Tuition vs. Intuition: Effects of Instruction on Naïve Theories of Evolution

Andrew Shtulman   *Occidental College*
Prassede Calabi   *University of Massachusetts–Boston*

Recent research suggests that a major obstacle to evolution understanding is an essentialist view of the biological world. The present study investigated the effects of formal biology instruction on such misconceptions. Participants ($N = 291$) completed an assessment of their understanding of six aspects of evolution (variation, inheritance, adaptation, domestication, speciation, and extinction) before and after one of five evolutionary-themed courses. Most participants demonstrated pervasive misconceptions at both pretest and posttest. A subset, however, demonstrated reliable pre-post gains, and they differed from their peers in that they (a) began the semester with significantly less accurate, yet significantly more consistent, views of evolution, and (b) ended the semester with significantly less consistent, yet significantly more accurate, views of evolution. These findings indicate that naïve theories of evolution, while generally resistant to change, are less resistant the more consistent they are, possibly because consistency highlights limitations in their explanatory power and inferential scope.

The Gallup organization recently released a document summarizing Americans' attitudes toward evolution over the past 30 years (Newport, 2010). Beginning in 1982, Gallup has asked the American public this question:

> Which of the following statements comes closest to your views on the origin and development of human beings? (1) Human beings have

developed over millions of years from less advanced forms of life, but
God guided this process, (2) Human beings have developed over millions
of years from less advanced forms of life, but God had no part in this
process, or (3) God created human beings pretty much in their present
form at one time within the last 10,000 years or so.

With every poll, the percentage of Americans endorsing statement 2—the
secular evolution option—has remained around 15%. The remaining 85%
tends to be split fairly evenly between statement 1 (the *theistic evolution*
option) and statement 3 (the *strict creationist* option). Even among those
with postgraduate degrees, the trends are not much different: 25% endorse
the secular evolution option, 49% endorse the theistic evolution option, and
22% endorse the strict creationist option.

   While Gallup has been assessing public attitudes toward evolution
over the past three decades, science educators have been assessing pub-
lic understanding of evolution. One of the first to address this topic was
Brumby (1984), who documented pervasive misconceptions about evolu-
tion among medical students, including (a) the belief that organisms are
born more adapted to their environment than their parents were at birth,
(b) the belief that biological changes accrued during an organism's lifetime
will be passed down to that organism's offspring, and (c) the belief that
organisms are more likely to survive and adapt than to die or go extinct.
Many subsequent studies have confirmed these findings, documenting mis-
conceptions of a similar nature in populations ranging from elementary-
school students (Berti, Toneatti, & Rosati, 2010) to middle-school students
(Lawson & Thompson, 1988) to high-school students (Settlage, 1994)
to college undergraduates (Nehm & Reilly, 2007) to doctoral students
(Gregory & Ellis, 2009) to preservice teachers (Crawford, Zembal-Saul,
Munford, & Friedrichsen, 2005) to in-service teachers (Deniz, Donnelly,
& Yilmaz, 2008) to natural history museum visitors (Evans et al., 2010) to
educational professionals (Nadelson & Sinatra, 2009).

   From where do such misconceptions arise? Their prevalence across
development and across education suggests that they are not the product
of a few bad teachers or a few bad textbooks but are instead the product
of a more general cognitive bias. Consistent with this observation, recent
research on evolution understanding has linked students' misconceptions
to the bias of *essentialism* (Evans et al., 2010; Nettle, 2010; Shtulman &
Schulz, 2008): the commonsense assumption that the observable properties
of an entity are determined by some unobservable property at its core—
its "essence." Applied to biology, essentialism dictates that each species is
characterized by a different essence and that each member of that species

is conferred its essence at birth. What makes a tiger a tiger, for instance, is not that it has stripes or eats meat or lives in the jungle but that its parents were tigers and that, if it reproduces, its babies will be tigers, as well. The surface properties of a tiger (striped fur, carnivorous diet, tropical habitat) are seen as *consequences* of "tigerhood," not its determinants, and possession of tigerhood is seen as an organism's birthright, not its developmental end point (see Gelman, 2003; Schwartz, 1979).

The results of recent work by cognitive anthropologists and developmental psychologists suggest that essentialist thinking is ubiquitous across cultures and across age. Young children in places as diverse as Israel (Diesendruck & Haber, 2009), Brazil (Sousa, Atran, & Medin, 2002), Mexico (Atran et al., 2001), and Madagascar (Astuti, Solomon, & Carey, 2004) have all been shown to hold essentialist intuitions, particularly in the context of two experimental paradigms: an unknown-property paradigm and a switched-at-birth paradigm.

In the *unknown-property paradigm,* preschool-aged children are taught a novel property of a familiar organism—say, that cats (a familiar organism) can see in the dark (a novel property, at least to children of this age; see Gelman & Markman, 1987). They are then shown a handful of novel organisms and asked which might also possess that property. Some of the novel organisms are of the same species as the familiar organism but differ in appearance (e.g., a cat that looks like a skunk), whereas others share the same appearance but are of a different species (e.g., a skunk that looks like a cat). Children of all ages tend to extend the novel property to the former (the skunklike cat) but not the latter (the catlike skunk), implying that they view species membership as a better predictor of shared properties than is mere appearance (Gelman & Coley, 1990; Gelman & Markman, 1987). They also assume that organisms retain their species kind across their life span, even as they undergo dramatic changes in size, shape, or complexity (Keil, 1989; Rosengren, Gelman, Kalish, & McCormick, 1991). A frog is a frog is a frog, even if it starts out life looking more like a fish than a frog.

In the *switched-at-birth paradigm,* children are presented with scenarios that resemble the tale of the ugly duckling. A baby animal is removed from its birth parents (e.g., cows) and raised by members of a different species (e.g., pigs), and the question is posed as to whose properties the animal will grow to possess as an adult: those of its birth parents (a straight tail and a diet of grass) or those of its adopted parents (a curly tail and a diet of slop). Children of all ages tend to select the properties of the birth parents (Gelman & Wellman, 1991; Waxman, Medin, & Ross, 2007). They also tend to justify their judgments by appealing to the continuity of species kind, arguing, for instance, that the baby will come to eat grass "because

it's a cow, not a pig!" (Johnson & Solomon, 1997). In short, young children assume that an organism's species kind is conferred at birth and remains constant across changes in appearance, environment, and upbringing.

This basic assumption serves us well in our everyday reasoning about the biological world, for an organism's species is, indeed, a reliable predictor of its properties. Knowing that an organism is a cardinal, for instance, enables us to make accurate predictions about how that organism should look (small and red), where that organism should live (in nests and trees), what that organism should eat (worms and insects), how that organism should reproduce (by laying eggs), and so forth. Yet, despite its utility for reasoning about the properties of individual organisms, essentialism can act as a major impediment for reasoning about population-level phenomena, such as evolution and natural selection. The problem is that essentialism, while true in spirit, is false in detail. Offspring resemble their parents but that resemblance is not exact. Every organism is unique, and every population is full of variation, yet essentialist biases lead one to focus on differences *between* species and overlook the differences that actually matter most to evolution: differences among individuals *within* species.

Historians of science have long bemoaned the influence of essentialist thinking on the development of evolutionary theory (Gould, 1996; Hull, 1965). Ernst Mayr (1982), for instance, has argued that all evolutionary theorists prior to Darwin conceptualized species as discrete, homogeneous units whose aggregate properties are true of all members of the species and are faithfully transmitted from one generation to the next. These theorists posited mechanisms of evolution that operated indiscriminately over all individuals within a population—mechanisms like the inheritance of acquired traits (Jean-Baptiste Lamarck's mechanism by which offspring inherited ontogenetically derived traits that proved useful to the parent) or the law of accelerated growth (Edward Cope's mechanism by which species evolved through the compression of early stages of embryonic development and the addition of new stages). Mayr (2001) termed these pre-Darwinian theories of evolution *transformational* theories because they construe evolution as the cross-generational transformation of an entire species, with every organism producing offspring better adapted to the environment than the organism itself was at birth. Not until Darwin did evolutionary biologists begin eschewing specieswide similarities for within-species differences. The result was a qualitatively different view of evolution—what Mayr termed a *variational* theory—in which evolution is (correctly) construed as the selective propagation of within-species variation, with some individuals possessing variations that are a better match for

the environment at that time and thus enable them to survive and reproduce more successfully than other individuals.

To illustrate exactly how essentialist biases give rise to evolutionary misconceptions, consider the six phenomena listed in Table 1: variation, inheritance, adaptation, domestication, speciation, and extinction. At a microevolutionary level, essentialist biases lead to a consistent

Table 1. Variational (V) and transformational (T) interpretations of six evolutionary phenomena

Phenomenon	Theory	Interpretation
Variation	V	Individual differences are fodder for selection.
	T	Individual differences are minor and nonadaptive.
Inheritance	V	Differences between parents and offspring are random and unpredictable.
	T	Differences between parents and offspring are adaptive and purposeful.
Adaptation	V	Adaptation is a product of differential survival and reproduction.
	T	Adaptation is a product of mutation in response to a species' needs.
Domestication	V	Species are domesticated via selective breeding.
	T	Species are domesticated via enduring changes to individual organisms.
Speciation	V	New species emerge when two populations diverge.
	T	New species emerge when one population transforms into another.
Extinction	V	Extinction is more common than adaptation.
	T	Adaptation is more common than extinction.

undervaluation of both individual differences within a population (variation) and randomly occurring differences between parents and their offspring (inheritance). As a result, adaptations are viewed not as the product of differential survival and differential reproduction but as the product of some endogenous process ensuring that the species *as a whole* will change in response to its environment. Inheritance thus takes the place of selection as the mechanism of adaptation, with every organism producing offspring better adapted to the environment than it was at birth. At a macroevolutionary level, essentialist biases cause confusion about how new species arise (speciation), how old species die off (extinction), or how humans can modify species, at a genetic level, to fit their needs (domestication). Those confusions are resolvable, however, if one views speciation as the metamorphosis of one discrete essence into another (vs. the divergence of two separate populations), extinction as sudden, widespread death (vs. selection gradually outpacing variation), and domestication as a seemingly enduring form of nurturing or conditioning (vs. selective breeding).

This analysis of how essentialist biases give rise to transformational thinking has been empirically validated by Shtulman and colleagues in a series of studies on evolutionary reasoning in high-school and college students (Shtulman, 2006; Shtulman & Calabi, 2012; Shtulman & Schulz, 2008). One of the most interesting findings is that students' essentialist misconceptions are highly intercorrelated. Students who hold essentialist misconceptions of some topics (say, inheritance) tend to hold essentialist misconceptions of other topics (say, speciation), and they appear to have retained those misconceptions across several years of biology instruction, even college-level instruction. From those data, Shtulman (2006) argued that students' evolutionary misconceptions represent more than just a handful of isolated, surface-level errors; they represent an *alternative theory* of evolution or a self-consistent network of causal-explanatory beliefs that defines both the phenomena to be explained and the mechanisms for explaining them.

In the present study, we extend this research by directly tracking changes in students' naïve theories of evolution in the context of formal biology instruction. Of interest was whether, and how, students' transformational misconceptions change in light of prolonged exposure to correct, variational principles of evolution. Other researchers who have measured evolution understanding before and after instruction have observed remarkably little change (e.g., Bishop & Anderson, 1990; Demastes, Settlage, & Good, 1995; Jensen & Finley, 1995). However, these researchers might have failed to observe changes that had actually occurred because their assessment tools were not sufficiently sensitive to the *transformational nature* of students' preinstructional misconceptions.

Here, we revisit the question of how evolution understanding changes with instruction using an assessment tool, developed by Shtulman (2006), that has proven effective at eliciting transformational interpretations of a wide range of evolutionary topics (variation, inheritance, adaptation, domestication, speciation, extinction) across a wide range of problem formats (closed-ended predictions, open-ended explanations, similarity judgments, figure completions, analogical mappings). We administered this assessment tool to students at the beginning and end of five different courses on evolutionary topics, variously taught between one and seven times across the period of study. Our intention was not to test the efficacy of one course relative to another but to track patterns of change in students' understanding on the whole. Two questions were of particular interest.

First, how does the consistency of students' reasoning influence their learning? Conceptual change researchers like Ohlsson (2009) and Chi (2008) have noted that a major obstacle to achieving conceptual change is *recoding* beliefs and experiences from the conceptual vocabulary of an earlier theory (e.g., transformationism) into the conceptual vocabulary of a new theory (e.g., variationism). Presumably, the more beliefs and experiences one has encoded in terms of the earlier theory, the more difficult conceptual change will be. On the other hand, the more beliefs and experiences that one has encoded in terms of the earlier theory, the more salient those beliefs are likely to be—in terms of their explanatory power, their inferential scope, or both—and thus the more clearly they might appear to conflict with the correct views on offer in the classroom. While the difficulty-of-recoding account leads to the prediction that students at the cusp of variational reasoning would benefit most from instruction, the saliency-of-belief account leads to the prediction that students with the strongest misconceptions would benefit most from instruction.

Second, how does learning affect the consistency of students' reasoning? As already noted, students have been found to exhibit highly consistent patterns of reasoning, adopting mainly transformational or mainly variational interpretations of the six phenomena listed in Table 1. Mixed patterns of reasoning were uncommon, accounting for less than a third of the 45 cases in observed by Shtulman (2006) and less than a quarter of the 45 cases observed by Shtulman and Calabi (2012). The relative scarcity of mixed patterns of reasoning suggests that students who do make the transition from transformational views of evolution to variational views tend not to linger at intermediate states but instead adopt variational interpretations of a whole host of interrelated phenomena. On the other hand, the scarcity

of mixed patterns of reasoning in previous studies might be a sampling artifact. Students who held mixed patterns of reasoning were, perhaps, less likely to complete a study on evolutionary reasoning than those who held more coherent views. We hoped to distinguish between these two possibilities by observing changes in students' evolutionary reasoning more directly.

Our approach to answering our two primary questions was to identify participants who made significant conceptual progress over the course of any semester (*learners*) and compare their preinstructional and postinstructional understanding of evolution to those who did *not* make significant conceptual progress during the same period (*nonlearners*). Specifically, we sought to determine whether learners' conceptions of evolution were significantly more coherent than nonlearners' conceptions (a) before instruction, which speaks to the question of how conceptual coherence affects learning, and (b) following instruction, which speaks to the question of how learning affects conceptual coherence. Moreover, both sets of findings speak to the broader question of how conceptual representation relates to conceptual change.

Method

Participants

The participants were 291 college undergraduates at the University of Massachusetts–Boston. The demographics of the sample were similar to the demographics for the institution as a whole: 60% women, 61% White, 52% first-generation college students, and 50% over the age of 22. Participants were enrolled in one of five courses: Evolutionary Biology (Bio 102, $n = 84$), Insect Life (Bio 338, $n = 17$), Animal Behavior (Bio 348, $n = 40$), Biological Foundations of Psychology (Psyc 105, $n = 99$), and Evolution and Behavior (Psyc 467, $n = 51$). The sampling of courses was intentionally broad because our objective was to explore the effects of instruction in general and not the effects of one particular curriculum. Consistent with this objective, participants came from both upper-level courses (Bio 338, Bio 348, Psyc 467) and lower-level courses (Bio 102, Psyc 105), and from both biology courses (Bio 102, Bio 338, Bio 348) and nonbiology courses (Psyc 105, Psyc 467). All courses, however, included evolution by natural selection as a core unifying principle of their curriculum, and none proved to be an outlier in any of the analyses reported here.

Materials

Participants completed a 30-item assessment of their understanding of evolution both before and after instruction. The assessment was

administered as an online questionnaire that participants could complete from any computer at any time (within a 10-day window around the first and last days of class). It consisted of six sections—variation, inheritance, adaptation, domestication, speciation, and extinction—with five items per section. Each item was designed to elicit either a correct, variational interpretation of the phenomenon at hand or an incorrect, transformational one. Responses indicative of variational reasoning were assigned a score of +1; responses indicative of transformational reasoning were assigned a score of −1; and responses potentially indicative of either form of reasoning (i.e., vague or ambiguous responses) were assigned a score of 0. The entire assessment and coding scheme can be found in the Shtulman (2006) appendix. Here, we provide a sample item from each section to illustrate both the nature of the assessment and the nature of the misconceptions it was designed to elicit. All sample responses were provided by actual participants in the study.

Variation. The critical difference between variationists and transformationists on the topic of variation is that only variationists view individual differences among members of the same species as relevant to evolution. Transformationists, on the other hand, view such differences as nonadaptive, or even maladaptive, deviations from the species' essential nature. One question designed to elicit this difference in interpretation was the following:

> During the 19th century, England underwent an Industrial Revolution that resulted in the unfortunate side effect of covering the English countryside in soot and ash. During this same period, the members of England's native moth species *Biston betularia* became, on average, darker in color. Assuming that darker coloration was adaptive, how might a change in the moths' environment have brought about a change in the moths' color?

Participants who referenced individual differences within the moth population that might have acted as fodder for natural selection (e.g., "moths who did not blend in got eaten and did not reproduce" or "the darker moths are less likely to be seen and eaten and so are able to reproduce more often") were coded as providing a variational response (+1 point). Participants who referenced only the needs of the species as a whole (e.g., "they needed a darker color to camouflage themselves" or "they needed to be able to blend into their environment") were coded as providing a transformational response (−1 point).

Inheritance. Because transformationists believe that species change if they need to change, they also believe that organisms will inherit the traits they need to inherit. Variationists, on the other hand, see heritability

as a strictly genetic matter. This difference was elicited with a series of questions, including the following, about the heritability of different traits:

> Suppose that a pair of woodpeckers migrates to a different island with fewer trees and more wind. As a consequence of flying in a windier environment, both woodpeckers develop stronger wing muscles. Compared to the offspring of the woodpeckers on the original island, the offspring of these two woodpeckers should have (a) stronger wing muscles; (b) weaker wing muscles; (c) either stronger wing muscles or weaker wing muscles; neither feature is more likely. Please explain your answer.

Participants who selected answer *c* and justified their selection by appealing to the distinction between genetic and acquired traits (e.g., "the stronger wing muscles are not genetic traits but rather developed throughout life in response to the environment" or "the physical musculature of the parents has no effect on the genetics of the children") were coded as providing a variational response (+1 point). Participants who selected answer *a* and justified their response by appealing to the utility of stronger wings (e.g., "the offspring will be raised in a windier environment and therefore will have to have stronger wing muscles" or "so they can survive in a windier climate") were coded as providing a transformational response (−1 point).

Adaptation. On the topic of adaptation, transformationists differ from variationists in that they are prone to conflate changes occurring at the level of the individual (ontogenetic changes) with changes occurring at the level of the species (evolutionary changes). We elicited this error with questions that assessed participants' ability to distinguish selection-based adaptation from other forms of adaptation, like the following:

> A youth basketball team scores more points per game this season than they did the previous season. Which explanation for this change is most analogous to Darwin's explanation for the adaptation of species? (a) Each returning team member grew taller over the summer. (b) Any athlete who participates in a sport for more than one season will improve at that sport. (c) More people tried out for the same number of spots this year. (d) On average, each team member practiced harder this season.

Explanation *c* was the only explanation that evoked a type of selection pressure and was thus the only explanation scored as variational (+1 point). All other explanations evoked organism-level changes in physiology or behavior and were thus scored as transformational (−1 point).

Domestication. Similar to the concept of adaptation, the concept of domestication is viewed by transformationists as a process that operates over individuals rather than groups or populations, rendering it responsive to factors other than selective breeding (e.g., weather, climate, food). We elicited this intuition with a series of questions about the domestication of corn from a wild grass known as teosinte. In one question, participants were asked whether they thought corn could be changed back into a plant like teosinte and why or why not? Participants who claimed it was possible through selective breeding (e.g., "by artificially selecting for the desired traits we could, over time, create a plant similar to teosinte" or "you just have to keep selecting for traits that are similar to the teosinte traits") were coded as providing a variational response (+1 point). Participants who claimed it was not possible because evolution is not reversible or manipulable (e.g., "evolution only happens in one direction" or "even man-made evolution cannot go backwards once it's proven beneficial") were coded as providing a transformational response (−1 point).

Speciation. Variationists view speciation as a consequence of prolonged reproductive isolation such that two groups of organisms that had once been able to interbreed are no longer able to do so. Transformationists, on the hand, do not have recourse to such an explanation because reproductive isolation would not change the fact that all members of the species share one, and only one, essence. Branching evolution (*cladogenesis*) is thus not a sensible outcome on a transformational view of evolution, leaving only the possibility of linear evolution (*anagenesis*). Applied to primate evolution, the difference between cladogenic and anagenic views of speciation is the difference between viewing chimpanzees as a *sister* species to humans and viewing them as a *parent* species to humans (see Catley, Novick, & Shade, 2010). One question used to elicit this difference in interpretation was

> As chimpanzees continue to evolve, they will become (a) more similar to modern-day humans; (b) less similar to modern-day humans; (c) either more or less similar to modern-day humans; neither outcome is more likely. Please explain your answer.

Participants who selected answer *b* or *c* and justified their selection by appealing to differences in selection pressures (e.g., "humans and chimps live in completely different environments and therefore have different selection factors acting on them" or "depends entirely on the constraints of the chimpanzees environment") were scored as providing a variational response (+1 point). Participants who selected answer *a* and justified their

selection by appealing to the linear or goal-directed nature of evolution (e.g., "if humans evolved from chimps then current chimps will continue to evolve to be humans" or "it can only go from human on; there's no turning back") were scored as providing a transformational response (−1 point).

Extinction. Extinction, for a variationist, is just selection writ large, but it cannot be such for a transformationist because transformationists attribute evolutionary change to a process of directed inheritance, not selection. Whatever that process is, it ensures that a species will adapt to is environment and should also ensure that the species will not go extinct. Transformationists thus greatly underestimate the frequency of extinction over time and the ubiquity of extinction across species (e.g., see Poling & Evans, 2004). We elicited this misperception with questions like the following:

> Compared to the number of living bacteria species, the number of extinct bacteria species is (a) greater; (b) smaller; (c) either greater or smaller; neither situation is more likely. Please explain your answer.

Participants who selected answer *a* and justified their selection by appealing to the ubiquity or inevitability of extinction (e.g., "since Earth was formed there have been hundreds of million of different strains of bacteria that have lived and died, so the amount of living bacteria today pails in comparison to once living bacteria" or "if you compare one year vs. the billions of years before that, it's only logical that overall there was a lot more bacteria that aren't alive anymore") were coded as providing a variational response (+1 point). Participants who selected answer *b* or *c* and justified their selection by appealing to the resilience or adaptability of living things (e.g., "more likely that most bacteria evolved rather than become extinct" or "bacteria does not become extinct; it just transforms into a different type of bacteria") were coded as providing a transformational response (−1 point).

Coding

Of the 30 items, 16 were closed-ended in nature and did not require any content analysis. The remaining 14 included an open-ended, verbal component that did require such analysis. This task was accomplished by two independent coders who read each verbal response and assigned it a score of either a 1, 0, or −1. The 8,148 verbal responses (14 responses × 291 participants × 2 assessments) were assigned numeric codes and then randomized so coders were blind both to the identity of the participant and the time of the assessment (pretest vs. posttest). Overall agreement between coders was 83% (Cohen's kappa = .75), and the vast majority of

disagreements involved responses that were assigned 0 by one coder and either 1 or −1 by the other. In fact, in only 129 (1.6%) of the 8,148 cases did the two coders assign contradictory codes (1 vs. −1). Disagreements were resolved by a third coder.

Results

Overall Patterns of Change

At pretest, participants' overall assessment scores ranged from −26 to +27 and averaged −8.4 ($SD = 11.3$). At posttest, they ranged from −25 to +28 and averaged −5.6 ($SD = 11.6$). This difference was statistically significant, $t(290) = 6.22$, $p < .001$, paired samples. To analyze whether participants' assessment scores varied by course, we submitted them to a $2 \times 2 \times 2$ analysis of variance (ANOVA), treating assessment period (pretest vs. posttest) as a within-participants factor and course department (psychology vs. biology) and course level (upper level vs. lower level) as between-participants factors. Not surprisingly, participants from upper-level courses outscored those from lower-level courses ($M = −1.9$, $SD = 12.4$, vs. $M = −10.0$, $SD = 8.4$), and participants from biology courses outscored those from psychology courses ($M = −5.3$, $SD = 11.4$, vs. $M = −8.6$, $SD = 10.0$). Both effects were significant: $F(1, 287) = 42.90$, $p < .001$, and $F(1, 287) = 8.29$, $p < .01$, respectively. There were no significant interactions, however, among assessment period, department of instruction, or level of instruction, as the *pre-post gains* for all five courses were roughly equivalent. In other words, participants from upper-level courses increased their score by approximately the same amount as those from lower-level courses ($M = 3.5$, $SD = 7.7$, vs. $M = 2.4$, $SD = 7.6$), and participants from biology courses increased their score by approximately the same amount as those from psychology courses ($M = 2.0$, $SD = 7.3$, vs. $M = 3.5$, $SD = 7.8$). Subsequent analyses were therefore collapsed across the five courses.

Differentiation of Learners From Nonlearners

To distinguish participants who demonstrated a statistically reliable increase in conceptual understanding over the course of the semester from those who did not, we compared each participant's response profile at pretest to his/her response profile at posttest by using a paired-samples *t* test (one per participant) and identified 58 participants who increased their score by a significant amount: $t(29) > 1.70$, $p < .05$, one-tailed. We used one-tailed tests, rather than two-tailed tests, because our prediction was

that instruction would increase understanding, not decrease it. This prediction was confirmed by the finding that participants who gained points from pretest to posttest were nearly twice as frequent as those who lost points, and participants who gained 10 or more points were nearly 7 times as frequent as those who lost 10 or more points. Our use of a one-tailed criterion was thus empirically justified as well.

Participants who did not meet the statistical criterion to be classified as *learners* were classified as *nonlearners* (n = 224), with the exception of nine participants who began the semester with a score of 22 or higher and were thus at ceiling on the instrument at hand. Those nine participants reported having taken an average of 3.2 biology courses (SD = 1.1) prior to their current course, which was significantly more than those taken by either the learners or the nonlearners, $t(65)$ = 3.11, p < .01, and $t(231)$ = 3.21, p < .01, respectively. Learners and nonlearners, on the other hand, reported having taken a comparable number of courses (M = 2.0, SD = 1.1, vs. M = 1.9, SD = 1.2). The nonlearners, as a group, earned a mean pretest score of −8.7 (SD = 9.4) and a mean posttest score of −8.5 (SD = 9.6), and most of the individuals (80%) within that group earned negative scores on both assessments.

Preinstructional Differences Between Learners and Nonlearners

Three analyses were conducted to gauge whether learners and nonlearners started the semester with comparable degrees of understanding or whether one group held consistently stronger transformational misconceptions: (a) an analysis of the overall frequency of transformational misconceptions, (b) an analysis of the consistency of transformational responding per section, and (c) an analysis of the number of sections for which the predominant response was transformational in nature.

Frequency of misconceptions. The mean number of variational and transformational responses provided at pretest can be seen in the left-hand panel of Figure 1. While there was no significant difference between learners and nonlearners in the number of variational responses provided at pretest, there was a significant difference in the number of transformational responses, with learners providing approximately 15% more transformational responses than nonlearners, $t(280)$ = 3.14, p < .01.

Consistency of misconceptions per section. Learners' and nonlearners' pretest scores are compared by section (i.e., by subtest) in the top panel of Table 2. Scores for each section could range from −5 to +5, and all mean scores were significantly less than zero with the exception of those for the Variation and Domestication sections for the nonlearners. More importantly,
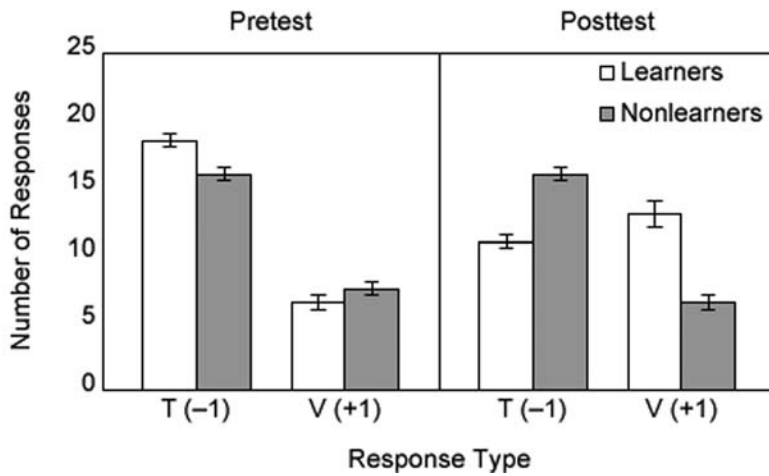
Figure 1.   Mean number of transformational responses (T) and variational responses (V) provided by learners and nonlearners at both pretest and posttest.

learners' scores were significant lower than nonlearners' scores for three of the six sections: variation, $t(280) = 2.50$, $p < .05$; domestication: $t(280) = 2.10$, $p < .05$; and extinction, $t(280) = 2.10$, $p < .05$. Thus, the overall difference in

Table 2.   Preinstructional and postinstructional comparison of learners' and nonlearners' mean section scores (range −5 to +5); standard errors are in parentheses

Test	Section	Learners	Nonlearners	Difference
Pretest	Variation	−1.0 (0.2)	0.2 (0.5)	−0.8*
	Inheritance	−1.8 (0.2)	−1.4 (0.4)	−0.4
	Adaptation	−3.5 (0.4)	−3.4 (0.2)	−0.1
	Domestication	−1.1 (0.4)	−0.3 (0.2)	−0.8*
	Speciation	−1.8 (0.3)	−1.7 (0.1)	−0.1
	Extinction	−2.8 (0.3)	−2.1 (0.1)	−0.7*
Posttest	Variation	1.6 (0.4)	0.6 (0.2)	1.0*
	Inheritance	0.4 (0.4)	−1.4 (0.2)	1.8**
	Adaptation	−1.2 (0.6)	−3.4 (0.2)	2.2**
	Domestication	1.9 (0.3)	−0.5 (0.2)	2.4**
	Speciation	−0.1 (0.3)	−1.6 (0.2)	1.5**
	Extinction	−1.0 (0.3)	−2.3 (0.1)	1.3**

*$p < .05$. **$p < .01$.

transformational responses documented in Figure 1 was not limited to a single section of the assessment but was more general in nature.

Consistency of misconceptions across sections. The previous analyses indicate that learners as a group held more preinstructional misconceptions than did nonlearners as a group. But how consistent was the reasoning of individual participants within those groups? To address this question, we tabulated the number of sections (out of 6) on which each participant provided more transformational responses than variational responses and compared those scores between groups. A score of 6 on this measure would indicate consistently transformational reasoning across sections, a score of 0 would indicate consistently variational reasoning across sections, and a score of 3 would indicate an even mixture of variational and transformational reasoning across sections (i.e., inconsistent reasoning).

On average, learners provided more transformational responses than variational responses on 4.5 sections ($SD = 1.5$), whereas nonlearners did so on 4.0 sections ($SD = 1.5$)—a statistically significant difference: $t(280) = 2.00$, $p < .05$. Both means, however, were significantly greater than 3.0: learners, $t(57) = 7.52$, $p < .001$; nonlearners, $t(223) = 10.33$, $p < .001$. This finding indicates that *all participants'* reasoning was consistently more transformational than predicted by chance or, more aptly, by a "knowledge-in-pieces" view of naïve conceptions (see diSessa, 1993). This finding also replicates previous findings regarding the consistency of students' evolutionary misconceptions (Shtulman, 2006; Shtulman & Calabi, 2012), albeit using a different measure of consistency than that used in previous studies (intercorrelations among section scores).

Postinstructional Differences Between Learners and Nonlearners

The same analyses conducted on learners' and nonlearners' pretest scores, to assess the degree to which preinstructional coherence was associated with learning, were conducted on their posttest scores, to assess the degree to which postinstructional coherence was associated with learning. Note that all such analyses are statistically independent of the classification of participants as learners and nonlearners, as this classification was done using a *within*-participants criterion, not a *between*-participants criterion. In other words, nothing in the statistical definition of a "learner" constrained the frequency or consistency of a learner's postinstructional misconceptions relative to those of a nonlearner.

Frequency of misconceptions. At posttest, learners provided significantly fewer transformational responses and significantly more variational responses than nonlearners: $t(280) = 6.53$, $p < .001$, and $t(280) = 6.41$,

p < .001, respectively. More interestingly, learners went from providing significantly more transformational than variational responses at pretest to providing approximately equal numbers of transformational and variational responses (see the right-hand panel of Figure 1). In other words, learners went from exhibiting consistently transformational reasoning at pretest—even more consistent than their nonlearning peers—to exhibiting a mixed, or inconsistent, pattern of reasoning at posttest.

This finding was further explored with an analysis of how participants increased their score on a question-by-question basis. There were, after all, three ways to increase one's score from pretest to posttest: (a) replacing a transformational response at pretest (scored −1) with an ambiguous response at posttest (scored 0), (b) replacing an ambiguous response at pretest (scored 0) with a variational response at posttest (scored +1), or (c) replacing a transformational response at pretest (scored −1) with a variational response at posttest (scored +1). Not surprisingly, learners were significantly more likely than nonlearners to exhibit each type of score increase. However, when we controlled for the overall differences in score increases by dividing the frequency of each type of increase by the total number of increases per participant, a different picture emerged (see Figure 2). Whereas nonlearners were more likely than learners to replace transformational responses with *ambiguous* responses (*M* = 0.59, *SD* = 0.25, vs. *M* = 0.48, *SD* = 0.22), learners were significantly more likely than nonlearners to replace transformational responses with *variational* responses (*M* = 0.32, *SD* = 0.19, vs. *M* = 0.22, *SD* = 0.25). Both differences were significant: *t*(280) = 1.97, *p* < .05, and *t*(280) = 2.57, *p* < .05, respectively. Thus, learners exhibited greater conceptual gains than nonlearners even when controlling for the total number of score increases per participant.

These findings help to explain why learners began the semester with significantly *lower* scores than nonlearners but ended the semester with significantly *higher* scores. When learners and nonlearners made conceptual progress, they did so differently. Nonlearners were more likely to move from a transformational understanding of a given topic to an ambiguous understanding, but learners were more likely to move all the way to a variational understanding. Learners thus made conceptual progress on both ends of the continuum, abandoning transformational views of evolutionary phenomena while also adopting variational views of those phenomena (vs. adopting no view at all). While the goal of sound instruction is typically to accomplish both tasks, it is important to recognize that they are logically (and here, empirically) distinct.

Consistency of misconceptions per section. Not surprisingly, learners exhibited fewer misconceptions than nonlearners following instruction.
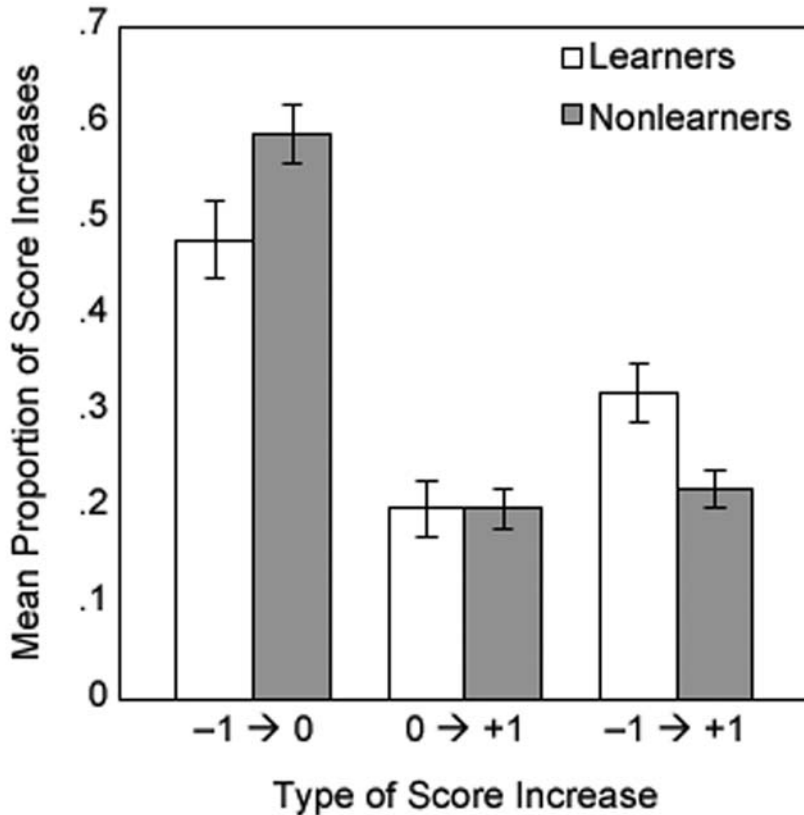
Figure 2. Mean proportion of score increases for learners and nonlearners that consisted of replacing a transformational response at pretest with an ambiguous response at posttest ($-1 \rightarrow 0$), replacing an ambiguous response at pretest with a variational response at posttest ($0 \rightarrow +1$), or replacing a transformational response at pretest with a variational response at posttest ($-1 \rightarrow +1$).

What was surprising, however, is that they did so on all six sections of the assessment (see the bottom panel of Table 2). This group-level pattern was observed at the individual level, as well; 44 (76%) of the 58 learners increased their score on at least four different sections of the assessment, and there were was no relationship between the number of sections on which a learner increased his/her score and the particular course from which that learner came. The effect of instruction was thus quite pervasive, at least for those who actually benefited from instruction.
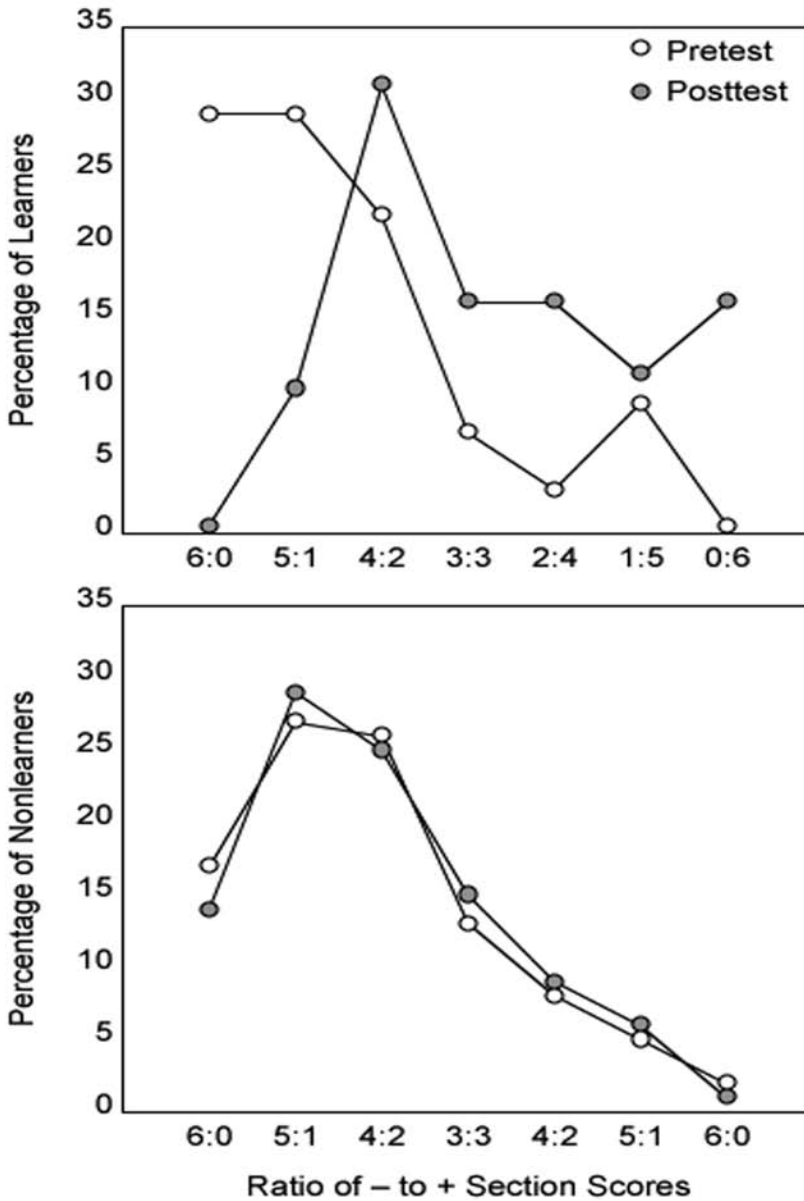
Figure 3. Percentage of learners (top panel) and nonlearners (bottom panel) who earned each of seven possible combinations of negative section scores (indicative of transformational reasoning) and positive section scores (indicative of variational reasoning).

Consistency of misconceptions across sections. Figure 1 shows that learners, as a group, ended the semester with a seemingly mixed view of evolution. This finding was confirmed at the level of the individual with an analysis of the ratio of negatively scored sections to positively scored sections for each individual learner (see Figure 3). At posttest, learners earned net negative scores on a mean of 2.7 sections ($SD = 1.7$), which was significantly less than the corresponding pretest mean of 4.5, $t(57) = 11.56$, $p < .001$. It was not, however, significantly less than the chance value of 3.0, $t(57) = 1.60$, *ns*. This pattern of results, which can be seen in the top panel of Figure 3, indicates that learners went from reasoning in a consistently transformational manner at pretest to reasoning in an inconsistent manner at posttest. Nonlearners, on the other hand, continued to reason in a consistently transformational manner, earning negative scores on an average of 4.0 sections (similar to pretest). A direct comparison of learners' and nonlearners' posttest scores on this measure revealed that, while learners reasoned more consistently (albeit more incorrectly) than nonlearners at pretest, nonlearners reasoned more consistently (albeit more incorrectly) at posttest: $t(280) = 5.98$, $p < .001$.

Discussion

Evolution by natural selection is a theory that, despite its seeming simplicity and vast utility, has proven extremely difficult to learn and to teach (Bishop & Anderson, 1990; Brumby, 1979; Jensen & Finley, 1995). We explored some of the reasons for this difficulty by tracking the conceptual understanding of 291 students across a semester of instruction on evolutionary topics, assessing the degree to which that understanding conformed to correct, variational principles (consistent with post-Darwinian views of evolution) or incorrect, transformational principles (consistent with pre-Darwinian views of evolution). Two main findings emerged.

First, participants who demonstrated robust learning over the semester held significantly more preinstructional misconceptions than did their non-learning peers, as measured both by their section scores and by the number of sections for which they earned a negative score. These findings cast doubt on the commonsense assumption that the more misconceptions one holds, the more difficult learning will be. Preinstructional misconceptions actually appeared to *facilitate* learning in the present study. Not only did learners exhibit more preinstructional misconceptions than did nonlearners, but the frequency of such misconceptions was a significant predictor of pre-post gains for the sample as a whole, $r(291) = .29$, $p < .001$. One reason preinstructional misconceptions might facilitate learning is that the contrast

between a student's understanding of evolution and the teacher's under-standing might become more obvious—to the student, to the teacher, or to both—as the misconceptions are encountered more often. Another reason is that the more misconceptions a student holds, the more likely he or she will discover the inferential limitations of the framework that licenses those misconceptions (in this case, transformationism). Either way, students would likely benefit from being prompted to articulate and elaborate their misconceptions, as this process could help highlight the manner and degree to which their intuitive understanding of the relevant phenomena deviates from a scientific understanding (see also Clement, 1993; Smith & Unger, 1997; Wiser & Amin, 2001).

Second, when learning occurred in the present study, it occurred incre-mentally for all forms of reasoning covered by the assessment. This finding is empirically consistent with the results of previous research using the same assessment tool (Shtulman & Calabi, 2012) and theoretically consistent with the *theory-theory* view of naïve conceptions, on which such concep-tions are organized in interrelated networks rather than isolated clusters (for a similar pattern of findings, see Au et al., 2008). Seemingly inconsistent with the theory-theory view, however, was the finding that learning resulted in *mixed* patterns of reasoning rather than *consistently accurate* reasoning. We label this finding as seemingly inconsistent because the theory-theory view, as described by most proponents of this view (Carey, 2009; Chi, 2008; Smith, 2008; Vosniadou, 1994), makes no stipulations about the nature of learning, only the start point and end point of learning—namely, that knowledge is organized in a theory-like manner at both points. While it is theoretically possible that learners could jump from one coherent theory to another, this pattern of learning is not typically observed in either sci-ence education (e.g., Schneider & Hardy, in press; Smith & Unger, 1997) or cognitive development (e.g., Siegal, Butterworth, & Newcombe, 2004; Slaughter & Lyons, 2003). Rather, theory change appears to be incremental, with learners adopting partial understandings or mixed patterns of reason-ing before adopting conceptions that are fully incommensurate with their initial conceptions. The learners in the present study clearly conformed to this pattern, showing evidence of *conceptual progress* but not full-fledged *conceptual change.*

In short, conceptual coherence appeared to increase learning, but learning appeared to decrease conceptual coherence (at least within the period of observation). Taken together, these findings shed light on an issue of importance not only to evolution education but also to sci-ence education and conceptual development more generally: the role of conceptual coherence in conceptual change (Murphy & Medin, 1985;

Thagard, 1989). As already noted, students' understanding of evolution is constrained by early emerging, essentialist biases in the interpretation of biological phenomena. Although those constraints appear to interfere with learning a correct, scientific view of evolution, the constraints themselves do not guarantee the generation of a *coherent alternative.* Consistent with this idea, students' preinstructional views of evolution vary widely in their coherence. And while much attention has been paid to the question of how coherent students' preinstructional views actually are (e.g., Gupta, Hammer, & Redish, 2010; Slotta, 2011), little attention has been paid to the question of how *differences* in coherence affect a student's receptivity to instruction and subsequent likelihood of undergoing conceptual change. Here, we found that coherence might actually facilitate learning, with participants who held more coherent forms of transformationism exhibiting greater learning than those who held less coherent forms. We also found that the learning itself tended to proceed in a coherent manner, with learners making conceptual progress on each and every topic covered by the assessment.

Our findings also highlight an unexplored tension in the notion of coherence itself—namely, a tension between coherence of reasoning and coherence of learning. *Coherent* learning, as indexed by pre-post gains on all sections of the assessment, actually appeared to result in *incoherent* reasoning, as indexed by equivalent amounts of variational and transformational responses at posttest. A state of incoherent reasoning—or, less pejoratively, mixed reasoning—might actually be necessary before one can build coherence among the concepts and principles of the to-be-acquired theory. It should be noted that this pattern of change—that is, moving from a conceptually coherent, yet scientifically, inaccurate theory toward a less coherent, yet more accurate, theory—is not particularly consistent with a "knowledge-in-pieces" view of learning (diSessa, 1993; Wagner, 2006). On that view, preinstructional knowledge consists of globally incoherent fragments of phenomenological experience, and learning consists of uniting and systematizing those fragments. Our findings are at odds with both ideas: not only were learners' preinstructional conceptions significantly more consistent than predicted by chance, but they were also significantly more consistent than their *post*instructional conceptions. Of course, if our learners were observed over additional semesters of instruction, the inconsistency in their postinstructional reasoning would most likely resolve into consistent *variational* reasoning. This picture is more consistent with the knowledge-in-pieces view of learning, but it is an incomplete picture insofar as it neglects the true starting point of the conceptual trajectory.

While our findings are theoretically provocative, we must acknowledge that they are retrospective in nature and thus vulnerable to certain confounds. For instance, learners might have earned significantly lower pretest scores than nonlearners because participants with lower pretest scores had more room for improvement than those with higher pretest scores. Although this possibility cannot be ruled out by the data at hand, there are at least two reasons to doubt it. First, learners did not simply "catch up" to their peers at posttest; they exceeded their peers by a significant margin, surpassing the mean rather than regressing to it. Second, the change in learners' responses from pretest to posttest was not a *quantitative* change; it was a *qualitative* change, with learners exhibiting forms of reasoning at posttest that are fundamentally incompatible with those exhibited at pretest. Still, a prospective study following the progress of both strong and weak transformationists at many different intervals throughout instruction would provide more conclusive evidence that coherence facilitates learning.

Another major limitation of the study is that it tracked changes in learning without also tracking the actual source of that learning: the particular curricula to which participants were exposed. While we did not happen to observe any effect of course level (upper level vs. lower level) or course department (biology vs. psychology) on participants' degree of learning, we must acknowledge that participants in some courses might have required more effective instruction to show the same degree of learning as those in other courses. Still, the study's major findings—regarding the effects of conceptual coherence on learning and the effects of learning on conceptual coherence—held true across the diversity of courses from which we sampled our participants. In fact, the ratio of learners to nonlearners (1:4) was roughly the same in all five courses, with a chi-square analysis revealing no association between course of instruction and the presence of learners, $\chi^2(4, N = 291) = 2.78$, *ns.* Further research needs to be done to determine which types of instruction, if any, are more likely to yield learners than nonlearners.

Despite its limitations, the present study provides some of the first evidence as to how naïve theories of evolution change over the course of instruction and how differences in the coherence of those theories lead to differences in the nature of that change. While more research is needed to determine which types of instruction are most effective at facilitating conceptual change, the present findings suggest that, whatever that instruction might be, it will be of greatest benefit to those who enter the science classroom with the most essentialist views of evolution.

References

Astuti, R., Solomon, G. E. A., & Carey, S. (2004) Constraints on conceptual development: A case study of the acquisition of folkbiological and folksociological knowledge in Madagascar [Monograph]. *Monographs of the Society for Research in Child Development, 69,* 1–135.

Atran, S., Medin, D., Lynch, E., Vapnarsky, V., Ek, E. U., & Sousa, P. (2001). Folkbiology doesn't come from folkpsychology: Evidence from Yukatek Maya in cross-cultural perspective. *Journal of Cognition and Culture, 1,* 3–42.

Au, T. K. F., Chan, C. K. K., Chan, T. K., Cheung, M. W. L., Ho, J. Y. S., & Ip, G. W. M. (2008). Folkbiology meets microbiology: A study of conceptual and behavioral change. *Cognitive Psychology, 57,* 1–19.

Berti, A. E., Toneatti, L., & Rosati, V. (2010). Children's conceptions about the origin of species: A study of Italian children's conceptions with and without instruction. *Journal of the Learning Sciences, 19,* 506–538.

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

Brumby, M. N. (1984). Misconceptions about the concept of natural selection by medical biology students. *Science Education, 68,* 493–503.

Carey, S. (2009). *The origin of concepts.* New York: Oxford University Press.

Catley, K. F., Novick, L. R., & Shade, C. K. (2010). Interpreting evolutionary diagrams: When topology and process conflict. *Journal of Research in Science Teaching, 47,* 861–882.

Chi, M. T. H. (2008). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In S. Vosniadou (Ed.), *Handbook of research on conceptual change* (pp. 61–82). Hillsdale, NJ: Erlbaum.

Clement, J. (1993). Using bridging analogies and anchoring intuitions to deal with students' preconceptions in physics. *Journal of Research in Science Teaching, 30,* 1241–1257.

Crawford, B. A., Zembal-Saul, C., Munford, D., & Friedrichsen, P. (2005). Confronting prospective teachers' ideas of evolution and scientific inquiry using technology and inquiry-based tasks. *Journal of Research in Science Teaching, 42,* 613–637.

Demastes, S. S., Settlage, J., & Good, R. J. (1995). Students' conceptions of natural selection and its role in evolution: Cases of replication and comparison. *Journal of Research in Science Teaching, 32,* 535–550.

Deniz, H., Donnelly, L. A., & Yilmaz, I. (2008). Exploring the factors related to acceptance of evolutionary theory among Turkish preservice biology teachers. *Journal of Research in Science Teaching, 45,* 420–443.

Diesendruck, G., & Haber, L. (2009). God's categories: The effect of religiosity on children's teleological and essentialist beliefs about categories. *Cognition, 110,* 100–114.

diSessa, A. (1993). Toward an epistemology of physics. *Cognition and Instruction, 10,* 105–225.

Evans, E. M., Spiegel, A. N., Gram, W., Frazier, B. N., Tare, M., Thompson, S., & Diamond, J. (2010). A conceptual guide to natural history museum visitors' understanding of evolution. *Journal of Research in Science Teaching, 47,* 326–353.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought.* Oxford: Oxford University Press.

Gelman, S. A., & Coley, J. D. (1990). The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children. *Developmental Psychology, 26,* 796–804.

Gelman, S. A., & Markman, E. M. (1987). Young children's inductions from natural kinds: The role of categories over appearances. *Child Development, 58,* 1532–1541.

Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understanding of the nonobvious. *Cognition, 38,* 213–244.

Gould, S. J. (1996). *Full house: The spread of excellence from Plato to Darwin.* New York: Three Rivers Press.

Gregory, T. R., & Ellis, C. A. J. (2009). Conceptions of evolution among science graduate students. *BioScience, 59,* 792–799.

Gupta, A., Hammer, D., & Redish, E. F. (2010). The case for dynamic models of learners' ontologies in physics. *Journal of the Learning Sciences, 19,* 285–321.

Hull, D. (1965). The effect of essentialism on taxonomy: 2000 years of stasis. *British Journal for the Philosophy of Science, 15–16,* 314–326.

Jensen, M. S., & Finley, F. N. (1995). Teaching evolution using historical arguments in a conceptual change strategy. *Science Education, 79,* 147–166.

Johnson, S. C., & Solomon, G. E. A. (1997). Why dogs have puppies and cats have kittens: The role of birth in young children's understanding of biological origins. *Child Development, 68,* 404–419.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development.* Cambridge, MA: MIT Press.

Lawson, A. E., & Thompson, L. D. (1988). Formal reasoning ability and misconceptions concerning genetics and natural selection. *Journal of Research in Science Teaching, 25,* 733–746.

Mayr, E. (1982). *The growth of biological thought: Diversity, evolution, and inheritance.* Cambridge, MA: Harvard University Press.

Mayr, E. (2001). *What evolution is.* New York: Basic Books.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289–316.

Nadelson, L. S., & Sinatra, G. M. (2009). Educational professionals' knowledge and acceptance of evolution. *Evolutionary Psychology, 7,* 490–516.

Nehm, R. H., & L. Reilly (2007). Biology majors' knowledge and misconceptions of natural selection. *BioScience, 57,* 263–272.

Nettle, D. (2010). Understanding evolution may be improved by thinking about people. *Evolutionary Psychology, 8,* 205–228.

Newport, F. (2010, December 17). Four in 10 Americans believe in strict creationism. *Gallup.* Retrieved December 20, 2010, from http://www.gallup.com/poll/145286/four-americans-believe-strict-creationism.aspx

Ohlsson, S. (2009). Resubsumption: A possible mechanism for conceptual change and belief revision. *Educational Psychologist, 44,* 20–40.

Poling, D. A., & Evans, E. M. (2004). Are dinosaurs the rule or the exception? Developing concepts of death and extinction. *Cognitive Development, 19,* 363–383.

Rosengren, K. S., Gelman, S. A., Kalish, C., & McCormick, M. (1991). As time goes by: Children's early understanding of biological growth. *Child Development, 62,* 1302–1320.

Schneider, M., & Hardy, I. (in press). Profiles of inconsistent knowledge in children's pathways of conceptual change. *Developmental Psychology.*

Schwartz, S. P. (1979). Natural kind terms. *Cognition, 7,* 301–315.

Settlage, J. (1994). Conceptions of natural selection: A snapshot of the sense-making process. *Journal of Research in Science Teaching, 31,* 449–457.

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

Shtulman, A., & Calabi, P. (2012). Cognitive constraints on the understanding and acceptance of evolution. In K. S. Rosengren, S. Brem, E. M. Evans, & G. Sinatra (Eds.), *Evolution challenges: Integrating research and practice in teaching and learning about evolution* (pp. 47–65). Cambridge: Oxford University Press.

Shtulman, A., & Schulz, L. (2008). The relationship between essentialist beliefs and evolutionary reasoning. *Cognitive Science, 32,* 1049–1062.

Siegal, M., Butterworth, G., & Newcombe, P. A. (2004). Culture and children's cosmology. *Developmental Science, 7,* 308–324.

Slaughter, V., & Lyons, M. (2003). Learning about life and death in early childhood. *Cognitive Psychology, 43,* 1–30.

Slotta, J. D. (2011). In defense of Chi's ontological incompatibility hypothesis. *Journal of the Learning Sciences, 20,* 151–162.

Smith, C. (2008). Conceptual change. In E. M. Anderman & L. Anderman (Eds.), *Psychology of classroom learning: An encyclopedia* (Vol. 1, pp. 250–256). Farmington Hills, MI: Thompson Gale.

Smith, C., & Unger, C. (1997). What's in dots-per-box? Conceptual bootstrapping with stripped down visual analogues. *Journal of the Learning Sciences, 6,* 143–181.

Sousa, P., Atran, S., & Medin, D. (2002). Essentialism and folkbiology: Evidence from Brazil. *Journal of Cognition and Culture, 2,* 195–223.

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences, 12,* 435–467.

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction, 4,* 45–69.

Wagner, J. F. (2006). Transfer in pieces. *Cognition and Instruction, 24,* 1–71.

Waxman, S., Medin, D., & Ross, N. (2007). Folkbiological reasoning from a cross-cultural developmental perspective: Early essentialist notions are shaped by cultural beliefs. *Developmental Psychology, 43,* 294–308.

Wiser, M., & Amin, T. (2001). Is heat hot? Inducing conceptual change by integrating everyday and scientific perspectives on thermal phenomena. *Learning and Instruction, 11,* 331–355.