JOURNAL OF MICROBIOLOGY & BIOLOGY EDUCATION, December 2016, p. 389-398
DOI: http://dx.doi.org/10.1128/jmbe.v17i3.1156

Research

Prevalence and Persistence of Misconceptions in Tree Thinking †

Tyler A. Kummer*, Clinton J. Whipple, and Jamie L. Jensen
Department of Biology, Brigham Young University, Provo, UT 84602

Darwin described evolution as "descent with modification." Descent, however, is not an explicit focus of most evolution instruction and often leaves deeply held misconceptions to dominate student understanding of common ancestry and species relatedness. Evolutionary trees are ways of visually depicting descent by illustrating the relationships between species and groups of species. The ability to properly interpret and use evolutionary trees has become known as "tree thinking." We used a 20-question assessment to measure misconceptions in tree thinking and compare the proportion of students who hold these misconceptions in an introductory biology course with students in two higher-level courses including a senior level biology course. We found that misconceptions related to reading the graphic (*reading the tips* and *node counting*) were variably influenced across time with *reading the tips* decreasing and *node counting* increasing in prevalence. On the other hand, misconceptions related to the fundamental underpinnings of evolutionary theory (*ladder thinking* and *similarity equals relatedness*) proved resistant to change during a typical undergraduate study of biology. A possible new misconception relating to the length of the branches in an evolutionary tree is described. Understanding the prevalence and persistence of misconceptions informs educators as to which misconceptions should be targeted in their courses.

INTRODUCTION

Darwin defined evolution as descent with modification. Often, when evolution is taught in university courses, the mechanisms of evolution are emphasized. While understanding the mechanisms of evolution is critical to the study of biology, all too often, the descent portion of Darwin's statement is neglected in undergraduate studies. This process of descent is most often depicted in primary literature and in textbooks as branching trees. Understanding how to interpret the information conveyed in these trees is an important skill that is used in nearly every field of research in biology (4).

A considerable amount of research has been done to help educators understand how students learn about evolutionary trees. Some researchers have attempted to identify and characterize common student misconceptions related to tree thinking (10, 12). Misconceptions range from naïve interpretations due to a lack of familiarity with this style of graphical representation to fundamentally flawed conceptions of how descent and evolution occur. A second area of research has focused on how students interpret different forms of evolutionary trees and which forms are most easily understood (6). A third area of research has focused on how to improve instruction related to tree thinking (8, 22). Despite all of this worthwhile study, little research has been done to determine how prevalent these misconceptions are among biology undergraduate students and how these misconceptions change as students progress in their studies.

We selected four major misconceptions as the focus of our study, each of which has been identified as commonly held (3). Previous studies on the prevalence of misconceptions among college students primarily focused on misconceptions that are based on unfamiliarity with the graphic (e.g., not understanding what the bifurcation means, not recognizing the axis of time, equating a straight line with no change) (18). While misconceptions based on the graphic are worthy of study, misconceptions based on the theoretical underpinnings of evolution are more concerning and perhaps more difficult for students to overcome. We chose to focus on two misconceptions related to reading a graphic and two related to the fundamental underpinnings of evolutionary theory.

The first misconception related to reading a graphic we refer to as *reading the tips*. Students with this misconception use the proximity of one tip to another to determine

*Corresponding author. Mailing address: 4102 LSB, Department of Biology, Brigham Young University, Provo, UT 84602. Phone: 801-422-2582. Fax: 801-422-0090. E-mail: tykummer@yahoo.com.
†Supplemental materials available at http://asmscience.org/jmbe

relatedness. The closer two taxa are in the tree, the more related they are. The second misconception related to reading a graphic we refer to as *node counting*. Students with this misconception use the number of nodes between two taxa to determine relatedness. The fewer the number of nodes between two taxa, the more related they are. The first misconception related to evolutionary theory we called *ladder thinking*. *Ladder thinking* can be manifested in many ways, but the common thread is teleological-based reasoning. One example of *ladder thinking* is stating that one extant group evolved or "advanced" up the tree by acquiring more complex traits and becoming another extant group that is in the tree. Another way this misconception manifests is when a student states that one group of organisms is more evolved than another "lower" in the tree. While the phrasing is different, the implication is the same. The organisms were the same and one advanced through evolution while the other remained primitive. The second misconception related to evolutionary theory we refer to as *similarity equals relatedness*. Students with this misconception determine relatedness based on how similar the physical traits are between various groups in the tree. For example, the more physical traits two groups share, the more closely related they are.

The purpose of this study was to determine both the prevalence and persistence of these four misconceptions among biology undergraduate students. We used a 20-question assessment to measure these misconceptions and compared the proportion of students who held these misconceptions in an introductory biology course with a senior level capstone course.

METHODS

Ethics statement

The hosting university's Institutional Review Board reviewed the design of this study and gave approval for use of human subjects. We obtained written consent from all participants.

Subjects

Subjects came from a highly selective large private institution in the United States. The student population was highly homogenous in terms of culture and ethnicity. The students from this university performed in the 96th percentile of all universities on the evolution section of the Educational Testing Service (ETS) Biology Field Exam and in the 99th percentile specifically on the Population Genetics and Evolution Assessment Indicator (AI7). This exam is administered at the end of a capstone evolution course, usually in the senior year, and is used by the university to evaluate the effectiveness of life sciences programs. To address the issue of how prevalent and how persistent tree thinking misconceptions are among undergraduates, we recruited participants from two undergraduate courses at the host institution: a traditionally freshman level course and a senior capstone course, each described below. We selected these courses to represent the totality of students' educational progression as students in the introductory course will be required to take the senior capstone course as part of their program of study, and vice versa.

We recruited 76 students from six sections of an introduction to biology for life science majors course (INTRO). This course is the first course in the curricula of several life science majors at the university. We selected this course to assess misconceptions at the beginning of an undergraduate life science major. Subjects were offered extra credit as an incentive to participate in the study and were recruited with a classroom announcement. Seventy-one percent of students in the INTRO course participated in the survey.

We recruited 39 students from two sections of an evolution course (EVO) that is intended as a capstone course to be taken by students nearing the end of their undergraduate studies. These students receive considerable instruction on tree thinking, including a lab designed to teach the basics of phylogenetic systematics. We included this course in the present study to assess misconceptions that persist until the end of a student's undergraduate career. Subjects were offered extra credit as an incentive to participate in the study and were recruited with a classroom announcement. Seventy-five percent of students in the EVO course participated in the survey.

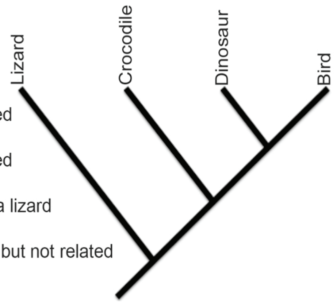
Measuring evolutionary tree understanding in students from these two courses allowed us to see how students entering the university compared with those who were near the end of their studies. Other studies have used a similar design to compare differences in student thinking (1, 2, 9, 15,18).

Study design

To measure misconception prevalence, we used an assessment that contained at least two items to elicit each of the misconceptions described earlier (*reading the tips*, *node counting*, *ladder thinking*, and *similarity equals relatedness*). To create a valid assessment, we used two independent researchers who study tree thinking and who have extensive experience in teaching these concepts to undergraduate students. Each researcher chose items from the previously published *Tree Thinking Quizzes I and II* that corresponded to misconceptions identified above (4). Eight items were selected. The authors wrote two additional items (17/18 and 19/20) that were included in the assessment. These items were based on previous student responses and interactions where misconceptions were demonstrated. Each question had the potential to elicit multiple misconceptions depending on the answer chosen and on the reasoning described (see Fig. 1 and Table 1). Reliability statistics are described below. We used a multiple-choice format with the goal of producing an easily scored objective assessment. One issue with using a

By reference to this tree, which of the following is an accurate statement of relationships?

A. A crocodile is more closely related to a lizard than to a bird.
B. A crocodile is more closely related to a bird than to a lizard.
C. A crocodile is equally related to a lizard and a bird.
D. A crocodile is related to a lizard, but not related to a bird.



Explain the reasoning you used to answer the previous question:

Student answers:

Reading the Tips (Selected answer: A)
"The distance between the two species are not equal and thus the crocodile is more related to a lizard."
Node Counting (Selected answer: B)
"Birds have 2 common ancestors with crocodiles but the lizard only has one common ancestor with the crocodile"
Ladder Thinking (Selected answer: B)
"From the way this tree is drawn, it appears that the entire mainline is denoted as 'bird'. In this case, both lizard and crocodile branched off of bird."
Similarity Equals Relatedness (Selected answer: A)
"Crocodiles split off after lizards, closer in morphology to lizard than to bird."

FIGURE 1. A sample item set from the assessment used to measure student misconception and examples of student responses to the item set.

multiple-choice format in determining the prevalence of misconceptions among students is that the same wrong answer can be selected due to several different misconceptions. For example, in Figure 1, students may incorrectly choose answer choice A using the *reading the tips* misconception or *similarity equals relatedness*. Similarly, students can choose the correct answer (answer choice B) using the wrong reasoning (*node counting* or *ladder thinking*). To overcome this issue, students answered the multiple-choice content-based question and then answered a follow-up free-response question explaining the reasoning behind their choice. Doing this allowed us to more accurately determine any misconception the subject held. This approach is similar to the pattern used on other assessments such as Lawson's Classroom Test of Scientific Reasoning (12, 16).

The assessment consisted of 20 paired items: 10 multiple-choice content-based items and 10 open-response follow-up reasoning items (see Appendix 1). We administered the assessment to students using an online survey system. Students had the opportunity to take the assessment during a one-week period after being recruited with an in-class announcement and an e-mail containing the link. Not all courses were surveyed at the same time. We administered the assessment to students in the INTRO course prior to the students receiving any formal instruction on tree

TABLE 1.
The misconceptions most commonly associated with selected answers for each item pair and the correct answer for each item pair.

Question Pair	Answer Option	Most Commonly Categorized Misconception		Correct Answer
		INTRO	EVO	
1/2	A	Reading the tips	Ladder thinking	B
	B	Similarity equals relatedness	Branch length	
	C	Reading the tips	Node counting	
3/4	A	Reading the tips	Reading the tips	B
	B	Ladder thinking	Branch length	
	C	Node counting	Node counting	
5/6	A, B	Ladder thinking	Ladder thinking	E
7/8	A, B, C, D, E	Ladder thinking	N/A	E
9/10	A, B, D, E	Ladder thinking	Ladder thinking	C
11/12	A	Reading the tips	Node counting	C
	B	Similarity equals relatedness	Ladder thinking	
	C	Ladder thinking	Branch length	
13/14	A, B, E	Ladder thinking	Ladder thinking	C
15/16	A, B	Ladder thinking	Ladder thinking	C
	D	Similarity equals relatedness	N/A	
17/18	A, B, D	Similarity equals relatedness	Similarity equals relatedness	C
	C	Similarity equals relatedness	Ladder thinking	
19/20	A	Ladder thinking	Ladder thinking	D
	D	Similarity equals relatedness	Ladder thinking	

N/A = not applicable.

thinking related topics. This was done to assess what level of misconception the students had at the beginning of their undergraduate study of life sciences. We administered the survey to the EVO course near the end of the course to assess the misconception levels of students near the completion of the capstone course.

We calculated a student's score on the exam using only the 10 multiple-choice content-based items. We used the follow-up reasoning items in conjunction with the multiple-choice items to diagnose the presence of common misconceptions held by students. To interpret the results of the reasoning questions, two raters individually evaluated each pair of items, with an emphasis on the written response. Raters were science education researchers and instructors of introductory biology. They were both familiar with the tree thinking assessment items and had experience with student responses to these items. Data were anonymous so that raters were unaware of student identities; however, they were aware of the populations (INTRO and EVO) from which the data came. Raters evaluated subject responses item by item. Raters classified each response as correct, correct with one of the described misconceptions, incorrect with one of the described misconceptions, or incorrect with no clear misconception. Misconceptions were identified by first looking at the answer selected; for many of the questions, a selected answer indicated a misconception was likely held. Raters would then evaluate phrasing in the written response to either confirm or identify a misconception not indicated by the selected answer. Example responses that would indicate each of the given misconceptions are shown in Figure 1; additional examples are provided in Appendix 1.

Raters evaluated written responses with no clear misconceptions for any commonalities between them in an effort to identify misconceptions not previously described. If a subject was deemed to have demonstrated a misconception on any one item, we classified them as holding that misconception. This approach allowed us to effectively measure the prevalence of each misconception but it did not give us an indication of how strongly each subject held a misconception. After individually evaluating each response, the raters met together and discussed differences in evaluation in an attempt to reach agreement on the evaluation.

Statistical analysis

We analyzed the persistence of each misconception by means of Mann-Whitney U tests using SPSS software v. 21. These analyses allowed us to evaluate the significance of any differences found between the groups. Mann-Whitney U tests were used because the data failed to meet the assumption of normality.

We ran a Spearman's correlation of the two raters' evaluations to measure the degree of correlation between them. We found that reviewer categorizations were significantly correlated with each other with an inter-rater reliability of 0.992.

We also used Cronbach's alpha to measure the internal reliability of the assessment for only the multiple-choice item responses. The analysis produced a Cronbach's alpha coefficient of 0.638, which is within the acceptable range for an assessment of this type (14). Cronbach's alpha functions as an equivalent of KR-20 when used with dichotomous data. Because multiple misconceptions could manifest from the same item, we did not evaluate the internal reliability of subsets that were intended to measure specific misconceptions.

RESULTS

Overall prevalence of misconceptions

The prevalence and persistence of tree thinking misconceptions varied among the four misconceptions we measured in this study. We compared student responses from the EVO course with student responses in the INTRO course (Fig. 2 and Table 2). The results of the Mann-Whitney U tests showed that significantly fewer students from the EVO course gave answers that were based on the *reading the tips* misconception. However, we found that students from the EVO course demonstrated significantly higher levels of the *node counting* misconception than those in the INTRO course. The most prevalent and persistent misconception we measured in this study was the *ladder thinking* misconception; students at both levels demonstrated a high level of this misconception. The *similarity equals relatedness* misconception was equally prevalent in both courses.

Detailed student response rates

To look for further evidence as to how tree thinking differs between students in the EVO course and those in the INTRO course, we compared the percentage of each

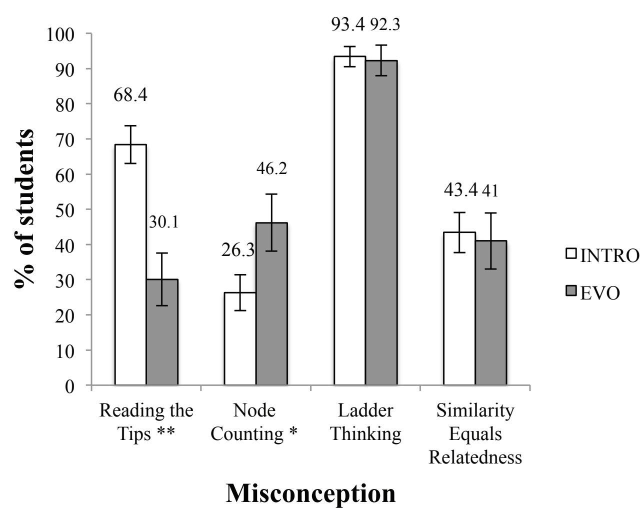


FIGURE 2. The proportion of students who gave answers indicating they held each of the misconceptions assessed in this study for the INTRO course and the EVO course. Error bars represent one standard error. ** $p < 0.01$; * $p < 0.05$.

TABLE 2.
The results from the Independent-Samples-Mann-Whitney U test comparing the proportion of students demonstrating each misconception in the INTRO course and EVO course.

Misconception	Subjects (*n*)	Mann-Whitney U	*P* value
Reading the tips	115	924.00	0.000
Node counting	115	1,776.0	0.033
Ladder thinking	115	1,465.5	0.825
Similarity equals relatedness	115	1,446.5	0.807

α = 0.05.

misconception used to justify student answers on each of the questions in the survey. Several interesting patterns were found in responses. Percentages pertaining to each question are displayed in Supplemental Table 1 (Appendix 1). The main findings are highlighted below.

Reading the tips. Looking at question pairs 1/2 and 3/4, we see that explanations using *reading the tips* reasoning were much more frequent in the INTRO course than in the EVO course. For example, in question 1, distractor 'C' was designed to elicit the *reading the tips* misconception and was chosen for this reason 42% of the time among the INTRO students while only 15% of EVO students did this. Likewise, in question pair 3/4, 55% of INTRO students chose distractor 'A' and used *reading the tips* reasoning compared with only 18% in the EVO course. The same trend is seen in item pair 11/12.

Unexpectedly, we saw a large proportion of students in the INTRO course using *reading the tips* when answering item pair 15/16, a question not specifically designed to elicit this misconception. In fact, 16% of students chose distractor 'D,' citing reasoning such as, "Student D described how the closer the species are on the tree, the more closely related they are" or "All of the branches are related to the branches next to them." Interestingly, distractor 'D' was designed to elicit *similarity equals relatedness*. Students in the EVO course almost exclusively used *similarity equals relatedness* when choosing this distractor (18%), using reasoning such as, "The alga are most closely related because the only thing differing between them is their color, and moss and pine are most closely related because their common ancestor developed into a multicellular organism."

Node counting. Contrary to *reading the tips*, the *node counting* misconception appears to be more common in EVO students than INTRO students. On item pair 1/2, we see that instead of choosing distractor 'C' using *reading the tips* reasoning like INTRO students, the majority of students in the EVO course who chose "C" did so using *node counting* reasoning, such as, "There is just one node between a trout and a coelacanth and there is just one node from a trout to a stingray." We see a similar shift on item pair 3/4. Only 3% of INTRO students chose distractor 'A' and 11% chose distractor 'C' using *node counting* reasoning, whereas 15% of EVO students chose distractor 'A' and 21% chose distractor 'C' using *node counting* reasoning.

Item pair 11/12 resulted in a similar pattern. Sixty-six percent of the students in the INTRO course and 62% of the students in the EVO course selected distractor 'A.' However, 45% of the students selected "A" using *reading the tips* in the INTRO course while only 7% used *node counting*. In the EVO course, 38% of students selected "A" using *node counting* compared with 21% using *reading the tips*.

Ladder thinking. The prevalence of *ladder thinking* was found to be equal and at high levels in both groups. Looking at responses in finer detail shows that there were differences between the two groups even though the overall prevalence was similar. In item pair 7/8, *ladder thinking* is the only misconception that manifests in student explanations. However, students in the INTRO course were nearly equally distracted by answers 'B' (12%) and 'D' (16%) and only slightly more by 'C' (28%), whereas EVO students overwhelmingly favored distractor 'C' (54%) while not entertaining 'D' at all. It appears that students in the EVO course were less distracted by words such as 'intermediate' or 'advanced' but still maintained 'ancient species' as a probable explanation. Interestingly when 'ancient' is not used as a distractor, as in item pair 15/16, students are more likely to consider the possibility that a species could be 'most advanced', i.e., 28% of students chose answer 'A' in the EVO course as opposed to only 15% in the INTRO course. Explanations like the following were used to justify this answer: "The pine is the latest to diverge and builds on the changes made to each previous species."

Similarity equals relatedness. On item pair 1/2, the use of *similarity equals relatedness* in student explanations is somewhat common among INTRO students (18%). In the EVO course we see that *similarity equals relatedness* is never used. We see a similar pattern on item pairs 3/4 and 11/12. Despite this small but convincing shift, the overall prevalence of *similarity equals relatedness* was not significantly different. Two item pairs can help to explain this phenomenon. On item pair 15/16, distractor 'D' suggested similarity as the basis for relatedness. In the EVO group, 28% of students selected D, with 18% indicating *similarity equals relatedness* as the reasoning. In comparison, 19% of the students in the INTRO class used this reasoning. We also see that on item pair 17/18, when students are asked to place a species on a tree, EVO students routinely (36%) cited similarity, regardless of the selected answer, as the basis for determining where in the tree it should fit (e.g., "I know that dolphins are fish-like, but they are also mammals so I put them in between fish and mouse").

Branch length. On the items that asked students to compare relatedness (item pairs 1/2 and 3/4), we noticed a common reasoning pattern among EVO students that seemed to be different than our four defined misconceptions. We labeled this misconception *branch length* because students were explaining their reasoning by describing the length of the branches connecting the two species. Interestingly, this misconception manifested despite the student getting the correct answer. For example, in item pair 1/2, 21% students chose the correct answer, 'B,' but did so by comparing branch lengths, as did this student: "The branch length is longer between stingray and trout than it is between trout and coelacanth, so the trout is more closely related to a coelacanth than to a stingray." The same pattern was seen on item pair 3/4, with 10% of students choosing the correct answer, 'B,' by using reasoning like the following: "The line drawn from Crocodile to Bird is shorter than the line drawn from Crocodile to Lizard." A similar pattern was seen on item pair 11/12. We compared the prevalence of these branch length-based responses between the courses. We used a Mann-Whitney U test and found that branch length-based responses were significantly higher in the EVO course (n = 115, U = 1,842.5, p = 0.000; see Fig. 3).

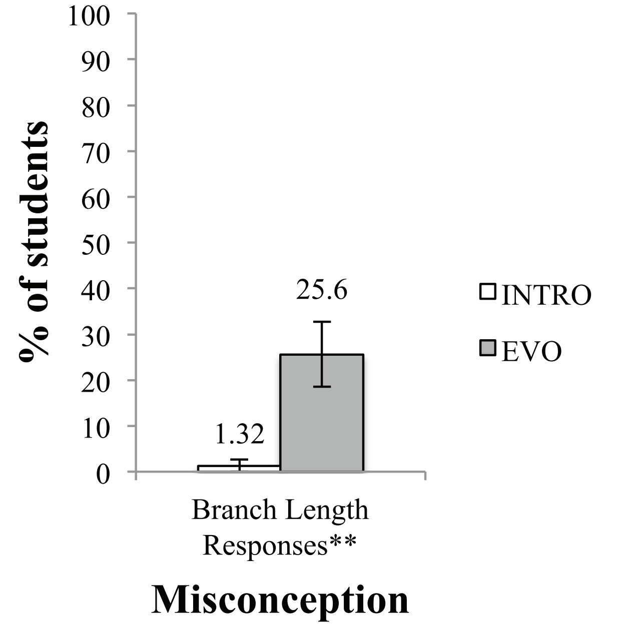


FIGURE 3. The proportion of students who gave responses based on branch length in the INTRO course and the EVO course. Error bars represent one standard error. ** $p < 0.01$.

Student performance

We compared student overall performance on the content component of the assessment between courses. The average score for both groups was below 50% on the assessment. Using a Mann-Whitney U test, we found no significant difference between the groups in assessment performance (n = 115, U = 1,635.5, p = 0.240). Despite changes in some misconception levels, student performance on the assessment overall remained low for both groups (Fig. 4).

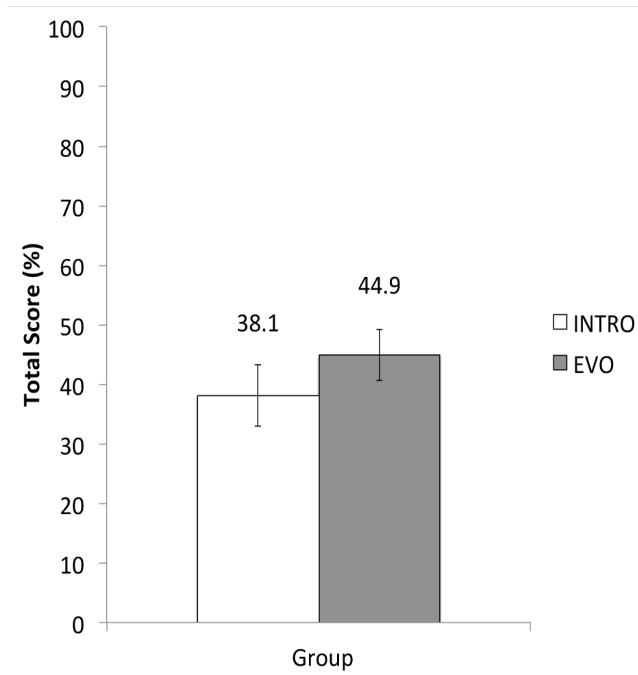


FIGURE 4. A comparison of overall performance on the assessment for each group in the study. Error bars represent one standard error.

DISCUSSION

The prevalence of the misconceptions in our study varied in interesting and informative ways. Based on our study populations being at the beginning of their education and at the end, we expected to see a decrease in all misconceptions due to specific instruction on this topic presumably throughout their biology degree, including a capstone course on evolution. This was indeed the case for some misconceptions but not for others. *Reading the tips* and *node counting* potentially demonstrate an inverse relationship. *Reading the tips* went from highly prevalent in the INTRO course to extremely low in the EVO course. This indicates that students are, for the most part, rejecting this misconception as they progress in their education, presumably due to course exposure. *Node counting*, however, appears to have the opposite trend. Students in the EVO course demonstrated significantly higher levels of *node counting* than the INTRO course students, indicating that instruction is not addressing this misconception and perhaps is introducing it.

One potential explanation for these two findings is that many students are abandoning *reading the tips* and replacing it with *node counting*. Students may be experiencing disequilibrium as they attempt to interpret relatedness using the closeness of the tips but are instructed that the nodes are what matter (5). Thus, they may be abandoning *reading the tips* as a method for interpreting evolutionary trees and instead adopting a *node counting* misconception. This potential explanation is supported by the lack of a significant difference in overall performance on the assessment.

If the *reading the tips* misconception had been replaced by a correct understanding, we would have likely seen a significant difference between the overall scores of the groups. Evidence from the question detail also supports this potential explanation. In all three question pairs that focused on having students interpret evolutionary relationships, we saw dramatic differences between the INTRO group and the EVO group when it came to *reading the tips* and *node counting*. EVO students in all three pairs used *node counting* at higher rates and used *reading the tips* at lower rates than INTRO students. However, certainly there may be other explanations that fit this data. Poor performance on interpreting relationships in this study is supported by another study that found no significant difference in student ability to interpret relationships from an evolutionary tree before and after instruction on evolutionary trees (7). The authors suspected that students were using the number of steps separating taxa to determine relatedness. The high levels of *node counting* held by EVO students found in this study appears to support their suspicion.

In a 2007 study that also includes *reading the tips* (tip proximity) and *node counting*, researchers saw a different pattern (18). They saw no significant differences between lower-level-course students and upper-level-course students for prevalence of either of these misconceptions. One potential explanation for this difference is in the study design. In this study, students from both groups received at a minimum one lesson of instruction on evolutionary trees. In contrast, our study compared students (INTRO) who had received no instruction on evolutionary trees with students who had received multiple lessons on the topic. It is possible that both groups in the 2007 study more closely represent those in the EVO course than they do those in the INTRO course even if they are enrolled in a lower-level course. This would explain why there was no significant difference between the upper- and lower-level courses. We do see one important commonality between their upper-level students and our EVO students: both groups have higher levels of *node counting* than of *reading the tips*. In other words, for these two misconceptions, we see a similar pattern in both studies for the advanced courses.

While the inverse relationship between *reading the tips* and *node counting* is potentially explained as described above, we also see a similar pattern with *reading the tips* and *branch length*. Indeed on the same three questions pairs (1/2, 3/4, 11/12) we see *branch length* used at a higher percentage and *reading the tips* used at a lower percentage by EVO students. A number of possible explanations may describe how the thinking of an INTRO student changes to that of an EVO student in light of the patterns we see with *branch length*. Not only might they adopt *node counting* as previously described, but they may also go from *reading the tips* to *branch length* or perhaps from *reading the tips* to *node counting* to *branch length*. A study consisting of student interviews, pre- and post- assessment, and tree thinking interventions that include phylogram introduction is needed to test these potential explanations. This would allow researchers to identify students who had *reading the tips* as a misconception and then after instruction demonstrated that they held *node counting* or *branch length*. After identifying these students, a series of interviews could be conducted to determine why students abandoned the *reading the tips* misconception and why they adopted *node counting* or *branch length*. Our current study can only provide us with patterns that allow us to generate hypotheses for future tests.

Perhaps the most troubling finding in this study is the high levels of the *ladder thinking* misconception found in both groups. When looking at the question detail, we do see some slight differences between the groups, though the overall prevalence remains unchanged. Students still held *ladder thinking* ideas, but the interpretation of the trees may have been different. For example, a student in the EVO course appears to be more likely to favor explanations declaring species to be advanced or primitive over explanations that one species on the tree was an intermediate on a path of transformation. *Ladder thinking* is still indicated by both types of explanation, but the view of what the tree is depicting is different. As mentioned previously, this misconception indicates a fundamental misunderstanding of evolution based in teleological reasoning. Teleological reasoning is a pervasive idea that contributes to misconceptions in many fields of science (13). The persistence of the *ladder thinking* misconception is perhaps understandable when we consider how entrenched the underlying reasoning is among students. Teleology is pervasive because it is fundamental to how most people see and interpret the world around them. Fundamental presuppositions, like teleology, are very difficult to overcome (24). If the fundamental presupposition is not addressed, a misconception will be persistent. We believe lessons and activities that seek to address *ladder thinking* must also address teleological reasoning to be successful.

Meir et al. (18) also looked at a subset of *ladder thinking*, that a strait line equals no change. This means that if no branching is occurring on the line to a particular species, then there must be no evolution occurring. Given that this is just one subset of responses that we would categorize as *ladder thinking,* it is expected that they saw lower levels of *straight line equals no change* (40%) than we saw of *ladder thinking* (92.3%). Gregory and Ellis (11) conducted a study to look at conceptions of evolution among science graduate students. They found that the most commonly held misconception among graduate students was teleological thinking; viewing evolution as a progressive process was the second most commonly held misconception. Given the ties of both of those concepts to *ladder thinking,* we believe our study, like that of Gregory and Ellis, shows that *ladder thinking* and teleology relating to evolution are misconceptions that still need to be addressed even with advanced students.

Similarity equals relatedness was shown to be persistent across both courses. This misconception can be difficult because of the role similarity plays in the building of evolutionary trees. Many instructors focus on tree reconstruction methods using character matrices when they teach students about evolutionary trees. It is possible that students are mistaking similarity as the basis for building the tree rather than parsimony. When a dominant characteristic (morphological or ecological) is due to an ancestral state (plesiomorphy) or convergence rather than synapomorphy, it could lead to incorrect interpretations of evolutionary relationships by students who mistake basic similarity as the basis for determining relatedness.

Novick, Cately, and Funk (21) found one potential explanation for why similarity can be such a challenge for students when looking at evolutionary relatedness. This study asked students to compare evolutionary relationships between three taxa. Researchers found that when familiar taxa were used in the comparison, students who could correctly determine relationships with unfamiliar taxa often gave incorrect responses. Students' familiarity with the physical characteristics of the taxa caused them to override their reading of the tree. Unfortunately, our study used taxa that were likely familiar to the students in each of the questions asking the students to compare relationships, so we were not able to see whether a similar pattern of obstruction was seen in our students when answering.

The question detail provided some interesting patterns that were not apparent from comparing the total prevalence from both groups. In the INTRO group, we saw students using *similarity* to determine relatedness on several questions including those that required students to compare evolutionary relatedness and those that required students to place a new taxon in a tree. Alternatively, in the EVO group, we saw that students on almost every evolutionary relatedness question did not use *similarity* as an explanation. One exception was on question pair 15/16. Answer 'D' specifically appeals to similarity as the basis for determining relationships. In the EVO group we saw *similarity equals relatedness* used to justify a determination of evolutionary relationships but only on answer 'D.' The overall trend from all the questions seems to indicate that EVO students do not rely on similarity when being asked to determine relatedness, but some may be open to using it if it is suggested by the answer choice, as was the case in pair 15/16.

Despite this difference in using similarity for analyzing an existing tree, we did not see a difference in the overall prevalence of this misconception, primarily due to question pair 17/18. This question asked students to place a new taxon in an existing tree. When placing the taxon in the tree, EVO students were just as likely to place it in the tree based on similarity as the INTRO students were.

In addition to the four misconceptions discussed above, we found that many students in the EVO course referred to the overall length of the branches (e.g., in millimeters) between the taxa in their written explanations as something that had meaning, indicating a new misconception. The reasoning was that the longer the branch lengths were between two taxa, the less related these taxa were. We believe this may be a potential misconception that forms when students are introduced to phylograms. Branch length in a phylogram is informative, but the branch length provides information on the rate of change not on the relationships between taxa. An alternative explanation is that this pattern is a special case of the *similarity equals relatedness* misconception. Branch length in a phylogram represents some sort of evolutionary change (DNA mutations) that occurred after the most recent branching. The higher the number of evolutionary changes, the longer the branch. It is possible that students are interpreting longer branch lengths between two taxa as dissimilarity and using that to infer less relatedness. Student responses were not extensive enough to distinguish the underlying reasoning. Further study and characterization of students using this explanation is needed.

Although our findings are compelling and informative, certain limitations should be considered. Our research subjects came from a highly selective institution for academics. Thus, the prevalence of these misconceptions may not necessarily reflect what would be seen at a less selective institution. In addition, the institution is a private religious institution with neutral views toward evolution, but cultural biases may exist that could potentially influence results (17). To account for these biases, we examined student responses to item pair 19/20, which discusses human evolution, to determine whether this item was more frequently missed than others. The data showed that this item had the highest number of accurate responses in comparison with all other items, suggesting that students were capable of responding scientifically to potentially controversial evolution statements. In addition, in a recently published study using subjects from the same institution, researchers found that student acceptance of evolution, as measured by the Measure of Acceptance of the Theory of Evolution (MATE), was over 80%, and knowledge of evolution, as measured by the Knowledge of Evolution Exam (KEE), was nearly 80% (19). This is in contrast to a study of students at the University of Minnesota, who scored an average of 54.2% on the KEE, and a study of life science faculty at a large Midwestern university, who scored 74.3% on the KEE and 87.6% on the MATE (20, 23). Thus, our student body does not appear to have different acceptance or knowledge of evolution than other student bodies.

Another potential limitation pertains to our assessment. Although we took most items from a previously published assessment, it had not been subjected to rigorous statistical analyses for reliability and was not necessarily intended to be a concept inventory (4). We added an additional two items and included an overall reliability analysis. This instrument appeared sufficient to solicit student misconceptions enough to quantify them. Since a published tree thinking concept inventory does not yet exist, we are limited to what experts have produced.

In addition, the assessment was only 10 questions. It is possible that students who have a strongly held misconception may also have other misconceptions that are not elicited by the limited number of questions. Given the choice of two answers on a given question, it may be that the strongly held misconception influenced the choice even though the other misconception was held. We attempted to overcome this issue by having targeted questions for each misconception and including free-response opportunities to explain reasoning, but our method does not preclude this possibility.

CONCLUSION

From this study, we conclude that tree thinking is a difficult skill for undergraduate students. Despite our students performing exceptionally well on the ETS Biology Field Exam (in the 96th percentile), they have a fundamental misunderstanding of, and inability to correctly interpret, phylogenetic trees. In addition, we see that the misconceptions we measured have different trajectories across a student's educational career. Misconceptions related to reading the graphic (*reading the tips* and *node counting*) proved to be tricky to address in that we may be able to correct one practice (i.e., *reading the tips*), but we may simply be replacing it with an alternative erroneous practice (i.e., *node counting*, *branch length*). This indicates that instructors should be meticulous in the way that they teach students to read a phylogenetic tree and ensure that students have indeed understood what each component of the tree represents (e.g., a branch, a node). Using formative assessment often is a way to ensure that students understood the concept in the way in which the instructor intended. It would also allow instructors to detect any new misconceptions that may have been introduced (e.g., measuring branch lengths). Using a tree thinking assessment would be useful to this end.

Alternatively, misconceptions related to the fundamental underpinnings of evolutionary theory (*ladder thinking* and *similarity equals relatedness*) proved nearly completely resistant to change, remaining equally prevalent in freshman and seniors. This indicates the need for instructors covering evolutionary topics at all stages to attempt to directly address these fundamental misconceptions. *Ladder thinking* is by far the most prevalent and persistent of all the misconceptions addressed in this study, indicating a significant need for the development and testing of educational interventions that address this problematic misconception. Ultimately, by understanding the prevalence and persistence of misconceptions of students in our classroom, we can make pedagogical decisions targeted for our course.

SUPPLEMENTAL MATERIALS

Appendix 1: Assessment, detailed student response table, and student quotes

ACKNOWLEDGMENTS

An NSF Grant (IOS-1253241) awarded to Clinton Whipple supported this project. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We thank a number of undergraduate researchers who aided in this project and the instructors who allowed us to recruit from their courses. The authors declare that there are no conflicts of interest.

REFERENCES

1. **Atman, C. J., M. E. Cardella, J. Turns, and R. Adams.** 2005. Comparing freshman and senior engineering design processes: an in-depth follow-up study. Design Studies **26**(4):325–357.
2. **Azizi-Fini, I., A. Hajibagheri, and M. Adib-Hajbaghery.** 2015. Critical thinking skills in nursing students: a comparison between freshmen and senior students. Nurs. Midwifery Stud. **4**(1):e25721.
3. **Baum, D., and S. Smith.** 2012. Tree thinking: an introduction to phylogenetic biology. Roberts, Greenwood Village, CO.
4. **Baum, D., S. Smith, and S. Donovan.** 2005. Evolution: the tree-thinking challenge. Science **310**(5750):979–980.
5. **Bransford, J., A. Brown, and R Cocking.** 2000. Learning: from speculation to science. How people learn brain, mind, experience, and school, p 14–16. The National Academies Press, Washington, DC.
6. **Catley, K., and L. Novick.** 2008. Seeing the wood for the trees: an analysis of evolutionary diagrams in biology textbooks. BioScience **58**(10):976–987.
7. **Dees, J., J. Momsen, J. Niemi, and L. Montplaisir.** 2014. Student interpretations of phylogenetic trees in an introductory biology course. CBE Life Sci. Educ. **13**(4):666–676.
8. **Eddy, S., A. Crowe, M. Wenderoth, and S. Freeman.** 2013. How should we teach tree-thinking? An experimental test of two hypotheses. Evol. Educ. Outreach **6**:13–24.
9. **Genco, N., K. Hölttä-Otto, and C. C. Seepersad.** 2012. An experimental investigation of the innovation capabilities of undergraduate engineering students. J. Eng. Educ. **101**(1):60–81.
10. **Gregory, T.** 2008. Understanding evolutionary trees. Evol. Educ. Outreach. **1**(2):121–137.
11. **Gregory, T., and Ellis, C.** 2009. Conceptions of evolution among science graduate students. BioScience **59**(9):792–799.
12. **Halverson, K., C. Pires, and S. Abell.** 2011. Exploring the complexity of tree thinking expertise in an undergraduate systematics course. Sci. Educ. **95**(5):794–823.
13. **Kampourakis, K.** 2007. Teleology in biology, chemistry and physics education: what primary teachers should know. Rev. Sci. Math. ICT Educ. **1**(2):81–93.
14. **Kline, P.** 2000. The handbook of psychological testing, 2nd ed., p 13. Routledge, London, England.
15. **Kögce, D., and C. Yıldız.** 2011. A comparison of freshman and senior mathematics student teachers' views of proof concept. Proc. Soc. Behav. Sci. **15**:1266–1270.
16. **Lawson, A.** 1978. The development and validation of a classroom test of formal reasoning. J. Res. Sci. Teach. **15**(1):11–24.

17. **Ludlow, D., and W. Evenson.** 1992. Encyclopedia of Mormonism, p 340–343. Macmillan Publishing, New York.
18. **Meir, E., J. Perry, J. Herron, and J. Kingsolver.** 2007. College students' misconceptions about evolutionary trees. Am. Biol. Teach. **69**(7):E71–E76.
19. **Manwaring, K. F., J. L. Jensen, R. A. Gill, and S. M. Bybee.** 2015. Influencing highly religious undergraduate perceptions of evolution: Mormons as a case study. Evol. Educ. Outreach **8**(1):1–12.
20. **Moore, R., S. Cotner, and A. Bates.** 2009. The influence of religion and high school biology courses on students' knowledge of evolution when they enter college. J. Eff. Teach. **9**(2):4–12.
21. **Novick, L., K. Catley, and D. Funk.** 2011. Inference is bliss: using evolutionary relationship to guide categorical inferences. Cog. Sci. **35**(4):712–743.
22. **Phillips, B. C., L. R. Novick, K. M. Catley, and D. J. Funk.** 2012. Teaching tree thinking to college students: it's not as easy as you think. Evol. Educ. Outreach **5**(4):595–602.
23. **Rice, J. W., M. P. Clough, J. K. Olson, D. C. Adams, and J. T. Colbert.** 2015. University faculty and their knowledge & acceptance of biological evolution. Evol. Educ. Outreach **8**(1): 1–15.
24. **Vosniadou, S.** 1994. Capturing and modeling the process of conceptual change. Learn. Instruct. **4**:45–69.