JOURNAL OF RESEARCH IN SCIENCE TEACHING VOL. 46, NO. 3, PP. 311–332 (2009)

Digging Deep: Exploring College Students' Knowledge of Macroevolutionary Time

Kefyn M. Catley, Laura R. Novick

Vanderbilt University, Nashville, Tennessee 37203-5721

Received 14 May 2007; Accepted 23 June 2008

Abstract: Some ability to comprehend deep time is a prerequisite for understanding macroevolution. This study examines students' knowledge of deep time in the context of seven major historical and evolutionary events (e.g., the age of the Earth, the emergence of life, the appearance of a pre-modern human, *Homo habilis*). The subjects were 126 students recruited from psychology, education, and biology classes at two universities. They were assigned to stronger and weaker background groups based on their college-level biology coursework. Subjects provided startlingly large time ranges for all questions, ranging over several orders of magnitude (e.g., from 1,000 to 600 billion years ago for when most dinosaurs became extinct), coupled with the strong tendency to underestimate how long ago the events occurred. Converting absolute time estimates to relative time estimates allowed subjects' knowledge of the spacing of the events to be examined and also provided a clearer picture of their patterns of over and underestimation. The results of this study suggest that many students are without an effective conceptual framework to make sense of very large time frames. Although there were no consistent differences in the accuracy of students' responses as a function of their biology background, the weaker background students showed greater variability, providing more time estimates at both the low and high extremes. We describe a pedagogical strategy that uses a relative approach presenting major evolutionary events as they unfolded in time and advocate a tool from professional practice to depict events in time and space. © 2009 Wiley Periodicals, Inc. J Res Sci Teach 46: 311–332, 2009

Keywords: macroevolution; cladistics; deep time; tree thinking; evolution education; cladograms

> "The influence of deep time is felt in a variety of scientific disciplines including geology, cosmology, and evolutionary biology. Thus, any scientist or student that wants to master any of these subjects must first have a good understanding of deep time" (Dodick & Orion, 2003, p. 708).

Understanding Macroevolution

It is impossible to have a scientifically literate public without a widespread understanding of the evolutionary principles that allow us to make sense of all facets of the natural world. Scientific literacy in this domain requires an understanding of both microevolutionary (e.g., variation in populations, natural selection) and macroevolutionary processes (processes that operate at and above the level of species, resulting in the formation, radiation, and extinction of higher groups of taxa). However, a glance through most biology textbooks will reveal that macroevolution per se is particularly poorly served, both at the secondary and undergraduate levels. Indeed, Catley (2006) has argued that as a result of the overwhelming emphasis placed on understanding microevolutionary processes, learners have real problems conceptualizing the historical processes that provide the pattern of biological diversity we see in the world around us.

Understanding evolution can be a daunting task. Arguably, understanding the macro component of evolution can be even more challenging. A number of barriers have to be negotiated before the full spectrum of evolution can be comprehended. Some of these difficult concepts are the mechanisms for creating variation, selection, adaptation, speciation, and cladogenesis. That many of these mechanisms operate in the

---

Correspondence to: K. M. Catley; E-mail: kcatley@wcu.edu or laura.novick@vanderbilt.edu
DOI 10.1002/tea.20273
Published online 9 January 2009 in Wiley InterScience (www.interscience.wiley.com).

context of deep time—the well-established theory that Earth is billions of years old—complicates matters even further.

Just as deep time is the yardstick of Earth science, so it is for evolutionary biology. Be it the slow accumulation of small changes over eons (gradualism; Darwin, 1859) or patterns of punctuated equilibrium (Eldredge & Gould, 1972), the conceptual framework for understanding macroevolutionary processes must be constructed around an understanding of deep time. For example, the homology and transformation of tetrapod limbs, the period and timing of extinction events, and arguably the most pressing contemporary issue, that of evaluating and contextualizing the unprecedented changes occurring to the planet and its biota (Eldredge, 2001), cannot be understood without reference to and appreciation of deep time.

Darwin himself recognized the issue of deep time as problematic for understanding his theory: ". . . the chief cause of our natural unwillingness to admit that one species has given birth to other and distinct species, is that we are always slow in admitting any great change of which we do not see the steps . . . The mind cannot possibly grasp the full meaning of the term of a hundred million years: it cannot add up and perceive the full effects of many slight variations, accumulated during an almost infinite number of generations" (Darwin, 1859, p. 499).

Yet, scant attention has been paid to issues surrounding temporal understanding as it relates to historical (i.e., macro) evolution. Indeed, there is virtually no research that documents people's understanding of deep time in the domain of macroevolution. In line with recent calls for recognizing and addressing issues surrounding macroevolutionary understanding (Baum, Smith, & Donovan, 2005; Catley, 2006; Catley, Lehrer, & Reiser, 2005; Dodick & Orion, 2003; Goldsmith, 2003; Meir, Perry, Herron, & Kingsolver, 2007; Novick & Catley, 2007), here we present data on university students' knowledge of evolutionary deep time.

Evolutionary Deep Time

A temporal framework for understanding the evolution of life on Earth should include a number of key macroevolutionary events. Clearly, there may be some disagreement about which events to include in this category, but at the very least they should span the entirety of evolutionary time and have been instrumental in driving the most prominent evolutionary radiations. We propose the historical events shown in Figure 1: the origin of the Earth approximately 4–4.6 billion years ago (bya); the first appearance of life (prokaryotic cells) about 3.5 bya; the appearance of eukaryotic (nucleated) cells roughly 1.75 bya; the appearance of multicellular life and the subsequent radiation of multiple body plans in the Cambrian explosion (chordates and arthropods in particular) about half a billion years ago; the evolution of tetrapods about 360 million years ago (mya); the development of the amniotic egg about 300 mya and the subsequent radiation of terrestrial tetrapods, including mammals (roughly 200 mya); the extinction of most dinosaurs 65 mya; and the establishment of the "hominid"[1] lineage 2 mya.

Figure 1 is an attempt to construct a conceptual scaffold that aligns taxa with markers of key macroevolutionary events (characters) and the time frames in which they evolved. The *Y*-axis of Figure 1 is a time scale expressed as a proportion of the Earth's age, while the cladogram drawn alongside it illustrates historical events in deep time supported with the characters that are the markers of such events. Cladograms, the preferred tool of biologists, are the product of phylogenetics, the science of reconstructing evolutionary relationships using cladistics. They depict the distribution of characters among taxa and can be interpreted as a hypothesized and testable record of the relationships among taxa over time. The characters that support such relationships are known as synapomorphies (shared, derived characters; Henning, 1966) and can be thought of as observable "markers" of the processes that resulted in these relationships. Cladograms are a particular way of representing the natural world as a series of hierarchically nested sets of taxa (clades) that contain the most recent common ancestor (MRCA) and all descendent taxa. By definition, each member of a clade defined by a synapomorphy (e.g., nucleus, amniotic egg, body hair, etc.) possesses that character.

Along with the taxa and events that were the focus of our study (together with the synapomorphies that support them), we also include other taxa in Figure 1 (i.e., arthropods, modern birds, non-primate mammals, non-"hominid" primates, and primates). This helps to align important characters ("markers") with significant events and to contextualize the phylogenetic relationships among the depicted taxa. As such, Figure 1 provides a rich representation of evolutionary relationships and processes throughout time.

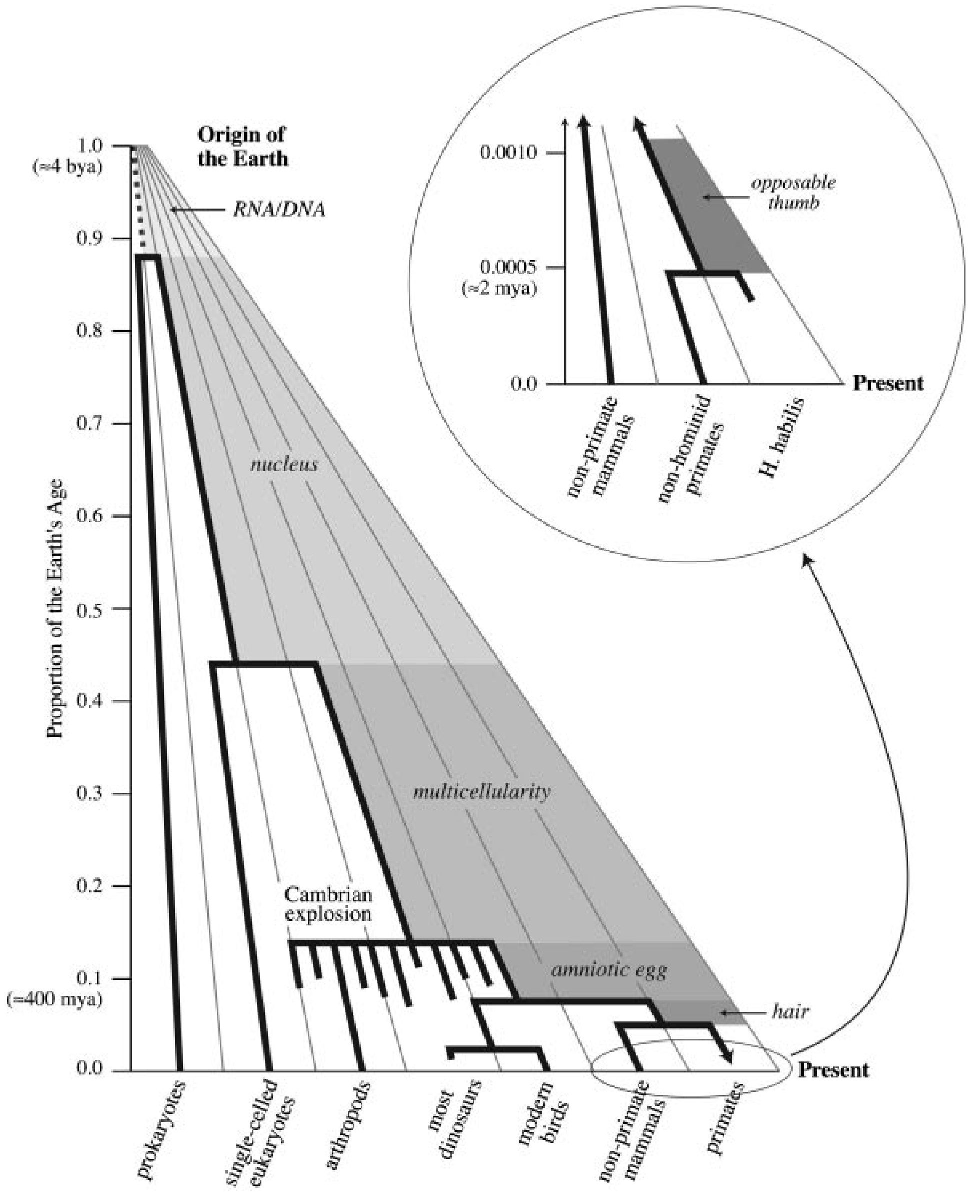

Figure 1. A cladogram showing phylogenetic relationships among taxa, including taxa that are the focus of questions 1–7, drawn alongside a time scale expressed as a proportion of the Earth's age.

What a Scientifically Literate Populace Needs to Know

We begin with the assumption that knowledge of the chronology of the major events in the history of life on Earth should be a significant component of science literacy. Such a chronology would include an appreciation of the following sequence of critical events, many of which are depicted in Figure 1: For approximately two billion years following the origin of Earth, there was little or no oxygen in the atmosphere. The accumulation of atmospheric oxygen was a result of photosynthesis by cyanobacteria (a prokaryote).

These are the earliest taxa with a fossil record, dating back to 3.5 bya. The build-up of toxic oxygen resulted in an ''oxygen holocaust'' (2.2 bya) that selected for those few prokaryotes able to utilize oxygen.

Random endosymbiotic events that occurred between prokaryotic cells are the most parsimonious explanation for the evolution of the nucleated eukaryote cell at approximately 1.75 bya. During the enormous amount of time that elapsed between the origin of life and the appearance of eukaryotic cells (fully 50% of the time life has existed on our planet), Prokaryota diversified greatly. The radiation of prokaryotes subsequent to the oxygen holocaust resulted in almost all the biological systems known today, albeit restricted to singled-celled taxa. In commenting on the period between the appearance of prokaryotes and eukaryotes, Margulis and Dolan (2002, p. 2) write: ''By the time marine algae and animals [eukaryotes] appeared, microbes had developed all the major biological characteristics: diverse energy-transforming and feeding strategies, movement, sensing, sex, and even sociality and predation. They had invented nearly everything in the modern repertoire of life except, perhaps, hallucinogens, language, and music.''

Multicellular life first appeared in the fossil record somewhere between 1,300–650 mya and later became very evident in the multiple body plans of the Cambrian explosion (about 550 mya). During the Carboniferous Period (354–290 mya), the amniotic egg was selected for in certain tetrapods, thereby allowing the ancestors of reptiles (which includes birds) and mammals to reproduce on land. Flowering plants and pollinating insects co-evolved some 125 mya, and most dinosaurs (all except the MRCA of modern birds) became extinct as a result of a meteor impact 65 mya. The hominoid lineage evolved from primate ancestors some 4 mya, with the first true ''hominid,'' *Homo habilis*, appearing 2 mya. Our own species, *Homo sapiens*, is less than one-tenth as old as *H. habilis*, having arisen a mere 150,000 years ago.

We contend that students who have been exposed to these events, in particular to their sequence and, perhaps more importantly, to their temporal spacing, will have the conceptual framework necessary to allow them to begin to make sense of the complexities of change over time and the pressures that drive it. On the other hand, a poor understanding of deep time poses a significant barrier when trying, for example, to understand changes to homologous structures over time. Such changes are the result of environmental selection acting on the variation present in organisms, conferring, or not, some advantage to the individual. Typically, changes in homologous structures, such as the five digit limbs of tetrapods (seen in, e.g., bats, whales, and frogs), take vast periods of time as traits become fixed in populations as a result of speciation over countless generations of selection.

Previous studies examining issues surrounding deep time, reviewed in the next section, have done so in the context of geology and Earth history. There is only minimal overlap between the events explored in these studies and the evolutionarily relevant events noted earlier and depicted in Figure 1—typically the age of the Earth, the origin of life, and dinosaur extinction.

Conceptions of Deep Time from Earth Science Education

Studies of deep time from the Earth science education literature can be grouped into those that probe for knowledge of actual dates of events, those that predominately ask students to sequence geological events as an indicator of at least relative knowledge of deep time, and those that examine students' ability to reason about deep time using tools from geological practice. Only the studies in the first two categories are relevant to the issues addressed here.

Sequencing Geological Events

Trend (1998) asked 10–11-year-olds in the UK to order 8–10 events (different subjects received different sets of events). The events were primarily geological (e.g., the formation of the sun, the Big Bang, the origin of the Earth, the Ice Age), although two evolutionary events were also included: the appearance of life and dinosaur extinction. Although subjects showed a general awareness of such events as the Ice Age and moving continents, there was little sense of a chronology that would allow them to appropriately order the events. Indeed, their answers fell into just two categories: extremely ancient and less ancient.

Trend (2000, 2001a,b) subsequently demonstrated that the conceptual difficulties of understanding geological deep time are not restricted to preteens. Using tasks that involved both sequencing geological events and estimating absolute time, he found that pre-service elementary teachers' and in-service primary school teachers' (in the UK) perceptions of deep time clustered into extremely ancient, less ancient, and geologically recent events (also see Hofstadter, 1985).

Estimating Absolute Time

One finding from the Earth science literature that is consistent across both ages and cultures is that people do not know the age of our planet. Marques and Thompson (1997) assessed knowledge of the origin and physical nature of the Earth and the development of life throughout geological time in two groups of Portuguese students (10–11-year-olds and 14–15-year-olds). None of these students had been exposed to any instruction in Earth science. Pertinent to our study were the extremely large range of dates given for the age of the Earth. The younger students generally gave answers in the 100s and 1,000s, whereas the older students tended to give answers in the millions, billions, or even trillions. Across all subjects, only 7.2% answered "billions" ($n \times 10^9$), the correct order of magnitude.

Oversby (1996) examined people's interpretations of fossil data and earth science concepts by administering a questionnaire to students aged 9–11 and 14–16 in the UK. When asked to give the age of the Earth, students were unable to distinguish between billions and millions of years in their answers; 36% of the younger children provided no answer at all to this question. Libarkin, Anderson, Science, Beilfuss, and Boone (2005) presented data on students' ideas about the Earth obtained from questionnaires and interviews with undergraduates from four institutions in the U.S. Fewer than half of their subjects (<10% in some institutions) knew that the Earth is 4–4.6 billion years old. The authors reported that "a significant number" of students from all four institutions either did not respond to this question or gave non-numerical answers such as "a long time ago" or "to dinosaur age."

A second intriguing finding is that many students from preteens through college age seem to be confused about the relative dating of the origin of the Earth and of life. Marques and Thompson (1997) found that approximately half of their subjects in each age group believed that these events occurred simultaneously. Similarly Libarkin et al. (2005) found that the majority of their subjects believed that some form of life existed when Earth first formed as a planet. Although this misconception was consistent across students from all four institutions, the form of this life varied widely, including dinosaurs, insects, plants, micro-organisms, and fish. Marques and Thompson opined that because both events occurred such a long time ago, their subjects may have had difficulty relating the two events on a suitably large time scale. This finding seems consistent with Trend's (1998) results that students tend to group all extremely ancient events together.

Hidalgo, Fernando, and Otero (2004) conducted the only study we have found that investigated students' knowledge of both geological and evolutionary deep time. Their subjects were Spanish secondary (16-year-olds) and post-secondary (19–20-year-olds) students who had completed several geology courses. Subjects temporally ordered four pictured scenarios of biological evolution and provided absolute time estimates: invertebrate marine taxa with a largely lifeless terrestrial background, a landscape containing dinosaurs, a landscape with birds and mammals, and a group of "hominids." Some 56% of the 16-year-olds and 83% of the 19–20-year-olds were able to correctly order the events. However, only 17% of the 16-year-olds and none of the 19–20-year-olds were able to provide correct time frames, even to an order of magnitude.

People's Misconceptions of Deep Time

The research in science education clearly shows that people from preteen school children to undergraduates and primary school teachers do not have a good working knowledge of geological deep time. For example, in the four studies that asked subjects to provide absolute time estimates, we see evidence that (a) subjects provided time frames that span unimaginably large periods of time, (b) the ability to discriminate between very large numbers was rare, (c) the frequency of time estimates to the correct order of magnitude was very low, and (d) a common misconception among both school children and college students was that life arose concurrently with the formation of Earth. This literature is less forthcoming, however, concerning the nature of people's misconceptions. For example, are absolute time estimates symmetrically distributed about the correct answer, or are they systematically biased to be too ancient or too recent? For evidence relating to this question, we must turn to other literatures.

Errors in Estimating Dates and Large Numbers

Psychological research on estimation is primarily concerned with the estimation of time on the order of seconds or minutes (roughly $n \times 10^{-7}$ to $n \times 10^{-5}$ years ago). However, a handful of studies from the psychological and survey literatures have examined people's ability to estimate the dates of events that happened in the past month to past year (i.e., $\leq 1 \times 10^0$ years ago). In addition, one study investigated people's ability to estimate the dates of events that happened approximately 2.7×10^0 to 5.0×10^1 years ago, and two studies examined people's ability to estimate populations of $n \times 10^6$ or greater.

Forward Telescoping

The consumer survey literature (e.g., Neter & Waksberg, 1964; Sudman & Bradburn, 1973) has examined errors in people's estimates of the dates of major expenditures (e.g., purchasing a new refrigerator, hiring a contractor to make residential repairs). Typically, respondents are asked to report whether they made a certain purchase within a particular time frame—the last month, the last 6 months, the last year. A consistent finding in this literature is a phenomenon known as ‘‘forward telescoping,’’ in which respondents estimate that events occurred more recently than they actually did. For example, they might report having purchased a refrigerator in the past 6 months when in fact it was purchased 8 months ago.

Forward telescoping is not restricted to estimating the dates of major expenditures. Loftus and Marburger (1983) found evidence for this phenomenon when subjects were asked to indicate whether they had been the victim of a crime during the preceding 6 months. Huttenlocher, Hedges, and Prohaska (1988) investigated forward telescoping within approximately this same time frame, asking students which movies they had seen, and when, among those shown by University of Chicago film societies during the preceding 2 months (i.e., spring quarter) or 8 months (i.e., the past academic year). They proposed a model to account for the forward telescoping of dates in memory that posited an important role for boundaries between time units at different levels (e.g., weeks, months, quarters, academic years). The model contends that it is the boundaries between these salient units in human time that provide the important markers for people when estimating time. While this model is restricted to very short time frames (i.e., those within human experience), it may still be useful in considering instructional strategies that provide ‘‘markers’’ of evolutionary events that would bracket periods of deep time. We consider the utility such an approach in Discussion Section.

Janssen, Chessa, and Murre (2006) asked respondents to a WWW survey (mean age of $\approx$44 years) how long ago major international news events occurred. For remote events, defined as those that occurred from 1,000 days (2.74 years) to about 50 years ago, they found a large forward telescoping effect. Respondents judged that these events happened approximately 520 days (1.42 years) more recently than they actually did.

Although these studies investigated the dating of events within the time frame of a human life, we hypothesize that the phenomenon of forward telescoping also applies to estimates of the age of events in Earth's history. After all, even underestimated dates of events such as the origin of life or the first appearance of mammals would likely be an unimaginably long time ago (see Hofstadter, 1985; Trend, 1998, 2000, 2001a,b).

There are also data on people's population estimates that support our hypothesis that forward telescoping applies to estimates of very large numbers. Brown and Siegler (1992) reported data from two studies of undergraduates' population estimates for 96 countries that had at least 4×10^6 people at the time the data were collected. The subjects were told that these were among the 100 most populous countries, and they were also told the current population of the U.S. Averaging the data across the two studies, the mean estimated population was only 70% of the true population.

Extreme Variability

Mirroring people's estimates of geological deep time, an October 13, 2006, survey conducted by National Public Radio (NPR) found a very wide range of estimates for the population of the U.S. On that day, the population of the U.S. hit 300 million (3×10^8). An NPR correspondent asked 80 people on the National Mall in Washington, DC, the question, ‘‘What's the current population of the United States?’’ (Kramer, 2006). The respondents came from all over the U.S. (and six came from other countries). Their estimates ranged from 17,000 (the size of a small town) to 20 billion (an order of magnitude larger than the current population of the world—6.5 billion). One-eighth said they had no idea how many people live in the U.S.

Overview of Our Research

The goal of our study is to advance understanding of the complexities of learning and teaching evolutionary time. Therefore, we interpret our data in the context of informing instruction within the larger framework of evolution education reform (Baum et al., 2005; Catley, 2006; Catley et al., 2005; Dodick & Orion, 2003; Goldsmith, 2003; Meir et al., 2007; Novick & Catley, 2007). More specifically, the overarching goal of our study is to investigate students' knowledge of the absolute dates and, critically, the relative spacing of key evolutionary events so as to inform the pedagogy of macroevolution.

Our study was motivated by a lack of data on people's knowledge of deep time through a consistent evolutionary lens. Accordingly, we presented subjects with a series of seven major events that encompassed the whole gamut of evolutionary time (see Figure 1): origin of the Earth, the first appearance of life, the appearance of eukaryotic cells, the Cambrian explosion, the appearance of mammals, the extinction of most dinosaurs, and the appearance of *H. habilis*.

The subjects in our study were university students. We assume that such students would have been exposed to these (or comparable) historical events in their high school biology classes. Therefore, we examined knowledge of evolutionary deep time as a function of students' college-level biology background. We wanted to determine whether students who had taken more advanced biology courses (including, possibly, evolution) were better able to situate events in evolutionary time than were students who had not taken such courses.

Method

Materials and Procedure

University students were asked to indicate how many years ago each of seven events in the history of life on our planet occurred. Table 1 shows the question as it was presented to subjects and, in italics, the correct answers. These particular events were chosen because of their importance to understanding macroevolution both in relative and absolute terms, as discussed earlier (see Figure 1).

This question was embedded in a multi-part problem booklet, developed by the authors, that involved data collection for several separate studies that addressed distinct conceptual and theoretical issues. The results for the other studies are presented elsewhere. In Part I of the booklet, subjects received eight cladograms involving familiar taxa and were asked several questions about the information presented in each one. In Part II, they were asked to translate cladograms from one format to another (Novick & Catley, 2007). In Part III, they answered one question about each of three cladograms involving familiar taxa. The deep time question appeared in the final section of the booklet, as part of a questionnaire that solicited background information (e.g., year in school, biology courses taken) and assessed knowledge of evolution through a series of multiple-choice, true-false, and short-answer questions (e.g., "Is variation a characteristic of species or

Table 1
The deep time questions used in this study. The correct answers (approximate dates) are given in italics.

Historical Event	How Many Years Ago?
Origin of the planet Earth	$4–4.6 \times 10^9$
Oldest rocks with indications of life (i.e., the first fossils)	3.5×10^9
Eukaryotic cells: (i.e., cells containing a nucleus and organelles)	1.75×10^9
Cambrian "explosion": (the "sudden" appearance of a large number of animals comprising many different body plans)	$5–6 \times 10^8$
Appearance of the first mammals	2×10^8
Most dinosaurs became extinct	6.5×10^7
Appearance of the first hominid (*Homo habilis*)	2×10^6

Listed above are seven events in the history of our planet. For each event, indicate how many years ago it occurred. The events are listed in order from oldest to most recent.

individuals?," "Feathers evolved so that birds could fly."—true or false?). Subjects received the same booklets regardless of the class from which they were recruited.

Subjects participated individually or in groups in a single session that took place outside of class in a laboratory room or classroom on campus. Each subject completed the booklet on his/her own without the use of any outside resources. Subjects completed the booklet at their own pace, which required about 45–70 minutes.

Subjects and Design

The subjects were 107 students recruited from a medium-sized, private, Research I university in the South and 19 students recruited from a small, private, science and engineering oriented college in the Midwest. The 107 students from the Research I university participated in partial fulfillment of course requirements for introductory psychology (20 females, 10 males) or for extra credit in the psychology (10 females, 1 male, 1 unknown sex), education (28 females, 4 males), or upper-level biology (21 females, 11 males, 1 unknown sex) class from which they were recruited. The 19 students from the science and engineering school (9 females, 9 males, 1 unknown sex) voluntarily participated at the very beginning of their evolution class, which is the third quarter of the three-quarter introductory biology sequence for science majors.

Sixteen of the 126 subjects (10 from the Research I university and 6 from the science and engineering school) did not provide any time estimates. Of the remaining 110 students, 97 provided a time estimate for all seven events, and 13 provided estimates for 1–6 events. The students at the Research I university who provided at least partial data included 28 students recruited from the introductory psychology subject pool, 11 recruited from an upper-level psychology class, 26 recruited from an education class, and 32 recruited from a biology class.

One of the questions on the background questionnaire asked subjects whether they had taken any of 10 biology (e.g., biology today, evolution, zoology) and 3 geology (environmental geology, the dynamic earth, life through time) classes at the Research I school (or similar courses at the science and engineering school). Students were assigned to *stronger* and *weaker* biology background groups based on these data. For the students at the Research I university, the minimum criterion for assignment to the stronger background group was having taken at least the first semester of the two-semester biology sequence for biology majors, pre-med majors, and other serious science students. The science and engineering school is on a quarter system rather than a semester system as is the Research I university. Those students who had taken the first two quarters of the introductory biology sequence for science majors (roughly equivalent to 1.33 semesters of the two-semester sequence at the Research I university) were assigned to the stronger background group. Looking at the missing data from this perspective, 7 of 61 stronger background and 9 of 65 weaker background students left all of the deep time questions blank.

Focusing on the 110 subjects from whom we have at least partial data, the 54 stronger background students had taken an average of 3.2 semesters of biology (or relevant geology) classes. In contrast, the 56 weaker background students had taken an average of only 0.5 semesters of such coursework. This sixfold difference in the average number of classes taken largely reflects the different majors selected by the students in the two groups. The 54 stronger background students consisted of 39 students with a biology-related major (e.g., ecology and evolutionary biology, molecular and cellular biology, biomedical engineering), 2 students with an education major, 4 students with a psychology-related major, and 9 students with a variety of other liberal-arts-related majors (e.g., history, chemistry, anthropology). The 56 weaker background students consisted of 5 students with a biology-related major, 25 students with an education major (e.g., early childhood education, elementary education, secondary education and math), 14 students with a psychology-related major (e.g., psychology, child development, cognitive studies), and 12 students with a variety of other liberal-arts-related majors (e.g., English, history, economics, human and organizational development). The stronger background students were approximately one semester farther along in their schooling than were the weaker background students (mean year in school of 2.72 and 2.13, respectively; ranging from 1 = first-year undergraduate to 6 = second-year masters student), $t(108) = 2.50$, $p < 0.02$, SE = 0.24.

Results

Apparent "Creationists"

The task of estimating the dates of events in Earth's history is different for people who believe the Biblical account of creation than for those who accept the scientific account of the origin of Earth and the evolution of life inhabiting it. In particular, creationists would be expected to give an artificially low estimate of the age of the Earth, coupled with a relatively flat pattern of time estimates for the remaining events. By this criterion, three of the 110 subjects who answered the deep time questions gave a pattern of estimates characteristic of a belief in the Biblical account of creation. Two of these subjects provided a time estimate for all seven events; one provided estimates for all events except the Cambrian explosion.

These three subjects estimated Earth to be 10, 6, and 4.5×10^3 years old. The next most recent estimate for this event was 4×10^4 years ago. To provide a simple assessment of the steepness or flatness of the pattern of time estimates, we computed a ratio of each subject's time estimate for the appearance of *H. habilis*, the most recent event, to his/her estimate of the age of the Earth. A proportion equal to 1 suggests a flat function across all events; smaller proportions suggest a steeper function. The proportions for the three apparent creationists ranged from 0.5 to 1.0. The largest proportional (relative time) estimate for *H. habilis* among the remaining 107 subjects was 0.25. Looking at the distribution of relative time estimates for this event across the entire sample, the break is clearly between 0.25 and 0.5, supporting our contention that the three subjects in question provided a very different set of time estimates than did the remaining subjects.

Absolute Time Estimates

Descriptive Data. The statistical analyses exclude data from the three apparent creationists. Table 2 provides descriptive data concerning the distribution of time estimates, along with the number of subjects providing estimates for each event. Immediately apparent are the startlingly large time ranges provided for all seven events. Indeed, the variation is so large (ranging over 8–9 orders of magnitude for each event) that presenting these data graphically is a challenge. The actual range of dates for these events is approximately 2×10^6 to 4×10^9 years ago. Subjects' answers ranged from 8×10^2 years ago (the Middle Ages—the time of the Magna Carta, St. Thomas Aquinas, and the mathematician Fibonacci) to 6×10^{11} years ago (586 billion years *before* the big bang). In the following subsections, we consider the results for each event in more detail.

Age of the Earth. Figure 2 shows the distribution of Earth age estimates for stronger and weaker background subjects. Note that the *X*-axis is on a logarithmic scale, so the actual time span is much larger than it appears in the figure. Although the distribution of Earth age estimates ranged over eight orders of magnitude—from 4×10^4 to 6×10^{11} years ago, most subjects (67.3%) indicated the correct order of magnitude ($n \times 10^9$ years ago). There was no difference between stronger and weaker background subjects for this measure of accuracy, $t(102) = -1.22$, $p > 0.23$. However, stronger background subjects were much more likely than weaker background subjects to give a more precisely correct time estimate in the range of $4–4.6 \times 10^9$ years ago: 41.2% versus 13.2%, respectively, $t(102) = 3.35$ $p < 0.001$. Nevertheless, fewer than half of the stronger background students gave a correct estimate.

Table 2
Summary of subjects' responses for the age of the seven historical events, showing N for each question and the minimum, 25th percentile, median, 75th percentile, and maximum.

Historical Event	*N*	Min.	25th percentile	Median	75th percentile	Max.
Age of the earth	104	4×10^4	2×10^9	4×10^9	6×10^9	6.0×10^{11}
First fossils	103	3×10^4	1×10^8	1×10^9	3×10^9	2.5×10^{11}
Eukaryotic cells	102	1×10^4	8×10^6	6×10^8	2×10^9	2.0×10^{11}
Cambrian explosion	101	6×10^3	2×10^6	1×10^8	5×10^8	1.0×10^{11}
First mammals	101	5×10^3	1×10^6	3×10^7	1×10^8	2.5×10^{10}
Dinosaur extinction	103	1×10^3	6×10^5	1.5×10^7	6.5×10^7	2.0×10^{10}
First "hominid"	103	8×10^2	1×10^4	3×10^5	4×10^6	5.0×10^9

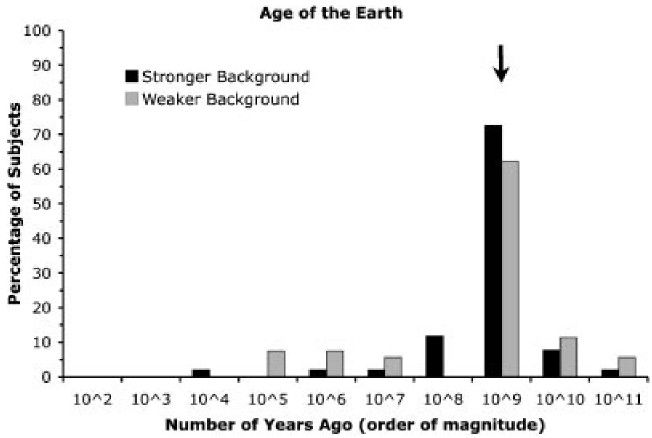

Figure 2. Histogram of the distributions of the time estimates of stronger and weaker background subjects for the age of the earth. Numbers of years ago are presented in orders of magnitude. An arrow indicates the order of magnitude that contains the correct date.

To test for differences in the distributions of the estimates given by subjects in the two knowledge groups, we collapsed the responses into four categories: 10^4–10^7, 10^8, 10^9, and 10^{10}–10^{11}. Categories were collapsed here and for the analyses of the distributional data for the remaining questions due to low expected frequencies. For each event we collapsed adjacent categories with low frequencies that showed similar patterns across knowledge groups. This solved the problem of low expected frequencies without obscuring patterns in the data. The correct order of magnitude was always maintained as a separate category. The results of the distributional analysis for Earth's age indicated that the distributions differed for stronger versus weaker background subjects, $X^2(3) = 11.90$, $p < 0.01$. As can be seen in Figure 2, the weaker background subjects had more estimates in the tails of the distribution.

Age of the First Fossils. Figure 3 shows the distribution of time estimates for the first appearance of life (i.e., the first fossils). This event, which happened approximately 3.5×10^9 years ago, was grossly underestimated by a large number of subjects. This underestimation provides evidence for the forward telescoping that we predicted. The median estimate was 1×10^9, and 75% of subjects gave a date of 3×10^9 years ago or earlier. The range encompassed an impressive amount of time from 3×10^4 to 2.5×10^{11} (see Table 2). Somewhat more than half of the subjects in each knowledge group gave an answer in the correct order of magnitude (55.5% overall), with no difference between the groups on this measure of ballpark accuracy, $t(101) = -0.66$, $p > 0.51$. However, just 5.9% (3 of 51) of stronger background subjects and only a single weaker background subject (of 52; 1.9%) knew that life appeared 3.5×10^9 years ago. As for the more lenient measure of accuracy, the difference between groups is not significant, $t(101) = 1.10$, $p > 0.30$. To test for distributional differences in the estimates, we collapsed the responses into four categories: 10^4–10^6, 10^7–10^8, 10^9, and 10^{10}–10^{11}. As for the age of the Earth, the estimates of weaker background subjects shower greater variability, $X^2(3) = 8.97$, $p < 0.03$, with the trend being largely underestimation.

First Eukaryotic Cells. Figure 4 shows the distribution of time estimates for the first appearance of eukaryotic cells (i.e., cells with nuclei), which range from 1×10^4 to 2×10^{11} years ago (see Table 2). The date of this very significant event in the evolution of Eukaryota (which comprises all taxa except bacteria and archebacteria) was grossly underestimated, again providing evidence for forward telescoping, as the median estimate of 6×10^8 is an order of magnitude below the correct date (1.75×10^9 years ago). Fewer than half of all subjects (44.1%) provided answers to the correct order of magnitude—$n \times 10^9$, with no significant difference between subject groups, $t(100) = -0.93$, $p > 0.36$. Because the correct date is not a "round" number, we counted answers in the range 1.5–2×10^9 years ago as correct. Sixteen subjects (15.7%) provided correct estimates by this measure, with no difference between the two groups, $t(100) = 1.25$,

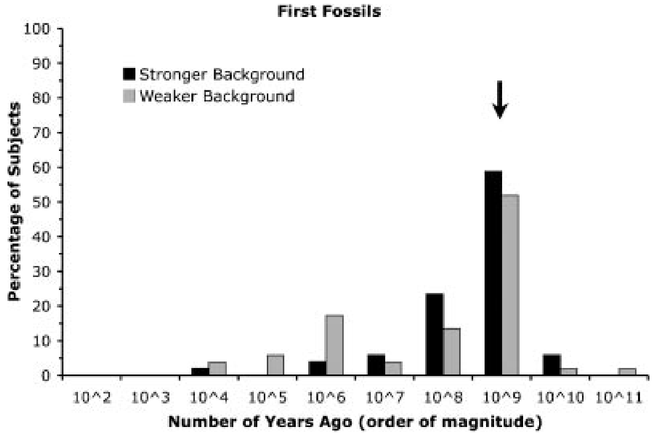

Figure 3. Histogram of the distributions of the time estimates of stronger and weaker background subjects for the age of the first fossils (i.e., the origin of life). Numbers of years ago are presented in orders of magnitude. An arrow indicates the order of magnitude that contains the correct date.

$p > 0.20$. To test for distributional differences in the estimates, we collapsed the responses into four categories: 10^4–10^5, 10^6–10^7, 10^8, and 10^9. (A single outlier, a weaker background subject whose time estimate fell in the $n \times 10^{11}$ category, was excluded from this analysis.) Unlike for the preceding two events, there was no difference in the distributions of the time estimates for the two knowledge groups for the date of the first eukaryotic cells, $X^2(3) = 2.82, p > 0.40$.

Cambrian "Explosion". Figure 5 shows the distribution of time estimates for the Cambrian "explosion." The estimates range from 6×10^3 to 1×10^9 years ago, with a median of 1×10^6 (see Table 2). As the correct date is approximately 5.5×10^8 years ago, we again find evidence for forward telescoping. There was no difference between the stronger and weaker background subjects in providing an answer either in the correct order of magnitude ($n \times 10^8$ years ago; 39% overall) $t(99) = -1.00, p > 0.32$, or in the more precisely correct time range of 5–6×10^8 years ago (19% overall), $t(99) = 1.42, p > 0.16$. To test for

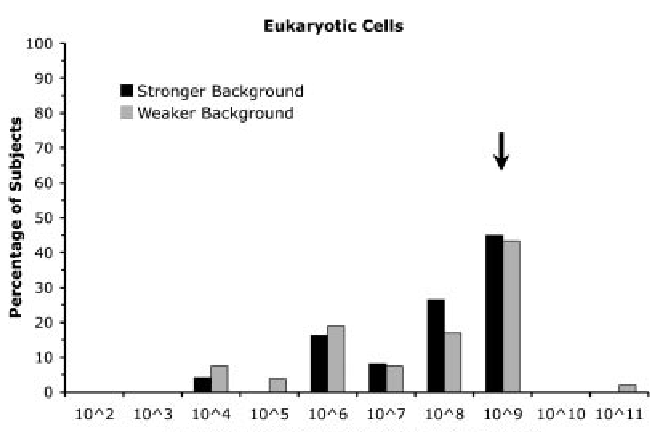

Figure 4. Histogram of the distributions of the time estimates of stronger and weaker background subjects for the age of the first eukaryotic cells. Numbers of years ago are presented in orders of magnitude. An arrow indicates the order of magnitude that contains the correct date.

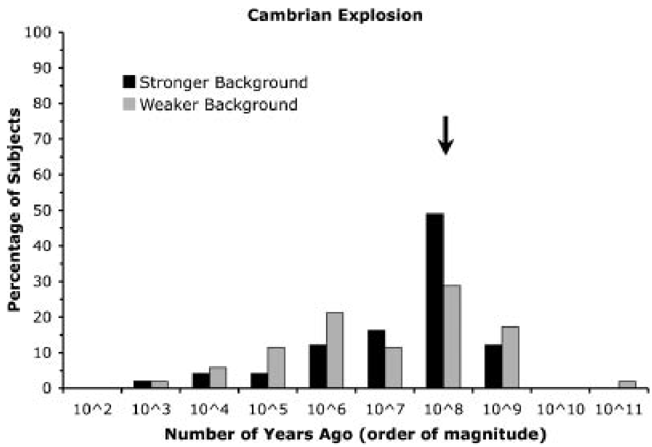

Figure 5. Histogram of the distributions of the time estimates of stronger and weaker background subjects for the age of the Cambrian "explosion." Numbers of years ago are presented in orders of magnitude. An arrow indicates the order of magnitude that contains the correct date.

distributional differences in the estimates, we collapsed the responses into five categories: 10^3–10^5, 10^6, 10^7, 10^8, and 10^9–10^{11}. As for the appearance of eukaryotic cells, there was no difference in the distributions for the two groups for the Cambrian explosion, $X^2(4) = 6.42$, $p > 0.10$. Approximately equal numbers of subjects gave answers in the categories $n \times 10^6$, $n \times 10^7$, and $n \times 10^9$ years ago. This pattern and the similarly wide distribution of estimates for the two subject groups suggest that this was an especially difficult event for our subjects to estimate.

First Mammals. Figure 6 shows the distribution of time estimates for the appearance of the first mammals. This question produced a wide distribution of times ranging from 5×10^3 to 2.5×10^9 years ago (Table 2). With a median time estimate of 3×10^7 years ago, compared with the correct date of 2×10^8 years ago, we again find evidence of forward telescoping. Only 23.8% of subjects gave answers that fell into the correct order of magnitude—$n \times 10^8$ years ago, with no difference between the knowledge groups, $t(99) = -1.10, p > 0.29$. A mere 5% of subjects provided the correct date, again with no difference between

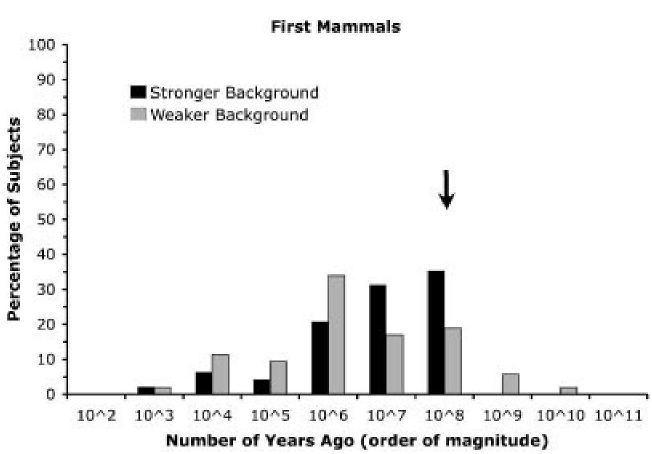

Figure 6. Histogram of the distributions of the time estimates of stronger and weaker background subjects for the age of the first mammals. Numbers of years ago are presented in orders of magnitude. An arrow indicates the order of magnitude that contains the correct date.

the two groups, $t(99) = -0.3$, $p > 0.73$. Note that approximately the same percentages of subjects provided answers that fell into the two categories earlier than the correct one—$n \times 10^6$ years ago (27.7%) and $n \times 10^7$ years ago (26.7%)—as provided answers of the correct order of magnitude. To test for distributional differences in the estimates, we collapsed the responses into five categories: 10^3–10^5, 10^6, 10^7, 10^8, and 10^9–10^{10}. As we found for several other events, the estimates of weaker background subjects were significantly more widely distributed, and in this case were roughly equally distributed in both tails of the histogram, $X^2(4) = 11.38$, $p < 0.03$.

Extinction of Most Dinosaurs. Figure 7 shows the distribution of time estimates for the extinction of most dinosaurs. Estimates range from 1×10^3 to 2×10^{10} years ago, with a median of 1.5×10^7 years ago (see Table 2). Some 35.0% of subjects gave answers that fell into the broadly correct category of $n \times 10^7$ years ago, with no significant difference between the two knowledge groups, $t(101) = -1.10$, $p > 0.31$. The remaining subjects seemed to have little idea of the correct date, with fully two-thirds underestimating (many to a large extent) the timing of this extinction event, again providing evidence for forward telescoping. Only 12.6% of subjects provided the correct date—6.5×10^7 years, with no significant difference between subject groups, $t(101) = 2.2$, $p > 0.30$. To test for distributional differences in the estimates, we collapsed the responses into six categories: 10^3–10^4, 10^5, 10^6, 10^7, 10^8, and 10^9–10^{10}. The stronger background subjects primarily gave estimates of the correct order of magnitude, whereas the weaker background subjects gave estimates that are more evenly distributed across a wider range, $X^2(5) = 15.08$, $p < 0.001$.

First "Hominid"—H. habilis. Figure 8 shows the distribution of time estimates for the appearance of the first "hominid," *H. habilis.* This question produced the lowest percentage of correct answers and a very wide range of estimates from 8×10^2 years ago (during the Middle Ages) to 5×10^9 years ago (roughly a billion years before the planet existed), with a median of 3×10^5 years ago. Only 21.4% of subjects gave estimates of the correct order of magnitude—$n \times 10^6$ years ago, with no difference between subject groups, $t(101) = -1.20$, $p > 0.24$. Absolute correct answers of 2×10^6 years were given by only 3.9% of subjects, with no difference between groups, $t(101) = 1.00$, $p > 0.30$. To test for distributional differences in the estimates, we collapsed the responses into six categories: 10^2–10^3, 10^4, 10^5, 10^6, 10^7, and 10^8–10^9. As we found for several other events, the estimates of weaker background subjects were somewhat more widely distributed, $X^2(5) = 10.31$, $p < 0.07$. Again, we see a pattern of gross underestimation indicative of forward telescoping: 25.2% of subjects reported $n \times 10^5$ years ago, 23.3% reported $n \times 10^4$ years ago, and 11.7 reported $n \times 10^3$ years ago.

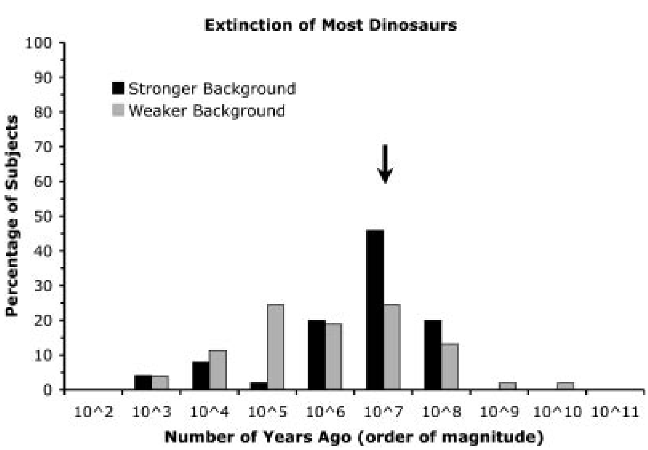

Figure 7. Histogram of the distributions of the time estimates of stronger and weaker background subjects for the age of the extinction of most dinosaurs. Numbers of years ago are presented in orders of magnitude. An arrow indicates the order of magnitude that contains the correct date.

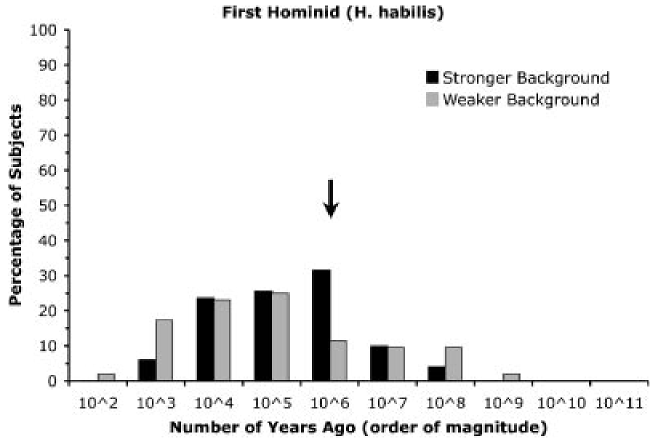

Figure 8. Histogram of the distributions of the time estimates of stronger and weaker background subjects for the age of the first ''hominid,'' Homo habilis. Numbers of years ago are presented in orders of magnitude. An arrow indicates the order of magnitude that contains the correct date.

Students Who Had Taken a Class on Evolution. The overriding impression from the preceding analyses is that students with stronger backgrounds in biology are hardly more knowledgeable about deep time than weaker background students. However, the stronger background group is far from homogeneous. From amongst this group we were particularly interested to see if those students who had taken a course in evolution gave more accurate time estimates as a result of their instruction in evolutionary theory. Based on responses to the background questionnaire, we identified 20 such students. Of these, 16 provided a time estimate for all seven events, 2 provided estimates for 5–6 events, 1 provided estimates for only two events (age of the Earth and *H. habilis*), and 1 provided no time estimates. All time estimates provided by these subjects are shown in Figure 9, with each horizontal line indicating an estimate given by an individual subject. The left and right panels show the individual data points for the first four events and final three events, respectively (note the somewhat different range of dates on the *Y* axes for the two panels). For each event, the large ''X'' represents the correct answer. Several important results are apparent from this figure: (a) a little more than half (55%) of these students knew the approximate age of the earth (11 of the 19 who gave estimates said the earth was 4–4.6 billion years old); however, (b) they did not know specifically when any of the other six evolutionary events occurred; (c) for each event, their range of estimates spans at least five orders of magnitude; and (d) dates were more likely to be underestimated, often drastically, than overestimated, again providing evidence for forward telescoping; a striking example is one student's claim that *H. habilis* appeared only 4,000 years ago (i.e., 2000 B.C.E.; for comparison, Stonehenge dates from 3000 to 1600 B.C.E.).

Summary. Apart from the age of the Earth, stronger and weaker background subjects did not differ statistically in providing estimates of the correct order of magnitude. However, differences between the distributions of stronger and weaker background subjects' time estimates were significant or marginally significant for five of the seven events (age of the Earth, age of the first fossils, extinction of most dinosaurs, appearance of the first mammals, and the first ''hominid''). In general, the weaker background subjects gave estimates that were distributed across a wider range (i.e., more estimates in both tails of the distribution). Moreover, as we predicted, subjects largely underestimated the age of these events, consistent with the phenomenon of forward telescoping.

Relative Spacing of the Events

To investigate students' knowledge of the relative spacing of the seven events, we converted each subject's absolute time estimate for all events after the appearance of the planet Earth to relative time estimates. This was done by dividing the absolute time estimate for each event by the subject's estimate of the

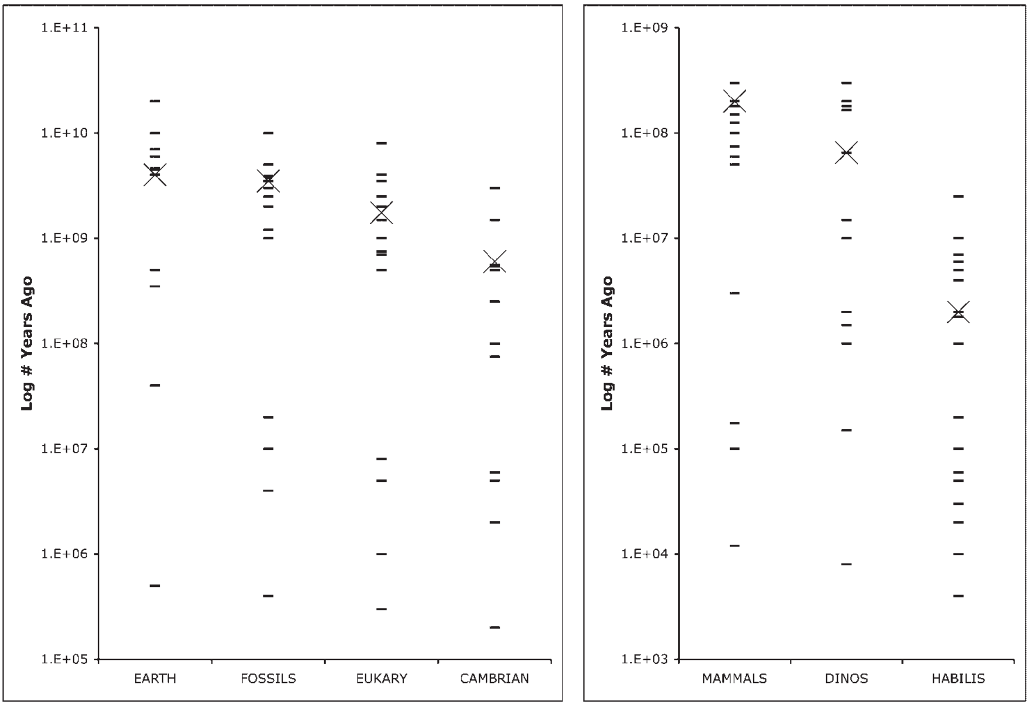

Figure 9. Distributions of time estimates given by those subjects who had completed a college course in evolution.

age of the Earth. Obviously, the three students who left the age of the Earth blank could not be included in this analysis. From these data, it is possible to compare subjects' relative spacing of the events to the correct function.

Such a comparison raises the question of whether the subjects had similar conceptions of the relative spacing of the seven events or whether different subjects (either correlated with biology background or cross-cutting those groups) produced discriminably different patterns of spacing. To answer this question, we submitted the relative time estimates (excluding those from the three apparent creationists) to the K means cluster analysis program in SPSS. Because this program only includes cases with data on all variables, the results are based on 95 of the 107 subjects (45 stronger and 50 weaker background). This program identifies clusters of cases in the data set. Unlike programs that cluster variables, the case-clustering program does not include an algorithm for determining the best-fitting number of clusters. Therefore, we ran the program multiple times, each time requesting a different number of clusters, from 3 to 8. The seven and eight cluster runs did not converge within 10 iterations (the program's default), so we presume that there are not that many discriminably different patterns of relative spacing in the data.

Figure 10 shows the results for the five-cluster solution; this solution seems the most illuminating and includes a reasonable number of subjects in each cluster. In addition, we computed the average relative spacing for the three apparent creationists by hand and included that function in the figure (see the grey line with an "X" marking each data point). It is obvious how different their function is from those of the remainder of the subjects. Finally, we included the correct function, labeled *REALITY* in the figure—the thick black line with no data point markers.

The graph clearly shows for each cluster the extent to which the date of each evolutionary event was displaced closer to or farther from the origin of Earth relative to reality. The largest cluster (#1), containing 35 of the 95 subjects (36.8%), greatly overestimated the amount of time that passed before life appeared. While life appeared in the fossil record at 88% of Earth's age, the subjects in this cluster estimated

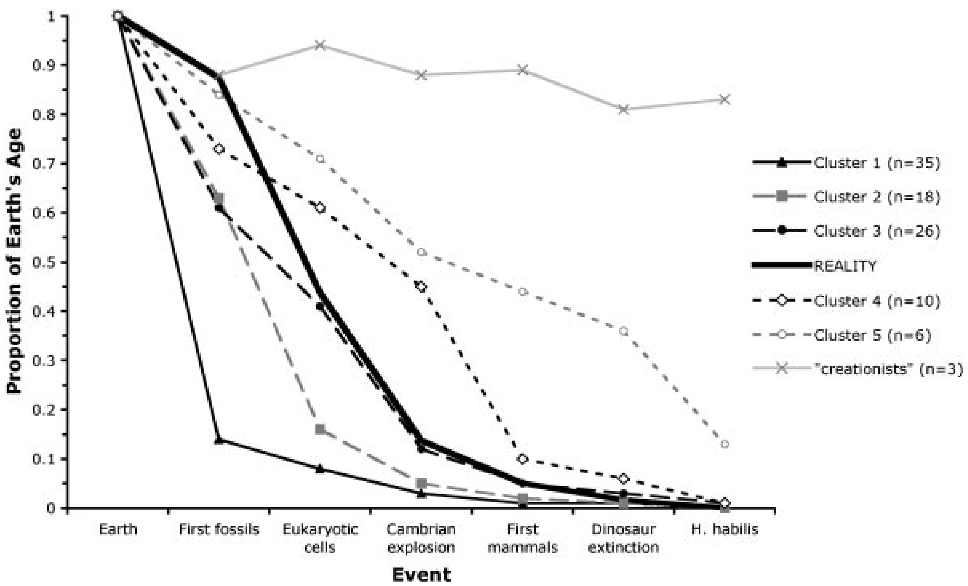

Figure 10. Results of a K means cluster analysis yielding five clusters. Note that the clusters derived by the program are numbered from left to right across the graph for ease of discussing the results rather than in the order in which they emerged from the analysis (i.e., as numbered in the SPSS output).

this event to be only 14% the age of the Earth. Subsequent events were also necessarily displaced too far from the starting point, as can be clearly seen in the figure.

The subjects in cluster 2 ($n = 18$; 18.9% of the sample), while still placing the events farther from the origin of the Earth than they should, were much more accurate about the relative spacing for the appearance of life: 63% of Earth's age versus 88%. The spacing of the subsequent events closely followed the pattern of cluster 1 but with slightly less compression toward the present. These two clusters, together comprising 56% of the subjects, displaced the major evolutionary events further from the starting point and bunched them too closely towards the present.

The other large cluster (#3), containing 26 subjects (27.4% of the sample), showed approximately the same relative spacing for the appearance of life as did the subjects in cluster 2 (61% of Earth's age). However, these subjects' relative spacing of the remaining five events was very close to reality, as can be clearly seen in Figure 10.

By comparison, the subjects in the other two much smaller clusters, comprising just 16.8% of the sample, generally underestimated the amount of time that elapsed from the origin of the Earth to the subsequent events, thereby compressing the spacing of the events too close to the starting point. The subjects in cluster 4 (10.5% of the sample) placed the appearance of life too far from the origin of the Earth, although not to the same degree as did those in clusters 1–3 (73% of Earth's age). The subsequent events, however, were placed too close to the starting point, especially the appearance of eukaryotes and the Cambrian "explosion." The subjects in cluster 5 (6.3% of the sample) were quite accurate at estimating the spacing between the origin of Earth and the appearance of life (84% of Earth's age vs. 88%), but they compressed the spacing of all of the subsequent events much too far back in the past (see Figure 10).

In a follow-up analysis, we examined whether the five clusters defined by patterns of relative spacing of the events differed with respect to the biology background of their members. The answer is *no*, $X^2(4) = 4.02$, $p > 0.40$. For each cluster, the proportion of stronger and weaker background subjects that were included in that cluster was approximately the same.

Finally, we examined the absolute time estimates for each of the five clusters. The median absolute time functions for clusters 1–4, when compared to the function depicting the true ages of the seven events, show a pattern similar to that seen in Figure 10. The six subjects in cluster 5, however, not only have a different relative spacing function than the remaining subjects, they also generally begin from a very different starting point. Whereas the median estimate for the origin of the Earth is $4–4.6 \times 10^9$ years ago for each of clusters 1–4, it is three orders of magnitude less for cluster 5—2.75×10^6 years ago. The median estimate for the origin of the Earth for the three apparent creationists is another three orders of magnitude smaller—a mere 6×10^3 years ago.

Discussion

Conceptualizing Evolutionary Deep Time

Overview. The difficulty of appreciating deep time is not hard to understand. A hundred million years is an unimaginable amount of time for most people to comprehend (Hofstadter, 1985), yet it only represents one-fortieth the age of the earth. From a more recent perspective, a hundred million years is 50 times longer than the time *H. habilis* was extant, and some 700 times longer than our own species, *Homo sapiens,* has existed. In terms of an individual person, this figure represents some 1.5 million human life spans. These numbers are so large as to be outside a normal frame of reference.

The difficulty of conceptualizing deep time, however, does not obviate the need to do so. Without an appreciation of both the relative spacing and absolute dates of key macroevolutionary events, the deep history of our planet cannot be understood. An understanding of this history, particularly its long length and interconnectedness, is vital in comprehending the gravity of the sixth mass extinction event currently in progress (Eldredge, 2001). In the next section, we consider what well-educated students ought to know about evolutionary deep time versus what they do know, and we reflect on the apparent deficiencies in their knowledge.

The Chasm Between Scientific Knowledge and Students' Knowledge. Earth's history begins with the appearance of the planet approximately 4–4.6 bya. Students should have this date memorized because knowledge of this anchor coupled with the relative spacing of major events would enable them to reconstruct the absolute dates that are difficult to remember. Our subjects were most accurate when estimating the age of the Earth, yet fewer than a third of them knew this date. Without this anchor, it is not surprising that students' estimates of the dates of the remaining six events examined in our study were characterized by both extreme variability (ranging across 8–9 orders of magnitude for every event) and generally gross underestimation. We discuss the relation between these patterns and prior research in the next section.

The origin of eukaryotic cells approximately 1.75 bya is arguably the most significant evolutionary event ever because it resulted in a staggering diversification of single- and multiple-celled taxa including plants and animals. As such it should be an anchor for understanding the relative timing of the events that followed it. It is critical, therefore, to know that a huge amount of time passed between the first appearance of life (prokaryotic fossils evident 3.5 bya; this is the second event we asked our subjects to date) and nucleated eukaryotic cells (the third event)—fully half of the time life has existed on our planet.

The Cambrian "explosion," which occurred approximately 500–600 mya, was another watershed event because it resulted in an unprecedented radiation of body plans (see Figure 1) that led to both arthropods—the most successful group (i.e., containing the largest number of species) our planet has ever seen—and chordates—the group in which our own taxon falls (with mammals appearing in the fossil record approximately 200 mya). As such, this event represents a vital piece of knowledge in building an understanding of the sequence and timing of evolutionary history.

The next-to-last event in our list—dinosaur extinction—was included because dinosaurs are a hot topic with kids of all ages, and the public is frequently exposed to dinosaur factoids at museums, on the Discovery Channel, etc. We expected that even if our subjects did not know the dates of the other events, they would know this one. We were thus surprised to find that the precise time of this extinction event (65 mya) eluded the vast majority ($>87\%$) of our subjects. Even if we adopt a more lenient criterion, considering estimates of 60–65 mya as correct, 83% of subjects got the date wrong. Students' failure to know even this date, to which they have presumably been exposed multiple times, highlights the difficulty of remembering these large numbers when they are presented as isolated facts rather than in some organizing framework.

Finally, one might have expected greater accuracy in dating *H. habilis,* our recent relative and the final (therefore most recent) evolutionary event in our list. Our results, however, again indicated an astonishing range of estimates, in this case from recorded history (the Middle Ages) to a billion years before the planet existed. These results reflect an astounding inability of students to link up evolutionary events with events from world history, unless one assumes that key figures in recorded history belonged to the species *H. habilis* rather than *H. sapiens*. We find this lack of knowledge of the history of our own lineage to be especially disturbing.

Not only were our subjects inaccurate in their absolute estimates of the ages of the evolutionary events just discussed (e.g., see Figures 2–9), but their knowledge of the relative spacing of these events was far from reality. In fact, six distinct spacing patterns were observed (see Figure 10). While there is some convergence towards the correct spacing for the three most recent events, there are major errors in estimating the spacing of the three earlier ''marker'' events—the emergence of life, the first eukaryotic cells, and the Cambrian explosion. When the shape of the curves for the large majority of subjects is so different from that of reality, it is difficult to see how subjects could possibly make sense of the timing of these events in relation to one another. Underscoring the deeply rooted difficulty of understanding the unfolding of events over deep time, there was no significant effect of biology background on the spacing functions.

Relating Our Results to Previous Research

Geological Deep Time. Although there has been very little research on people's knowledge of evolutionary deep time, researchers have documented significant misconceptions concerning geological deep time among 10-year-olds through high school students, pre-service teachers, and primary school teachers (Dodick & Orion, 2002b, 2003; Marques & Thompson, 1997; Trend, 1998, 2000, 2001a,b). Our results extend these findings by documenting significant problems in university students' knowledge of evolutionary deep time, including students who have stronger backgrounds in biology (and even those who have taken a college class on evolution—see Figure 9). These problems include: (a) low levels of accuracy, even when assessed in terms of the correct order of magnitude (see the bars marked by arrows in Figures 2–8); (b) extreme variability in students' estimates; and (c) forward telescoping of dates toward the present. These latter two aspects of our results are discussed in the following two sections.

There is one inconsistency between our results and those from the Earth science education literature. Half of our subjects indicated that the appearance of life was a much more recent phenomenon than the origin of the Earth. This is not in accordance with the results of the studies of Marques and Thompson (1997) and Libarkin et al. (2005), which reveal a pattern of half or more subjects conflating the time of these two events. We can think of two reasons for this discrepancy: (a) We told our subjects that fossils came after the origin of Earth, and (b) students may think that life existed on Earth as soon as the planet was formed but that it took time for conditions to be right for the formation of fossils in rocks, which is the specific event we asked about in our study. Resolution of this discrepancy awaits further research.

Extreme Variability. Marques and Thompson (1997) and Oversby (1996) found that 9- to 16-year-olds' estimates of the dates of events in geological deep time were characterized by extreme variability. Similarly, Kramer (2006) found that adults' estimates of the current population of the United States (300 million) ranged over seven orders of magnitude (from 17,000 to 20 billion). Our results replicate these findings with respect to university students' estimates of evolutionary deep time, which ranged over at least eight orders of magnitude for each of the seven events (see Figures 2–9). This extreme variability suggests that many (perhaps most) college students are without an effective conceptual framework to make sense of the very large time frames inherent in understanding the evolution of life on Earth.

Forward Telescoping. Psychological research and consumer surveys have investigated adults' estimates for time frames ranging from several months to about 50 years (Huttenlocher et al., 1988; Janssen et al., 2006; Loftus & Marburger, 1983; Neter & Waksberg, 1964; Sudman & Bradburn, 1973). These studies found that respondents consistently underestimated the amount of time that had past since the occurrence of a queried event, a phenomenon known as *forward telescoping*. We predicted that, even though the time periods are not comparable, we would find evidence for forward telescoping (i.e., underestimation) in our subjects' estimates of deep time. This prediction was supported, as can clearly be seen in the asymmetrical distributions of subjects' estimates around the correct order of magnitude in Figures 2–9: Underestimation—by as much as five orders of magnitude—was the rule rather than the exception.

Effects of Biology Background

One of the prime motivators of this study was to ascertain whether university students who had undertaken serious coursework in biology, perhaps including a course in evolution (see Figure 9), had better knowledge of evolutionary deep time than students who had not undertaken such coursework. We found no consistent differences in accuracy between our stronger and weaker background groups, nor were there differences between the two groups in the relative spacing of events. One would have to conclude, therefore, that whatever macroevolutionary knowledge the stronger background students (including those who had taken evolution) obtained from their coursework, it was insufficient to help them make sense of the absolute timing or relative spacing of the key evolutionary events examined in our study.

When considering the distributions of time estimates for the two groups, we did find consistent differences. The weaker background subjects gave more estimates in the tails of the distribution (i.e., their estimates were more variable). This pattern is evident for most of the evolutionary events investigated in this study (see Figures 2–8). It is little consolation that the estimates of stronger background subjects are less variable, however, because (a) their estimates still range over as many as eight orders of magnitude and (b) their order of magnitude accuracy ranges from 31% to 73% correct ($M = 48\%$).

Limitations of This Study

While our sample was more than adequate to draw meaningful statistical conclusions ($N = 126$) and was not gender biased, we do recognize its socio-economic homogeneity and the selective nature of the two universities. However, by the same token, we might expect these subjects, who largely have had access to exceptional educational opportunities, to have a better knowledge of and perspective on deep time than most college students. Consequently, we feel generalizing these data to the student population at large is justified.

With respect to our definitions of stronger versus weaker backgrounds in biology, we recognize that splitting the subject sample approximately in half is a less powerful design than selecting extreme groups. We should note, however, that the absence of differences in accuracy between the two groups in this study is unlikely to reflect a weak manipulation of biology background, as we have found strong differences between these groups on tasks involving cladogram understanding (e.g., Novick & Catley, 2007).

Thoughts on How to Teach Evolutionary Deep Time

Overview. It seems clear that our subjects had little to support them in their task of estimating these particular evolutionary times frames. Our data point to the conclusion that a working knowledge of deep time is not being provided in high school or even undergraduate biology classes when macroevolution is being considered. In light of our results, and especially in response to recent calls to introduce *tree thinking* into the biology classroom (Baum et al., 2005; Catley, 2006; Catley et al., 2005; Dodick & Orion, 2002a, 2003; Goldsmith, 2003; Meir et al., 2007; Novick & Catley, 2007), we consider here an instructional strategy that uses a phylogenetic tree to introduce key evolutionary events in terms of their ‘‘markers’’ and spacing.

Background. Trend (2001b) argues that time lines may be appropriate for historical dates but likely not for deep time. Similarly, the time-honored method of learning a mnemonic to remember the geological epochs does nothing to provide students with a conceptual framework. Indeed, Steven J. Gould once admitted in an interview that as an undergraduate he was never able to recall the geological time scale! Presumably, however, practicing evolutionary biologists and geologists do not have difficulty conceptualizing deep time. It seems reasonable to suggest, therefore, that exposing students to the ways of thinking inherent in these fields may facilitate their comprehension of deep time.

Looking at the strategies experts use when working across many orders of magnitude of size (from atomic to galactic), Tretter, Jones, and Minogue (2006) are struck by the consistent necessity for some linkage to the ‘‘concrete’’ when conceptualizing very large or very small sizes. For example, linking spatial scales to personal experiences such as using instruments to visualize such scales was of great importance. Without some ‘‘marker,’’ understandable at the human scale, experts often perform no better than non-experts in estimating extreme sizes. How then do we move from the abstract to the concrete when working in the world of deep time?

Dodick and Orion (2003) argue that because evolution is so often approached from a genetic or molecular perspective, focusing on abstract processes such as mutation that students are unable to ‘‘see,’’ students have difficulty understanding or even accepting evolution. In their evolution curriculum, *From Dinosaurs to Darwin*, Dodick and Orion (2002a) focus on reconstruction of evolutionary changes rather than on mechanisms, which are the current focus in Israel’s (and most other countries’) high school biology programs. They suggest that evolutionary processes in geological time might be better represented by a phylogenetic tree, one that emphasizes both the temporal scale and the events (represented by characters) that support them.

A Tree-Thinking Approach. Building on the work of Dodick and Orion (2002a,b, 2003), as well as responding to recent calls to introduce tree thinking into the biology classroom (Baum et al., 2005; Catley, 2006; Catley et al., 2005; Goldsmith, 2003; Meir et al., 2007; Novick & Catley, 2007) and to the appeal in the National Science Education Standards (National Research Council, 1996) that students be exposed to scientific tools and ways of thinking, we suggest that deep time be taught using a relative approach. The cladogram in Figure 1, which is integrated with a (proportional) time scale to highlight the relative spacing of major evolutionary events, is illustrative of this approach. Each hierarchic set of relationships is based on a shared, derived character (a synapomorphy) or ‘‘marker’’ (e.g., having an opposable thumb defines the group of primates). This tree-thinking framework provides a conceptual basis that situates the evolution of characters and selection pressures in historical time and clearly shows patterns of diversification during different stages in the history of planet Earth. It is the way evolutionary biologists think, and it is a tool used by all biologists who work in historical or comparative frameworks.

Combining evolutionary and deep time perspectives might be achieved by using a diagram like Figure 1 in conjunction with the ‘‘reality’’ curve for relative time shown in Figure 10. Rather than presenting students with a piece-meal series of events in disembodied time, an approach that gives students a holistic trajectory of events in relative time might be an effective tool in the classroom. Such a framework might scaffold knowledge-building by presenting significant evolutionary events not only by their sequential appearance in time, but also aligned with the observable characters that are the markers of these events. If students are provided with knowledge of the absolute timing of a small number of critical events, along with a visualization of the relative spacing of these events linked to concrete observable characters (synapomorphies) such as those depicted in Figure 1, we might expect to see a very different pattern of estimates than those presented in Figure 10.

Concluding Remarks

Referencing Gould (1990, p. 24), who makes the case that we need a ‘‘proper scale for our environmental crisis,’’ one that the ‘‘measuring rod of a human lifetime’’ cannot provide, Trend (2000) argues convincingly for the critical importance of a framework of deep time to make sense of the patterns of drastic change we observe on our planet. Elsewhere, Trend (2001b) makes the case that the classroom teacher must face the challenge of confronting the many confusions surrounding deep time, thus allowing a clearer perspective to emerge on contemporary environmental issues and crises, such as the current sixth mass extinction (Eldredge, 2001).We also make a plea for further research into the complexities that surround understanding deep time and into finding ways to incorporate such thinking into biological, earth science, and environmental studies. Only by understanding the importance of deep time in historical biology and by being exposed to the ways of thinking inherent in earth science and geobiology will our students see the relevance of deep time to their contemporary world. This is a non-trivial task, yet one with powerful implications; there is a real necessity and great utility in this approach. Two unique and highly significant insights result from studying deep time: That over vast periods of time very improbable events do indeed occur; and that very small, seemingly insignificant changes do accumulate, often with major impacts. Living during a period of unprecedented assault on the environment, while at the same time constricted by miniscule life spans, we would do well to heed such insights.

We would like to thank Ella Ingram in the Department of Applied Biology and Biomedical Engineering at the private science and engineering oriented college for collecting data from the students in her evolution class. We also thank Courtney Shade for her help in collecting data from the students at the Research I university and for her help in scoring the data. Finally, we thank Mark Putnam for help with designing and drawing Figure 1.

[1]Although classification issues involving the genus *Homo* are far from resolved, we use the term ‘‘hominid’’ throughout in this article to refer to members of the genus *Homo*.

References

Baum, D.A., Smith, S., & Donovan, S.S. (2005). The tree-thinking challenge. Science, 310, 979–980.

Brown, N.R., & Siegler, R.S. (1992). The role of availability in the estimation of national populations. Memory & Cognition, 20, 406–412.

Catley, K.M. (2006). Darwin’s missing link: A new paradigm for evolution education. Science Education, 90(5), 767–783.

Catley, K., Lehrer, R., & Reiser, B. (2005). Tracing a prospective learning progression for developing understanding of evolution. Paper Commissioned by the National Academies Committee on Test Design for K–12 Science Achievement. http://www7.nationalacademies.org/bota/Evolution.pdf

Darwin, C.R. (1859). On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life (1st ed.). London: John Murray.

Dodick, J., & Orion, N. (2002a). From dinosaurs to Darwin: Evolution from the perspective of ‘‘Deep Time’’. A curriculum for the Science for All Program. Rehovot, Israel: Weizmann Institute of Science.

Dodick, J., & Orion, N. (2002b). Introducing evolution to non-biology majors via the fossil record: A case study from the Israeli high school system. The American Biology Teacher, 65, 185–190.

Dodick, J., & Orion, N. (2003). Cognitive factors affecting student understanding of geologic time. Journal of Research in Science Teaching, 40, 415–442.

Eldredge, N. (2001). The Sixth Extinction. Retrieved Jan 15, 2008 from http://www.actionbioscience.org/newfrontiers/eldredge2.html.

Eldredge, N., & Gould, S.J. (1972). Punctuated equilibria: An alternative to phyletic gradualism. In: T.J.M. Schopf (Ed.), Models in paleobiology. San Francisco: Freeman and Cooper.

Goldsmith, D.W. (2003). Presenting cladistic thinking to biology majors and general science students. American Biology Teacher, 65, 679–682.

Gould, S.J. (1990). The golden rule: A proper scale for our environmental crisis. Natural History, Sept., 24–30.

Henning, W. (1966). Phylogenetic systematics. Urbana, Illinois: University of Illinois Press.

Hidalgo, A.J., Fernando, S.I.E.S., & Otero, J.I.C.E. (2004). An analysis of the understanding of geological time students at secondary and post-secondary level. International Journal of Science Education, 26, 845–857.

Hofstadter, D.R. (1985). Metamagical themas: Questing for the essence of mind and pattern (Essay #6, ‘‘On number numbness’’, pp. 115–135). NY: Basic Books.

Huttenlocher, J., Hedges, L., & Prohaska, V. (1988). Hierarchical organization in ordered domains: Estimating the dates of events. Psychological Review, 95, 471–484.

Janssen, S.M.J., Chessa, A.G., & Murre, J.M.J. (2006). Memory for time: How people date events. Memory & Cognition, 34, 138–147.

Kramer, M.J. (2006, October 16). Three hundred million? Depends on who you ask. *NPR*. Retrieved October 19, 2006, from http://www.npr.org/templates/story/story.php?storyId=6262965.

Libarkin, J.C., Anderson, S.W., Science, J.D., Beilfuss, M., & Boone, W. (2005). Qualitative Analysis of College Students’ Ideas about the Earth: Interviews and Open-Ended Questionnaires. Journal of Geological Education, 53, 17–26.

Loftus, E.F., & Marburger, W. (1983). Since the eruption of Mt. St. Helens, has anyone beaten you up? Improving the accuracy of retrospective reports with landmark events. Memory & Cognition, 11, 114–120.

Margulis, L., & Dolan, M.F. (2002). Early Life. Evolution on the PreCambrian Earth (2nd ed.). Boston, MA: Jones and Bartlett Publishers.

Marques, L.F., & Thompson, D.B. (1997). Portuguese students' understanding at age 10/11 and 14/15 of the origin and nature of the Earth and the development of life. Research in Science and Technology Education, 15, 29–51.

Meir, E., Perry, J., Herron, J.C., & Kingsolver, J. (2007). College students' misconceptions about evolutionary trees. The American Biology Teacher, 69(7), e71–e76.

National Research Council. (1996). National science education standards. Washington, DC: National Academy Press.

Neter, J., & Waksberg, J. (1964). A study of response errors in expenditures from household interviews. Journal of the American Statistical Association, 59, 18–55.

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.

Oversby, J. (1996). Knowledge of earth science and the potential for its development. School Science Review, 78, 91–97.

Sudman, S., & Bradburn, N.M. (1973). Effects of time and memory factors on response in surveys. Journal of the American Statistical Association, 68, 805–815.

Trend, R.D. (1998). An investigation into understanding of geological time among 10- and 11-year old children. International Journal of Science Education, 20, 973–988.

Trend, R.D. (2000). Conception of geological time among primary teacher trainees, with reference to their engagement with geoscience, history and science. International Journal of Science Education, 22, 539–555.

Trend, R.D. (2001a). Deep time framework: A preliminary study of U.K. primary teachers' conceptions of geological time and perceptions of geoscience. Journal of Research in Science Teaching, 38, 191–221.

Trend, R.D. (2001b). An understanding of geological time among 17-year-old students, with implications for the subject matter content knowledge of future teachers. International Research in Geographical and Environmental Education, 10, 289–321.

Tretter, R., Jones, M.G., & Minogue, J. (2006). Accuracy of scale conceptions in science: Mental maneuverings across many orders of spatial magnitude. Journal of Research in Science Teaching, 43(10), 1061–1085.