The Effects of Diagram Format on Students' Interpretations of Evolutionary Diagrams

Courtney K. Shade

Vanderbilt University

April 4, 2008

Running Head: Reasoning from Evolutionary Diagrams

Abstract

The present study researches university students' understanding of different forms of evolutionary diagrams. It is important to look at students' understanding of evolution so that teachers can use the most effective instructional method possible. College students demonstrate a very poor understanding of evolutionary processes, particularly when asked to interpret evolutionary relationships. Evolutionary biologists traditionally employ cladograms to demonstrate such relationships, but textbooks often give students other types of diagrams, which I label as "textbook" diagrams. The diagrams to be used in this study were found in contemporary high school and college textbooks. Subjects' ability to reason from these "bad" diagrams was compared to their ability to reason from two types of cladograms. Subjects were presented with three sets of evolutionary relationships, each in a different format- either as a textbook diagram, a cladogram tree, or a cladogram ladder and asked to reason from them. Subjects were divided into two groups based on the strength of their biology background. It was hypothesized that subjects with a stronger biology background would exhibit a better understanding of evolutionary relationships than subjects with a weaker biology background.

Textbooks use a variety of formats of biology diagrams to depict evolutionary relationships among taxa. These different diagrams appear in textbooks used in lower-level through university-level biology courses (Catley & Novick, 2008). Many of these diagrams are misleading because they can lead students to misinterpret the relationships presented and can suggest processes that are not in line with current evolutionary understanding (Catley, Novick, & Shade, 2008). These diagrams can be especially misleading because most students (even students in universities) do not have a firm grasp on the basic concepts of evolution (Bishop & Anderson, 1990; Catley et al., 2008; Settlage, 1994). These diagrams only serve to reinforce these misconceptions and to prevent students from correctly interpreting the important concepts and information.

Textbook diagrams, for the purpose of this study, are evolutionary diagrams that are not cladograms and tend to lead to a misinterpretation of the relationships represented (Catley & Novick, 2008). One such misinterpretation is the idea of the evolutionary process as anagenesis. Anagenesis is the incorrect reasoning that one species turns into another species and then ceases to exist. There is no branching in anagenesis—it is simply a change within a species that leads to the demise of the previous species while creating a new one (Catley et al., 2008; Wiley, Siegel-Causey, Brooks, & Funk, 1991). Cladogenesis, on the other hand, is the process by which species branch into two separate species through speciation and is the currently accepted understanding of evolution. Cladograms, which are the accepted and preferred diagram format among evolutionary biologists, represent evolutionary relationships through a series of branching events. Students often think of evolution through anagenesis (Bishop & Anderson, 1990), and these textbook diagrams help to guide students who are still in the process of internalizing the concepts of evolution (Settlage, 1994) to such misconceptions (Catley et al., 2008). See Figure 1 for an example of a set of evolutionary relationships (*hominid* relationships) represented in both cladogram and textbook formats.

Cladograms can be presented as either trees or ladders (see Figures 1b and 1c). Both formats present the relationships in roughly the same way (and are conceptually equal) but college students have a more difficult time interpreting ladder diagrams (Novick & Catley, 2007). This is understandable because of the Gestalt principle of good continuation, which states that a continuous line will be interpreted as one unit. This leads to an incorrect interpretation of the ladder diagram because the diagram is actually a hierarchy of levels rather than one single level even though it may not be intuitive to think of the ladder diagram in this way (Novick & Catley, 2007). Professional evolutionary biologists seem to prefer the tree format over the ladder format because over 80% of cladograms in a recent biology publication were presented as trees (Novick & Catley, 2007). Despite this obvious preference, ladder cladograms appear in textbooks used in classes from middle school through university level more often than tree cladograms (Catley & Novick, 2008).

Several studies have addressed college students' understanding of evolutionary concepts. Bishop and Anderson (1990) found that undergraduate students' (not biology majors) understanding of evolution differs from contemporary evolutionary understanding in several important ways. These students misunderstood the terms "adaptation" and "fitness" as they relate to evolution. Variability of characteristics within a population is also widely misunderstood, which is a very important component of contemporary understanding of the processes of evolution. Another misconception that is often found in undergraduate evolutionary thinking is the tendency to think of evolution as acting on individuals rather than species. Evolution is a process that acts on the species as a whole; an individual within a species does not evolve. The idea that evolution acts on individuals is inconsistent with contemporary evolutionary understanding and previous biology instruction has little to no effect on performance in this area (Bishop & Anderson, 1990).

Despite these large misunderstandings, it is possible to change students' misconceptions of evolution, although with great difficulty (Bishop & Anderson, 1990). Settlage (1994) studied college students' understanding of evolution and found that it is possible to correct students' misconceptions. In addition, he argued that extended exposure to evolutionary principles will help students to truly learn these concepts. It follows that appropriate evolution education should begin early in schooling and that children should be taught with proper diagrams that can be used to aid their understanding. Textbook diagrams surely serve to at least reinforce students' misconception, so the utilization of cladograms should lead to an increased ability to correctly understand evolutionary relationships. This study seeks to determine the type of diagram that students are best able to understand and interpret. By determining this, I hope to facilitate the best possible evolutionary understanding by influencing the type of diagrams that students encounter as they learn vital evolutionary concepts.

There has not been a lot of research in the area of evolutionary understanding as it relates to students' interpretations of various types of evolutionary diagrams. The bulk of the literature regarding evolution addresses either students' misunderstanding of evolutionary concepts (Bishop & Anderson, 1990; Demastes, Good & Peebles, 1995; Settlage, 1994) or students' belief in the truth and validity of the theory of evolution (Bishop & Anderson, 1990; Shtulman, 2006). Because of the paucity of research devoted to students' ability to interpret evolutionary diagrams, the bulk of the background research for this study is derived from the two experiments upon which this study is expanding.

The first of the two studies investigated the effect of textbook diagrams on college students' understanding of evolutionary relationships. This study presented undergraduate students with two of four possible textbook evolutionary diagrams and asked them to define specific aspects of the evolutionary relationships presented (Catley et al., 2008). The main goal of this study was to determine if students were able to correctly interpret evolutionary diagrams that are often encountered in evolution education. The diagrams in question were misleading because they could be interpreted in a way that is inconsistent with current evolutionary understanding, such as interpreting evolutionary relationships through an anagenic rather than a cladogenic perspective, which is how many subjects interpreted the relationships. These diagrams were found in contemporary textbooks and in the popular press and were found to be especially hazardous for students because these diagrams perpetuate common misconceptions that students hold as previously mentioned. It was hypothesized that students with stronger biology backgrounds would provide fewer responses that are in conflict with correct evolutionary understanding, but this was not the case. The additional biology experience did not seem to help stronger background subjects while interpreting evolutionary relationships from the textbook diagrams. This supports the conclusion that textbook diagrams reinforce common misconceptions even if the student has a competent understanding of the material (Catley et al., 2008). The present study plans to extend these results by presenting subjects with both textbook diagrams and cladograms and analyzing differences in performance to determine if students are able to reason effectively reason from cladograms or if students just do not understand evolutionary relationships.

The second of the two previous studies compared students' abilities to reason from ladder and tree cladograms. This study found that undergraduate students perform better with tree cladograms than ladder cladograms and that weaker background subjects have a much more difficult time with ladder cladograms than do stronger background students (Novick & Catley, 2008).

The present study extends and combines the past work by comparing students' abilities to reason about specific evolutionary relationships from textbook diagrams as well as corresponding tree and ladder cladograms. Although even students with stronger backgrounds in biology (who, for the purposes of this study, have completed at least the first semester of a two-semester introductory biology course that is intended for serious science students) perform poorly when reasoning from textbook diagrams, it is my hypothesis that they do so because these diagrams are so misleading that they interfere with the knowledge that students already possess about evolution and evolutionary processes. If stronger and weaker background students are asked to reason from appropriate cladograms, weaker background students should perform poorly while the performance of stronger background students should increase (Novick & Catley, 2008). This is because stronger background students are not distracted by the confusion of the diagram and are therefore able to focus on the information presented. In addition to this hypothesis, I also expect to find that students perform better on tree cladograms than ladder cladograms regardless of biology background. This would suggest that the information provided by and/or inferred from evolutionary diagrams contributes to a student's interpretation of the relationships presented. It would also suggest that certain diagrams provide information that is more likely to be correctly interpreted than other diagrams. In other words, certain diagrams lead students to reason more correctly about evolutionary relationships than others (Catley & Novick, 2008) and I hope to determine which diagrams are the most beneficial for college students.

I hope that the present study will provide evidence that students perform significantly differently on the three diagrams formats, in particular that students interpret evolutionary relationships more accurately from cladograms than from textbook diagrams that are often found in textbooks. I expect to find that students will perform the best on the tree diagram and slightly less well when asked to reason from ladder diagrams. Hopefully these findings will bring about a change in the types of diagrams in textbooks that are used to teach evolutionary relationships and processes. If students are taught to read and understand the proper evolutionary diagrams, cladograms, they will be better able to glean the proper information from the diagrams (Baum, Smith & Donovan, 2005; Catley & Novick, 2008; Wiley et al., 1991). The overall goal of this study is to determine which diagram format leads to the most accurate understanding by students with both stronger and weaker backgrounds in biology.

Method

Subjects

The subjects were 86 students from Vanderbilt University. The students participated in partial fulfillment of course requirements for introductory psychology (10 females, 5 males) or evolutionary biology (12 females, 18 males), or for extra credit in the psychology (21 females, 2 males) or education (15 females, 3 males) class from which they were recruited.

Subjects completed a background information questionnaire which asked about their biology coursework. Subjects were asked if they had taken any of 12 primarily organismal biology and 3 relevant geology courses at Vanderbilt. Based on these responses, subjects were assigned to either the stronger or weaker background condition. Stronger background subjects had taken at least BSCI 110a, the first semester of the two-semester introductory biology sequence for biology majors, pre-med majors, and other serious science students; most had taken more. Weaker background subjects, on the other hand, had not taken BSCI 110a and had little exposure to biology in general. The 33 stronger background students (14 females, 19 males) had taken an average of 3.62 semesters of biology (or relevant geology) classes, and the 53 weaker background students (44 females, 9 males) had taken an average of 0.45 semesters of such coursework. The responses of one additional subject (not included in the 86 subjects previously mentioned) are excluded from the analyses because he or she did not fill out the background questionnaire and therefore could not be assigned to a biology background group.

Materials

The textbook diagrams presented in this study were found in contemporary high school and college textbooks. Each diagram presents the taxa in a way that could potentially lead to a misinterpretation of the evolutionary relationships among the depicted taxa, which will be discussed in more detail later. Each diagram was then translated into both tree and ladder cladogram formats.

Each set of relationships—horse, primate, and hominid—was followed by a set of 3-5 questions. Within each set, subjects received the same questions regardless of diagram format. The only exception was an additional question given to subjects who received the horse taxa on the textbook diagram.

Hominid diagrams. Diagrams depicting hominid evolution abound in textbooks and are often presented in non-cladogram formats (Catley &Novick, 2008). The diagram used in the present study was found in two high school textbooks (Johnson, 1998; Johnson & Raven, 2001). This diagram seems to lead students to an anagenic interpretation of the taxa (Catley et al., 2008). Figure 2 shows the textbook, ladder, and tree formats. The questions following each diagram were: (1) "Which taxon—*H. habilis* or *H. sapiens*—is the closest evolutionary relation to *H. erectus*? What evidence supports your answer?"; (2) "The diagram splits at the point indicated by the dashed circle. How do you interpret this split?"; (3) "*A. robustus* and *A. boisei* are joined together. What does this tell you about the evolutionary relationship between these two taxa?"; and (4) "What is the evolutionary relationship between *A. afarensis* and all the rest of the taxa?"

Primate diagrams. The diagram depicting primate evolutionary relationships was found in a biology textbook intended for non-majors (Campbell, Reece, & Simon, 2004). From a brief glance, this diagram almost looks like a ladder cladogram, but there are some important differences (see Figure 3). The lines that extend to the taxa are bent, and there is a bend in the "main line," which is marked with an "X" in order to probe students' understanding of this feature. The questions following each diagram were: (1) "Which taxon—lemurs or gibbons—is the closest evolutionary relation to new world monkeys? What evidence supports your answer?"; (2) "Draw a circle around the two taxa on the diagram that are most closely related to each other. What evidence supports your answer?"; and (3) "An "X" marks a place along one of the lines on the diagram. What does this point on the diagram represent?"

Horse diagrams. The diagram depicting the evolution of horse taxa was found in a textbook used for a Human Biology course at Vanderbilt University (Chiras, 2002). It depicts the taxa in a way that could lead students to interpret the relationships as anagenesis. The taxa in this diagram are presented in a way that could seem to be a progression of one species turning into another, such as the line connecting *Eohippus* and *Miohippus*. The diagram includes many "nodes" and "branches" whose purposes are unclear. Several of these branches end abruptly, which could be confusing for students. Figure 4 includes each diagram format. The questions following each diagram were: (1) "Which taxon—*Equus* or *Miohippus*—is the closest evolutionary relation to *Merychippus*? What evidence supports your answer?"; (2) "Notice that multiple lines extend up from the location marked by the dotted circle on the diagram. How many lines extend up from this location? What is the relationship among these lines?"; (3) "The lines leading from *Merychippus*, *Z*, and *Pliohippus* all meet at a common place. What does this tell you about the relationship among these three taxa?"; (4) "Scientists have recently discovered that *Miohippus* had a novel shaped bone in its middle ear. Is *Eohippus* or *Pliohippus* more likely to also have had this bone in its middle ear? What evidence supports your answer?"; and, for the textbook diagram only, (5) "Near the bottom of the diagram is an arrow. How do you interpret the indicated part of this diagram?"

Design

There were two independent variables in this experiment—biology background and diagram format. As described earlier, the format manipulation was implemented with three different sets of taxa. Subjects were presented with three diagrams, each depicting a different set of evolutionary relationships among taxa—i.e., horse, hominid, and primate taxa—and a different diagram format—textbook, tree cladogram, and ladder cladogram. This resulted in six possible diagram conditions. Each of the six diagram conditions was arranged in the six possible orderings, yielding 36 booklet outcomes. For example, subjects in one condition received the tree diagram with horse taxa, then the ladder diagram with primate taxa, and finally, the textbook diagram with hominid taxa.

The 36 possibilities were randomly assigned to subjects. As a result of this random assignment, the number of stronger background subjects who received the textbook diagram ranged from 8-14 for the three sets of taxa. Between 8 and 15 subjects received the ladder diagram and 7-15 subjects received the tree diagram. For weaker background subjects, 15-20 received the textbook diagram, 16-20 received the ladder diagram, and 16-19 received the tree diagram.

Procedure

This study was Part II or Part III of a larger, multi-part problem booklet. Subjects completed this paper and pencil booklet at their own pace and required between 50 and 70 minutes for completion. Each part of the booklet addressed different theoretical questions. After completing the three problem booklets, subjects were asked to fill out the background questionnaire. The questionnaire provided information about subjects' biology background as well as information such as year in school and career plans.

Results

Overview of the Analyses

The results from the first question for each diagram (the evolutionary distance question) will be presented first because they are equivalent. The results of the hominid, primate, and horse diagrams will follow in this order. For each set of taxa, each question will be presented individually in order of appearance. It was predicted that stronger background subjects would be more accurate than weaker background subjects and that stronger background subjects would provide stronger justifications for their answers. It was also predicted that subjects would be more accurate with cladograms than textbook diagrams and that subjects would perform better on tree rather than ladder cladograms.

As a result of the varying formats of questions posed to subjects, responses were coded in several ways. The objective questions were coded as either correct (1) or incorrect (0). All of questions asked subjects to explain or justify their answers in a free response format. These responses were content-coded based on pre-determined codes, which were then given a score of

0, 0.25, 0.5, or 1. This score was based on the quality of the justification given and helped to analyze the quality of reasoning that subjects offered. These two quantitative scores were averaged into a combined score. This combined score represents subjects' accuracy as well as the strength of their reasoning. This is helpful because some subjects may answer the questions correctly but are using incorrect thought processes in order to get to their answers. This combined score represents subjects' overall understanding more completely than either individual code. For each question, these combined scores were evaluated by a 3 (textbook vs. ladder vs. tree diagram) X 2 (stronger vs. weaker biology background) between subjects ANOVA.

As previously mentioned, each free response answer received a code based on the content of the response. The codes for each question were determined by the responses most often provided by subjects and were ranked according to accuracy and relevance. Each response received only one code. Responses may have mentioned multiple possible codes but were given the code that was ranked highest. Most codes appear in many questions and will be explained with the question in which they first appear and then will be referenced in each succeeding question when applicable. These data were evaluated by a justification code X biology background chi-square analysis and a justification code X diagram type (textbook vs. ladder vs. tree) chi-square analysis. Where possible, an analysis of subjects' accuracy and justifications will be reported through a justification code X response accuracy chi-square test. The number of codes differs for each question, but the same test was run for each.

Evolutionary Distance Question

Overview. This is the only question that was asked of each set of taxa. Subjects were given a reference taxon (such as *homo erectus*) and asked which of two taxa (such as *homo* *habilis* and *homo sapiens*) is a closer evolutionary relation to the reference taxon. Subjects were then asked to provide evidence to support their answer. This question required subjects to interpret the relationships represented in a way that requires more reasoning than simple proximity of the taxa on the diagram. The closest relation was generally farther (in a strictly linear distance sense) from the reference taxon on the diagrams. Unfortunately, for the textbook hominid diagram the correct answer was minutely closer to the reference taxon, which may have led subjects to the correct answer through incorrect reasoning.

Codes Used. Seven codes were used in order to classify subjects' justifications for this question. All codes are listed and defined in Table 1 for easy reference. If a subject mentioned a most recent common ancestor he or she was coded as MRCA and received full credit because it was the most correct answer. An example of this code is "They share a more recent common ancestor" (Subject 7F062, stronger background). Subjects received the DIV code if they mentioned splitting, branching, or diverging through the evolutionary process and received half (0.5) credit. For example, "They split more recently than did Miohippus and Merychippus" (Subject 7F055, weaker background).

The rest of the codes received no credit. They are described hear in decreasing order of quality. If a subject referred to a degree of commonality between taxa, the COM code was applied. For example, "shares some trait that lemurs do not" (Subject 7F075, stronger background). The ANAG code was given to any response that referred to evolution as an anagenic process, such as "Because Lemurs form into new world monkeys." (Subject 7F049, weaker background). The TIME code was given to subjects who referred to evolution occurring in time or to the ordering of evolution in time, such as "It comes before it in the evolutionary progression." (Subject 7F012, weaker background). If subjects referred to proximity on the diagram they received the DIST code, such as “Miohippus is closer to Merychippus than Equus” (Subject 7F030, weaker background). If they referred to similarity of the pictures in the diagram, such as “The size and shape of h. habilis looks more similar than h. sapiens” (Subject 7F008, weaker background), they received the PICT code. A final code, OTH, was given if a response did not fall within the previously mentioned categories.

Hominid Diagram. A 3 (textbook vs. ladder vs. tree) X 2 (stronger vs. weaker biology background) between-subjects ANOVA on subjects’ combined scores (the average of each subject’s accuracy and evidence quality) found a main effect for biology background, $F(1, 80) = 36.56$, $p < .001$, $MSE = 0.07$. Stronger background subjects ($M = .66$) performed better than weaker background subjects ($M = .35$).

Subjects also performed differently across the diagram conditions, $F(1, 80) = 7.89$, $p < .001$, $MSE = 0.07$. Subjects performed poorly when reasoning from the textbook diagram ($M = .36$), better with the ladder diagram ($M = .48$), and even better with the tree diagram ($M = .56$). In addition, there was a significant interaction between biology background and diagram condition, $F(1, 80) = 4.41$, $p < .05$, $MSE = 0.07$ (see Figure 5). Stronger background subjects performed very well when reasoning from cladograms ($M = .88$ and $M = .80$ for ladder and tree diagrams). Stronger background subjects in the textbook condition ($M = .43$), however, were more similar to weaker background subjects in all three conditions ($M = .35$) than to stronger background subjects reasoning from cladograms.

Stronger and weaker background subjects provided different justifications, $X^2(6, N = 86) = 29.70$, $p < .001$. The data are presented in Table 2. Weaker background subjects provided 78.6% of anagenic and 90% of distance responses. Stronger background subjects provided 100%

of the most recent common ancestor codes and 60% of codes mentioning divergence. Subjects' justifications did not differ as a function of diagram condition, $X^2(12, N = 86) = 19.93, p > .05$.

An analysis of the content of subjects' reasoning found that their justifications differed based on accuracy, $X^2(6, N = 86) = 19.91, p < .01$. The data are presented in Table 3. Of subjects who answered incorrectly, 42.9% provided anagenic responses, compared to a total of only 7.7% among those who answered correctly. Divergence and most recent common ancestor were only mentioned by subjects who answered this question correctly.

Primate Diagram. Stronger background subjects' combined scores ($M = .49$) were higher than those of weaker background subjects ($M = .20$), $F(1, 80) = 16.64, p < .001, MSE = 0.09$. In addition, there was a main effect of diagram condition, $F(2, 80) = 6.05, p < .01, MSE = 0.09$. Surprisingly, subjects in the textbook condition ($M = .47$) performed better than subjects in either cladogram condition ($M = .26$ for ladder and $M = .18$ for tree). The interaction between biology background and diagram condition was not significant, $F(2, 80) = 0.01, p > .05, MSE = 0.09$.

The main effect of biology background difference was primarily due to stronger background subjects providing better justifications for their answers, $X^2(6, N = 86) = 45.61, p < .001$ (see Table 2). The ANAG and PICT code were both only given to weaker background subjects, while only stronger background subjects mentioned divergence or a most recent common ancestor. Weaker background subjects were much more likely to reason according to proximity (86.7%) than stronger background subjects (13.3%). Subject's justifications did not differ as a function of diagram condition, $X^2(12, N = 86) = 16.35, p > .05$.

In addition, subjects' explanations differed according to accuracy, $X^2(6, N = 86) = 18.51, p < .01$ (see Table 3). Explanations referring to proximity on the diagram accounted for 44.7% of the explanations for subjects who answered incorrectly. A most recent common ancestor was mentioned by 26.3% of subjects who answered this question correctly and only 4.3% of subjects who answered incorrectly.

Horse Diagram. Similar to the results for the primate diagram, biology background had an effect for subjects' combined scores, $F(1, 80) = 16.41$, $p < .001$, $MSE = 0.08$. Stronger background subjects received higher scores ($M = .56$) than weaker background subjects ($M = .26$), which fits the pattern found previously. There was also a main effect of diagram condition on combined scores, $F(2, 80) = 9.00$, $p < .001$, $MSE = 0.08$. Similar to performance on the hominid diagram, subjects performed worst when reasoning from the textbook diagram ($M = .18$), better with the ladder diagram ($M = .39$), and best with the tree diagram ($M = .55$). There was not an interaction between biology knowledge and diagram condition, $F(2, 80) = .45$, $p > .05$, $MSE = 0.08$.

As can be seen from the difference in combined scores, biology background had an effect on the quality of justifications that subjects offered, X^2 $(6, N = 86) = 28.82$, $p < .001$ (see Table 2). Weaker background subjects mentioned 75% of the total anagenic responses, 91.7% of the total picture responses, and 80% of the total proximity responses, while stronger background subjects accounted for all of the responses that mentioned a most recent common ancestor (12 total).

Subjects' justifications did not differ as a result of diagram condition, X^2 $(12, N = 86) = 14.01$, $p > .05$. Although not statistically significant, some of the patterns from the hominid and primate diagrams were found. Subjects in the cladogram conditions accounted for 91.7% of the total MRCA responses and 50% of the PICT responses came from subjects in the textbook condition.

Similar to the previous diagrams, subjects' explanations differed according to accuracy, $X^2(6, N = 86) = 13.46, p < .05$ (see Table 3). Subjects who answered incorrectly accounted for 80% of the total DIST codes and 75% of the total ANAG codes, whereas subjects who answered correctly accounted for 91.7% of the DIV codes and 58.3% of the MRCA codes. Although the latter difference is not incredibly large, it is in the same direction as was found in the previous questions.

Hominid Diagram

Question 2. This question directed subjects' attention to a circle drawn around a split on the diagram and asked subjects to interpret the meaning of the split. The codes used to describe subjects' justifications were, in decreasing order of quality: MRCA, ANC, SPEC, REL, SC, AH, and OTH. As noted earlier, only the new codes are described here. Definitions of all codes can be found in Table 1. Subjects received the ANC code if they mentioned a common ancestor—for instance, "The common ancestor at this node evolved into two lineages most likely due to some kind of selective pressure" (Subject 7F066, stronger background). Because of the small number of subjects receiving the MRCA and ANC codes, these related codes were combined for the analyses. Subjects received the SPEC code if they referred to the development of a new species, for example: "Something led to the formation of a new species" (Subject 7F068, stronger background). The REL code was applied if subjects mentioned the degree of evolutionary relationship between two taxa. An example of this code is "the taxa that are included in the split are are closely related and more similar than the H. sapiens" (Subject 7F014, weaker background). The mention of shared characters or characteristics received the SC code. An example of this code is "some characteristic is only shared by those of each new line" (Subject 7F075, stronger background). The REL and SC codes were also combined for the analyses because they received the same quality score (0.25). The AH code was applied if subjects made a distinction between apes and humans—for instance, "the split between monkey and human" (Subject 7F054, weaker background).

Subjects' responses did not differ based on diagram condition, $X^2(8, N = 86) = 4.35, p > .05$, but stronger and weaker background subjects provided different justifications for their answers, $X^2(4, N = 86) = 19.97, p < .01$ (see Table 4). Of the stronger background subjects, 30.3% received either the MRCA or ANC code, but no weaker background subjects did. Also, 17% of weaker background subjects justified their answer with the AH code, whereas only 9.1% of stronger background subjects received this code.

Question 3. This question directed subjects to the fact that two taxa (*A. robustus* and *A. Boisei* for the cladograms and *A. robustus* and *A. africanus* for the textbook diagram) are joined together and asked subjects to explain the evolutionary relationship between those taxa. The codes used for this question were MRCA, ANC, REL, SC, ANAG, and OTH. As for Question 2, REL and SC were combined for the analyses of this question. Biology background had an effect on subjects' responses, $X^2(4, N = 86) = 19.22, p < .01$ (see Table 5). Eleven of the 15 subjects who provided anagenic responses (73.3%) were in the weaker background condition while 66.7% of ANC responses and 88.9% of MRCA responses were made by stronger background subjects.

Diagram condition also had an effect on subjects' responses, $X^2(8, N = 86) = 23.20, p < .01$. Of subjects in the textbook condition, 41.4% provided anagenic responses, compared with only 10.7% and 0% subjects in the ladder and tree conditions. Of subjects in the tree condition, 17.2% received the MRCA code, compared to only 6.9% and 7.1% of subjects in the textbook and ladder conditions. In addition, 42.9% of subjects in the ladder condition mentioned the degree of relationship between taxa or shared characters while only 24.1% of subjects in the textbook condition and 27.6% of subjects in the tree condition provided such a response.

Question 4. The final question for this set of taxa asked subjects to describe the relationship of *A. afarensis* to the rest of the taxa on the diagram. The codes used for this question were OUTG, MRCA, ANC, ANAG, TIME, and OTH. Subjects received the OUTG code if they mentioned the word *outgroup,* which is a critical concept in phylogenetics. The outgroup is a taxon at the base of the cladogram that is less derived than all of the other taxa on the diagram. In the cladograms used here, *A. afarensis* is the outgroup. An example of this code is “It is the outgroup. Less related than any others are to each other.” (Subject 7F046, stronger background). OUTG and MRCA were combined for the analyses because they both represent a sophisticated understanding of evolutionary processes (and both received a quality score of 1).

There was no difference in responses based on diagram condition, $X^2(4, N = 86) = 10.57$, $p > .05$, but stronger and weaker background subjects provided different responses, $X^2(4, N = 86) = 36.75$, $p < .001$ (see Table 6). Of the weaker background subjects, 32.1% received the anagenesis code, compared with only 6.1% of stronger background subjects. Looking at the data another way, weaker background subjects received 89.5% of the anagenesis codes given. These subjects also received 93.3% of time codes. Although there were no statistical differences for diagram condition, it is worth noting that the two stronger background subjects who mentioned anagenesis received the textbook diagram, and 46.7% of weaker background subjects who received the anagenesis code were in the textbook condition and 35% were in the ladder condition. The OUTG and MRCA codes were given only to stronger background subjects and 76% of subjects who received the ANC code had a stronger biology background.

Primate Diagram

Question 2. This question required subjects to circle the two taxa on the diagram that are the most closely related and to provide evidence to support their answer. The codes used to analyze this question were MRCA, ANC, DIV, REL, CONN, PICT, TIME, KNOW, and OTH. The CONN code was applied if subjects said that the taxa are linked or connected together—for example, "branch off from same spot" (Subject 7F027, weaker background). Subjects received the KNOW code if they referred to knowledge that was not present in the diagram, such as "Research on chimp and human brains shows direct similarities and patterns between the two." (Subject 7F011, weaker background).

The correct answer to this question was humans and chimpanzees because they share a most recent common ancestor, but 51.2% of subjects answered incorrectly. Gorillas, chimpanzees, and humans stem from the same point on the textbook diagram, which could easily be misinterpreted. Biology background had an effect on subjects' combined scores, $F(1, 80) = 29.23$, $p < .001$, $MSE = 0.06$. Stronger background had higher scores ($M = .53$) than weaker background subjects ($M = .23$). Diagram condition also had an effect on subjects' combined scores, $F(2, 80) = 12.93$, $p < .001$, $MSE = 0.06$. Subjects performed very poorly on the textbook diagrams ($M = .16$), better on the ladder diagrams ($M = .43$), and even better on the tree diagrams ($M = .46$). There was no interaction between biology knowledge and diagram format, $F(2, 80) = 2.74$, $p > .05$, $MSE = 0.06$.

Biology background also had an effect on subjects' justifications, $X^2(8, N = 86) = 26.86$, $p <.01$ (see Table 7). All eight subjects who answered the question by reasoning from the pictures were weaker background and answered the question incorrectly. Six of the seven subjects (85.7%) who reasoned from prior knowledge were weaker background subjects.

Stronger background subjects accounted for 70% of the total divergence responses and weaker background subjects accounted for only 30% of such responses. See Table 6 for the percentages of stronger and weaker background subjects' responses.

Diagram condition had an effect on the type of mistakes that subjects made, $X^2(4, N = 44) = 13.82, p <. 01$. Of the 44 subjects who answered this question incorrectly, the most common mistake was for subjects to circle chimpanzees and gorillas rather than chimpanzees and humans. Subjects in the textbook condition accounted for 76% of subjects who circled chimpanzees and gorillas. The next most common response was for subjects to circle new world monkeys and old world monkeys. Subjects in the textbook condition accounted for 70% of these responses and the final 30% were in the ladder condition. Responses other than these two common mistakes accounted for 20.5% of the incorrect answers. Subjects in the ladder condition make up 55.6% of these other responses. Overall, 61.4% of the incorrect answers occurred in the textbook diagram condition, 25% in the ladder condition, and 13.6% in the tree condition. The most common type of reasoning (behind OTH responses) for subjects who answered this question incorrectly was to reason according to similarities in the pictures (see Table 7). Each subject who received this code ($n = 8$) answered this question incorrectly. This difference in reasoning is again highlighted when looking at subjects who received the MRCA code. Of these subjects, 71.4% answered the question correctly. From these data, it appears that subjects who reason according to evolutionary accurate principles are much more likely to correctly understand these diagrams than subjects who reason in other ways.

Question 3. This question directed subjects' attention to an "X" that was placed on the diagram and asked to explain what the "X" represented. The codes used for this question were NCHAR, MRCA, ANC, DIV, CHANGE, HUM, and OTH. Subjects received the NCHAR code if they mentioned the acquisition of a new characteristic, such as "The X marks a place in which a new character was formed." (Subject 7F072, stronger background). NCHAR and MRCA were combined for the analysis of this question because they are both high quality responses. The CHANGE code was given if a subject's response mentioned the word 'change'. An example of this is "A distinct change in the evolution of orangutans." (Subject 7F038, weaker background). If a subject mentioned humans or the process of becoming human-like, he or she received the HUM code—for instance, "The X marks where evolution away from "monkeys" and more towards humans began." (Subject 7F041, weaker background).

Diagram condition did not have an effect on subjects' responses, $X^2(8, N = 86) = 6.10, p > .05$, but stronger and weaker background subjects answered this question differently, $X^2(4, N = 86) = 31.39, p < .001$ (see Table 8). Of stronger background subjects, 21.2% referred to a common ancestor or divergence and 39.4% of stronger background subjects mentioned the most recent common ancestor or the acquisition of a new character. In addition, weaker background subjects accounted for all of the CHANGE ($n = 8$) and HUM codes ($n = 9$).

Horse Diagram

Question 2. This question asked subjects to count the number of lines that extend from the dotted circle on the diagram. Subjects were then asked to explain the relationship among these lines. The codes used for this question were MRCA, SG, ANC, DIV, REL, SC, ANC, and OTH. The SG code was given if subjects said that the taxa are sister groups, such as "They are sister groups." (Subject 7F072, stronger background). MRCA and SG were combined for the analysis of this question because they both received full credit. REL and SC were also combined for the purposes of this analysis because they both received quarter (0.25) credit.

Four lines extend from the circle on the textbook diagram and three lines extend from the cladograms. Stronger background subjects' combined scores were higher than weaker background subjects' ($M = .60$ and $M = .35$, respectively), $F(1, 80) = 18.63$, $p < .001$, $MSE = 0.06$. Subjects' combined scores also differed as a result of diagram condition, $F(2, 80) = 8.04$, $p < .01$, $MSE = 0.06$. Subjects in the textbook ($M = .54$) and tree conditions ($M = .52$) performed better than subjects in the ladder condition ($M = .28$). There was not an interaction between biology background and diagram condition, $F(2, 80) = .11$, $p > .05$, $MSE = 0.06$. See Table 9 for the pattern of subjects' responses.

Biology background had an effect when comparing subjects' justifications, $X^2(5, N = 86) = 40.31$, $p < .001$ (see Table 10). All seven of the subjects who provided anagenic responses were in the weaker background condition, whereas all eight of the subjects who referred to a most recent common ancestor or a sister group were in the stronger background group. Stronger background subjects represented 70.6% of responses that mentioned a common ancestor and 83.3% of responses that referred to divergence.

Subjects' justifications also differed based on diagram condition, $X^2(10, N = 86) = 20.06$, $p < .05$ (See Table 10). For the textbook diagram, anagenesis was mentioned in 21.4% of responses and in only 3.7% of responses for the ladder diagram and 0% of responses in the tree diagram. Also, 85.7% of the total anagenic responses were in the textbook condition. Subjects in the tree condition mentioned a common ancestor 29% of the time and 12.9% mentioned a most recent common ancestor or sister group. Shared characters and the degree of evolutionary relationship accounted for 25.9% of responses in the ladder condition.

Question 3. This question referred subjects to three taxa—*Merychippus, Z,* and *Pliohippus*—that are joined together and asked subjects to describe the relationship among these taxa. The codes used for this question were MRCA, ANC, REL, SC, and OTH. REL and SC were combined for the analysis of this question. Subjects' biology knowledge produced an effect on subjects' justifications, $X^2(3, N = 86) = 33.76, p < .001$ (see Table 11). Subjects with a stronger biology background mentioned a common ancestor in 48.5% of the responses and the mention of a most recent common ancestor accounted for 24.2% of stronger background subjects' responses. Of subjects with a weaker background, 37.7% mentioned the degree of evolutionary relation or shared characters, while none mentioned a most recent common ancestor and only 13.2% mentioned a common ancestor. There was not a significant difference between diagram conditions, $X^2(6, N = 86) = 6.59, p > .05$.

Question 4. This question asked subjects *Eohippus* or *Pliohippus* would be more likely to have a novel shaped bone that was found in *Miohippus*. The codes used to analyze this question were MRCA, DIV, REL, ANAG, TIME, and OTH. There was a main effect of biology knowledge for subjects' combined scores, $F(1, 80) = 22.17, p < .001, MSE = 0.08$. Stronger background subjects ($M = .59$) did better than weaker background subjects ($M = .25$). Subjects' combined scores, $F(2, 80) = 1.59, p > .05, MSE = 0.08$, did not differ as a function of diagram condition and there was not a significant interaction between biology knowledge and diagram condition for combined scores, $F(2, 80) = 1.99, p > .05, MSE = 0.08$.

Stronger and weaker background subjects justified their answers differently, $X^2(5, N = 86) = 36.78$ (see Table 12). Stronger background subjects accounted for each of the 11 subjects who mentioned MRCA and 87.5% of subjects who referred to divergence. Twenty one (39.6%) weaker background subjects mentioned time, compared to only two stronger background subjects. Subjects' justifications did not differ across diagram conditions, $X^2(10, N = 86) = .29, p > .05$.

Question 5. This question was asked only of subjects who received the textbook diagram. Subjects were asked to interpret the meaning of an arrow that was located near the bottom of the diagram. The codes used to analyze this question were CHANGE, OTH, and TIME. Because of the small amount of subjects who received this question, the results will be discussed descriptively (see Table 13). Of the 28 subjects who answered this question, there were eight stronger background and 20 weaker background subjects. Three stronger background subjects (30%) and seven weaker background subjects (70%) mentioned change; one stronger (16.7%) and five weaker background subjects (83.3%) mentioned time. The remaining four stronger background subjects (33.3%) and eight weaker background subjects (66.7%) received the OTH code.

Discussion

Although the population size for this study was not ideal (especially when considering stronger background subjects), there are consistent patterns in the data that support my hypotheses. The data from this study support the hypothesis that non-cladogenic textbook diagrams contribute to students' misunderstandings of evolution. In general, subjects performed best when reasoning from tree cladograms and slightly less well when reasoning from ladder cladograms. Subjects produced the least sophisticated responses when reasoning from the textbook diagrams. This tendency affected both stronger and weaker background subjects. Subjects who have a strong biology background are susceptible to the effects of the textbook diagram even when they demonstrate higher quality understanding when reasoning from cladograms.

In addition to these differences, diagram condition had an effect on the accuracy and quality of subjects' responses. Stronger background subjects consistently gave higher quality justifications when compared to weaker background subjects, as predicted. Stronger background subjects were more likely to provide high quality justifications—such as mentioning a most recent common ancestor, divergence, and the acquisition of a new character. They also consistently performed better with cladograms than with the misleading textbook diagrams both in accuracy and justification quality. For instance, subjects in the cladogram conditions were much more likely to circle the correct taxa on Primate Question 2. Also, subjects reasoning from cladograms provided stronger justifications and mentioned a most recent common ancestor much more often. Subjects reasoning from textbook diagrams mentioned anagenesis more often than subjects in reasoning from cladograms. Although there are exceptions to this pattern, it is pervasive enough to make such a statement.

Weaker background subjects were consistently less accurate when answering questions about these diagrams and provided lower quality justifications such as anagenesis, distance on the diagram, and reasoning from similarities in the pictures. For example, for six of the seven questions for which anagenesis was coded, a higher proportion of weaker than stronger background subjects' responses received this code. It would appear that if students do not have a firm grasp on evolutionary concepts and exposure to proper evolutionary diagrams, they are unable to correctly reason from evolutionary diagrams and instead resort to superficial explanations such as those just mentioned. This result is consistent with research from the cognitive literature showing that novices in a domain tend to focus on superficial features to support their reasoning and problem solving (e.g., Feltovich, Prietula, & Ericsson, 2006).

The diagrams that offered a cladogeneic representation of relationships received many fewer anagenic interpretations, which supports the hypothesis that cladograms prevent the inappropriate interpretation of evolutionary processes. If the goal of evolutionary diagrams is to improve students' interpretations of relationships, then cladograms are the best option for educators to effectively teach this material. Tree thinking should be a priority among science educators in order to facilitate the best possible understanding for students who are learning the often difficult concepts of evolution (Baum et al., 2005; Catley et al., 2008).

Because it is so difficult to change students' preconceived notions of evolution (Settlage, 1994), it is important to make sure that students are learning the concepts correctly at an early age (Settlage, 1994). An integral part of ensuring that students properly understand the concepts of evolution is to master the concept of cladogenesis, in which diagrams play an integral role. If students do not have a complete understanding of evolutionary processes, they are likely to misinterpret all diagrams, especially diagrams that are misleading, such as the diagrams used in the previous and present study (Catley et al., 2008). As shown by the present results, an appropriate diagram that encourages students to draw conclusions that are consistent with current evolutionary understanding does not lead its readers to interpret evolution as a process of anagenesis, but rather through cladogenesis.

Although it may seem impossible for young students can learn and master the difficult concepts of evolution, it has been found that they are able to do so when instructed properly (Bishop & Anderson, 1990; Settlage, 1994). It is more effective to teach difficult concepts such as evolutionary processes to students of a younger age (Gelman & Brenneman, 2004), and proper evolutionary diagrams are very important to an accurate understanding of such processes (Catley & Novick, 2008). When teaching evolution, cladograms need to be given as much emphasis as the evolutionary processes themselves. By providing students with the tools to properly interpret evolutionary diagrams, science educators are giving students a solid base to further understand more difficult evolutionary processes in the future (Baum et al., 2005).

References

Bishop, B. A. & Anderson, C. W. (1990). Student conceptions of natural selection and its role in evolution. *Journal of Research in Science Teaching, 27(5)*, 415-427.

Baum, D. A., Smith, S. D. & Donovan, S, S. (2005). The Tree-Thinking Challenge. *Science, 310*, 979-980.

Campbell, N. A., Reece, J. B., Simon, E. J. (2004). *Essential Biology* (2nd ed.). San Francisco (CA): Pearson.

Catley, K. M., & Novick, L. R. (2008). *Seeing the Wood for the Trees: An Analysis of Evolutionary Diagrams in Biology Textbooks.* Manuscript under revision for an invited resubmission.

Catley, K. M., Novick, L. R., & Shade, C. K. (2008). *Students' Inerpretations of Common Tupes of Evolutionary Diagrams.* Manuscript under revision.

Chiras, D. D. (2002). Human biology (4th Ed.). Sudbury, MA: Jones and Bartlett.

Demastes, S. S., Good, R. G. & Peebles, P. (1995). Students' conceptual ecologies and the process of conceptual change in evolution. *Science Education, 79 (6)*, 637-666.

Feltovich, P. J., Prietula, M. J., & Ericsson, K. A. (2006). Studies of expertise from psychological perspectives. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 41-67). New York, NY: Cambridge University Press.Johnson G. B. (1998). Biology: Visualizing life. Austin, TX: Holt, Rinehart and Winston.

Johnson G. B., & Raven, P. H. (2001). Biology: Principles & explorations. Austin, TX: Holt, Rinehart and Winston.

Novick, L. R. & Catley, K. M. (2008). *Understanding the Tree of Life: Reasoning from Cladograms Depicted as Trees and Ladders*. Manuscript in preparation.

Novick, L. R. & Catley, K. M. (2007). Understanding phylogenies in biology: The influence of a Gestalt perceptual principle. *Journal of Experimental Psychology: Applied*, 13, 197-223.

Settlage, J. (1994). Conceptions of natural selection: A Snapshot of the sense-making process. *Journal of Research in Science Teaching, 31 (5)*, 448-457.

Shtulman, A. (2006). Qualitative difference between naïve and scientific theories of evolution. *Cognitive Psychology, 52*, 170-194.

Wiley, E. O., Siegel-Causey, D., Brooks, D. R., & Funk, V. A. (1991). The Compleat Cladist: A Primer of Phylogenetic Procedures. *The University of Kansas Museum of Natural History Special Publication, No. 19.*

Table 1

Codes used to analyze free response questions

Code	Definition	Credit	Questions
NCHAR	acquisition of a new character; leads to the development of distinct species	1	Primate: 3
OUTG	*A. afarensis* is an outgroup; must say 'outgroup'	1	Hominid: 4
MRCA	they share a more recent common ancestor; must say MRCA or must clearly convey that the shared ancestor is the most recent; taxa constitute a clade	1	All questions
SG	the taxa are sister groups	1/0.5	Hominid: 2; Horse: 2
ANC	the taxa share a common ancestor	0.5	Hominid: 2, 3, 4; Primate: 2, 3; Horse: 2, 3
DIV	mention splitting/diverging/branching through the evolutionary process- must say either 'split' or 'diverge' or 'branch', but cannot just say the word; must convey the idea of multiple branches coming off from a single point	0.5	Hominid: 1; Monkey: 1, 2, 3; Horse: 1, 2, 4
SPEC	speciation; the formation of a new species or taxon; evolution in two different directions	0.5	Hominid: 2
REL	taxa are very closely related; closest evolutionarily; must mention degree of relationship	0.25	Hominid: 2, 3; Monkey: 2; Horse: 2, 3, 4
SC	they share certain traits/characteristics (eg: skull size; can be generic similarities)	0.25	Hominid: 2, 3; Horse: 2, 3
COM	X & Y are separated by fewer distinguishing characters than are X & Z; or they have more in common; or reference to amount of evolutionary change between taxa; use if subject talks about # of characters shared but not the specific character names	0	Hominid 1; Monkey 1; Horse 1

ANAG	anagenesis; one taxon became/evolved into/evolved from another taxon	0	Hominid 1, 3, 4; Monkey 1; Horse 1, 2, 4
CONN	linked together, branch from the same point; structural aspect	0	Monkey 2
CHANGE	use the word "change", eg: characteristics or species; or a statement with a specific change that occurred	0	Monkey 3; Horse 5
AH	distinction between *Austrolapithecus*/A/apes and *Homo*/hominid/human; must mention something specific, can't just say that two taxa go in different directions	0	Hominid 2
TIME	mention time at which evolutionary event happened or taxa existed; ordering of taxa in time; one taxon came earlier or later than another without use of evolutionary terms such as "evolved from	0	Hominid 1, 2, 4; Monkey 1, 2; Horse 1, 4, 5
KNOW	any reference to prior knowledge (correct or not) about taxa or common usage of taxa names	0	Monkey 2
DIST	taxon is (taxa are) closer or farther on diagram to *Merychippus*	0	Hominid 1; Monkey 1; Horse 1
PICT	pictures on diagram look similar	0	Hominid 1; Monkey 1, 2; Horse 1
HUM	any mention of humans or being human-like; path to becoming humans	0	Monkey 3
OTH	any other explanation	0	All questions

Table 2

Percentages of Codes for Biology Background across Hominid, Primate, and Horse for Evolutionary Distance Question

	Hominid Diagram		Primate Diagram		Horse Diagram	
	Stronger Background	Weaker Background	Stronger Background	Weaker Background	Stronger Background	Weaker Background
MRCA	39.40%	0.00%	36.4%	0.0%	36.4%	0.0%
DIV	9.10%	3.80%	18.2%	0.0%	12.1%	15.1%
COM	6.10%	9.40%	15.2%	5.7%	15.1%	5.7%
ANAG	9.10%	20.80%	0.0%	9.4%	3.0%	5.7%
DIST	3.00%	17.00%	12.1%	49.1%	6.1%	15.1%
PICT	6.10%	18.90%	0.0%	9.4%	3.0%	20.8%
OTH	27.30%	30.20%	18.2%	26.4%	24.2%	37.7%

Note: Numbers sum to 100% (within rounding error) for each column.

Table 3

Percentage of Codes for Accuracy across Hominid, Primate, and Horse for Evolutionary Distance Question

	MRCA	DIV	COM	ANAG	DIST	PICT	OTH
Hominid diagrams							
Correct	20.0%	7.7%	9.2%	7.7%	10.8%	12.3%	32.3%
Incorrect	0.0%	0.0%	4.8%	42.9%	14.3%	19.0%	19.0%
Primate diagrams							
Correct	26.3%	7.9%	7.9%	2.6%	21.1%	13.2%	21.1%
Incorrect	4.3%	6.4%	10.6%	8.5%	44.7%	0.0%	25.5%
Horse diagrams							
Correct	14.9%	23.4%	10.6%	2.1%	4.3%	14.9%	29.8%
Incorrect	12.8%	2.6%	7.7%	7.7%	20.5%	12.8%	35.9%
Average							
Correct	20.4%	13.0%	9.2%	4.1%	12.1%	13.5%	27.7%
Incorrect	5.7%	3.0%	7.7%	19.7%	26.5%	10.6%	26.8%

Table 4

Percentages of Codes for Biology Background for Hominid Question 2

	Stronger Background	Weaker Background	Total
MRCA or ANC	30.3%	0.0%	11.6%
REL or SC	6.1%	7.5%	7.0%
SPEC	18.2%	13.2%	15.1%
AH	9.1%	17.0%	14.0%
OTH	36.4%	62.3%	52.3%

Note: Numbers sum to 100% (within rounding error) for each column.

Table 5

Percentages of Codes for Biology Background and Diagram Condition for Hominid Question 3

	Biology Background		Diagram Condition			
	Stronger Background	Weaker Background	Textbook	Ladder	Tree	Total
MRCA	24.2%	1.9%	6.9%	7.1%	17.2%	17.40%
ANC	24.2%	7.5%	13.8%	14.3%	13.8%	14.00%
ANAG	12.1%	20.8%	41.7%	10.7%	0.0%	10.50%
REL or SC	27.3%	34.0%	24.1%	42.9%	27.6%	26.70%
OTH	12.1%	35.8%	13.8%	25.0%	41.4%	31.40%

Note: Numbers sum to 100% (within rounding error) for each column.

Table 6

Percentages of Codes for Biology Background for Hominid Question 4

	Stronger Background	Weaker Background	Total
OUTG	15.2%	0.0%	5.8%
ANC	57.6%	11.3%	29.1%
ANAG	6.1%	32.1%	22.1%
TIME	3.0%	26.4%	17.4%
OTH	18.2%	30.2%	25.6%

Note: Numbers sum to 100% (within rounding error) for each column.

Table 7

Percentages of Codes for Biology Background and Accuracy for Monkey Question 2

	Biology Background		Accuracy	
	Stronger Background	Weaker Background	Correct	Incorrect
MRCA	21.2%	0.0%	11.9%	4.5%
ANC	12.1%	9.4%	11.9%	9.1%
DIV	21.2%	5.7%	14.3%	9.1%
REL	9.1%	3.8%	4.8%	6.8%
CONN	9.1%	20.8%	16.7%	15.9%
TIME	9.1%	5.7%	9.5%	4.5%
KNOW	3.0%	11.3%	9.5%	6.8%
PICT	0.0%	15.1%	0.0%	18.2%
OTH	15.2%	28.3%	21.4%	25.0%

Note: Numbers sum to 100% (within rounding error) for each column.

Table 8

Percentages of Codes for Biology Background for Monkey Question 3

	Stronger Background	Weaker Background	Total
NCHAR or MRCA	39.40%	1.90%	16.30%
ANC or DIV	21.20%	9.40%	14.00%
CHANGE	0.00%	15.10%	9.30%
HUM	0.00%	17.00%	10.50%
OTH	39.40%	56.60%	50.00%

Note: Numbers sum to 100% (within rounding error) for each column.

Table 9

Subjects responses to Horse Question 2, which asked to indicate the number of lines that extend from an indicated location.

	2	3	4	5	6	7	8	9
Textbook Diagram	0	2	23*	0	0	2	0	1
(*n=28)*	0.0%	7.1%	82.1%	0.0%	0.0%	7.1%	0.0%	3.6%
Ladder Diagram	18	9*	0	0	0	0	0	0
(*n*=27)	66.7%	33.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Tree Diagram	3	24*	0	0	1	0	3	0
(*n=31)*	9.7%	64.9%	0.0%	0.0%	3.2%	0.0%	9.7%	0.0%

Note. Cell entries are the frequency and percentage of students in each group who answered in this manner.

* indicates the correct answer for the given diagram

Table 10

Percentages of Codes for Biology Background and Diagram Condition for Horse Question 2

	Biology Background		Diagram Condition			
	Stronger Background	Weaker Background	Textbook	Ladder	Tree	Total
MRCA or SG	24.2%	0.0%	7.1%	7.4%	12.9%	9.3%
ANC	36.4%	9.4%	10.7%	18.5%	29.0%	19.8%
DIV	15.2%	1.9%	7.1%	7.4%	6.5%	7.0%
REL or SC	9.1%	13.2%	7.1%	25.9%	3.2%	11.6%
ANAG	0.0%	13.2%	21.4%	3.7%	0.0%	8.1%
OTH	15.2%	62.3%	46.4%	37.0%	48.4%	44.2%

Note: Numbers sum to 100% (within rounding error) for each column.

Table 11

Percentages of Codes for Biology Background for Horse Question 3

	Stronger Background	Weaker Background	Total
MRCA	24.20%	0.00%	9.30%
ANC	48.50%	13.20%	26.70%
REL or SC	9.10%	37.70%	26.70%
OTH	18.20%	49.10%	37.20%

Note: Numbers sum to 100% (within rounding error) for each column.

Table 12

Percentages of Codes for Biology Background for Horse Question 4

	Stronger Background	Weaker Background	Total
MRCA	33.30%	0.00%	12.80%
DIV	21.20%	1.90%	9.30%
REL	3.00%	11.30%	8.10%
ANAG	9.10%	9.40%	9.30%
TIME	6.10%	39.60%	26.70%
OTH	27.30%	37.70%	33.70%

Note: Numbers sum to 100% (within rounding error) for each column.

Table 13

Percentages of Codes for Biology Background Horse Question 5

	Stronger Background	Weaker Background	Total
CHANGE	37.5%	35.0%	35.7%
TIME	12.5%	25.0%	21.4%
OTH	50.0%	40.0%	42.9%

Note: Numbers sum to 100% (within rounding error) for each column.

Figure Captions

Figure 1. *Homo sapien, H. erectus,* and *H. habilis* evolutionary relationships represented in (a) an anagenic, (b) ladder, and (c) tree representation.

Figure 2. Hominid relationship questions with textbook, ladder, and tree diagram formats

Figure 3. Primate relationship questions with textbook, ladder, and tree diagram formats.

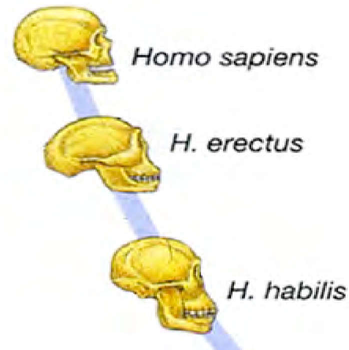
Figure 4. Horse relationship questions with textbook, ladder, and tree diagram formats.

*Question only appears for textbook diagrams

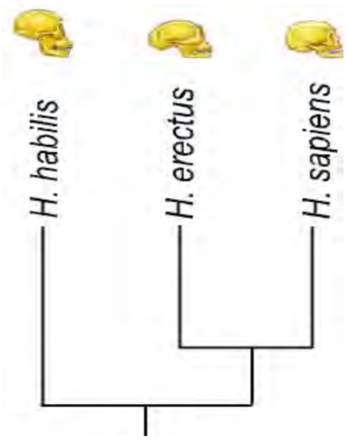
Figure 5. Mean scores on the *hominid* evolutionary distance question according to diagram condition.

Figure 1

(a)


Homo sapiens
H. erectus
H. habilis

(b)


H. habilis
H. erectus
H. sapiens

(c)

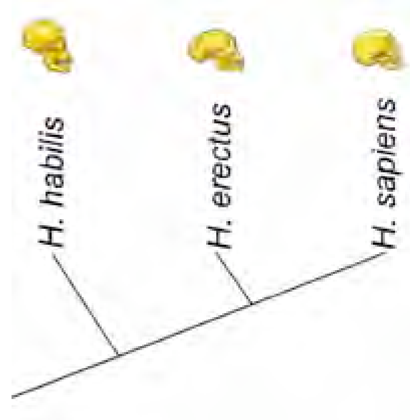

H. habilis
H. erectus
H. sapiens

Figure 2

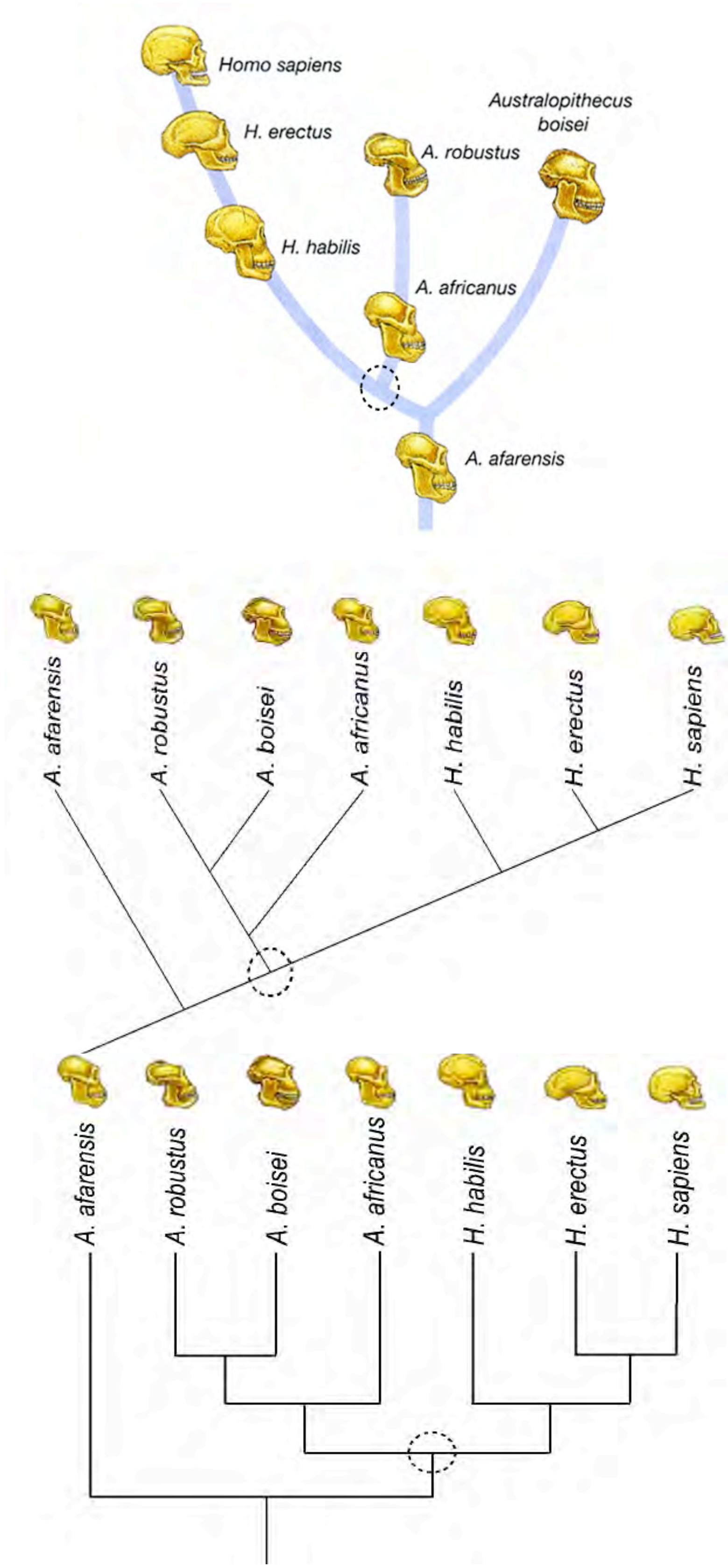

Homo sapiens
H. erectus
H. habilis
A. robustus
Australopithecus boisei
A. africanus
A. afarensis
A. afarensis
A. robustus
A. boisei
A. africanus
H. habilis
H. erectus
H. sapiens
A. afarensis
A. robustus
A. boisei
A. africanus
H. habilis
H. erectus
H. sapiens

Figure 3

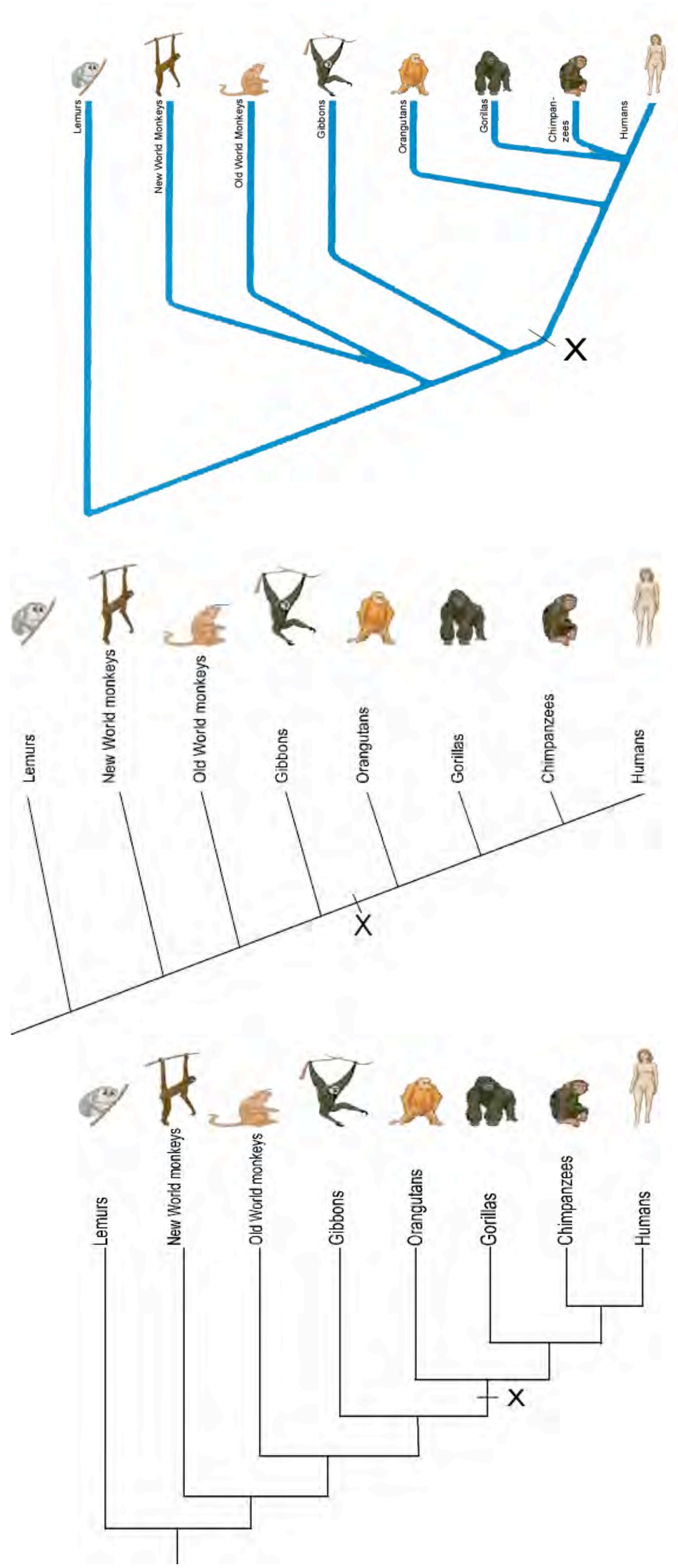

Lemurs
New World Monkeys
Old World Monkeys
Gibbons
Orangutans
Gorillas
Chimpan-
zees
Humans
X
Lemurs
New World monkeys
Old World monkeys
Gibbons
Orangutans
Gorillas
Chimpanzees
Humans
X
Lemurs
New World monkeys
Old World monkeys
Gibbons
Orangutans
Gorillas
Chimpanzees
Humans
X

Figure 4

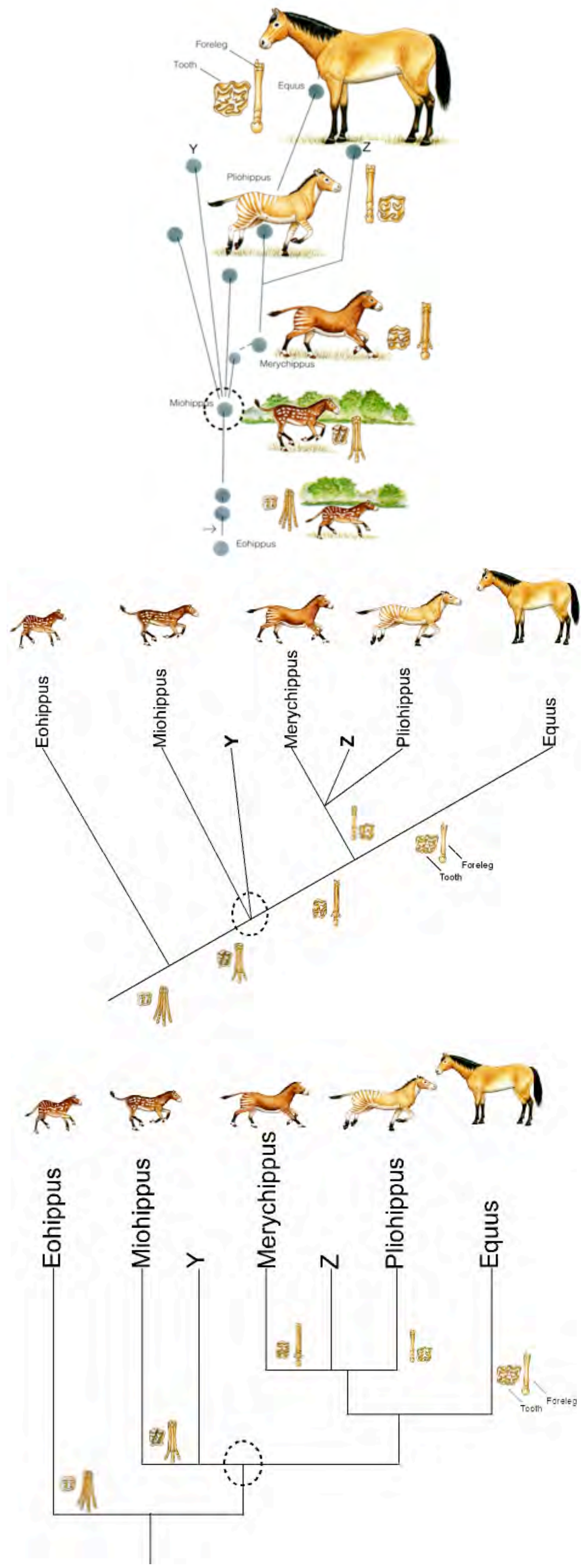
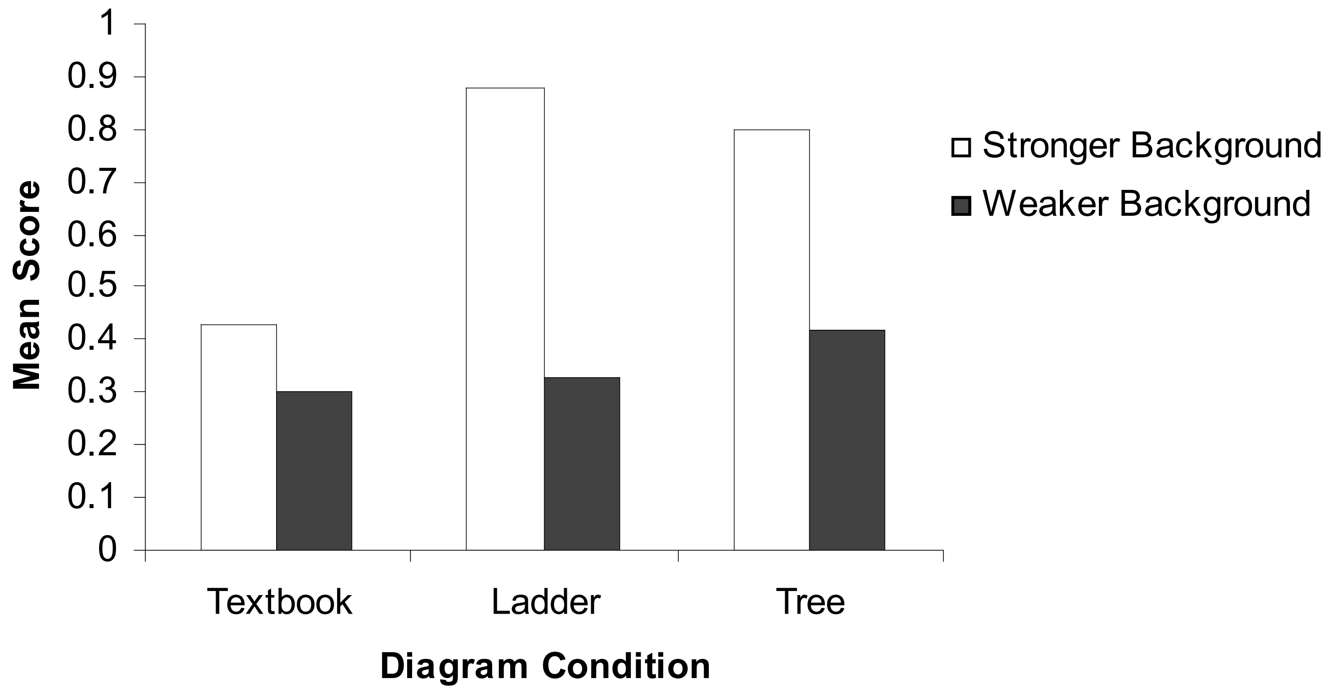

Foreleg
Tooth
Equus
Y
Z
Pliohippus
Merychippus
Miohippus
Eohippus
Eohippus
Miohippus
Y
Merychippus
Z
Pliohippus
Equus
Foreleg
Tooth
Eohippus
Miohippus
Y
Merychippus
Z
Pliohippus
Equus
Tooth
Foreleg

Figure 5


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Mean Score
Textbook
Ladder
Tree
Diagram Condition
Stronger Background
Weaker Background