novices worked backward and explained the data rather than producing it.

Slack and Stewart (1990) say that their finding corroborate those of Smith and Good (1984) and Hackling (1984) and further extend the research knowledge by making the participants generate their own data as well as conclusions. Slack and Stewart (1990) suggest that genetic instruction should include predicting data, redescribing the problem quantitatively, generating and testing both an original and alternative hypotheses, and checking results. It is necessary to teach both genetic concepts and problem-solving skills if students are to have a better understanding of the nature of science.

In contrast, Angelo Collins (1986) also used the GCK program but was concerned with the performance of experts rather than novices. She also produced a profile of expert problem solvers (Ph.D. geneticists) and found them to be the exact
opposite of novices. Her profile includes the following characteristics of expert genetics problem solvers:

1. Experts have extensive knowledge of genetics and inheritance patterns.

2. Their knowledge-base of genotype - phenotype mapping allowed them to move easily between genotype and phenotype.

3. Experts had a well-planned problem agenda that included: (a) redescribing the field populations in terms of number of traits and variations; (b) generating tentative inheritance pattern hypotheses; (c) using their hypotheses to synthesize solutions, make crosses, and interpret results; (d) redescribing data from crosses; (e) considering other causes to produce the same data; and (f) checking the results. However, in spite of the superiority of their attack, the experts did not solve all of the problems correctly. This was because of the fact that their overall plan, once made, was not changed until the entire problem space had been searched (Collins, 1986). During this time cues to other paths of inheritance were overlooked. The result was a failed solution that was typical of novice problem solvers.

Problem Solving in Evolution

A study entitled *An Account of Expert Phylogenetic Tree Construction From the Problem Solving Research Tradition in Science Education* was conducted by Brewer (1996). This study was an account of how experts, persons with advanced degrees who practiced phylogenetic analysis in their occupations, constructed phylogenetic trees using a computer program called *Phylogenetic Investigator* (PI).
Brewer (1996) developed the PI program by reducing phylogenetic diagrams into their component parts and comparing them with the objects and actions needed to create cladograms and phylogenetic trees as stated in Eldredge and Cracraft (1980). After working a type problem using PI, Brewer (1996) presented each participant with a series of problems to be solved either using PI or paper and pencil methods. As the participants were solving the problems, their think-aloud statements were tape recorded. Using a combination of PI printouts, think-aloud protocols, and any notes or drawings the participant may have made, Brewer (1996) analyzed these data with an eye toward finding similarities and differences in solution sequences.

One of the primary goals of this study was to construct a model that describes the consistent and regularly employed strategies used by experts in constructing phylogenetic trees. Several models were produced as the data from each participant was subsequently analyzed. A final model of expert solution was produced by forming an amalgamation of the individual participant models. Brewer (1996) discovered three primary strategies or global approaches to the construction of phylogenetic trees by experts. The most commonly employed strategy was **inclusion/exclusion analysis**. This strategy involves looking at the data matrix and taking one character or a pair of characters (usually the most inclusive for all taxa) and dividing the remainder of the characters into three categories: inclusive, exclusive, or conflicting. The number of inclusive, exclusive, and conflicting characters are compared and, using the most parsimonious pathway (the one with the least number of character assignments to satisfy the matrix) a phylogenetic tree is constructed from the common ancestor(s).
While this strategy worked well across all problem types and was used by all participants, it requires a well developed pattern of recognition skills and substantial working memory (Brewer, 1996).

Another strategy, although much less commonly employed, was the use of duplicated taxa. This strategy involves drawing a separate tree for each of the taxa and placing each newly considered character on one of more of the several trees until the character have been mapped. Similarly distributed characters are then joined making a combination tree. This process continues until all of the characters have been placed on a combination tree or have been left off and placed on an incompatible tree. The two resulting trees, both of compatible characters and incompatible characters, are resolved into one tree that is the most parsimonious (has the least number of steps or character assignments to adequately explain the matrix.). The advantage of using the duplicated taxa strategy is that it frees working memory and each character decision is immediately reflected in the diagram. The weakness of this strategy is that it involves making decisions without considering the order in which the characters are combined leading to solutions that are not parsimonious (Brewer, 1996).

The last strategy, again not commonly used, was the use of order of divergence. This strategy involves counting the number of apomorphies (derived) characters for each taxa across the entire matrix. Once this is accomplished, the solver orders the taxa by placing the taxa with the fewest apomorphies at the bottom and the taxa with the most apomorphies at the top. Identical and similar taxa are
recognized and grouped together and the tree is subsequently constructed. (The order of the taxa in the matrix sometimes serves as an important heuristic for causing side branches to be constructed from the main lineage.) The strength of this strategy lies in the fact that reorganization of the matrix uses features of display-based reasoning to reduce the need for working memory (Brewer, 1996). This method (strategy) also frees working memory space. The weakness of this strategy lies in the fact that, with more difficult problems, the number of apomorphies does not strongly correlate with the order of divergence. This is especially true when homoplasy is involved. (Homoplasy being the placement of a character in more than one position on a phylogenetic tree.) In addition, this strategy does not lend itself to the finding of alternate topologies for the same matrix.

Brewer’s (1996) study confirms the findings of previous studies on the problem-solving activities of experts. Experts make extensive use of working memory to represent and organize information. They also possess strategies and heuristics that aid them in achieving a solution more efficiently. Chase and Simon (1973) attributed problem-solution success by experts to be the result of holding data in memory as “chunks”. This chunking activity allows for more efficient data processing that would enhance the consideration of each individual piece of data. The idea of chunking allows for a smoother employment of the inclusion/exclusion strategy where each character must be evaluated and order into a matrix. This, of course, means a more efficient ordering of taxa into a drawing field and the subsequent construction of the tree. Larkin (1983), in comparing experts and novices, found novices
to be focused on objects in the real world whereas experts combined real world objects with abstract ideas such as force and momentum. Phylogenetic trees contain elements of both the real world (fossil and extant taxa) and abstract ideas such as common ancestors and character transitions. While studies of novice construction of phylogenetic trees is lacking (until this study), it is assumed that novices would simply link taxa together without postulating a common ancestor or character transitions. Stewart and Dale (1989) described a series of general problem-solving heuristics and advanced the argument that these become more powerful when utilized in domain specific problem-solving strategies. Reif (1983) described three categories of strategic problem-solving strategies: (1) data redescription, (2) solution synthesis, and (3) solution assessment. Experts solvers used all three. As examples, data redescription may be considered as the use of inclusion/exclusion strategy to generate inclusion/exclusion character groups, solution synthesis may be considered as the translation of groups into taxa and the subsequent construction of a tree, and solution assessment may be considered the use of parsimony to decide on a single topology when multiple topologies are generated. At any rate, Brewer (1996) found that experts employed memory aids such as chunking as well as general and domain specific heuristics to generate phylogenetic trees from encoded data.

Brewer’s (1996) study used research problems that were constructed with the idea of studying the variability of behaviors as phylogenetic tree construction becomes more complex. Complexity means an increase in both the number of taxa and the number of characters assigned. Brewer (1996) found that all expert
participants used the same pattern strategies across all problem types. Experts had no difficulty determining the most parsimonious tree solution but, after they had found it, they did not look for multiple topologies or other tree possibilities. Errors occurred when experts did one or more of four things: (1) did not find all of the most parsimonious trees, (2) did not find a single parsimonious tree (did not solve the problem), (3) accepted a less parsimonious tree as the most parsimonious tree, and (4) had an incompatibility between the tree and the data. The errors, while varying by the participant in question, are pointers to anyone planning instruction in the construction of phylogenetic trees.

Brewer (1996) presents a Model of Desired Performance for Phylogenetic Tree Construction that includes eight steps for the proper solution to phylogenetic tree problems. These eight steps include: first, the organization of characters on the matrix; second, organize the taxa into the drawing field; third, postulate an ancestor for each character group; fourth, link the ancestors to the terminal taxa; fifth, distribute homoplasious characters; sixth, consider the order of homoplasy starting with the character that requires the most steps; seventh, consider alternate tree topologies; and eighth, consider alternate arrangements (convergence or reversal) for each homoplasious character. This model is the guide that educators should be aware of in order to construct curricula designed to give instruction in phylogenetic tree construction.

Brewer (1996) ends his study by indicating that it would be useful to have a model of novice phylogenetic tree construction in order to compare both novice and expert performance in phylogenetic tree construction.
Recapitulation on Problem-Solving

This section begins with a general procedure for solving problems. All of the research indicates that experts and novices solve problems very differently. This is due to several factors, the greatest of these being experience. Because experts bring a broader knowledge base and range of experience into the problem setting, they are better prepared to achieve a correct solution. Experts are able to redescribe a problem in terms of variables. They plan a forward approach by considering how each of these variables influences the problem and then proceed to eliminate or solve for each variable, in a systematic way, in order to reach a final goal. Novices, lacking both a broad knowledge base and experience, tend to look at a problem and try anything at their disposal to solve it. They do not plan their approach but, rather, do the problem in a piecemeal fashion, trying first this and then that. This lack of a planned solution causes them to make many errors and often contributes to both frustration and an incorrect solution.

Generally, two types of problems have been used in problem-solving research. The first problem type embodies those problems that have a simple solution through the employment of a formula or algorithm. These problems offer little challenge to the solver other than the guarantee of a correct solution if the algorithm is properly executed. The second problem type is not solved by a formula and, in addition, may not have a single pathway of solution. This problem type lends itself to the application of heuristics for making decisions on how to proceed. These heuristics often involve: problem redescription, the division of the solution into several
sub components and solving for each one separately, or the application of an analogy from a related domain. At any rate, heuristics apply to problem-solving across domains and are often employed by experts when solving problems.

Problem-solving in the sciences has most frequently been studied in the areas of physics and genetics. In such studies, expert and novice solution performance has been compared. Research in both of these domains confirms the expert-novice difference in solving problems. It further reveals that when novices are given instruction in solving a given problem type, their success rate in achieving a correct solution is greatly increased.

Teaching and Learning Evolution

Introduction

The purpose of this section is to inform the reader of the types of instruction that have occurred in the teaching of evolution during the past ten years. Accordingly there are two sections: (1) conceptual change studies, and (2) learning evolution other than by conceptual change.

The teaching of evolution, as is true of many other domains, has been influenced by the work of Posner, Strike, Hewson, and Gertzog (1982). Posner et al. (1982) introduced a new theory or approach to instruction that was applicable to any subject matter domain. In essence Posner’s method involved three important ideas: (1) measure a student’s preinstructional concepts about a given domain, (2) design instruction that would address any wrongly held conceptions in such a way that the
student would reject his or her wrongly held conceptions and accept the correct conceptions, and (3) retest the student to make sure that the student's current conceptions agree with the accepted conceptions. These three activities have come to be known as conceptual change and the teaching of evolution during the 90's was greatly influenced by this idea. This presented a new way of looking at evolution other than the presentation of time lines, fossil remains, and Darwinian principles. Instructors (researchers) who practiced conceptual change in their teaching found that about half of their students responded to conceptual change (gave up alternate concepts for the scientifically-accepted concepts) and half did so to varying or lesser degrees.

In 1996, further research revealed that there are several factors that determine whether or not a student will successfully learn evolutionary principles. All of these factors have to be considered. These researchers also concluded that the teaching of evolution is absolutely vital and should take place whether or not a student responds to wholesale or partial conceptual change. Evolution, they feel, is the common thread that binds together all the diverse organisms that are observed in the biological world. They believe that it is the only scientific theory that explains the world's diversity.

Conceptual Change Studies

The Bishop and Anderson study (1990) is one of the first research studies to study conceptual frameworks of students in the domain of evolution. This study designed instructional material to address the alternate conceptions of students and then tested the effectiveness of these materials. The study involved college students.
in a general biology course who were pretested using an exam with both open-ended and multiple-choice questions. The students were asked about their beliefs in the domain of evolution as well as the extent of any prior course work in biology. The students were then involved with instruction in natural selection that had been designed from earlier studies about a student's prescientific conceptions concerning the evolutionary phenomena. The instruction was based on the theory of conceptual change and was designed to allow students to confront their evolutionary misconceptions in order to achieve a more accepted scientific understanding. After instruction, students were post-tested to assess their degree of conceptual change. Researchers identified three student misconceptions of natural selection. The first misconception concerned the origin and survival of new traits in populations. Students did not recognize the processes of random genetic variation and natural selection operating on that variation in order to perpetuate it or destroy it. Students, instead, believed that only one process operated in bringing about species change: the environment. Students saw the environment as creating a need which influences variation through use, disuse, and adaptation. According to Bishop and Anderson (1990), the main stumbling block to proper understanding was the inability to distinguish between the origin of a trait and the subsequent selection of that trait. The second misconception of students was with the role of species variation within a population. Students placed little importance on the role of variation within members of a population. Students rather view evolution as a change in a trait within a homogeneous population and not as change within a heterogeneous population. The third student
misconception was concerned with the target of evolution. Students saw evolution as a gradual change in the traits themselves and not as an increase or decrease in the number of individuals in a population with such a trait.

The fact that most of the participants in the Bishop and Anderson study (1990) had completed one year of high school biology prior to their college course, made little difference in their prescientific conceptions of natural selection. The study shows that college students have a poor understanding of how change occurs in a population, the role of variation in a population, or the idea of evolution changing populations. After instruction, more than half of the students understood these ideas. Bishop and Anderson (1990) conclude that natural selection is far more difficult to understand than most educators realize and that students can change their conceptions only when instructors identify student’s alternative conceptions and are prepared to confront them.

Evolutionary Misconceptions

Several studies have identified misconceptions concerning evolution: Renner, Brumby and Sheperd (1981), Jungwirth (1975), Clough and Wood-Robinson (1985), Deadman and Kelly (1978), Hallden (1988), and Greene (1990). Many of these studies found that students believe that adaptation is need driven in response to environmental change. They have difficulty believing that change is random or that it just happens. There has to be a reason for it. Further, they believe that evolution is always focused toward some grand end. Everything must be teleologically useful
and further, more often than not, students believe that evolution is anthropomorphic in that its end is to develop human qualities in nonhuman things. The idea is that there is a spiritual direction to evolution that culminates in the crowning of both the human form and behavior. Man is at the top and the progression to the pinnacle of man is what evolution is all about. Students also have trouble dealing with the fact that evolution occurs in populations and not within individuals. The idea that evolution does not occur within the lifetime of a single individual is difficult for many to see. Studies have also found that when new conceptual frameworks of evolution are formed, students tend to combine them with their earlier misconceptions rather than give them up altogether (Hallden, 1988). The conclusion of many researchers is that Lamarkian ideas of “need” created by environmental change is the main block to learning Darwinian explanations. There are also some educators who feel that the concept of evolution is too abstract to be taught to the concrete thinkers that fill most high-school classrooms. However, there are studies that say otherwise and conclude that the central ideas of evolution are attainable by high-school students.

Learning Evolution Other Than by Conceptual Change

Demastes, Good, and Peebles (1996) conducted research where the central question was how do students come to know evolution? Their research involved high school students in a university laboratory school at Louisiana State University. These students were in their second year of studying biology and were exposed to a ten-day instructional unit on evolution although the ideas of evolution had been liberally
sprinkled throughout the course. Using daily observations, pretests and posttests, and journal entries as data, students were interviewed and asked to answer five questions with accompanying diagrams. (The questions were both open- and close-ended.) They found that the Conceptual Change Model (CCM) of Posner et al. (1982) was not the model of all learning. The intent of CCM is to model wholesale change of major, paradigmatic conceptions. This study indicates that holistic change is not the only pattern to be expected in a learner’s conceptual restructuring. The presence of incremental changes implies that the learner is subtly changing major conceptions. Thus, there are no competing conceptions or abrupt cognitive disharmony like the CCM model describes. Learning is viewed as a much more gradual and less dualistic event. Thus, one should measure learning only in terms of complete conceptual change. Some learning pathways may follow CCM while other learning pathways may not (Demastes et al., 1996). These researchers determined that the conceptual ecology for understanding evolution has six facets: (1) prior conceptions related to evolution, both alternative and scientific; (2) the degree to which the participant has organized his/her life around scientific activities (understood the natural world through physical causation; (3) the view of the nature of scientific knowledge (realists, relativist, pragmatist, etc.); (4) the view of the biological world (one of competition, causes, aesthetics, or order?); (5) the degree to which the participant has organized his or her life around religious beliefs; and (6) the acceptance of evolutionary theory as closely tied to religious beliefs. Posner et al. (1982) speak of an interactive complex that exists between conceptual ecologies and associated instances
of conceptual change. Desmastes, Good, and Peebles (1995) find that this is difficult
to tease out but that several patterns emerge. The patterns are as follows:

1. Conceptual change can occur in the absence of corresponding belief (A
learner who personally rejects the truthfulness of evolution can experience consider­
able change toward a scientific conceptual framework).

2. There is no statistical association between the stated belief of students and
their ability to apply a scientific concept for an issue of evolutionary theory.

3. Conceptual change is not wholesale but can include instances when two
competing theories are held and applied.

4. Students who are not mechanistic in their approach to natural phenomena
and apply the creationist understanding of evolutionary phenomena have the most
difficulty in constructing a conceptual framework for evolution.

The net result of Desmastes et al.’s (1995) study is that evolution can and should be
taught. Even though there are obstacles to the teaching of evolution, the concept is
important enough to the biological domain that all students should be exposed to it.
Whether or not a complete conceptual change occurs as Posner et al. (1982)
advocate, the effort is not lost on any student because he or she can accept a more
limited view than total conceptual change. Evolution is a way to tie together the
diversity observed in the biological world. That fact, in combination with DNA and
its four nucleotides makes the biological world more manageable than it appears to
be.
CHAPTER III

METHODOLOGY

Introduction

The purpose of this research is to describe the strategies used by novices to construct phylogenetic trees from coded data placed in a matrix. This chapter describes the novice research population and how they were selected, a description of the research site, the nature of the pre-research instruction, the nature of the research problems, data gathering, and data analysis.

The bulk of this chapter is concerned with the research problems themselves. There are five research problems and each one follows the pattern of Steve Brewer’s (1996) model problems. Each of these research problems is described below along with a correct solution or solutions.

Description of the Research

The purpose of this research project is to determine how novices, high school students, construct phylogenetic trees and to, subsequently, create a model of novice phylogenetic tree construction. The research was conducted at Lee High School, Wyoming, Michigan where the researcher is a faculty member. The research involved two sophomore level biology classes and it was conducted over a two-month interval that began on April 1, 1997 and concluded May 30, 1997. After the
recruitment of potential participants through an announcement in the two biology
classes, a two week instructional unit on phylogenetic tree construction was planned
and delivered. This instruction did not involve computers or the program *Phylo-
genetic Investigator*, but was concerned with the analysis of trees that were already
constructed. It was only on the last day of instruction that the computer and the pro-
gram *Phylogenetic Investigator* were introduced. Following the instructional unit,
thirteen randomly chosen participants were selected to solve five research problems
involving the construction of phylogenetic trees. The research problems were solved
while each participant met one on one with the researcher either before or after
school. As students solved the problems, a tape recorder was used to record think-
aloud protocols of what each participant was thinking as he or she solved the research
problems. The problems were solved on a Macintosh computer using the program
*Phylogenetic Investigator*. At the conclusion of the problem-solving sessions, the
protocols were transcribed and analyzed along with the computer printouts and any
notes or drawings that the participant may have made. These data, protocols, trees,
notes, and drawings, were used to create a model of novice phylogenetic tree con-
struction. The participants were allowed to solve no more than three research prob-
lems at a session and the number of sessions needed for each participant to solve all
five problems ranged from two to five sessions. Three sessions, the third often being
shorter, were the average time spent. During the one-on-one sessions the researcher
did not answer any questions except to clarify the operation of the computer program.
The researcher’s task, during each session, was to keep the participant talking as well
as making sure the computer program operated correctly.

Selection of Participants

The researcher visited two biology classes, prior to instruction, in order to recruit participants. During the visitation, the nature of the research, the instructional unit, the one-on-one protocols, and the time commitment were discussed. Students who expressed an interest in participation had consent letters mailed to their parents. (These students are minors and range between 15 and 17 years of age.) On the basis of the returned and signed consent letters, 15 participants were randomly chosen for the research project. It was further determined that each participant should individually sign an additional consent form indicating their willingness to participate. All forms were signed and returned prior to any contact with a research participant. All participants were volunteers and none received any reward other than the thanks of a grateful researcher.

Description of the Research Site

Lee High School is a small high school located in the metropolitan Grand Rapids area. Grand Rapids is the second largest city in Michigan and reflects, as a whole, strong middle-class values. The school is unique in that it is the center of the surrounding community with many of the student's parents having attended it as well. While there is a strong family tradition for some students, many come from broken homes where only one parent is present on a regular basis. (That parent is
often the mother.) While the school district, at one time, reflected the Dutch reformed religious tradition, there has been a distinct movement away from it as children grew up and new families moved in. Many of these new families have a different religious affiliation or even none at all.

There have been several attempts to consolidate this school district with one of the larger surrounding districts but, each time this has been proposed, the community votes it down being willing to raise the tax dollars necessary to keep “their” school district in operation. The school, at one time, was predominately Caucasian but now has a 35% minority composed of African Americans, Hispanics, and Asians. About 25% of the student body goes on to higher education while the remainder of the student body considers a high school diploma the termination of their formal education.

Nature of Instruction

The instruction lasted for a two-week period and extended from April 14, 1997 to April 25, 1997. The instruction began by measuring the preconceptions of evolution by the creation of evolutionary histories of 12 selected Caminacules or computer generated organisms (Sokal, 1983). Students were asked to construct evolutionary histories that involved structure alone and also histories involving the same organisms that had been assigned an arbitrary age. The idea was to measure how students conceive the evolutionary history of the organisms to have taken place. These evolutionary histories were created by the students on the initial day of instruction.
 Following the initial day, all instruction was designed to emphasize cause to effect problems. That is to say, students were instructed in what a phylogenetic tree represents, what characteristics are used in the construction of it, how to select the best tree from several competing trees, incidences of homoplasy involving both the processes of convergence and reversal, and, during the second week, instruction in more complex trees involving the problem models as established in the research study of Steven Brewer (1996). (These model problems represent the simplest problems that the experts in Brewer's 1996 study were asked to solve.)

In all cases the instruction involved going from the phylogenetic tree to the data matrix putting each problem in the cause to effect category. During the first week, the data matrix had actual characteristic names placed on it rather than the symbols used to stand for a character state. The character states were either ancestral (present in the ancestor) or derived (not present in the ancestor). The students were told the difference between the two categories (ancestral and derived) and how to recognize them. During the second week, the tree not only became more complex but the data matrix began to use the encoded data of “0” for ancestral and “1” for derived character states. Students were given homework assignments as well as practice in class. None of the research problems, with their associated species and character states, were covered in class. (In other words, problems involving blood flukes, sea gulls, quails, and mustard plants were not part of the instruction). The fifth research problem was one that the researcher constructed. It involved three extinct (fossil) species and three extant (living) species. A new twist to this problem was the
fact that the fossil species could not be moved vertically on the computer model of the tree. There was only a possibility of horizontal movement. This meant that the fossil species would actually be part of the branching tree and not just at the base of the tree as in problems 1-4.

There was no contact with the computer and the program *Phylogenetic Investigator* until the last day when the instructor demonstrated the use of the program to the entire class. (None of the student in these classes actually worked the program; they only observed how it was operated.) The computer demonstration involved common activities such as moving taxa, adding nodes, adding links, and adding character states along the links. There was also instruction in removing nodes, removing links, and removing character states as well as how to show homoplasy by the methods of convergence and reversal. The period of instruction in the operation of *Phylogenetic Investigator* was held on Friday, April 22, 1997. The first one-on-one session with a research participant began on Monday, April 25, 1997.

Nature of the Research Problems

There were five research problems. Each of these problems was presented to the participant in three ways. First, there were diagrams of the species in question showing the structural characteristics or behaviors that were being used. Second, there was a list telling the subject which of these character states were ancestral and which were derived. Third, there was a data matrix encoded with “0’s” and “1’s” to denote which of the taxa possessed each of the character states. Using these three
pieces of information and the computer program *Phylogenetic Investigator*, each participant constructed a phylogenetic tree for each problem. The research problems, except for problem 5, were designed using Steven Brewer's (1996) model problems. These model problems were selected because they deal with a small enough number of taxa and character states so that the resulting trees do not exceed the level of problem-solving complexity typically expected from high school students. While the goal was not a scientifically accepted solution for every research problem, it was for students to be able to construct some type of solution whether or not it was correct.

**Research Problem Number 1**

Research problem 1 follows Steven Brewer's (1996) model problem 1.1 and involves blood flukes. There is only one correct solution and every derived character state appears on the phylogenetic tree in a single location. (In other words, there is no homoplasy.) In looking at the matrix, it is observed that character number one is present in the derived state for all of the taxa except the common ancestor (whole group synapomorphy). Further, character number five is nested within two identical characters, numbers three and four, and character number two is exclusive from characters one, three, four, and five. (In other words, character number two has a different distribution than the other four character states). The phylogenetic tree and data matrix for research problem number one appears in Figure 3. R05 and R11 form a group, R09 and R13 form a group and R07 is an outgroup with a single unique character state. Character 1 is placed at the base of the tree because all of the taxa possess
66

Figure 3. Problem 1, Blood Flukes, Matrix.

it (whole group synapomorphy). Derived character state number 2 is placed along the link PA-PB because the taxa group R05 and R11 is the only one to possess it. Derived character states 3 and 4 are placed on the link PA-PC because the two taxa groups above it share those character states. Derive character state 5 is placed on the link PC-PD because the taxa group R09 and R13 is the only group to possess it. There is only one problem solution (see Figure 4) and the problem is solved in 5 steps or character assignments.

Research Problem Number 2

Research problem number 2 is concerned with the behavior of sea gulls and follows Steven Brewer’s (1996) research problem 1.2. This problem has one solution
Figure 4. Problem 1, Blood Flukes, Single Solution.

but, since it involves homoplasy, there are two equally parsimonious solutions, one involving the convergence option and one involving the reversal option. In looking at the matrix for research problem 2, it may be observed that character states 1 and 2 have an identical distribution. Character state 4 is nested in identical character states 1 and 2 while character state 5 has a distribution unlike any of the other 4 character states. Character state 3 is homoplasious and may be mapped onto the tree as a convergence or a reversal. Figure 5 illustrates the matrix for problem 2, and Figures 6 and 7 show the convergence and reversal options respectively. There are six steps to
Figure 5. Problem 2. Gull Behavior, Matrix.

the solution of this problem.

In solution 1 a single group is formed by taxa R1 and R3, another group is formed by taxa R4 and R5 and a single taxon, R2, forms the outgroup. In the convergence option, solution 1, character states 3 and 5 are placed between F73 and PA. Character states 1 and 2 are placed between F73 and PB. Character state 4 is placed between PB and PC. Character state 3, the homoplasious character is placed on the link PB and R2. By being placed on the tree a second time, the method of convergence is used.

In solution 2, the reversal option, the taxa groups remain the same but the character states are placed on the tree in different positions. Character state 3 is placed at the bottom of the tree between F73 and PA. Character state 5 is placed on
Figure 6. Problem 2, Gulls Behavior, Convergence Option.

the link PA-PB. Characters 1 and 2 are placed on the link PA-PC to show that all of
the taxa above them possess those character states. Character state 4 is placed on the
link PC-PD while character state 3 is placed on the link PC-PD as a reversal because,
at this point, it has reverted back to the ancestral state as denoted by 3 1>0.

Research Problem Number 3

Research problem number 3, quails (Figures 8, 9, and 10), has no whole
group synapomorphy. Characters 3 and 4 are identical and exclusive from character
Figure 7. Problem 2, Gulls Behavior, Reversal Option.

2. Characters 1 and 4 conflict with each other and this makes for two equally parsimonious solutions. In solution 1 (Figure 9), character 1 defines the taxa group R10 and R06 and in solution 2 (Figure 10), character 4 defines the taxa group R10 and R08. In each solution the conflicting character can be visualized as either two gains or a gain and a loss. The two most parsimonious solutions require 6 steps.

Research Problem Number 4

Research problem number 4 is concerned with mustard plants and there are
three solutions, each with seven steps. The three solutions are given in Figures 11-14. There is no whole group synapomorphy. Characters 3 and 5 are identical and exclusive from character 2. Character 4 conflicts with character 1. Character 2 conflicts with all other characters. Three topologies are possible. In the first topology, solution 1 (Figure 12), homoplasy is shown in both characters 1 and 4. Character 4 is present in R11 and the R17 and R19 group. Character 1 is present in R15 and R19. Both homoplasious characters, 1 and 4 respectively, are assigned by the method of convergence.

In the second topology, solution 2 (Figure 13), characters 2 and 4 are homoplasious but are assigned using two different methods. Character 2 is assigned by the method of convergence and is present in both R11 and R13. Homoplasious character
Figure 9. Problem 3, Quail, Solution 1.

4 is assigned by the method of reversal and is found in R11, R17, and R19, but is lost in R15. In this topology R13, R11, and R17 all form separate taxa groups while R19 and R15 are grouped together.

In the third topology, solution 3 (Figure 14), character 4 is homoplasious and appears in three different locations on the tree. Character 4 is assigned by the method of convergence. Character 4 is present in R11, R17, and R19. With the grouping of taxa in this topology, character 4 must appear on the tree in three different locations. All three topologies are equally parsimonious and there are seven steps to each
Figure 10. Problem 3, Quail, Solution 2.

solution.

Research Problem Number 5

Research problem number 5, caminacules (Figures 15, 16 and 17), is unique in that it displays more than one fossil taxa. All of the other problems have a single fossil taxon. The three fossil taxa are F73, F40, and F09 respectively. These fossil taxa actually become a part of the tree and are not merely found at the base of the tree as a single fossil taxon would be. The problem is done with a set of computer
generated animals known as caminacules. The animal figures have different structural characteristics such as a neck or no neck, feet or no feet etc. All of these characters are assigned as ancestral characters, denoted as a “0” on the problem matrix, or as derived characters, denoted as a “1” on the problem matrix. Characters 1, 3, 2, and 4 are rather unproblematic and are assigned to the link F73-PA and the link PA-PB respectively. However, character 4 is homoplasious and may be assigned by the method of reversal in solution 1 or by the method of convergence in solution 2. Character 5 is also unproblematic and is assigned to all of the recent taxa groups, namely R11, R13, and R15. There are six steps to each of the two most parsimonious solutions.

Figure 11. Problem 4, Mustards, Matrix.
Data Analysis

The three types of research data are collected from each participant, namely transcribed think-aloud protocols, computer printouts of the phylogenetic trees, and any notes or drawings made by the participant, were analyzed both problem by problem and participant by participant. Flow charts of tree construction activity were made for all thirteen participants across all five of the research problems. Once all of the individual flow charts were constructed there was an analysis made of all five
problem activities to form a composite model for each student. (The five separate flow charts were constructed into a single flow chart for each participant.) Thirteen summary flow charts were made in all. From these thirteen composite models an overall model was constructed. The construction of this model was an important step and several overall models were made until there was a single model that could explain the actions that any participant took in solving the five research problems. The data from participant number 7 were not included in the analysis because this participant did not solve all of the problems.
Figure 14. Problem 4, Mustards, Solution 3.
<table>
<thead>
<tr>
<th>Taxa</th>
<th>Characters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20</td>
</tr>
<tr>
<td>R11</td>
<td>1 1 1 0 1</td>
</tr>
<tr>
<td>R13</td>
<td>1 1 1 0 1</td>
</tr>
<tr>
<td>R15</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>F09</td>
<td>1 1 1 1 0</td>
</tr>
<tr>
<td>F40</td>
<td>1 0 1 0 0</td>
</tr>
<tr>
<td>F73</td>
<td>0 0 0 0 0</td>
</tr>
</tbody>
</table>

Problem: 5, Caminacules, Solution 1

Figure 15. Problem 5, Caminacules, Matrix.
Figure 16. Problem 5, Caminacules, Solution 1.
Figure 17. Problem 5, Caminacules, Solution 2.
CHAPTER IV

RESULTS

Introduction

This study used a computer program called *Phylogenetic Investigator* to determine the courses of action taken by novices in the construction of phylogenetic trees. Phylogenetic tree construction may be viewed as a series of transitions between a data matrix and the tree itself. The transitions are the courses of action taken in order to change the encoded data matrix of taxa and character states into a branching tree that explains the same data in diagrammatic form. Participants always created a single tree for each problem and the actions of each participant conform to a flowchart of procedures known as an "Overall Model of Novice Phylogenetic Tree Construction". This chapter is concerned with an explanation of that model and how participants employed it in the solution of phylogenetic tree construction problems. The data presented are organized by problem with a section devoted to each of the five research problems and an analysis of both their correct and incorrect solutions.

Problem references are given in Arabic numerals P1, P2, P3, P4, and P5 with each number representing a unique type of problem. Each problem, except P5, is composed of five living taxa and one fossil taxon. Taxa that are prefixed by the letter "R" are recent or living taxa and are denoted by the numbers R1-R19. Taxa that are
prefixed by the letter "F" are fossil taxa and are denoted by the numbers F70-F75. (F21-F69 have been omitted because, in the initial setup, the program, Phylogenetic Investigator, would not place the fossil taxa low enough on the screen to make the problem workable). Characters are single digit numbers, C1-C5, and are used to represent physical or behavioral characteristics of the taxa involved. Participants are referred to with a capital letter "S" (for student) followed by a numeral (e.g., S8). The numerals range from 1-13 with the numeral indicating the order in which the student appeared for his or her first problem solution session. There were 12 actual participants because S7 solved only one of the problems and was, therefore, omitted from the study. Each of these 12 participants solved five research problems by changing the encoded matrix into a phylogenetic tree. Thus, there were 60 trees created in all.

The purpose of this chapter is to both report and analyze the results of novice phylogenetic tree construction. To that end this chapter begins with the components of good phylogenetic tree construction problem solution together with the overall descriptive results of this study presented in two tables. This is followed by a presentation of the Overall Model of Novice Phylogenetic Tree Construction and an explanation of each of the four major areas of the model: (1) matrix analysis, (2) build tree topology, (3) assign characters, and (4) checking. The table which follows explains, in terms of computer moves and verbal dialogue, how S8 solved research problem number 3. This is presented in order to demonstrate what the steps or courses of action taken in a typical problem solution might look like. Next is a detailed, by
Components of a Good Solution

A good solution should include the proper grouping of taxa, an adequate number of links and nodes to properly connect the taxa, and the proper placement of characters along the nodes to express where the characters change their state between the fossil and the living taxa. The order of character assignment may vary but each character must be placed in such a way that it appears on as few links as possible in order to satisfy the matrix. This practice is called parsimony. For a good solution, all five characters must be parsimoniously assigned. After solving the problem, the solver should check to make sure that his or her solution correctly addresses the matrix in all respects. Further, a good solution will also seek to develop any alternative trees that
are possible for the same data matrix. There is often more than one solution to a
typical problem in phylogenetic tree construction.

Descriptive Statistics of the Results

The following two tables give the reader an overall picture of the results
obtained across all five of the research problems. Table 1 simply lists the results of
problem solution by each student. There are two categories represented by two verti­
cal columns, “correct solution” and “incorrect solution”. The student number is also
listed and the problems are categorized as P1, P2, P3, P4, and P5. A correct solution
is defined as properly grouped taxa, the proper number of nodes and links to map the
data matrix, and the most parsimonious placement of all of the characters. Incorrect
solutions indicate a violation of taxa grouping, improper placement of nodes and
links, or characters omitted, placed in a non-parsimonious fashion with the matrix
mapping or not mapping, or entirely wrongly placed. (Often, incorrect solutions will
have improper grouping of taxa as well as characters either wrongly or nonparsimoni­ously placed.)

Table 2 is presented below in order to give the reader an overall picture of the
results of this study by problem. The table is organized by problem number and pro­
vides data in two categories of problem solution: (1) correct solution; (2) incorrect
solution.
Table 1
Results of Problem Solution by Student

<table>
<thead>
<tr>
<th>Student</th>
<th>Correct Solutions</th>
<th>Incorrect Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>P1, P2, P3</td>
<td>P4ad, P5a</td>
</tr>
<tr>
<td>S2</td>
<td>P2, P3</td>
<td>P1, P4, P5d</td>
</tr>
<tr>
<td>S3</td>
<td>NONE</td>
<td>P1ac, P2, P3a, P4a, P5a</td>
</tr>
<tr>
<td>S4</td>
<td>NONE</td>
<td>P1, P2, P3ad, P4ac, P5a</td>
</tr>
<tr>
<td>S5</td>
<td>P1, P2, P3, P4</td>
<td>P5</td>
</tr>
<tr>
<td>S6</td>
<td>P4</td>
<td>P1d, P2ad, P3ad, P5a</td>
</tr>
<tr>
<td>S8</td>
<td>P1, P2, P3, P4, P5</td>
<td>NONE</td>
</tr>
<tr>
<td>S9</td>
<td>NONE</td>
<td>P1ac, P2, P3a, P4d, P5a</td>
</tr>
<tr>
<td>S10</td>
<td>P3, P4</td>
<td>P1d, P2a, P5a</td>
</tr>
<tr>
<td>S11</td>
<td>P4, P5</td>
<td>P1, P2, P3ad</td>
</tr>
<tr>
<td>S12</td>
<td>P2</td>
<td>P1c, P3a, P4, P5</td>
</tr>
<tr>
<td>S13</td>
<td>NONE</td>
<td>P1a, P2d, P3a, P4ad, P5</td>
</tr>
</tbody>
</table>

(a = improper grouping of taxa)
(c = character(s) omitted)
(d = matrix maps but parsimony violated)
(no letter = parsimony violated and matrix does not map)
(* S7 solved only one problem and does not appear on the table)

A Procedural Model for Novice Phylogenetic Tree Construction

After analyzing the final trees created as well as the verbal transcripts of each student solving all five problems, a procedural model of novice phylogenetic tree construction was constructed (Figure 18). The model explains the pathways that...
Table 2
Results of Problem Solution by Problem

<table>
<thead>
<tr>
<th>Problem</th>
<th>Students With Correct Solutions</th>
<th>Students With Incorrect Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>S1, S5, S8</td>
<td>S2, S3ac, S4, S6d, S9ac, S10d, S11, S12c, S13a</td>
</tr>
<tr>
<td>P2</td>
<td>S1, S2, S5, S8, S12</td>
<td>S3, S4, S6ad, S9, S10a, S11, S13d</td>
</tr>
<tr>
<td>P3</td>
<td>S1, S2, S5, S8, S10</td>
<td>S3a, S4ad, S6ad, S9a, S11ad, S12a, S13a</td>
</tr>
<tr>
<td>P4</td>
<td>S5, S6, S8, S10, S11</td>
<td>S3a, S4ac, S9d, S12, S13ad</td>
</tr>
<tr>
<td>P5</td>
<td>S8, S12</td>
<td>S2d, S3a, S4a, S6a, S9a, S10a</td>
</tr>
</tbody>
</table>

(a = improper grouping of taxa)  
(c = character(s) omitted)  
(d = matrix maps but parsimony violated)  
(no letter = parsimony violated and matrix does not map)  
(* S7 solved only one problem and does not appear on the table)

every student used to solve all five of the research problems. While there are several ways of solving the problems, there are three processes that all students performed. These processes, which are listed on the left side of the model, include: analyzing the matrix, building a tree topology, and assigning characters. In addition, some students added the fourth component of checking all steps in problem solution. If one follows the vertical axis through the center of the model, the most common pathway of solution is revealed. The most common pathway includes the following processes: looking at the matrix; looking at the taxa by character; identifying similar, identical, and different characteristics of taxa; arranging the taxa; adding nodes; adding links; build-
Figure 18. A Procedural Model of Novice Phylogenetic Tree Construction.
ing a tree topology; looking at characters; arranging characters by prior grouping of taxa; assigning homoplasious characters first or assigning homoplasious characters last; and considering parsimony in character assignment.

The boxes to the left and/or right of the center axis under the processes of matrix analysis and character assignment, represent the less common, although not necessarily unfruitful, pathways of solution. Thus, one can see that the chief variability lies in the areas of matrix analysis and character assignment. When analyzing the matrix in order to group the taxa, it is also possible to count the number of apomorphies either by character or by taxa. After considering the order of apomorphies, (number of derived characters denoted by 1’s on the matrix), the taxa are then grouped. When assigning characters, it is possible to do so by assigning the characters singly and in numerical order or by counting the group of character apomorphies. Characters are then assigned and the remainder of the model follows in order. The procedural model is a road map to explain how and in what order students solved the research problems. As such, it serves as a guide as to how students went about the task as well as a pointer as to which strategies were successful and which strategies were not successful. Since this is essentially a nonstatistical study, the model also serves as a way of both organizing and explaining the data.

Success of the Model Pathways/Strategies Chosen

**Matrix Analysis**

The variability within the model pathways lies in two areas: matrix analysis
and character assignment. In term of matrix analysis, redescription of the data occurred in 5 five instances and was done by only two of the participants, S2 and S6. Two of these redescribed matrix problems were correctly solved and three were incorrect. (Both S2, and S6 had one correct solution using this method of matrix analysis.) Counting the number of apomorphies and considering order, another strategy of S2 and S6, was used in two instances. Both of these instances turned out to be correct solutions of P1 and P4. By far the most common method of matrix analysis was looking at the taxa and determining which taxa possessed identical, similar, and different characteristics. There were 44 instances of matrix analysis using this strategy. In 10 of those instances, there were correct problem solutions. S1, S5, S8, S10, and S11 all achieved correct problem solutions with this method. The data indicates that success in phylogenetic tree construction does not favor any particular method of matrix analysis. Grouping taxa according to identical, similar, and different characteristics achieved the greatest number of correct solutions but it was also the method used to solve most of the problems. (It was used 44 times out of sixty problems solved.)

Character Assignment

There are three methods of character assignment shown in the model. The one used the least is the category of singly and in numerical order. In other words, the characters were assigned in order from 1-5 without considering any other factor. This was done in six instances, three of which were correct. S5, S6, S8, and S10
used this method. The second most commonly used method was the assignment of character or group apomorphies. This happened in 14 instances, 5 of which produced a correct solution. S1, S2, S5, S6, S8, S9, S10, S11, and S12 used this method one or more times. Correct solutions were produced by S1, S5, S8, and S11 using this method. The third method of character assignment involves placement of characters by the prior grouping of taxa. The use of assigning characters by the prior grouping of taxa was used in 40 instances with 12 of those instances resulting in a correct solution. Correct solution were achieved by S1, S2, S5, S6, S8, S10, and S12. Again, more successful solutions were achieved using the prior grouping of taxa but more students chose to use that method thereby resulting in a better success rate.

The data indicates that there is no one method any more successful than another when it comes to matrix analysis and character assignment. All of the methods achieved at least two or more correct solutions. While the majority of correct solutions were achieved by the central pathway of analyzing the matrix by looking at identical, similar, and different characteristics, and then assigning the characters by the prior grouping of taxa, it is also true that most of the participants chose this particular pathway through the model. The fact remains that students either knew how to solve the problems or did not know how to solve them. The method used does not make a difference in the number of correct solutions.

Explanation of the Major Sections of the Novice Model

The procedural model of novice phylogenetic tree construction (Figure 18)
has four major sections: (1) matrix analysis, (2) building a tree topology, (3) assigning characters, and (4) checking each of these sections corresponding to the activities that participants performed as they constructed phylogenetic trees. Each of the model sections will be analyzed and each analysis is preceded by a flowchart of activities depicting that particular section of the overall model. The specific problem addressed in this analysis is research problem number 3. Where there is variability in approach, as in matrix analysis and character assignment, all of the possible variations are discussed. The purpose of this analysis is not only to dissect the overall novice performance model, but to also set the tone for the presentation of an actual solution of research problem 3. This analysis involves participant S8 and provides not only a verbal dialogue but the accompanying computer moves and tree diagrams as well.

Matrix Analysis

While each problem presents a unique matrix composed of six taxa arranged horizontally and five character traits arranged vertically, the matrix for P3 has been selected as a typical example (see Table 3). The characters appear at the top and are denoted by the numbers 1-5. The taxa appear along the left side and are prefixed by the letter “R” for extant (living) and “F” for fossil, extinct (nonliving). The taxa numbers are randomly chosen with F74 being a higher number to make it appear lower on the screen when using the program Phylogenetic Investigator. The “1’s” indicate derived character states that will appear on the tree as 1 $\Rightarrow$ 1, 2 $\Rightarrow$ 1, 3 $\Rightarrow$ 1,
Table 3
Data Matrix for Problem #3

<table>
<thead>
<tr>
<th>Characters</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>R02</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>a</td>
<td>R04</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>x</td>
<td>R06</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>a</td>
<td>R08</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>R10</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>F74</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

etc. The “0’s” indicate ancestral character states that will not appear on the tree unless a form of homoplasy known as reversal is operating. When it does, the ancestral character states are shown on the tree as $1 \rightarrow 0, 2 \rightarrow 0, 3 \rightarrow 0$, etc.

Some students chose to redescribe the matrix by listing all of the taxa on paper and then writing every characteristic under that taxa. It was their way of reorganizing the data in a manner that is interpreted as being more congruent with their concrete way of thinking. Another participant, S5, chose to write all of the fossil characteristics at the bottom of the paper and then write, under each of the taxa, the words “absent” or “present”, as the case happened to be. Perhaps this was another way of making the data more concrete and thus more readily accessible to their thought process than the abstract “1’s” and “0’s”.

After considering the data matrix and looking at the characters, a participant
may group the taxa in one of two ways (see Figure 19). Firstly, he or she may look at
the taxa and identify similarities, differences, or identical character states among taxa.
In looking at the taxa shown in Table 3, it may be noted that R02 and R04 have iden-
tical character distributions while R06, R08, and R10 have different character distri-
butions. R06 and R10 have similar, except for character #4, while R08 and R10 have
similar character distributions except for character #1. Therefore, one might decide
to group R06 and R10 together while leaving R08 as a separate group or one might
decide to group R08 and R10 together and leave R06 as a separate group. (Either
arrangement will satisfy the matrix and both are equally correct.) F74, the common
ancestor, is placed at the bottom of the tree to show that all of the other taxa were
derived from it.

Secondly, he or she may look at the character columns and count the number

Figure 19. Matrix Analysis for Problem #3.
of apomorphies. Apomorphies indicate the derived state of a characteristic and are denoted by a “1” on the matrix. To count the number of apomorphies simply means to count the number of “1’s” in a column or row. Once this is done, the subject must consider the order of these derived characteristics. The student may say something like “R06, R08, and R10 have more derived characteristics than R02 and R04”. So they then group R02 and R04 together and either R06, R08, and R10 together, or R08 and R10 together with R06 by itself, or R06 and R10 together with R08 by itself. (R06 and R08 should not be grouped together because their distribution of characters is not at all the same.) This second approach can be problematic in that if one is careless in counting the “1’s”, some may be omitted or even duplicated yielding an incorrect tree for a given matrix. The count of apomorphies becomes less problematic if one is careful to check each of the character assignments on the created tree against the number of apomorphies in the matrix.

Building a Tree Topology

Tree building follows a similar pattern for all participants. This pattern consists of adding nodes, adding links, and then building the branching pattern of the tree (see Figure 20). The nodes, denoted in Phylogenetic Investigator by PA, PB, PC, PD, etc., represent speciation events which sometimes bring about a new character state that is not found in the ancestor. For example, in birds, if the ancestor had a gray breast and the recent taxa have a red breast, the red breast would be considered a derived or changed character that would be separated from the ancestor by one or
Figure 20. Tree Building Topology.

more nodes depending on the distribution of the other characters. Thus, nodes represent a convenient way of indicating a kind of fork in the path where one taxa group branches in one direction and another taxa group in the other direction.

The links are the lines that connect the nodes together. The number of links used depends on the number of taxa groups present in the problem. The importance of links is not the fact that they serve to connect a node with a single taxon or group of taxa, but that they represent the passage of time from one part of the tree to another and, as such, represent microevolution. It so happens that the program Phylogenetic Investigator places character transitions along these links. The program allows the solver to chose the link and it assigns the character to that link at a predetermined position. (The solver has no control of the location of a character along a link.) Once the links are in place with the associated nodes and taxa groups, the tree topology is constructed and the solver, having now constructed the tree, proceeds with character
assignment.

Without exception, participants always assigned a single node to each of the taxa groups. Whether there were one, two, or three taxa in a group made no difference. Each taxa group was connected to a single node. (None of the correct solutions have three taxa in a group although many of the incorrect solutions were grouped in that way.) Each taxa group was then connected to the common ancestor. Sometimes this was done directly and sometimes it was connected to another node first, that node being just above the common ancestor (see Figure 21). There was little variation in the construction of the tree topology. All of the variation was in the selection of taxa groups and in the assignment of characters to those groups.

Figure 21. Showing Connection, With and Without a Node to the Common Ancestor.
Character Assignment

Students used three separate methods for assigning characters (see Figure 22). The first and most common method is found in the center of the model. It involves assigning character states by the prior grouping of taxa. In other words, once the taxa have been grouped and placed at the top of the tree, the character states are assigned to each one accordingly. For example, in P3 (see solution of P3 in chapter 3) once the taxa were grouped as R02 and R04, R10 and R6, and R08 as a separate group, the appropriate characters are placed under each group. R02 and R04 share character #2 and it is placed on the link under them. R06 and R10 share character #1 and it is

Figure 22. Character Assignment.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
placed under them. R08 and R10 both share character #4 but it cannot be placed under each one because there is a conflict with character #1. Therefore, both R08 and R10 have character #4 placed under each one separately making character 4 homoplasious. Finally, characters #3 and #5 are shared by R08, R10, and R06 and so it is placed on a link under each of those three taxa. The assignment of characters is shown in Figure 23.

Once the characters have been matched to the taxa groups, it is necessary to remember that the fossil taxa, which for problem 3 is F74, goes at the base of the tree. Each taxa group is connected by a node to the common ancestor. R02 and R04 link directly to the common ancestor through a node known as PA. R08 links to the common ancestor through a node known as PB and R10 and R06 link to the common ancestor by a node known as PC. The characters are now placed along the appropriate links and the tree is constructed. There are two characters that are homoplasious

![Figure 23. Character Assignment for Problem 3.](image-url)
in this problem depending on how the taxa are grouped. If the taxa are grouped as
R02 and R04, R10 and R06, and R08 separately, the homoplasious character is char-
acter number 4. If the taxa are grouped as R02 and R04, R10 and R08, and R06 sep­
arately, the homoplasious character is number 1. (Homoplasy means that the charac-
ter appears more than once on the tree.) Both of the homoplasious characters are
placed on the tree using the convergence option. Figure 24 shows what the com-
pleted tree for problem 3.

There are two other alternative methods of character assignment. Neither of
these methods was widely used by the participants in this study. The first of these
alternative methods is to simply assign the characters singly and in numerical order
by character (e.g., 1, 2, 3, 4, 5). With this method any homoplasious character (one
appearing on the tree more than once) is handled randomly with no predetermined

Figure 24. Complete Tree Problem 3 Character 4 Convergent.
notion of its being homoplasious. A distinct advantage of this method of character assignment is that it prevents the omission of characters and gives an orderly way to proceed with character assignment. If there were problems with more than five characters to map, assigning characters singly and in numerical order would be a wise way to proceed.

The second alternative method is to assign characters by the number of character or group apomorphies. With this method the solver simply looks for the distribution of derived characters which is denoted by "1's" in the matrix. In using this method the solver is always comparing the derived characters to those of the common ancestor or fossil taxa. Again, homoplasious characters are randomly assigned. In P3, one would choose to assign characters #3 and #5 first since those characters appear as derived three times and at identical locations in the vertical columns of the matrix. Character #2 would be assigned next since it is distributed either vertically or horizontally in the top two taxa, R02 and R04 respectively. Character #1 would be assigned next in order because there are two apomorphies in a single taxa group, namely R10 and R06. Character #4 would be assigned last since it is the homoplasious character and must appear under both R10 and R08 which are both separate groups. The assigning of characters by counting the number of apomorphies has an advantage in that it prevents the assigning of ancestral characters to the links on the tree. Ancestral characters are assigned to the tree only in cases of homoplasy by the method of reversal. The counting of apomorphies also gives an order to character assignment with the solver organizing the apomorphies from the
greatest to the least number of distribution or vice versa. Characters are not easily omitted with this type of character assignment and often it is easier to see where incidences of homoplasy might occur.

Parsimony must also be considered when assigning characters. Parsimony has to do with limiting the number of character transitions on a tree to the smallest possible number that adequately explains the data matrix. It is assumed that nature always takes the most parsimonious route in the evolution of derived character states. Therefore, the most parsimonious trees or, the ones with the fewest character transitions explaining the matrix, are the correct problem solutions. The model also identifies categories of parsimony, both complete and partial. Complete parsimony indicates that all of the characters are correct with the fewest number of transitions to map the data matrix. In partial parsimony one of more of the characters are parsimoniously placed but the remaining characters are assigned without regard to the number of transitions and appear more often on the tree than the matrix dictates. Parsimony must be considered for each character state in any given problem.

Checking

Checking forms the last section of the model and is based on whether or not the participant verbally indicated that the problem was checked (see Figure 25). If no verbal reference was made to checking, it was assumed that the problem was not checked. Checking, when it was done, was indicated by verbally going over the characters under each taxa group or by the character itself with no reference to the
Advantages and Disadvantages of the Three Methods of Character Assignment

The advantage of assigning characters singly and in numerical order is that none of the characters will be skipped or omitted. It provides a systematic way of assigning each character without regard to the order of assigning the homoplasious character(s). The homoplasious character is handled in numerical order and there is no preoccupation to look for it. This method works equally well across all research problems and would be particularly useful in any problem that had a complex matrix involving ten, fifteen, twenty or more characters.

When assigning characters by the prior grouping of taxa, which many of the successful solvers did, you are dividing the matrix into a series of three or more subtrees, depending on the complexity of the matrix. The solver may choose to list the characters under each of the taxa groups or may even create a subtree for each taxa
group on which all of the characters for a given taxa group are placed. Using this method and checking back against the matrix, it is easier to see which characters are homoplasious (appear more than once on the tree) and which characters are not homoplasious (appear only once on a tree). This works particularly well for research problems 2, 3, and 4 where there are one or more alternative trees all of which involve homoplasious characters. To use this method is to subdivide the tree into its component parts and then, omitting duplicated characters where not needed, joining the parts together into a single tree. This method also helps to more easily detect cases of multiple homoplasy because you can quickly tell which groups are going to require a multiple placement of characters. Further, assignment of characters by prior grouping of taxa makes the achievement of the most parsimonious solution a much more likely event.

The advantage of counting the number of apomorphies provides another systematic way of viewing the problem and allows an easy separation of the ancestral from the derived characters. (This prevents the mapping of ancestral characters, something that two of the participants did.) Counting the number of apomorphies also helps to order the characters from those with the greatest number to those with the least number. Generally speaking, characters with the greatest number of apomorphies are placed near the base of the tree. Characters with the least number of apomorphies are often placed under a single taxon group. For example, in P1-character #1, which has the greatest number of apomorphies (also called whole group synapomorphy since all of the taxa have the derived character state) is positioned at
the bottom of the tree, right above the common ancestor. Characters #3 and #4, which have three synapomorphies each, are positioned on the link above character #1. The characters with the least number of apomorphies, #2 and #5, are placed under a single taxon group for character #2 and at the top of the tree, under another taxa group for character #5. Counting the apomorphies will also help find the most parsimonious solution by eliminating the duplication of characters on a given tree. If you order the characters from the greatest number of apomorphies to the least number or vice versa, homoplasious characters will be more easily detected.

The bottom line, however, is that it really does not make any difference which of the three methods of character assignment you use as long as you have a concern for and do not violate the rule of parsimony.

The focus will now by placed on S8 solving P3. The purpose of this section is to share with the reader what a typical problem solving session was like from the participant’s point of view.

Student 8 Solving Research Problem 3

A detailed analysis of S8 solving P3 is presented here because this participant consistently chose the most common, central pathway through the model of novice performance in phylogenetic tree construction. P3 represents relative sophistication in phylogenetic tree construction since it deals with homoplasious characters as well as two different types of taxa groupings, and therefore, serves as a good measure of the more detailed type of analysis needed for novice phylogenetic tree construction.
S8 was chosen because this participant was the only one to achieve a correct solution on all five of the research problems. This participant was also very verbal in problem solving and the verbal dialogue presented here is judged as typical of what a novice, who really understands phylogenetic tree construction, might say. The participant's actions are preceded by Table 4 which explains what the symbolic computer moves mean. Table 4 is then followed by Table 5 which has three columns. The first column of the table represents the computer operations with the time, operation, and location of the action. The second column gives the detailed verbalization of the student's think-aloud protocol. The third column explains the alignment of the activity with the actions listed in the model of novice performance.

The next section of this study is designed to report what actually happened in the solution of the problems. This section is organized by problem and begins with the correct solutions of those problems along with the participants who achieved the

Table 4

A Key to Program Operations

<table>
<thead>
<tr>
<th>Code</th>
<th>Operation</th>
</tr>
</thead>
<tbody>
<tr>
<td>MON</td>
<td>Move Node</td>
</tr>
<tr>
<td>ADN</td>
<td>Add Node</td>
</tr>
<tr>
<td>ADL</td>
<td>Add Link</td>
</tr>
<tr>
<td>REL</td>
<td>Remove Link</td>
</tr>
<tr>
<td>ACT</td>
<td>Add/Change Character</td>
</tr>
<tr>
<td>EXT</td>
<td>Exit Program</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
### Table 5

**Problem Three Solved by S8 Using the Overall Model of Novice Phylogenetic Tree Construction**

<table>
<thead>
<tr>
<th>Operation</th>
<th>Verbalization</th>
<th>Model Application</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&quot;Right now I’m evaluating drawings and the matrix&quot;</td>
<td>Look at Matrix</td>
</tr>
<tr>
<td></td>
<td>&quot;I notice that R2 and R4 have the same characteristics, so I’m going to put those two together at the top of the tree.&quot;</td>
<td>Look at Taxa by character</td>
</tr>
<tr>
<td></td>
<td>&quot;I’m going to connect R2 and R4.&quot;</td>
<td>Identify Similar Taxa</td>
</tr>
<tr>
<td></td>
<td>&quot;I’m going to see what characteristics R10 and R8 share. I note they’re relatively the same so I’m going to put those two together.&quot;</td>
<td>Arrange Taxa</td>
</tr>
</tbody>
</table>

6:18:46 MON R8
6:18:49 MON R10
Table 5—continued

<table>
<thead>
<tr>
<th>Operation</th>
<th>Verbalization</th>
<th>Model Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>6:19:03 MON F74</td>
<td>“I’m going to put F73 at the bottom because it’s the ancestor.”</td>
<td>Arrange Taxa</td>
</tr>
<tr>
<td>6:19:49 ADN PA</td>
<td>“I’m going to put some nodes on the screen and I’m going to connect them.”</td>
<td>Add Nodes</td>
</tr>
<tr>
<td>6:20:13 ADL PA R2</td>
<td></td>
<td>Add Links</td>
</tr>
<tr>
<td>6:20:21 ADL PA R4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6:21:52 ADN PB</td>
<td>“Right now I’m connecting the links to the node-R8 to the same mode.”</td>
<td></td>
</tr>
<tr>
<td>6:22:09 ANL PB R10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6:22:16 ADL PB R8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6:22:24 AND PC</td>
<td>“Right now I’m adding a node and I’m going to connect R6 to that node on the main stream of the tree.”</td>
<td>Add Nodes</td>
</tr>
<tr>
<td>6:22:43 ADL PC R6</td>
<td></td>
<td>Add Links</td>
</tr>
<tr>
<td>6:22:51 ADL PB PC</td>
<td>“Now I’m going to connect the nodes with the links.”</td>
<td></td>
</tr>
<tr>
<td>6:22:58 ADN PD</td>
<td>“Whoops—now I’m going to add a node and I’m just putting in the links.”</td>
<td>Add Nodes</td>
</tr>
<tr>
<td>6:23:13 ADL PA PD</td>
<td></td>
<td>Add Links</td>
</tr>
<tr>
<td>6:23:20 ADL PC PD</td>
<td></td>
<td>Build Tree</td>
</tr>
<tr>
<td>Operation</td>
<td>Verbalization</td>
<td>Model Application</td>
</tr>
<tr>
<td>----------------</td>
<td>------------------------------------------------------------------------------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>6:23:29 ADLF74 PD</td>
<td>“Now I’m trying to clean it up a bit so I can look more easily.”</td>
<td></td>
</tr>
<tr>
<td></td>
<td>“Now I’m looking at the matrix to see if there’s anything they all share and I’m not seeing anything.”</td>
<td>Look at Characters</td>
</tr>
<tr>
<td></td>
<td>“I’m going to go to R2 and R4 and I notice their derived characteristic spotted flanks and I notice #2 is. I’m putting that between PA and PD to show that it’s found in R2 and R4.”</td>
<td>Look at Characters by Prior Grouping</td>
</tr>
<tr>
<td>6:24:26 ACT PA PD 2</td>
<td>“Now I’m looking at R10 and R8 and I notice they are present in gray breast so I’m going to place gray breast on the link R10 and PB.”</td>
<td>Character 2 is added</td>
</tr>
<tr>
<td>6:25:04 ACT PB R10 1</td>
<td>Character 1 is added</td>
<td></td>
</tr>
</tbody>
</table>
Table 5—continued

<table>
<thead>
<tr>
<th>Operation</th>
<th>Verbalization</th>
<th>Model Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>6:25:47 ACT PC R6 1</td>
<td>&quot;Now I'm looking at R6 and I notice that is also present in gray breast so I'm going to place that on the link R6 and PC.&quot;</td>
<td></td>
</tr>
<tr>
<td>6:26:19 ACT PC PD 3</td>
<td>&quot;And I notice that R6's beak is small and I also notice that R10 and R8's beak is small so I'm going to put that on the link below all three of those specimens.&quot;</td>
<td>Character 3 is added</td>
</tr>
<tr>
<td>6:26:54 ACT PC PD 5</td>
<td>&quot;And I also notice that cream throat is also in R6, R8 and R10 so I'm going to put that on the same link as the beak.&quot;</td>
<td>Character 5 is added</td>
</tr>
<tr>
<td></td>
<td>&quot;Right now I'm wondering if I did something wrong—this tree is supposed to show reversal so I'm going to go over my matrix and see if I did everything right.&quot;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&quot;Right now I'm thinking if there's some way I can put characteristics for R6 and R10 on the main stem.&quot;</td>
<td>S8 is troubled by homoplasy</td>
</tr>
</tbody>
</table>
Table 5—continued

<table>
<thead>
<tr>
<th>Operation</th>
<th>Verbalization</th>
<th>Model Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>6:29:22 REL PC R6</td>
<td>“I’m going to erase the link that connects R6, R8, and R10 and PB to put characteristic #1 between PC and PD.”</td>
<td></td>
</tr>
<tr>
<td>6:29:37 REL PB R10</td>
<td>“And now I’m going to replace the link of R6 and PC and R10 and PB.”</td>
<td></td>
</tr>
<tr>
<td>6:29:49 ADL PC R6</td>
<td>“I’m going to put the derived characteristic for gray breast on the link connecting PC and PD.”</td>
<td></td>
</tr>
<tr>
<td>6:29:56 ADL PB R10</td>
<td>“This is my derived characteristic and I want to show it changed back to the ancestral characteristic.”</td>
<td>S8 recognizes homoplasy in character #1.</td>
</tr>
<tr>
<td>6:30:58 ACT PC PD 1</td>
<td>“I just showed the characteristic for gray breast was reversal.”</td>
<td></td>
</tr>
<tr>
<td></td>
<td>“Right now I’m wondering if I did that wrong and if I should switch R6 and R8 in their places.”</td>
<td>S8 troubled by placement of character #1.</td>
</tr>
</tbody>
</table>
Table 5—continued

<table>
<thead>
<tr>
<th>Operation</th>
<th>Verbalization</th>
<th>Model Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>6:35:20 REL PC PD</td>
<td>“I’m going to erase the link PC and PD which has characteristics 1, 3, and 5 on it.”</td>
<td>S8 removes the link with #1 character on it</td>
</tr>
<tr>
<td>6:35:35 ADL PC PD</td>
<td>“I notice that the bill for R8 is small so I’m putting that on the link PC PD and it’s also small for R6 and R10.”</td>
<td>S8 relinks PC and PD</td>
</tr>
<tr>
<td>6:35:45 ACT PC PD 3</td>
<td>“I also notice that for R8 and R10 red back is present so I’m going to put #4 on the same link as #3.”</td>
<td>S8 reassigns character #3</td>
</tr>
<tr>
<td>6:35:46 ACT PC PD 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6:35:48 ACT PC PD 5</td>
<td>“The throat being cream color is also the same in R6, R8, and R10 and I’m going to put that on the link connecting PC and PD.”</td>
<td>S8 assigns character #5</td>
</tr>
</tbody>
</table>
Table 5—continued

<table>
<thead>
<tr>
<th>Operation</th>
<th>Verbalization</th>
<th>Model Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>6:35:49 ACT PC PD 1</td>
<td>“Now I’m looking at R6 and R10 and I notice that they have a gray breast so I’m going to put that on the link PB and PD.” “I notice that R6’s back is not red but R10’s is, so I’m wondering if I need to show reversal with the back being red.”</td>
<td>S8 reassigns character #1 to a new link</td>
</tr>
<tr>
<td>6:43:08 ACT PB R6 4</td>
<td>“I’m going to place character #4 on the link R6 &amp; PB to show reversal.” “And I think my tree is all right and I’m going to go through and look over the whole thing.” “Right now I’m going to go back and look at the matrix to see if I made a mistake.” “I notice that R2 is absent in gray breast, present in spotted flanks, the bill is normal, the back is not red and the throat is cream—that reads. That also reads correct for R4.”</td>
<td>S8 makes character #4 homoplasious Checking S8 checks each of the taxa separately</td>
</tr>
</tbody>
</table>
Table 5—continued

<table>
<thead>
<tr>
<th>Operation</th>
<th>Verbalization</th>
<th>Model Application</th>
</tr>
</thead>
</table>
|           | "Right now I’m looking at R6 and I notice that gray breast is present, spotted flanks absent, the bill is small, red back is absent and the throat should be cream color—it is."
|           | "Now I’m looking at R8 and I notice the gray breast absent, it is—spotted flanks absent, it is—the beak should be small, red back should be present and cream throat should be present—it is."
|           | "And now for R10—gray breast should be present, spotted flanks absent—it is, the bill should be small—it is, red back and cream throat should be present—it is." |
| 6:54:02 EXT | "I’m going to look over one last time—I’m pretty sure this tree is right with reversal." |

S8 is sure characters are correct using reversal

End
correct solutions and what they did during the solution of the problem.

Correct Solutions Organized by Problem Type

Introduction

Correct solutions are characterized by three important criteria: the proper grouping of taxa; the correct tree topology of nodes and links; and the correct assignment of characters. The correct grouping of taxa and the placement of nodes and links are absolutely vital to the proper assignment of characters. In addition, the characters must be placed in the most parsimonious fashion. This means that the character must be placed in such a way that it appears in the fewest number of places to correctly map the matrix. If the character transitions need to be placed on the tree in more than one place, a condition defined as homoplasy, it is necessary that the character(s) appear in the fewest number of places to correctly map the matrix. All of the correct solutions have met the criteria just described. In each case the taxa were correctly grouped, the tree topology was correctly constructed, and all of the characters were assigned in the most parsimonious fashion. The greatest variety occurs in the assignment of characters. For each of the research problems, 1-5, there is a statement indicating which participant achieved a correct solution along with a brief description of how he/she solved the problem. The actions of each student followed a similar pattern.
Correct Solutions of Problem One

Only three participants, S1, S5, and S8, solved P1 correctly. All three of these participants shared a common method of solution as judged by their verbal dialogues during problem solution. Each one began his/her solution by the grouping of identical, similar, and different taxa. All of the participants grouped R5 and R11 together, R9 and R13 together, and R2 separate from the other taxa groups. (The proper grouping of taxa is crucial to a good solution because character assignment will always be flawed when taxa are improperly grouped.) Each of the participants assigned a node to each taxa group (three in all) and used the proper number of links to connect each of the taxa groups to the common ancestor. However, all three of the participants assigned the characters differently. S5 chose to assign the characters in numerical order (1-5). S1, however, chose to assign character #2 to the taxa group R5 and R11. (Character #2 is the only character present in this taxa group.) Character #1 was assigned next and placed at the very base of the tree (the link just above the common ancestor) because all of the taxa, except the common ancestor, possess character #1. Following the assignment of these two characters, S1 proceeded to assigned characters #3 and #4 to the remaining taxa groups, namely R7, R9, and R13. Finally character #5 was assigned to the R9 and R13 taxa group and tree construction was complete.

S8 also chose to assign characters in numerical order and began with character #1 stating that all of the recent taxa shared character #1. Next character #2 was assigned to the R5 and R11 taxa group followed by characters #3 and #4 being
assigned to R7, R9, and R13 taxa groups. S8 could not, at first, decide what to do with character #5. S8 noticed that R7 does not have character #5 while R9 and R13 do have character #5. Accordingly, character #5 was placed on the link directly beneath R9 and R13. S8 kept wondering if a reversal was operating with character #5 but decided that it was not since none of the characters (1-5) was homoplasious. S8 was the only one of the problem solvers with a correct solution to check the problem solution. This was done by verbally indicating what character numbers were under a given taxa group and then checking to determine whether or not they were present. This strategy proved successful for S8 since this participant achieved the correct solution to all five of the research problems.

Correct Solutions of Problem Two

P2 was correctly solved by five participants: S1; S2; S5; S8; and S12. P2 was the first of the five research problems to show homoplasy or the display of a character in more than one spot on a phylogenetic tree. Homplasy may be shown as a reversal or a convergence. All of the solvers of P2 chose the convergence option.

S1 and S12 solved P2 in similar ways. Both participants grouped taxa by noting that R1 and R3 shared identical characteristics, that R4 and R5 shared identical characteristics, and that R3 shared similar and one different characteristic. Each of the taxa groups was assigned a single node and the necessary links to connect each taxa group to the common ancestor. Character assignment was identical for both S1 and S12. Characters #3 and #5 were assigned to the R1 and R3 taxa group first.
Both S1 and S12 assigned character #1 next followed by character #3 to R2. (With the assignment of character #3 in a second spot, character #3 was homoplasious and placed on the tree by the process of convergence.) Finally, characters #4 and #5 were assigned. There was no checking.

S5, using the computer as the only problem-solving medium (no paper and pencil diagrams), grouped the taxa by stating that R4 and R5 shared characters, that R1 and R3 share characters, and that R2 shared character #3 with R1 and R3. (This participant recognized the homoplasious nature of character #3 early on.) After assigning a node to each taxa and the necessary number of links to connect each taxa group to the common ancestor, the characters were assigned in the following order: #3, #5, #1, #2, #4, and #3 again. This means that the homoplasious character, #3, was assigned last. There was no checking.

S2 also solved P2 by looking for similarities and differences among the taxa. R2 and R4 were found to share characters #1, #2, and #4 while R1 and R3 share character #3, and #5. R2 has character #3. Nodes and links were added with one node for each taxa group and the appropriate number of links to connect each of the three taxa groups to the common ancestor. Characters were assigned in the following order: characters #4, #2, and #1 to the R4 and R5 taxa group and characters #3, and #5 to the R1 and R3 taxa group. The last character to be assigned was character #3 made homoplasious and assigned to taxa group R2 by the method of convergence.

S8 used an approach similar to that employed by other successful participants stating specifically that the method was to look for similarities and differences among
the taxa. S8 noticed first that R1 and R3 share characters #3 and #5 and that R2 has character #3 and is in its own category of grouping. R4 and R5 share characters #1, #2, and #4. Nodes and links were placed and characters were assigned in the following order: #3, #5, #1, #2, and then #3 again making #3 the homoplasious character. S8 wondered if convergence was present and, of course, it was. Character #4 was assigned last. S8 was the only one of the successful participants to check P2.

Correct Solutions of Problem Three

Five of the participants correctly solved P3. They were S1, S2, S5, S8, and S10. There are two ways to group the taxa in P3. Three of the participants used one way and two of the participants used the other way. S1, S8, and S10 grouped the taxa in the same way. Each made three groups of taxa: R2 and R4; R1 and R3; and R8 was in a separate taxa group. Each of these groups had a single node and the links necessary to connect each of the taxa groups to the common ancestor. S1 added characters in the following order: #2, #3 #5, #4, #4, and #1. Character #4 was the homoplasious character and was placed on the tree by the method of convergence. (It is interesting to note that R6 was grouped with R10 on the basis of character #1 - this was determined by S1's verbal dialogue.

S8, while grouping the taxa the same as S1, assigned the characters in a different order. The characters were assigned in the following sequence: #2, #1, #3, #5, and #1 again. The links were changed between node PC and R6 and between PB and R8. This of course meant a reassignment of characters. The new character
assignment was as follows: #2, #1, #3, #4, #5, #4, and #4. This, of course, means that character #4 was homoplasious and assigned by both the methods of convergence and reversal. S8 did check the solution of the problem.

S10 grouped the taxa in the same manner as S8 but assigned character in a different order. The characters were assigned in the following order: #1, #2, #3, #4, #5, and then character #4 again. Thus, character #4 was the homoplasious character and it was made homoplasious by the method of reversal. There was no checking of the results.

S2 and S5 grouped the taxa in an alternative arrangement namely: R2 and R4, R8 and R10, and R6 was assigned to a separate group. S2 began the solution of the problem by listing the characteristics under each of the taxa before going to the computer. The characters were assigned in the order of: #4, #5, #3, #2, and #1. Character #1 was the homoplasious character and it was assigned by the method of convergence. S2 did not check the problem. S5, while using the same taxa grouping and homoplasious character (i.e., #1) assigned the character in the order of: #2, #1, #1, #4, #3, and #5.

Correct Solutions of Problem Number Four

P4 was correctly solved by the following participants: S5, S6, S8, S10, and S11. S5 and S10 grouped the taxa in an identical way namely R13 and R11, R15 and R19, and R17 each in a separate group. Each of these participants added four nodes and the links necessary to connect all of these taxa groups to the common ancestor.
S5 began in a rather uncertain manner and assigned character #1 to the link just above the common ancestor but reconsidered it and removed the links this time adding character #2 to this link and one other link above the ancestor making character #2 a reversal. Character #4 was added next and then characters #1, #5, #3, and #4 again. Character #4 is also handled as a reversal. S10 assigned characters a little differently than S5 using the order #2, #4, #3, #5, #4, #4, and #1. Character #4 was homoplasious and handled by both the methods of convergence and reversal. S10 did check the solution to this problem.

S6, S8, and S11 grouped the taxa in the following manner: R11 and R13 in one group, R17 and R19 in another group and R15 in a separate group. After adding four nodes and the necessary links to connect the taxa to the common ancestor, each participant assigned the character differently. S6 assigned the characters in the following order #2, #4, #3, #5, #1, #4, and #1. Both characters #4 and #1 were made homoplasious by the process of convergence. S8 opted for the following order of character assignment: #2, #4, #1, #3, #5, #4, and #1. Character #4 was made homoplasious by the process of convergence and character #1 was made homoplasious by the process of reversal. S11 chose to assign characters in the following order: #1, #1, #4, #5, #3, #2, and #4. Both character #4 and #1 were made homoplasious by the process of convergence.

Only S8 and S10 checked their problem solutions.
Correct Solutions for Problem Five

Only two participants solved P5 correctly, S8 and S11. This problem was not difficult to solve in terms of characters assignment, however it did present multiple fossil groups that could not be moved up or down on the computer screen. This seemed to present a problem for the participants who did not know what to do with only three living taxa. S8 began the process of solution by noting that the taxa groups should be composed of R11 and R13 with R15 in a separate group. These two taxa groups were then connected to F9, F40, and F73. A single node was used to connect R11 and R13 to the F9 fossil. The characters were added in the following order: #1, #3, #2, #4, #5 and #4. Character #4 was made homoplasious by the process of reversal. S11 followed exactly the same format by grouping R11 and R13 together and placing R15 in a separate group. Two nodes were added, one for each taxa group with the second node linking to F9, F40, and F73, the three fossil taxa. Characters were added in the following order: #1, #3, #2, #4, #5, and #4. Character #4 was made homoplasious by the process of reversal. S8 checked the problem results but S11 did not.

After having discussed the correct solutions to the five research problems, it is necessary to consider the errors as well. The next section will do that by first identifying and explaining the common errors and then discussing each of the research problems and how these errors were made problem by problem. The section begins with Table 6 which details the error categories along with the participants who made the errors. This is followed by an explanation of each of the errors and then a
Table 6
Common Mistakes by Problem and Participant

<table>
<thead>
<tr>
<th>Problem</th>
<th>Improper Taxa Groups</th>
<th>Omission of Characters</th>
<th>Parsimony Violated</th>
<th>Ancestral Characters Plotted</th>
<th>Excessive # Homoplasies</th>
<th>Matrix Does Not Map</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem 1</td>
<td>S2, S3, S9, S13</td>
<td>S2#3, S3#4 S9#1</td>
<td>S3, S2, S4, S9, S10, S11 S13</td>
<td>S2</td>
<td>S2/1, S3/4, S4/2, S6/2, S9/2, S10/2 S11/1, S13/1</td>
<td>S2, S3, S4, S9 S12, S13</td>
</tr>
<tr>
<td>Problem 2</td>
<td>S3, S4, S6 S9, S11, S13</td>
<td>S3, S6, S10 S11, S4, S13</td>
<td>S9, S13</td>
<td>S3/5, S4/2, S6/2, S9/2, S10/2 S11/2 S13/2</td>
<td>S3, S4, S9, S11 S13</td>
<td></td>
</tr>
<tr>
<td>Problem 3</td>
<td>S3, S4, S6 S9, S11, S13</td>
<td>S3, S4, S6 S11, S12, S13</td>
<td>S3/2, S4/3, S6/2, S9/1, S11/2, S12/1 S13/1</td>
<td>S3, S6, S9, S12 S13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problem 4</td>
<td>S1, S3, S4 S9, S13</td>
<td>S4#2</td>
<td>S1, S3, S4, S9 S12, S13</td>
<td>S4</td>
<td>S1/5, S3/4, S9/3 S12/1, S13/2</td>
<td>S2, S3, S4, S9 S12</td>
</tr>
<tr>
<td>Problem 5</td>
<td>S1, S2, S3, S4 S6, S9, S10 S12, S13</td>
<td>S2, S3, S5, S6 S10, S12, S13</td>
<td>S3/5, S4/2, S6/4 S9/2, S10/2, S12/2 S13/1</td>
<td>S1, S2, S5, S6 S9, S12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
problem by problem account of who made the errors and what analysis they did or
did not make.

Table 6 gives a complete listing of the mistakes made by unsuccessful prob-
lem solvers. The table is arranged with the mistakes in the vertical columns and the
problem numbers in the horizontal rows. The students who made the error in that
particular problem are listed in the appropriate cell of the table. The vertical columns
list type of error made and I will now discuss each of the column headings.

**Improper Grouping of Taxa**

This category of mistake was a common one and also one that seemed to be
quick as far as the problem-solver was concerned. The improper grouping of taxa, in
every instance, involved grouping two taxa that were identical in character distribu-
tion and then lumping the remaining three taxa into one large taxa group without
regard to its appropriateness. As a result, characters were wrongly assigned and a
correct solution was not available. The improper grouping of taxa was closely related
to the violation of parsimony. Parsimony is the placing of characters onto the tree in
such a way as to insure that they are in the fewest places possible in order to map the
matrix. There were 30 cases of improper grouping of taxa and 28 cases of parsimony
violations.

**Omission of Characters**

Although this was not a common mistake, it prevented 4 students from
achieving correct problem solutions. This mistake occurred in only two of the research problems, P1 and P4. It is interesting to note that the four students who made the error all did so with different character numbers and each of the four students only made the mistake in a single instance.

**Parsimony Violation**

Parsimony violation means that the characters were not placed on the tree in the fewest number of locations but were placed on the tree in order to “make the matrix work”. As a result, many of the characters were placed on the tree more than once. This made for a quicker solution but also contributed to errors in the use of homoplasy. There were 28 cases of errors in parsimony and 27 cases of errors in homoplasy. Violation of homoplasy can take two forms: convergence and reversal. In the case of convergence the character is simply placed on the tree in more than one spot and this was the most common form of homoplasy violation. In the case of reversal, the character is placed near the base of the tree in its derived form but is placed again at some point higher on the tree to indicate a reversal back to the ancestral state. It is, of course, called reversal to indicate that, for a given taxa group, the character became derived and then reverted back to its ancestral state. This was the less common form of a homoplasy error and it was found in 11 cases.

**Ancestral Characters Mapped**

Ancestral characters are never placed on a phylogenetic tree unless there is a
case of the reversal of a homoplasious character. This was not a common mistake but happened in five cases. With the exception of S2 who made this mistake twice, the remaining students made the same mistake once.

**Excessive Number of Homoplases**

This error means that homoplasy was used more often than was necessary. It is, to be sure, a violation of parsimony. The characters are simply placed on the tree too many times. The average number of homoplasy errors was two with a range from one to five. (Five would indicate that every character assigned was made homoplasious.) There were 34 cases where excessive numbers of homoplasy were used. The column that addresses this in Table 6 gives both participant number and then indicates, after each participant, how often homoplasy was used.

**Matrix Does Not Map Correctly**

In this mistake the solver does not translate the matrix into the proper phylogenetic tree. The derived and ancestral distribution of characters in the matrix are not correctly placed on the phylogenetic tree. One can simply say that the matrix and the tree are not congruent. Another way of looking at it is that the tree does not make sense when comparing it to the matrix. This error can result when one of two things happen: either the taxa are improperly grouped, the characters are improperly assigned, or both. There were 27 cases of this error.
Mistakes in Solving Problem 1

S2, S3, S4, S6, S9, S10, S11, S12, and S13 solved P1 incorrectly. Three of these participants, S3, S9, and S13, made the major error of improperly grouping the taxa. This was done by assigning two of the taxa with identical character distributions to one group and then placing the remaining three taxa into another group. This is a serious error because it guarantees that character assignment will be incorrect and, more often than not, parsimony will be violated. (The characters will be placed on the tree in more places that the matrix requires.)

All of the incorrect P1 solvers violated the use of homoplasy. P1 does not require the placement of any homoplasious characters. Each character in this problem should appear on the tree in only one spot. All of the incorrect solvers, except S2, made one or more of the characters homoplasious. S3, S4, S6, S9, S11, and S13 assigned the homoplasious characters by the method of convergence. They placed as few as one and as many as three of the characters on the tree in more than one location. S10 assigned the homoplasious characters by the method of reversal and, as a result, did manage to make the matrix map onto the tree but at the cost of parsimony. (The two extra steps needed to accomplish reversal added to extra steps to the problem solution and the creation of the tree involved 7 character assignments instead of the correct five.) S2 used homoplasy by mapping one character, #3, as an ancestral character. This was not a case of reversal since character #3 was not assigned in the derived state at some point earlier on the phylogenetic tree. All of the incorrect solvers made characters homoplasious when there was no need to do so.
Character omissions presented problems for S3 and S9. S3 omitted two characters and S9 omitted one character. Although not a problem for the majority of the incorrect problem solvers, it nevertheless prevented S3 and S9 from obtaining a correct solution.

S12 did not use homoplasy at all but still achieved an incorrect solution. The characters were simply assigned to the wrong taxa and, even though parsimony was not violated, character placement was totally incorrect. (S12 was the only participant to do this on PI.)

S2, S10, S11, and S13 made the homoplasious character(s) the last to be assigned. S9 assigned the homoplasious character(s) as the first character on the tree, and S3, S4, and S6 assigned the homoplasious character(s) as the middle character on the tree.

Only two of the participants with incorrect solutions, S6 and S10, managed to make the matrix map. In other words, all of the characters were assigned to the proper taxa but not in the most parsimonious fashion. One of these participants, S6, used homoplasy by convergence to make the matrix map and the other participant, S10, used homoplasy by reversal in order to make the matrix map.

All of the tree lengths, except that of S12, were longer and had more steps in character assignment than the prototype of PI. The correct tree length for PI is 5. Most of the incorrect solvers had tree lengths between 7 and 10. They simply added characters, in too many places, on the phylogenetic tree.

S3 was the only incorrect solver to check the solution of the problem. This
checking was done verbally and recorded on audio tape.

Mistakes in Solving Problem 2

S3, S4, S6, S9, S10, S11, S12, and S13 all had incorrect solutions. All but two of these incorrect solutions, namely those by S6 and S10, had incorrect taxa groupings. The most common mistake was to group two taxa, each with identical character distributions, together and then place all of the remaining taxa in a single group. The result is one dichotomous group and one trichotomous group. The usual result of such taxa groupings is that one of the two taxa groups will have the correct character assignment while the other taxa group has flaws in the assignment of characters. This lack of precise character assignment results in not only a nonparsimonious tree but one in which character distribution does not agree with the matrix.

Only two of the incorrect solvers, S6 and S13, managed to correctly map the matrix. This means that all of the taxa groups had the correct characters assigned to it but not in the most parsimonious fashion. This was done by making one or more of the characters homoplasious. Both S6 and S13 chose convergence in order to assignment the homoplasious character. S6 did this for four of the characters when only 1 character required homoplasious placement. S13 assigned one of the characters, #4, as an ancestral character with no prior assignment of that character (#4) as derived. In other words, character #4 was handled convergently even though it was assigned to the tree, in two places, as ancestral.
All of the incorrect solvers added extra steps to their trees by adding characters in more than one place on the tree. The correct tree length for P2 is 6 meaning that the matrix is correctly mapped with 6 individual character assignments. The incorrect problem solutions had tree lengths ranging from 7 to 12 steps with lengths of 7 steps being the most common.

S3, S4, S6, S9, S11, and S13 all handled the homoplasious character(s) by the method of convergence. S10 handled the homoplasious character(s) by reversal. Both S9 and S13 assigned character 3 and 4 as ancestral characters without showing them as derived characters first. Therefore, characters 3 and 4 were not true reversals but were placed in the ancestral state in order to indicate an absence of these character numbers for a particular taxa group. The range of homoplasious character assignment was as few as two to as many as five. (Only one character, #3, is actually homoplasious in P2.)

Only two of the solvers, S6 and S13, made the matrix map. All of the rest had character assignment errors in one or more of the taxa groups which means that the matrix was neither correctly followed nor correctly mapped. There was no order to the assignment of the homoplasious character 3 that all participants followed. Some participants assigned this character first, some assigned it last, and few even assigned it as the middle character.

Mistakes in Solving Problem 3

Seven of the participants, S3, S4, S6, S9, S11, S12, and S13, failed to solve
P3 correctly. Trouble began, for all of the participants except S12, with an improper grouping of taxa. S12 actually grouped the taxa correctly. All of the others grouped two of the recent taxa, R2 and R4, together because both of these taxa have identical character distribution. The remainder of the taxa were placed in a single trichotomous group. S11, however, grouped recent taxa R6 and R8 together and R10 in a group all alone. This is an error because R10 must be grouped with either R6 or R8 for a correct solution. R10 cannot stand alone if the matrix for this problem is to be correctly followed.

Parsimony was compromised by all of the incorrect solvers except S13. However, every incorrect solver was off by only 1 tree length. (All incorrect tree lengths were 7 instead of the correct 6.) This means that in P3 there were fewer instances in the use of homoplasious characters and, without exception, they were always characters 1 and 4. Five of the solvers, S3, S4, S9, S12, and S13, chose to make characters 1 and 4 convergent while S6 and S11 chose reversal for characters 1 and 4. (It so happens that characters 1 and 4 are homoplasious in P3.)

All of the incorrect solvers, except S6 and S9, began with character #2 and placed the homoplasious characters last in the order of assignment. S6 handled the homoplasious characters at the beginning while S9 placed one of the homoplasious characters at the beginning of the character assignment order and the other at the end.

Three of the solvers, S4, S6, and S11, did manage to correctly map the matrix while all of the remainder were unsuccessful. (In other words, the characters were not placed on the tree under the correct taxa groupings.)
S3 was the only participant to check the phylogenetic tree that was created.

Mistakes in Solving Problem 4

P4 presented some unique challenges in tree construction in that there were three different possibilities for taxa groupings and each one calls for a different assignment of characters, two of which involve multiple homoplasies. S1, S2, S3, S4, S9, S12, and S13 all achieved incorrect solution to P4. For three of these solvers, namely S2, S9, and S12, the taxa were correctly grouped. However, for S1, S3, S4, and S13 the taxa were not correctly grouped. All of these participants chose to group the taxa on one dichotomous group and one trichotomous group. The usual procedure was to group two of the taxa with identical character distributions into one group and lump the remaining three taxa into another group. This approach is a problem since the improper grouping of taxa leads to improper character assignment which usually means that both parsimony and homoplasy will be compromised. For S1, S3, and S4 that is exactly what happened. All of these participants made four of the assigned characters homoplasious by the process of convergence. This, of course, leads to a tree with too many steps (# of character assignments) in its solution. Two of the solvers, S1 and S3, had eleven steps in their problem solution. S9 had nine steps, S4 and S13 had eight steps, and S2 had only six steps in their problem solution. Since the correct number of steps for the solution of P4 is seven, all of the incorrect solvers, except two, S1, and S12, sacrificed parsimony in an attempt to map the matrix. It is interesting to note that only three of the solvers, S1, S9, and S13, were
successful in making the matrix map to the tree. All of the other solvers had more fundamental errors that prevented a correct mapping no matter how many characters were made homoplasious.

Convergence was, by far, the most commonly used method for assigning homoplasious characters. All of the incorrect solvers used convergence except S9, S12, and S13 who also combined convergence with reversal. Every unsuccessful participant made character #4 a homoplasious character and, indeed, that is what a correct solution of P4 requires. The order of assigning the homoplasious character varied from participant to participant with S1, S2, S3, S9, S12, and S13 assigning it as the last of five characters while S4 assigned it as the first of five characters. S2 assigned two characters, #1 and #5, as ancestral. This was not a case of reversal since these two characters were not assigned in the derived state at some location earlier on the tree. The purpose of these ancestral characters was to indicate that characters #1 and #5 were no longer present in a given taxa. This is an error in the way ancestral characters are assigned. Ancestral characters do not appear on the tree unless there is a reversal in operation. It is also interesting to note that S4 omitted character #2 completely. This error of omission guarantees an incorrect solution.

Only two of the solvers, S1 and S2, checked the problem. This is unfortunate since a process of checking may have reduced the possibility of error.

Mistakes in Solving Problem 5

Problem 5 had the greatest number of students with incorrect solutions: nine.
This problem is not particularly difficult in terms of character assignment. However, it does present three fossil taxa whereas all of the other problems have one such fossil taxa. This situation proved to be difficult for many. S1, S3, S4, S6, S9, and S10 all grouped the taxa incorrectly. The most common error was to group all of the three recent taxa into a single group and somehow connect that group to the three fossils. (This method is doomed to failure because the three recent taxa do not belong to the same group. Two go into one group and the third must be placed in a group by itself.) The fossil taxa were also handled in different ways. Three of the participants handled the fossils as a separate group that was a separate branch off from the common fossil. Others would connect one of the fossils to one of the recent taxa groups and the other fossil in a group by itself. All of these groups were connected to the common ancestor. This failure to group properly accounted for the failure in these situations. Strangely enough, parsimony was not a source of common error. All of the incorrect solvers but three, had the proper tree length and the correct number of character placements: seven. The three who did not have parsimonious solutions were S6 and S9 who had a tree length of nine steps each and S3 who had a tree length of thirteen steps. Only one of the incorrect solvers, S2, achieved a correct mapping of the matrix. All of the others made fundamental errors (usually of taxa grouping) that prevented a correct one-to-one mapping of the matrix.

Convergence was, once again, the most common method of assigning homoplasious characters. S1, S3, S4, S6, S9, and S10 all made one or more of the characters convergent. All of the incorrect solutions did this to character #4 and it is
character #4 that must be made homoplasious in order to correctly solve P5. S5 and S13 chose to solve the problem by using both convergence and reversal to handle the homoplasious character #4. S2 used reversal as the exclusive way to handle homoplasy in character #4. The order of assigning the homoplasious character was varied with most of the participants assigning it last in the order of the five assigned characters.

None of the problem-solvers with incorrect solutions checked their solutions. This is unfortunate because in checking the character assignment of each of the taxa one can often eliminate errors that would otherwise remain. Checking, however, must also involve making sure that the taxa are correctly grouped. If this is not done, a correct solution will never be attained. Therefore, checking must be at two levels, taxa grouping and character assignment.

Summary of Common Errors in Problem Solution

The common errors made by participants often began with an improper grouping of taxa. The usual way that improper grouping was done was to group together two taxa that shared identical characteristics and then lumping the remaining three taxa into a group by themselves. With such an error, the assignment of characteristics is bound to be faulty. Taxa will not have the proper characteristics assigned to them if they have been grouped incorrectly in the beginning. Because of improper grouping, parsimony and homoplasy were also frequently violated.

The usually mistake in the area of character assignment was to make the
matrix map by making several of the characters, sometimes as many as all five, homoplasious. In other words, characters were assigned to every taxa group without regard to parsimony and the number of times they had already been placed on the tree. The usual way of showing homoplasy was by the method of convergence where the character appears in multiple places along the tree. This usually took the form of putting the same character on the tree in as many as three different locations. While this practice of multiple homoplasy might make for a successful mapping of the matrix, it certainly eludes the assignment of characters in the most parsimonious fashion. Although some of the participants also used the method of reversal to assign homoplasious characters, it was not as widely used. There were a few instances where both homoplasy and convergence were used together in the same problem.

Three of the participants omitted characters altogether. This always involved only a single character but, when this error is made, a correct solution is impossible.

It is interesting to note that all of the incorrect participants had at least a rudimentary knowledge of parsimony because all of them had one, two, or three individual characters parsimoniously assigned. Many made verbal references to the fact that in a given taxa group both members shared a given characteristic and therefore they could place it under both of the taxa rather than under each of the taxa individually.

It is unfortunate that more students did not check their work. That is not to say that all of the participants who checked their work obtained correct solutions. However, if the participants had made the effort to check the grouping of the taxa and the assignment of each individual character to the taxa groups, more correct solutions
could have been achieved.

Another factor in making of errors was that some of the participants did not solve the problems until 6 weeks after the instruction was given. This occurred due to scheduling problems and numerous cancellations of problem-solving sessions. (I suspect that when 6 weeks had elapsed between instruction and problem solving, their thought patterns were not as well synchronized as they might otherwise have been.) All of the incorrect solutions had some elements of their trees that were correct. Unfortunately, they were also flawed by some rather fundamental mistakes.
CHAPTER V

DISCUSSION

Introduction

Because this research project involving phylogenetic tree construction is an extension of the research done by Steven Brewer (1996) with experts, it is necessary to describe the expert model of phylogenetic tree construction and situate the overall model of novice performance with respect to that expert model. This, of course, means that the expert and novice models will be compared for similarities and differences. While this is of value in itself, there is an additional reason for such a comparison. If one is to design curricula aimed at instructing novices in phylogenetic tree construction, it is necessary to know not only what the desired expert model is but also to know what the novice model is so that an instructional bridge may be constructed between them. The expert model, of course, provides the desired performance criteria while the novice model provides a picture of how novices actually accomplish the task. In a manner of speaking, the expert model is the objectives and the novice model shows the starting point in the curriculum building process. Once an understanding of these end points is established, it is necessary to give some pedagogical strategies to indicate what must be done to make novices perform more like experts. This is not to say that all novices will become experts but that they can learn to perform more like experts if given the right opportunities to do so.
This chapter begins with an elaboration of the expert model of overall performance as well as a recapitulation of the novice model of overall performance. Both models are then compared in order to determine similarities and differences between them. This is followed by the prescriptive Model of Desired Performance which Steven Brewer (1996) distilled from the Overall Model of Expert Performance. This model lists the strategies that are necessary to properly construct the most parsimonious phylogenetic tree from a given data matrix. The prescriptive model of desired performance then serves as a template for instruction designed to teach novices how to properly construct phylogenetic trees. Accordingly, a series of problems designed to teach phylogenetic tree construction is presented along with the reasons for choosing these particular problems. The chapter concludes by arguing that phylogenetic tree construction is a good introduction to problem solving in the high school biology curriculum.

Elaboration of the Expert Model of Phylogenetic Tree Construction

The procedural model of expert phylogenetic tree construction (see Figure 26) has a double headed arrow going down the left hand side which divides the model into two parts. The upper half of the model deals with observations, inferences, and actions involving the matrix while the lower half deals with observations, inferences, and actions involving the tree itself. Connecting these two halves are a series of multiple path activities that involve three main strategies. The three main strategies are: consider order; consider inclusion/exclusion; and consider parsimony.
Figure 26. A Procedural Model of Expert Phylogenetic Tree Construction.
The first strategy, consider order, begins the minute that the matrix is inspected. In the inspection an expert problem solver will examine both the columns and rows of the matrix in order to make inferences about them. These inferences are usually centered around counting the number of apomorphies (derived character states) and then identifying identical, similar, and different character distributions among the taxa present. Once this has been done and the taxa have been analyzed, the solver begins to reorder the matrix so that those taxa that are the first to branch away from the common ancestor are the lowest on the matrix. Those taxa that are the last to branch away from the common ancestor are placed at the top of both the matrix and the tree. Once the reordering of taxa has occurred, the characters are mapped onto the tree in order to correctly map the matrix. This strategy works well when there is little or no homoplasy. If homoplasy is low or not operating, there is a close correspondence between the number of apomorphies of a taxon and its relative position on the tree.

The second strategy, consider inclusion/exclusion, was the most fruitful and was used by all of the experts who generated multiple-tree topologies. When using this strategy, a solver seeks to group character states into chunks so they may be more readily assessed. An expert would start by finding the most inclusive character or pair of characters and then place the nested characters within that group. (Characters are inclusive when the distributions of 1's for one character is a subset of another character.) If the solver found an exclusive character, it would be used to start a separate group. (Characters are exclusive when the distributions of 1's does not
overlap). Once an inclusion/exclusion hypothesis has been determined, the expert will translate this hypothesis into sets of taxa. Each taxa is based on a node or a postulated common ancestor and supported by a character or characters. The translation process usually involves listing the taxa across the top of the drawing field with the most apomorphic characters on the right and the least apomorphic characters on the left. There are three useful distinctions to be made between inclusion and exclusion:

1. Characters that are identical to or nested within characters can be placed in the same group.

2. Characters that are exclusive from a group should start a new group.

3. Characters that conflict with any other group should be set aside in a third group.

In the study with experts, the inclusion/exclusion strategy was found to offer the best potential to serve as a starting point for a model of desired performance. It was the only strategy consistently successful in constructing multiple tree topologies.

The third strategy, consider parsimony, acts as a filter allowing problem solvers to evaluate how the parsimony of a tree is influenced by placing a character or taxon on the tree. If the problem is not too complex, the characters may be placed unproblematically on the links that make up the tree. In other cases, usually those involving homoplasy, the effects of other characters and taxa need to be considered before modifying or constructing a tree. In the case of competing character possibilities, the implications of placing each one must be considered and compared in order to decide which one(s) to select. This strategy often is extrapolated in order to run
through the permutations of other ways the taxon or character could be placed on the
tree and what effects this arrangement would have on the rest of the tree. Homoplasi-
ous characters often have multiple optimizations (ways they can be placed on the
tree) and imply multiple arrangements of taxa. All of these factors must be con-
sidered when using the strategy of parsimony.

Another strategy, used by only one of the experts, was the use of duplicated
taxa. This was done by pencil and paper solutions that did not involve the program
*Phylogenetic Investigator*. With this strategy, the characters were placed singly on
small trees. When all of the characters were thus assigned, the smaller trees were
combined into larger ones according to similarities and differences in character distri-
bution. The result was two combination trees which were either exclusive or which
overlapped with respect to a single taxon. Any incompatibilities were then resolved
into a parsimonious tree, often one of the most parsimonious trees. The duplicated
taxa strategy makes good use of display-based reasoning to free working memory.
Each inference is immediately reflected in the diagram and each step is largely inde-
pendent of other steps. All of the decisions are parsimony decisions which depend on
an efficient use of working memory. The drawback to this strategy lies in the fact
that two characters can be placed together because they are mutually inclusive but
one of which is wrongly placed for the most parsimonious solution. It can actually
result in solutions that are not parsimonious.
Elaboration of the Novice Model of Phylogenetic Tree Construction

The novice model of phylogenetic tree construction (Figure 18, p. 87) is divided into four categories that separate the model into discrete activities. These activities are found on the left side of the model and include matrix analysis, building a tree topology, character assignment, and checking. These are activities except checking that every student used as a part of his or her solution. The center pathway of the model represents the order of procedure for most of the participants. The activities that branch to the right of the center for matrix analysis and to the right and left of the center for character assignment represent the procedures for a small number of the participants. These deviations from the central pathway proved to be neither more nor less advantageous than the central pathway since there were correct solutions for all three pathways. However, most of the participants used the central pathway and more correct solutions were recorded for it than for the two deviated pathways.

Pathways through the model were determined using the transcribed audio tapes made during problem solution sessions. The activities begin with the examination of the matrix and looking at the taxa by character. This was usually done by determining which of the taxa had identical characteristics, which of the taxa had similar characteristics and which of the taxa had different characteristics. Taxa with identical character states were placed in one group, those with similar character states in another group, and those with different character states in another group. (This means that in most of the research problems there were three taxa groups.) An
alternative way of doing the same thing was to count the number of apomorphies or derived character states that were present in the matrix. (These are denoted as 1's and may be counted either vertically or horizontally when examining the matrix.) The counted apomorphies are then placed in some kind of order and assigned to taxa thus dividing the taxa into groups.

Building the tree topology was very straightforward and involved, without exception, the placement of nodes (one per taxa), the adding of links, and the connecting of those links to the common ancestor in order to build the tree. When completed, all of the trees created had each of their taxa groups connected, by node and link, to the common ancestor.

Character assignment was the next activity and proceeded, with most of the participants, according to the prior grouping of taxa. The participants, having already grouped the taxa according to similarities and differences in character distribution, simply assigned the characters according to their predetermined grouping. The two alternative ways of assigning characters involved either doing so singly and in numerical order, or by again counting the number of character or group apomorphies. All of these character assignment techniques were successful even though most of the participants used the prior grouping of taxa as their assignment technique. While there was some variation in the order of placement of homoplasious characters, by far the most common method was to assign the nonhomoplasious characters (those that appear once on the tree) first and the homoplasious characters (those that appear more than once on the tree) last. There were examples of the homoplasious
characters being assigned in other orders but correct solutions did not depend on the order of homoplasious character placement. Whenever character assignment happened, it was necessary for the participants to consider parsimony. (Parsimony refers to placing the characters on the tree so that they appear in as few places as possible to correctly map the matrix.) Parsimony was either complete or partial. Complete parsimony meaning that the characters were properly placed in the fewest locations possible and the tree, thus constructed, was considered correct. If some of the characters were not parsimoniously placed but others were, then partial parsimony occurred. Most of the participants, whether solving the problem correctly or not, had at least one or more of the characters parsimonious placed. (The placing of characters according to partial parsimony did not result in a correct solution.)

The final activity, checking, was done by only a few of the participants. This usually involved verbally stating the characters that were present under each of the taxa and checking to see that they were properly placed or simply stating the character by name and number and looking to see it was properly placed whether or not the taxa groups were actually mentioned. Not all of the correct solutions involved checking but more than half of them did.

A Comparison of Expert and Novice Performance

While experts and novices approach problem solving in very different ways, the respective models of overall expert and novice performance have striking similarities as well as differences. This comparison between the expert and novice per-
formance models will consider their general topography, methods of matrix analysis and taxa grouping, character assignment, parsimony, and a review of any assumptions that may have been made.

The expert model of performance is more circuitous and not as simple and straightforward. It is composed of five clusters one of which involves matrix analysis and four of which involve major activities that are necessary in order to construct a most parsimonious tree. Between the activity clusters, as well as within them, there is a reevaluation of hypotheses in order to modify or change any activity previously performed. The activity clusters include matrix modification, considering inclusion and exclusion of characters, considering parsimony, and making inferences about the constructed tree.

The novice model of performance, while more linear and direct in topography, has four major activities that must be performed in order to construct a most parsimonious tree for any given set of data. Those activities include matrix analysis, building a tree topology, character assignment and checking. Each activity is placed along a central pathway which all of the participants followed with the only deviations occurring in taxa grouping and character assignment. Even though the novice performance model appears to be very different from the expert model, there is a remarkable similarity in tree construction procedures although on a less sophisticated plane.

Novices made no changes to the matrix whatsoever. Experts, however, felt free to move taxa and character states about in order to reorder the data and reorgan-
ize the matrix in order to better interpret the data. This modification of the matrix made it easier to group taxa especially when more than five taxa were present to be grouped. In reordering the matrix, and regrouping the taxa, experts used an inclusion/exclusion strategy. This strategy involves counting the apomorphies (derived character states) and determining which ones are inclusive and which ones are not. Exclusive characters were either identically distributed among taxa or were nested within the character distribution of a larger taxa group. Exclusive characters were those whose distribution did not overlap and, therefore, the taxa were placed in separate groups. Novices also used an inclusion/exclusion approach but did so by scanning the matrix and determining which taxa had an identical, similar, or different distribution of derived character states. (Since novices only worked with five taxa and five character states, their matrix analysis was much less time consuming and direct). A few of the novice participants actually counted the apomorphies before grouping the taxa. This was evidenced by the use of the words "derived" or "ancestral" for every character state and the conclusion that an identical distribution of the same character state meant that two taxa belong to the same group. Those novices, who were unsuccessful in the grouping of taxa, had no trouble grouping two taxa with identical character distributions in the same group but often, after having done this, they would group the remaining taxa into one large group regardless of any similarities or differences. (Of course, those participants who solved the problems correctly did not fall into this common trap.)

The expert model involves the evaluation of hypotheses in two different
places: taxa grouping and parsimonious character assignment. Two of the expert participants each used different hypotheses to construct their trees. One participant used a duplicated taxa strategy to convert each character systematically into a drawing. All of the characters were assigned to taxa on separate trees and then the trees were combined whenever possible to make trees which were exclusive or trees which overlapped with respect to one or more taxon. The result gave a very parsimonious tree, often one that was most parsimonious. Another expert used an order of divergence strategy and counted the number of apomorphies for each taxon. Taxa with the fewest apomorphies were placed at the bottom and taxa with the most apomorphies were placed at the top. Identical and similar taxa were then recognized and grouped. This strategy often allowed side branches to be constructed from the main lineage and again, allowed the construction of a most parsimonious tree. No matter which of the strategies were used by experts to group taxa (inclusion/exclusion, duplicated taxa, or order of divergence) their hypotheses were repeatedly reevaluated to allow them to both add and remove characters from a group as well as modify the taxa composing the group. Novices did little evaluating of hypotheses. So, once the taxa were grouped, it was on to the next task of adding the nodes and links in order to construct the branching topology of the tree. Every taxa group was given a link and every group of taxa was connected to the common ancestor by a series of links.

To consider parsimony is to focus on placing each of the derived character states in as few positions as is possible on the phylogenetic tree. Experts used the
parsimony strategy to assign characters, to modify the tree for missing and duplicated characters, and to postulate alternate arrangement of characters and taxa. Novices, in most cases, did not consider parsimony to be that important. They recognized that certain characters were shared by both taxa of a group and therefore could be placed under both of the taxa rather than on separate links under each of the taxa but they were not overly concerned with parsimony. Their goal was to see that all five of the character states were assigned and, in a few cases if it helped map the matrix, characters were assigned to multiple positions (made homoplasious) without regard to any kind of overall parsimony. It is important to note, however, that all of the novice participants placed at least one or two of the characters parsimoniously.

While the expert model indicates that the experts made inferences about their trees and thus modified them for correctness, novices used a process called checking. Checking was usually done by character and sometimes by taxa. Novices simply checked to see if each character was properly placed or that each of the taxa had the proper characters underneath it. Often, the checking resulted in correct solutions but more than half of the novice participants made no effort to check their trees.

Novices did not create any alternative trees. Experts, of course, did. While the creation of alternative trees is not a part of either the expert or the novice performance model, being more careful to consider an hypothesis of character inclusion and exclusion among taxa might have generated more alternative trees. Table 7 gives a specific comparison of the expert and novice solution by specific activity.

Given a comparison between expert and novice performance, it is helpful to
<table>
<thead>
<tr>
<th>Experts</th>
<th>Novices</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. One common pathway of activity using similarities/differences and identical characters to group taxa.</td>
<td>1. One common pathway using similarities/differences and identical characters to group taxa.</td>
</tr>
<tr>
<td>2. Main method of assigning characters is the inclusion/exclusion strategy.</td>
<td>2. Main method of assigning characters is by the prior grouping of taxa.</td>
</tr>
<tr>
<td>3. Two alternate pathways used by a small number of participants - considering order and duplicated taxa.</td>
<td>3. Two alternate pathways used by a small number of participants - counting apomorphies &amp; assigning characters singly and in numerical order.</td>
</tr>
<tr>
<td>4. Matrix was manipulated and reordered to facilitate the grouping of taxa.</td>
<td>4. Matrix was not changed in any by any of the novice problem solvers in order to group taxa.</td>
</tr>
<tr>
<td>5. Postulated and found alternate topologies for many of the problems.</td>
<td>5. None produced an alternate tree topology for any of the problems.</td>
</tr>
<tr>
<td>6. Made inferences about the tree involving missing characters and missing taxa as well as duplicated characters and taxa.</td>
<td>6. When checking was done, it was done either by checking each character for proper placement or listing each taxa placing the characters underneath.</td>
</tr>
<tr>
<td>7. Worked with groups of characters via the process of “chunking” data making for greater sensitivity to patterns in character position.</td>
<td>7. Worked with single characters and, and, when once assigned, went on to the next character.</td>
</tr>
<tr>
<td>8. No information given about the order of homoplasious character placement.</td>
<td>8. With exception of three participants all assigned the non-homoplasious characters first and dealt with homplasy last.</td>
</tr>
<tr>
<td>9. Used paper and pencil to solve problems using the two alternate strategies and the computer when the inclusion/exclusion strategy was used.</td>
<td>9. Used paper and pencil as well as the computer to solve problems indiscriminately no matter what strategy was used.</td>
</tr>
<tr>
<td>10. 93% of the experts solved all of problems by generating at least one most parsimonious topology.</td>
<td>10. 38% of the novices solved all of the problems by generating at least one most parsimonious topology.</td>
</tr>
</tbody>
</table>
analyze why participant from both groups made mistakes in their solutions. The next section will explain how the descriptive model of expert performance has given rise to a model of desired performance that is prescriptive and serves as a guide for constructing curricula designed to teach novices how to construct phylogenetic trees.

Prescriptive Model of Desired Performance

Steven Brewer (1996) constructed what he called a prescriptive model of desired performance in phylogenetic tree construction. This model was constructed by taking what Brewer (1996) termed the most powerful aspects of the descriptive model of expert performance and transforming them into a set of steps that should be followed in order to properly construct phylogenetic trees. The model has eight steps that embody both the activities and the order of activities in phylogenetic tree construction. The prescriptive model of desired performance is shown in Figure 27. The desired performance model is prescriptive because it sets forth the parameters or benchmarks necessary for designing curricula that are aimed at teaching phylogenetic tree construction from a problem-solving approach. The activities described form the basis for a curriculum that has at its foundation the salient features of the ways experts created their phylogenetic trees. This is not to say that novices are to be taught to mimic the actions of experts, but novices should be exposed to expert methods with the idea that expert methods form a suitable template to insure construction of one or more parsimonious trees for each data matrix. Brewer (1996) says that the inclusion/exclusion strategy (one of the three strategies that experts
1. Organize the characters mentally, in the matrix or on paper, to find the largest inclusion/exclusion character group hypothesis (hypotheses).

2. Translate the hypothesis into taxa by organizing the taxa into the drawing field.

3. Postulate an ancestor for each character or group of identical characters in the inclusion/exclusion hypothesis.

4. Link the most inclusive ancestors to the next less inclusive ancestors and continue until reaching a terminal taxa.

5. Distribute homoplasious characters.

6. Consider each homoplasious character starting with the character that requires the most steps.
   a. Evaluate whether options that improve a character’s distribution always result in matching losses in other characters.
   b. Evaluate whether other homoplasious characters have similar distributions that reinforce each other.

7. Construct other topologies based on additional inclusion/exclusion hypotheses from (1) or parsimony hypothesis from (6a) or (6b).

8. For each topology consider alternate optimizations for each homoplasious character.

Figure 27. A Prescriptive Model of Desired Performance in Phylogenetic Tree Construction.

employed) was the only strategy successful for finding multiple-tree topologies across all problem types. This strategy was also successfully employed by experts using Phylogenetic Investigator as well as paper and pencil methods. Therefore, the inclusion/exclusion strategy is the only strategy used in the prescriptive model of desired performance. The other two expert strategies, duplicated taxa and ordering, were not successful in generating alternate topologies and were also used by only two of the expert participants. Therefore, each of these methods is not considered in the prescriptive model of desired performance.

The Prescriptive Model of Desired Performance and High School Students

The Prescriptive Model of Desired Performance is appropriate for high school students because it encompasses all of the activities necessary to properly construct a phylogenetic tree regardless of the number of taxa or character transitions to be mapped. Tree construction must begin with an analysis of the matrix with its taxa and associated character states. No matter what strategy is used to do this, the solver must group the taxa together according to whether or not they share identical characters, similar characters, or different characters. The taxa, once grouped, are placed into a tree topology that links each taxa group with the common ancestor. The assignment of characters is the next procedure and should proceed with the nonhomoplasious characters first and the homoplasious characters last. The desired model addresses the procedure for assigning homoplasious characters. You begin with the homoplasious character that requires the most steps (occupies the most
positions) in order to map the matrix. This is followed by the character with the next lowest number of steps (positions on the tree) and keep doing so until all of the homoplasious characters have been assigned. The model then speaks of constructing alternate trees involving alternate optimizations of homoplasious characters and or alternate groupings of taxa. All of these procedures are good solid ways to properly construct phylogenetic trees. Of course, all of these procedures are given with an eye toward parsimony. Parsimony simply assumes that nature will take the shortest possible route in displaying character transitions. A parsimonious tree (one with the fewest number of steps or character transitions) is always preferred against any other possible character mapping. If the tree can be correctly constructed with five character transitions instead of seven, keep it at five and do not map it with seven transitions. All of these steps, matrix analysis, taxa grouping, character assignment, and parsimony, are necessary steps in planning any instruction in phylogenetic tree construction, no matter what the level of sophistication and or expertise. High-school students may thus benefit from the desired model in both their overall thinking about forming an hypothesis as well as actual construction of phylogenetic trees.

Instruction in Phylogenetic Tree Construction

The instruction of novices in phylogenetic tree construction should include a series of trees that build upon each other. With this idea in mind, trees showing automorphy, synapomorphy 1, 2, and 3, homoplasy 1, 2, 3, and 4, a tree with two fossil taxa, a tree showing both convergence and reversal, are presented. Each of these
trees shows not only the place of taxa and character states but also provides the matrix as well. These trees may be used by employing *Phylogenetic Investigator*, in which they are shown, or with paper and pencil diagrams. In either case, these problem-trees give a broad spectrum of what instruction in phylogenetic tree construction should include.

The simplest example of a phylogenetic tree is one with only two taxa groups, one extant (living) taxon and one extinct (nonliving) taxon. These two groups are connected by a single link to which one derived character is attached. Such a tree is known as *autopomorph* and is shown in Figure 28. For this specific example, R84 is the extant or living taxon and F95 is the extinct or nonliving taxon. R84 has a single derived character state which F95 lacks. Therefore, the link connecting the two taxa shows the derived character state 1 above F95 and below R84.

In the next sample problem, known as *synapomorph 1*, two extant taxa form a clade. The clade is connected by a single node and link to the extinct fossil taxa. In the example problem the extant (living) taxa are R82 and R87. These two identical taxa form a clade that is connected by the node PA to a link joining with the extinct (nonliving) taxa F98. This problem example introduces the concept of parsimony. Because R82 and R87 both share the derived character 1, that character is placed on the link below the two taxa rather than separately between each taxon and the node PA. Parsimony is a very important concept and needs to be stressed early on in the instruction process. This problem, of course, only has one solution. Synapomorphy 1 is pictured in Figure 29.
Figure 28. Example Tree and Matrix Showing Autapomorphy.
Figure 29. Example of Matrix and Tree Showing Synapomorphy 1.
In the next sample problem the concept of a clade and an outgroup are introduced. This example is synapomorphy 2 and is shown in Figure 30. In this example three extant (living) taxa are joined to a single extinct (nonliving) fossil taxon. The clade with the two identical taxa share two derived character states while the outgroup taxon has only one of the derived character states. In the sample problem, R87 and R84 make up the clade that shares both of the derived characters 1 and 2. R82 makes up the outgroup which only shares derived character 2. F97 is the fossil taxon at the very base of the tree. This tree requires two nodes, PA and PB, and 5 links. Derived character 2 is placed at the base of the tree and derived character 1 is placed above the node PA since only the taxa above it, R87 and R84 share derived character 1. The tree has a single solution that requires 2 steps or two character assignments in order to solve it.

In the next problem example there are 4 extant (living) taxa and one extinct (nonliving) taxon. This problem is known as synapomorphy 3 and is shown in Figure 31. The four extant taxa form two unique taxa groups or clades that share identical character states. One of the taxa groups is R84 and R80 and the other taxa group is R86 and R87. Each of these taxa groups is connected by a single node, PA and PB respectively, that connect to a third node, PC, that is just above the fossil taxon F96. Derived character state 2 is shared by both of the taxa groups (clades) and so it is placed on the link F96-PC. Character 1 is present only in the R84-R80 taxa group so it is placed on the link PA-PC. Character 1 is present only in the R86-R87 taxa group so it is placed on the link PB-PC. There is only one solution to this problem and it...
Figure 30. Example of Matrix and Tree Showing Synapomorphy 2.
Figure 31. Example of Matrix and Tree Showing Synapomorphy 3.
requires three steps or three separate character assignments.

The concept of homoplasy is introduced in the next sample problem known as homoplasy 1. Homoplasy means that a given character state appears on the tree in more than one location. In the homoplasy 1 problem, the method of convergence is used. In this method of assigning a homoplasious character, the character appears on more than a single link on the tree. In the sample problem, there are 3 extant (living) taxa and 1 extinct (nonliving) taxon. Two of the taxa, R84 and R86, form a single taxa group because each shares three of four character states. The third extant taxon, R87, shares only 2 of the 4 character states. The problem has two nodes, PA and PB respectively, and 5 links. Character 1 is placed on the link F94-PA because it is shared by all of the extant taxa. Characters 2 and 4 are placed on the link PA-PB because they are shared by the taxa group R84-R86. Character 3 is the homoplasious character and it must be placed on two separate links in order to solve this problem. It appears once on the PA-R87 link and must appear again on the PB-R86 link. There is only one solution to this problem and it requires 5 steps or character assignments. (Characters 1, 2, and 4 appear once and character 3 appears twice making a total of 5 character assignments.) Homoplasy 1, the convergence option, is shown in Figure 32.

The next example problem shows the second method of handling homoplasious character and that method is reversal. In the assignment of a character by the method of reversal, the character appears along two separate links one in the tree, one time in the derived state and another time in the ancestral state. In other words, one
Figure 32. Example of Tree and Matrix Showing Homoplasy 1, Convergence Option.
of the derived characters reverted back to its ancestral state and this must be reflected on the phylogenetic tree. In the example problem, homoplasy 2, the reversal option, shown in Figure 33. There are three extant taxa, R87, R85, and R88, one extinct taxon, F98. There are also two nodes, PA and PB respectively, and four character states. Derived character state 4 is shared by all of the taxa and is placed at the base of the tree on the link F98-PA. Characters 1 and 2 are placed on the link PA-PB because both of the extant taxa R85 and R88 share them. Character 3 becomes the homoplasious character and it is assigned by the method of reversal. It appears on the link F98-PA in the derived state (3 0>1) and on the link PB-R85 in the ancestral state (3 1>0). There is a single solution to this problem and there are 5 steps or character assignments. (Characters 1, 2, and 4, appear on one link each while character 3 appears on two links, once in the ancestral state and once in the derived state.)

In homoplasy 3, which is the next sample problem, the homoplasious character is assigned to two separate clades or taxa groups. This problem and its associated tree have 4 extant taxa forming two separate taxa groups and 1 extinct taxon. There are three nodes, PA, PB, and PC respectively, and 4 character states, one of which is homoplasious. This problem and tree is called homoplasy 3, convergence in two clades, and is shown in Figure 34. One of the taxa groups, R80-R82, share two of the character states and the other taxa group, R81-R85, share two of the character states. Since derived character 1 is found in all of the taxa, it is placed on the link PA-F94 at the base of the tree. Character 4 is shared by the R82-R80 taxa group and is placed on the link PA-PB. Character 2 is shared by the R81-R85 taxa group and is placed
Figure 33. Example of a Tree and Matrix Showing Homoplasy 2, Reversal Option.
Figure 34. Example of a Tree and Matrix Showing Homoplasy 3, Convergence in Two Clades.
on the link PA-PC. Character 3, the homoplasious character, is found in one of the taxa of the two clades but not in the other. (In other words, character 3 is present in R80 but not R82 and is present in R81 but not R85). As a consequence, character 3 must be placed on two separate links, PB-R80 and PC-R81 respectively. There is one solution to this tree and it requires 5 steps or 5 character assignments. (Characters 1, 2, and 4 are assigned once and character 3 is assigned twice making a total of 5 character assignments.)

In the next type of tree sample problem, known as homoplasy 4, the homoplasious character is assigned by the method of reversal and is present in only one taxa group. This problem is very similar to homoplasy 2, the reversal option, but a second example of this problem type is important to an overall understanding of homoplasious characters. In this problem there are three extant taxa and one extinct taxon. Homoplasy 4, the reversal option in one clade, is shown in Figure 35. R84 and R86 form a single taxa group while R88 forms an outgroup. There are two nodes, PA and PB respectively, and there are three character states. Derived character 3 is shared by all three of the extant taxa and it is placed on the link F97-PA at the base of the tree. Character 2 is shared by the R84-R86 taxa group and it is placed on the link PA-PB. Character 1 is the homoplasious character and it must be placed on two separate links of the tree. Character 1 is placed on the link F97-PA because it is present in two of the extant taxa, namely R88 and R86. However, character 1 is not present in R84 and, therefore, the method of reversal must be used. Character 1 is assigned first as a derived character on the link F97-PA and then as an ancestral character on the link.
Figure 35. Example of a Tree and Matrix Showing Homoplasy 4, Reversal Option in One Clade.
PB-R84. There is one solution to this problem and there are 4 steps of character assignments. (Characters 2 and 3 appear once but character 1 is homoplasious and appears twice.)

It is important to demonstrate the fact that both methods of assigning homoplasious characters, convergence and reversal, can be handled in the same problem and tree. The next problem demonstrates that convergence and reversal can coexist in the same problem. The problem is titled convergence and reversal and appears in Figure 36. This problem has 5 extant taxa and one extinct taxon. There are three nodes, PA, PB, and PC, respectively, and 5 character states. The extant taxa R05 and R03 form a single taxa group. The extant taxa R02, R01, and R04 form separate groups, each with a single taxon. Characters 1, 3, and 5 appear on the tree in only one location while characters 2 and 4 are homoplasious. Character 2 is handled by the method of convergence and character 4 is handled by the method of reversal. Character 4 is placed on the link PA-F98 because all of the taxa except R02 and R03 share it. Characters 3 and 5 are placed on the link PA-PB because R04, R05, and R03 all share it. Character 1 is placed on the link PB-PC because R05 and R03 share it. Character 2 is assigned to two separate links by the method of convergence. Character 2 is placed on the link R02-F98 and the link R01-PA. Character 4, already assigned in the derived state on the link F98-PA, is also placed on the link PC-R03 but is placed there in the ancestral state (4 l>0). There are seven steps or character assignments to this problem. (Characters 1, 3, and 5 are assigned once each and characters 2 and 4 are assigned twice each making a total of 7 character assignments.)
Figure 36. Example of a Tree and Matrix Showing Convergence and Reversal.
So far, all of the tree-problems have had a single extinct (nonliving) fossil taxon. This next sample tree-problem has two fossil taxa. One of the unique aspects here is that the program, Phylogenetic Investigator, will not allow these fossil taxa to be manipulated vertically but only horizontally. In this tree problem, known as two fossil taxa and shown in Figure 37. There are 4 extant taxa, R80, R83, R81, and R84, and 2 fossil taxa, F96 and F98. There are 4 nodes, PA, PB, PC, and PD respectively, and 4 character states. There is no homoplasy in this problem and each character is placed on the tree in a single location. The difference between this problem and synapomorphy 3 is that there is more than one fossil taxon. Character 2 is found in all of the taxa except F96 and so it is placed on the link PA-PC. Character 4 is present only in F96 and so it is placed on the link PA-F96. Characters 2 and 3 are shared by separate taxa groups and are placed on separate links. Character 3 is placed on the link PC-PD and character 1 is placed on the link PB-PC. There is one solution to this problem and there are 4 steps or 4 separate character assignments.

Now that example tree problems have been given in order to plan a unit of instruction on phylogenetic tree construction, it is necessary to discuss a procedure for teaching this material to novices, including high school students.

Suggested Methods for Use in a High School Biology Course

In order to teach problem solving in a high school biology course, instruction should involve three phases: modeling, coaching, and fading. The goal should be to learn to apply the simple rules of phylogenetics in the construction of phylogenetic
Figure 37. Example of a Tree and Matrix Showing Two Fossil Taxa.

Problem: Problem (Two Fossil Taxa)
trees using a data matrix. In this strategy the teacher models the problem(s) solving each one completely. All of the example problems explained in the preceding section should be done in this way. The purpose here is to model the types of activities that are needed to construct all types of trees and the instructor should use whatever instructional method he or she desires in order to do this. Once the distinct problem types have been modeled, the coaching phase can begin. The instructor now initiates a series of practice problems that may include examples just like the modeled problems or combinations of problem types. As the problem-solving progresses, the instructor suggests strategies that may be used to construct and analyze phylogenetic trees. Problem-solving activities should include both going from the matrix to the tree and also going from a fully constructed tree to writing a possible matrix. This practice of solving sample problems should continue until the instructor feels that students are ready to handle problems on their own. At this point, the teacher fades out of the picture and allows the students to solve further problems using any method they choose. During fading, the instructor only gives information when asked and he or she attempts to create a learning environment where this is kept to a minimum. (At this point, not every question asked is necessarily answered.) The student is often asked to construct his or her own way of solving the problems.

The use of the program *Phylogenetic Investigator* is optional but if the goal of the instruction is to practice problem solving while constructing phylogenetic trees, the software is ideal because it is open ended and forces the student to make choices during the process of tree construction. These choices may lead to a fruitful or an
unfruitful solution but the student is required to make a choice. Making the correct choices to properly construct phylogenetic trees is the outcome of this software. The instructor also will find that the software is an aid to making the printed trees more uniform and easier to read.

To achieve the desired outcome instruction requires four steps: (1) a combining of conceptual knowledge and the act of tree construction, (2) problems that encompass the range of phenomena studied, (3) models of problem-solving that lead to success, and (4) strategies that can guide tree construction across a full range of problem types. The results of this research can guide instruction in phylogenetic tree construction by providing insight into the nature of the problems, providing a model of the novice tree construction process, and providing important strategies to aid in the construction process. It is not necessary that the students have a wide conceptual view of evolution in order to construct phylogenetic trees. The problem-solving experience can be done whether or not one knows the doctrines of Malthus, Lamarck, and Darwin. Of course, a broad view of evolution and the doctrines that are key to understanding it would enhance instructional in phylogenetic tree construction.

Conclusion

This research provides insight into how novices create phylogenetic trees from a data matrix. It provides an overall model that includes the courses of action taken by participants while solving the research problems. Novices can be successful in this problem-solving activity provided they are given the proper instruction.
Further, this research can provide data that may be employed in the construction of curricula aimed at teaching evolution from a problem-solving perspective. It can also provide groundwork for subsequent research involving the construction of phylogenetic trees.
Appendix A

A Primer of Phylogenetic Terms Taken From
Steven Brewer's (1996) Dissertation
## A Primer of Phylogenetic Terms

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ancestor</td>
<td>A taxon, previous in time to a second taxon, from which the second taxon is descended.</td>
</tr>
<tr>
<td>Apomorphy</td>
<td>An evolutionary character, usually coded as “1”, that represents an evolutionary novel state.</td>
</tr>
<tr>
<td>Autapomorphy</td>
<td>The transition of a character that is uniquely evolutionarily novel (apomorphic) for a taxon.</td>
</tr>
<tr>
<td>Character</td>
<td>A recognizable feature that varies among taxa. For example, among ladybugs, the characters might include the presence or absence of spots. Characters are numbered, polarized, coded, and presented in a data matrix.</td>
</tr>
<tr>
<td>Clade</td>
<td>A monophyletic taxon.</td>
</tr>
<tr>
<td>Cladogram</td>
<td>A form of a phylogenetic tree that can only show sister-group relationships.</td>
</tr>
<tr>
<td>Conflict</td>
<td>A quality of characters that contain incompletely over-lapping distributions of apomorphies.</td>
</tr>
<tr>
<td>Convergence</td>
<td>A form of homoplasy whereby two taxa share a character that has appeared independently in separate lineages.</td>
</tr>
<tr>
<td>Data Matrix</td>
<td>A summary table of states with taxa in rows and characters in columns.</td>
</tr>
<tr>
<td>Descendant</td>
<td>A taxon which is the genealogical product of an earlier taxon.</td>
</tr>
<tr>
<td>Exclusive</td>
<td>Characters whose distributions of apomorphies do not overlap.</td>
</tr>
<tr>
<td>Homology</td>
<td>The quality of characters that are shared as the result of a common ancestry.</td>
</tr>
<tr>
<td>Homoplasy</td>
<td>Characters that are shared due to causes other than homology (evolutionary convergence or reversal).</td>
</tr>
<tr>
<td>Inclusive</td>
<td>When one character’s distribution of apomorphies is a superset of another character’s distribution of apomorphies.</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Ingroup</td>
<td>The group of taxa currently being studied using phylo-genetic inference.</td>
</tr>
<tr>
<td>Link</td>
<td>A line in between nodes in Phylogenetic Investigator that represents lines of ancestor/descendant relationships.</td>
</tr>
<tr>
<td>Monophyletic</td>
<td>A taxon that includes only the complete set of descendant taxa of an ancestral species.</td>
</tr>
<tr>
<td>Node</td>
<td>A circle in Phylogenetic Investigator used to represent a taxon.</td>
</tr>
<tr>
<td>Optimization</td>
<td>The process or product of distributing a homoplasious character on a phylogenetic tree.</td>
</tr>
<tr>
<td>Outgroup</td>
<td>A group of taxa used to polarize the character states.</td>
</tr>
<tr>
<td>Parallelism</td>
<td>A convergence.</td>
</tr>
<tr>
<td>Paraphyletic</td>
<td>A grouping of taxa that does not reflect the underlying evolutionary relationships by removing taxa from a monophyletic taxon.</td>
</tr>
<tr>
<td>Parsimony</td>
<td>A principle used to justify selecting the hypothesis that requires the fewest transitions. By assuming homology, one also selects the hypothesis that minimizes the number of assumptions of homoplasy.</td>
</tr>
<tr>
<td>Phylogenetic tree</td>
<td>A branching diagram that can illustrate both sister group and ancestor/descendant relationships among a set of taxa.</td>
</tr>
<tr>
<td>Phylogeny</td>
<td>The set of ancestor/descendant relationships that form the genealogy of a set of taxa. A phylogenetic tree is a hypothetical representation of these relationships.</td>
</tr>
<tr>
<td>Plesiomorphy</td>
<td>A form of a character (state) which is evolutionarily preexisting for the group of taxa under study (the ingroup).</td>
</tr>
<tr>
<td>Polarity</td>
<td>Whether a form of a character (a state) is considered apomorphic (evolutionarily novel) or plesiomorphic (evolutionarily preexisting). This is usually done through comparison with an outgroup.</td>
</tr>
<tr>
<td>Polyphyletic</td>
<td>A grouping of taxa that does not reflect the underlying evolutionary relationships by adding unrelated taxa to a monophyletic taxon.</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Reversal</td>
<td>The transition of a character that is apomorphic in some ancestor, changes polarity back to the plesiomorphic state resulting in a descendant taxa which are plesiomorphic for that character.</td>
</tr>
<tr>
<td>Sister group</td>
<td>The most closely related taxon to another taxon.</td>
</tr>
<tr>
<td>State</td>
<td>A form of a character that is polarized as either apomorphic or plesiomorphic and coded as “1” or “0”. For example, among ladybugs, the absence of spots might represent the plesiomorphic state and the presence of spots might represent the apomorphic state.</td>
</tr>
<tr>
<td>Steps</td>
<td>The number of transitions required to explain a character or characters.</td>
</tr>
<tr>
<td>Synapomorphy</td>
<td>The transition of a character that is homologously shared in the evolutionary novel (apomorphic) condition.</td>
</tr>
<tr>
<td>Taxon</td>
<td>A group of organisms that is given a name. The complete set of taxa descended from a common ancestor from a common ancestor is a monophyletic taxon. Incomplete sets are paraphyletic and sets with extra unrelated taxa are polyphyletic.</td>
</tr>
<tr>
<td>Topology</td>
<td>An arrangement of sister-group or ancestor/descendant relationships among a group of taxa.</td>
</tr>
<tr>
<td>Transition</td>
<td>A point in time in a lineage at which a character is hypothesized to have changed in state.</td>
</tr>
<tr>
<td>Treelength</td>
<td>The steps, or number of transitions, required to explain the data matrix using a phylogenetic tree.</td>
</tr>
</tbody>
</table>
Appendix B

Protocol Clearance from the Human Subjects

Institutional Review Board
Date: 28 March 1997

To: Robert Hafner, Principal Investigator
   Terrance Brisbin, Student Investigator

From: Richard Wright, Chair

Re: HSIRB Project Number 97-03-11

This letter will serve as confirmation that your research project entitled "An Account of Novice Phylogenetic Tree Construction for the Problem-Solving Research Tradition" has been approved under the exempt category of review by the Human Subjects Institutional Review Board. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now begin to implement the research as described in the application.

Please note that you may only conduct this research exactly in the form it was approved. You must seek specific board approval for any changes in this project. You must also seek reapproval if the project extends beyond the termination date noted below. In addition if there are any unanticipated adverse reactions or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the HSIRB for consultation.

The Board wishes you success in the pursuit of your research goals.

Approval Termination: 28 March 1998
Appendix C

Parental Consent Form
I understand that my child has been invited to participate in a research project entitled "Novice Phylogenetic Tree Construction From the Problem Solving Research Tradition". The purpose of this study is to describe what knowledge and strategies high school students use to construct phylogenetic trees. I further understand that the purpose of this project is to fulfill Mr. Brisbin's doctoral dissertation requirement.

My consent for my child to participate in this project means that my child, after two weeks of instruction in biology class, will meet with Mr. Brisbin for a two to four hour period, for no longer than one hour at a time, to solve seven problems involving phylogenetic tree construction and think aloud protocols. After each problem is solved, the researcher may ask clarifying questions.

I am aware that while my child is solving these problems, his or her actions will be computer recorded and his or her think aloud protocol will be tape recorded.

The benefits of this study lie in the fact that it will add a new dimension to the biology curriculum at Lee High School and participants will gain practice in the use of problem solving strategies.

I understand that no risks, hazards, or discomforts are foreseen as a consequence of this study. As in all research, there may be unforeseen risks to the participant. If accidental injury occurs, appropriate emergency measures will be taken; however, no compensation or treatment will be made available to the student except as otherwise noted in this consent form.

I understand that I may withdraw my child from this study at any time without any negative effect on the services to my child. If I have any questions about this study, I may contact Mr. Brisbin at 534-1038 or Dr. Robert Hafner at (616) 387-5844. I may also contact the Chair of the Human Subjects Institutional Review Board at (616) 387-8293 or the Vice President for Research at (616) 387-8296 with any concerns I may have.

My signature below indicates that I give my permission for __________________________ (child's name) to become a part of this research project by meeting, one on one, with Mr. Brisbin to solve phylogenetics problems with think aloud protocols.

__________________________  __________________________
Signature                     Date
Appendix D

Participant Consent Form
I have been invited to participate in a research project entitled: "Novice Phylogenetic Tree Construction From the Problem Solving Research Tradition". I understand that this study seeks to describe the knowledge and organization that high school students use to construct phylogenetic trees. I recognize that this study is Mr. Brisbin's dissertation project.

I agree to meet with Mr. Brisbin in the biology room or the computer lab for a two to four hour period to solve seven problems involving in the construction of phylogenetic trees. I understand that I will be asked to think aloud while solving these problems on the computer and that my "think aloud" protocol will be tape recorded. I further understand that I will be asked follow up questions after each problem is solved. I will be required to spend no more than one hour at a time with Mr. Brisbin and additional time will be scheduled until I solve all seven problems.

I understand that no risks, hazards, or discomforts are foreseen as a consequence of this study. As in all research, there may be unforeseen risks to the participant. If accidental injury occurs, appropriate emergency measures will be taken; however, no compensation or treatment will be made available to the student except as otherwise noted in this consent form.

I may benefit from this study by learning how to use the computer program Phylogenetic Investigator and how to construct phylogenetic trees from encoded data. I am also aware that the knowledge gained from this study may contribute to the teaching of high school biology.

I understand that information collected by Mr. Brisbin is confidential and my name will not appear on any papers on which this information is recorded. Students participating in the project will be identified by a code and a master list that shows the corresponding data will be kept in Dr. Robert Hafner's files. At the end of the project, the master list will be destroyed but the collected information will remain in Dr. Hafner's files for a period of three years.

I understand that I may refuse to participate or quit at any time, during the study, without penalty. If I have questions I may contact Mr. Brisbin at 534-1038 or Dr. Robert Hafner at (616) 387-5844. I may also contact the Chair of the Human Subjects Institutional Review Board at (616) 387-8233 or the Vice President of Research at (616) 387-8238. My signature below means that I understand the purpose of the study and agree to participate.

__________________________    __________________________
Signature                        Date

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
BIBLIOGRAPHY


