Secondary School Students’ Reasoning About Evolution

Cheryl To,1,2 Harriet R. Tenenbaum,2 and Henriette Hogh3

1London South Bank University, SE1 0AA London, United Kingdom
2School of Psychology, University of Surrey, Guildford, GU2 7XH Surrey, United Kingdom
3Institute of Sport, University of Chichester, Chichester, United Kingdom

Received 1 September 2015; Accepted 6 August 2016

Abstract: This study examined age differences in young people’s understanding of evolution theory in secondary school. A second aim of this study was to propose a new coding scheme that more accurately described students’ conceptual understanding about evolutionary theory. We argue that coding schemes adopted in previous research may have overestimated students’ grasp of evolutionary concepts. A total of 106 students aged 12, 14, and 16 took part in individual interviews investigating their understanding of evolution. Using the new coding scheme, we found that while 16-year olds were more likely than 12-year olds to endorse scientific concepts when answering a question about finches, their understanding of natural selection, however, did not generalize to the other four questions. Furthermore, students began to incorporate relevant terminology (e.g., adapt, evolve, etc.) and structure their explanations using relevant language at around age 14. Students often used relevant terminology without having a more advanced understanding of evolutionary theory. Instead, they used the relevant terms in a colloquial rather than a scientific sense. Implications of the current findings for teaching and theory are discussed.

Students’ use of multiple epistemologies in scientific reasoning has been well established (Bang & Medin, 2010; Chinn & Samarapungavan, 2001; Duit & Treagust, 2003; Evans, Legare, & Rosengren, 2011; Gelman & Rhodes, 2012; King & Kitchener, 2004; Kuhn, 1993; Rosengren & Evans, 2012; Shtulman & Calabi, 2013). When explaining scientific phenomena, students often simultaneously evoke both intuitive and scientific ideas (Evans et al., 2011; Evans & Lane, 2011; Hammer & Elby, 2003; Harris & Koenig, 2006; Harrison & Treagust, 2001; Hofer & Pintrich, 1997; Legare, Evans, Rosengren, & Harris, 2012). For example, in physics, 8-year-old children dropped a ball early to reach a target (implicitly endorsing a parabolic function) even when endorsing an incorrect straight-down trajectory (Krist, 2000, study 2), which shows that they may endorse different epistemologies for the same object.
Similarly in biology, when asked to explain processes of species change, students often assume a need-based view where evolution is assumed to occur at the individual level rather than the population level (Kelemen, 2004, 2012; Poling & Evans, 2002; Southerland, Abrams, Cummins, & Anelmo, 2001), invoking both intuitive (need-based reasoning) and scientific reasoning (random mutation). However, because past research on students’ understanding about biological evolution has not allowed for students’ use of multiple epistemologies, it thus may have overestimated the coherence of their understanding of this theory. For example, in a study examining Spanish secondary school students’ understanding of biological evolution, Banet and Ayuso (2003) categorized students’ mental frameworks as either Lamarckian or Darwinian. Similarly in an intervention study where Dutch students were given guided instructions over 2 weeks (two lessons, 50 minutes each) to reinvent evolutionary theory, Geraedts and Boersma (2006) described students’ pre-instruction understanding as Lamarckian and post-instruction understanding as Darwinian. Other studies have likewise categorized students’ understanding of evolution as either consisting of naïve conceptions or scientific ones (Bishop & Anderson, 1990; Demastes, Settlage, & Good, 1995; Jensen & Finley, 1995, 1996). However, people’s reasoning about biological evolution is rarely based either on intuitive or scientific knowledge alone (Evans et al., 2010; Spiegel et al., 2012). In a seminal study investigating American adults’ reasoning about evolution prior to visiting a natural history museum, Evans et al. (2010) found that 72% of adult museum visitors tended to invoke scientific and intuitive explanations simultaneously for species change. A further 28% of participants endorsed a reasoning framework that also included creationist explanations in addition to scientific and intuitive explanations. In this study, we investigate English secondary students’ reasoning about biological evolution, with a focus on multiple epistemologies.

Theoretical Framework

Many people represent multiple and seemingly contradictory epistemologies in their understanding of evolution (Chinn & Brewer, 1993; Evans et al., 2011; Gelman & Legare, 2011). Evans et al. (2010) propose different models for how people may allow these seemingly contradictory beliefs to co-exist. For example, people may explain species change using different epistemologies, they may invoke them in different contexts, or the epistemologies may be fused in a non-systematic manner.

Naïve Theories. One epistemology frequently cited in students’ reasoning about biological evolution is the use of naïve or folk biological and psychological theories (Evans, 2000; Evans, Rosengren, Lane, & Price, 2012; Gelman & Kremer, 1991; Kelemen, 1999, 2004, 2012; Kelemen & DiYanni, 2005; Murphy & Medin, 1985; Poling & Evans, 2002; Rosengren & Evans, 2012). Naïve theories are general rules that children learn about the physical world through personal and direct experiences with the world. These biases tend to be fairly well ingrained in children’s problem solving repertoires such that children frequently draw on these naïve explanatory frameworks regardless of the framework’s appropriateness (Chinn & Brewer, 1993; Gelman & Legare, 2011). These naïve theories give rise to different types of reasoning, such as essentialism, which comes from a naïve biology perspective. Essentialism is the assumption that there are biologically determined essences that differentiate one species from another; this innate essence is assumed to be passed on through biological reproduction (Coley & Tanner, 2012; Gelman, 2004; Gelman & Rhodes, 2012). The concept of essentialism accepts that living organisms may change, but such changes are deemed superficial, and that the core essences of organisms remain unchanged as before (i.e., caterpillar → butterfly; Gelman & Rhodes, 2012). While essentialist beliefs are not entirely at odds with evolutionary theory, because lateral species change truly
cannot occur (i.e., a change from a cat to a dog), strong commitments to species stability without consideration for individual variability may hinder people’s understandings of the theory (Coley & Muratore, 2012; Gelman & Rhodes, 2012; Shtulman & Calabi, 2013).

Also contributing to students’ epistemological beliefs about biological evolution is the idea of teleology, which comes from intuitive psychology. This is the assumption that every part or property of a living thing has a specific purpose (Poling & Evans, 2002). Children between the ages of 6 and 7 years generously apply this type of reasoning to both animate and naturally occurring inanimate objects (Kelemen, 1999). Not until 10 years of age do children restrict their teleological thinking to animate objects (Poling & Evans, 2002). Kelemen (1999) argues that over attribution of teleological thinking is an extension of younger children’s intentional thinking. Attributing mental states to all objects (intentional reasoning), both animate and inanimate, is typical of (but not exclusive to) 6- to 7-year-old children (Poling & Evans, 2002). More recently, Kelemen and Rosset (2009) found that teleological thinking continues to remain in adults’ conceptual search space as default mechanisms, even among professional scientists (Kelemen, Rottman, & Seston, 2013).

National Science Curriculum in England. In addition to intuitive theories, people may also invoke theories dominant in their socio-cultural milieu, such as scientific theories or religious ideas. One of these cultural beliefs includes evolutionary theory (Evans et al., 2010). To understand the role of formal education in the development of English students’ understanding of evolutionary theory, the National Curriculum in England must be understood.

In the National Curriculum1 for Science in England, students are introduced to the topic of evolution between the ages of 11 and 16 years (Key Stages 3 and 4; Department for Education, 2013a, b). Between the ages of 11 and 14 years, students learn about inheritance, chromosomes, DNA, and genes. The specific content covered includes learning about heredity as a genetic process; that differences between and within species can be attributed to differences in genetic information; that variation between individuals within a species is either continuous or discontinuous; the role of variation in natural selection; organism-environment fit and extinction; and the need for biodiversity (Department for Education, 2013a). For example, a unit on evolution for students between the ages of 11 and 14 years introduces students to evolutionary theory by exploring the ideas proposed by Darwin and Lamarck about evolution. Students further explore the concept of natural selection before moving on to the concept of extinction (the dodo bird and dinosaurs are often used as examples of extinct species). Students learn that extinction is caused by changes in the environment (e.g., a new disease, a new predator, a change in the physical environment, climate change, etc.), leading to a particular species being less able to compete and reproduce successfully. Students also learn that endangered species (e.g., pandas, gorillas, etc.) are on the brink of extinction. Species may become extinct because of the critically low level of available habitats for these animals or because the population of a species has fallen below a critical level. Finally, the unit of evolution ends with having students learn about biodiversity, and the conservation efforts to protect endangered species.

In grades with students aged 14–16 years, the curriculum continues to build on the earlier content to teach students that genetic mutation causing variation occurs at the gene level; that sexual reproduction contributes to variation within a population; monohybrid inheritance occurs when there are dominant and recessive alleles; the evolution of new species ensues over time through natural selection; genetic variation and environmental factors contribute to evolution (e.g., bacterial resistance to antibiotics, human evolution). They also learn about adaptation, evidence for evolution from geology and other fields, common descent, the three-domain model based on DNA analysis, Darwin and Mendel, and vocabulary specific to genetics and evolutionary

Journal of Research in Science Teaching
theory (e.g., dominant allele, recessive allele, homozygous, heterozygous phenotype, genotype; Department for Education, 2013b). Thus, all English school-aged children are exposed to the cultural epistemology of evolutionary theory in some detail.

**Creationism.** Another cultural epistemology, albeit from a supernatural perspective, is creationism (Evans et al., 2010). Compared to US secondary school students, however, English secondary school students (Tenenbaum, To, Wormald, & Pegram, 2015) and British science museum visitors (Abraham-Silver & Kisiel, 2008) rarely invoke creationism. For this reason, we will not discuss this theory further. Such variation in invoking creationism demonstrates that the use of different epistemologies vary with culture (Evans et al., 2010).

**Role of Culture**

Indeed, ample evidence suggests that culture and socialization play an important role in the way people reason about biological entities (Atran, Estin, Coley, & Medin, 1997, Atran, Medin, Lynch, Vapnarsky, Ek’, & Sousa, 2001, Atran, Medin, & Ross, 2004; Bang & Medin, 2010; Sousa, Atran, & Medin, 2004; Ross, Medin, Coley, & Atran, 2002; Waxman, Medin, & Ross, 2007). In a study examining three distinct populations that varied depending on the amount of contact they had with the natural world, Ross et al. (2002) found that 6- to 10-year-old children from the three groups adopted very different ways of reasoning when thinking about biological entities. While anthropomorphic thinking was specific to urban children who had the least contact with natural world, ecological thinking (reasoning based on relationship between two and more unique entities) was most characteristic of Native American children, who had the most contact with natural world. Furthermore, while there is evidence of anthropomorphic thinking in children from rural majority cultures, this type of reasoning ceases to exist in their thinking by the time children are 10 years old.

Cultures need not be drastically different from one another for effects of culture and socialization to take hold. Kelemen (2003) examined British and American children’s use of teleological reasoning when reasoning about natural phenomena. Given that both the UK and the US are similar in terms of industrialization (Micklethwait & Wooldridge, 2005) and that teleological thinking is universal among children from western urbanized cultures, one would expect little difference in the way teleological thinking is endorsed. However, what Kelemen (2003) found was that while children from both countries endorsed teleological reasoning, they did so in a slightly different manner. Specifically, while American children tended to regard body parts as possessing both biological self-serving and artifact-like social functions, British children tended to favor self-serving survival-enhancing function for selected body parts (e.g., neck and feet). Kelemen (2003) argued that the different ways in which British and American children use teleological reasoning may be because British adults tended to be less open about publically endorsing religious explanations. Furthermore, the nuanced difference between the two cultures may mean that children are either more or less likely to be exposed to statements supporting intelligent design depending on the majority culture where they are raised (Kelemen, 2003). Because there are differences between British and American children’s foundational knowledge about biological entities, we expect that in extrapolating their early understandings about biological entities, British young people’s reasoning about evolution will also differ.

**Conceptual Conflict and Conceptual Change**

How students integrate new information learned in the science classroom into their pre-existing epistemologies depends on whether or not students detect the inconsistencies between those epistemologies and the one proposed by the scientific community (Chinn & Brewer, 1993; Posner, Strike, Hewson, & Gertzog, 1982). In instances where students do not detect a discrepancy
between their alternative understanding and the normative explanation, learners’ knowledge may likely comprise a collection of individual pieces of knowledge (see diSessa, Gillespie, & Esterly, 2004). However, if students detect the discrepancy between their personal understanding and the normative explanation, students can respond to the information by: (i) ignoring the contradictions and continuing to use pre-existing conceptions in evolutionary reasoning; (ii) maintaining pre-existing conceptions and using new and old conceptions in parallel, such as in co-existence models (Evans et al., 2010); or (iii) constructing a new conceptual framework incorporating both naïve and scientific ideas into a single conception (Kuhn, 1989; Nehm & Ha, 2011). However, because past research on participants younger than 18 years has not generally coded students’ response to allow for their use of multiple epistemologies, it is uncertain to what extent students adopt each strategy when reasoning about biological evolution. Thus, our study extends work on multiple epistemologies conducted by Evans et al.

The Current Study

The aim of the current study is to determine the ways in which young people in England reason about biological evolution. Though past research has indicated that there are commonalities in the way people reason about evolution that cut across cultures (Abraham-Silver & Kisiel, 2008), there are also differences in children’s reasoning about biology in these culture (Kelemen, 2003) that could lead to differences in the way they reason about biological evolution. Indeed, a recent study exploring English secondary school students’ reasoning about evolution before and after a visit to a natural history museum suggested that students very rarely invoked creationist explanations when reasoning about evolution (Tenenbaum et al., 2015). This finding is in contrast to research conducted in the American Midwest/Southern US which suggests that many adult museum visitors endorse creationist explanations when reasoning about evolution (Evans et al., 2010). Though the study conducted by Tenenbaum et al. (2015) is not a comparison study, it highlights the need for taking a new look at how young people in England reason about evolution today.

Unlike previous work that has focused on students’ endorsement of specific evolutionary concepts (Beardsley, 2004; Clough & Wood-Robinson, 1985; Deadman & Kelly, 1978), or students’ endorsement of either Lamarckian or Darwinian theory of evolution (Banet & Ayuso, 2003; Geraedts & Boersma, 2006), this study examines developmental trends in secondary school students’ reasoning about evolution, in particular, their use of target-dependent reasoning. Past studies have not coded students’ answers allowing for the use of multiple epistemologies and as a result, may have overestimated the coherence of young people’s understandings of the evolutionary theory. The present study remedies this by using a different type of coding scheme. In addition, past coding schemes, such as the one used by Evans et al. (2010), have coded the use of words like evolution and adaptation as evidence that people correctly understood evolutionary theory. However, one could use these words to denote a prior conception based on naïve concepts (Demastes, Good, & Peebles, 1996). For this reason, we used a stricter coding scheme where the mere mention of these terms was not considered evidence of evolutionary understanding. Furthermore, the majority of studies investigating students’ emerging understandings about evolution have been conducted in the US (Beardsley, 2004; Bishop & Anderson, 1990; Clough & Wood-Robinson, 1985; Demastes et al., 1995; Jensen & Finley, 1995, 1996; Kampourakis & Zogza, 2008; Nehm & Reilly, 2007; Nehm & Ridgway, 2011; Settlage, 1994; Shtulman, 2006). Our study extends the work of Evans et al. (2010) by specifically focusing on participants between the ages of 12 and 16, a time when based on the National Curriculum in England (Department for Education, 2013a, 2013b), students would need to negotiate their understandings of evolution. Understanding how students in England develop knowledge about evolution is necessary to for a more complete view.
Based on past research on secondary school students’ (Banet & Ayuso, 2003; Beardsley, 2004; etc.), and adult museum visitors’ reasoning about evolution (Diamond & Evans, 2007; Evans et al., 2010), it is expected that: (i) intuitive concepts will be endorsed by students at all age levels; (ii) older students will have access to more scientific concepts about evolution, and thus will incorporate these concepts more frequently into their explanations about species change than younger students; (iii) students will exhibit target-dependent reasoning in that they will alter the way they reason about evolution based on the entity discussed. Specifically, (iv) students will be more likely to reject evolution when discussing human evolution than other the evolution of other species. Finally, we explore differences between our coding scheme and a previous one based on Evans et al. (2010).

Method

Participants and School Characteristics

One-hundred and six students from four state comprehensive schools and one independent school in London and Surrey, UK were recruited. There were 39 students at age 12 (M = 12 years 4 months, SD = 2.7 months; 18 girls, 21 boys), 31 students at age 14 (M = 14 years 5 months, SD = 3.7 months; 17 girls, 14 boys), and 36 students at age 16 (M = 16 years 6 months, SD = 3.5 months; 17 girls, 19 boys). Since both science and religious education are part of the mandatory National Curriculum in England as set out by the Qualifications and Curriculum Authority (Qualifications and Curriculum Authority, 2004), we sampled students from these two subjects. No age group was selected from one school only. We interviewed ten 14-year olds and nineteen 16-year olds from School A, eleven 14-year olds and seventeen 16-year olds from School B, sixteen 12-year olds and eight 14-year olds from School C, ten 12-year olds and two 14-year olds from School D, and thirteen 12-year olds from School E. The percentage of students achieving five GCSEs2 grade A+ to C for each of the schools respectively are 56%, 65%, 53%, 83%, and 84%. The catchment areas of the schools also varied in socio-economic status. Thus, the academic performance of students in each school varied and we are confident that students from a wide variety of academic and socio-economic backgrounds were included in this study.

Information about students’ religious affiliation was not collected, because the schools in which we interviewed students did not want us to ask about religiosity. The schools considered this question to be a personal question.

Procedure and Protocol

Consent to approach students in school was obtained first from the principal, then parental opt-out letters were sent home requesting parents to return letters if they did not want students to participate. There were no cases of parental opt-out. Finally, student verbal assent was also required for participation. Each participant engaged in a one-to-one semi-structured interview that lasted approximately 15 minutes.

Interview Protocol. Upon meeting with the students, the researcher introduced herself, and gave a brief overview of the study. Each participant gave further verbal assent before commencing the interview. One student did not agree to take part in the study. Students who agreed to take part answered all questions asked by the researcher. Each participant was asked 15 questions by the researcher, of which five questions are addressed in this study (see Table 1). The remainder of the questions will be investigated in a separate report, and therefore will not be reported in this study. The five questions explored in this study investigate students’ reasoning about species origins (how did the Tasmanian tigers come to be in Tasmania?, why
Table 1

Examples of most common student responses

<table>
<thead>
<tr>
<th>Interview Question</th>
<th>Examples of Student Responses and Codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>The last Tasmanian tigers died in the 1930s, the species is now extinct. This was a marsupial animal, like Kangaroos are marsupials for example. This animal also looked like a wolf and had stripes on its back. Why do you think the wolf like marsupials could only be found in Tasmania?</td>
<td>“Maybe they have adapted to that place, so they can only really survive there . . . Um, they have kind of grown up living with different plants and stuff so they know where they have to be during the day, where they have to be during the night they eat, what they can not eat.” (Alex, age 14) CODE: novice naturalistic reasoning</td>
</tr>
<tr>
<td>In the next hundreds of years, because of global warming, the ice caps are going to melt. The Arctic will be much warmer than the seals are used to, what will happen to the seals do you think?</td>
<td>“Um, I think they will probably start to run out, they would have to swim for longer because there is no ice. I think they would either find somewhere else, but I think most of them, they will not survive.” (Kenny, age 14) CODE: mixed informed (extinction)/novice naturalistic reasoning</td>
</tr>
<tr>
<td>The Galapagos Islands are located off the coast of South America. On one of these islands, scientists have been studying one kind of finch. They measured the size of the finches’ beaks. On the first trip to the island, they found that the beak of this finch was on the small side. Then a severe drought occurred on the island and it wiped out most of the plants that make the small seeds that the finches feed on. The only seeds that were really common were the tough ones that require a large beak to open. Then the scientists came back a few years later and measured the beaks again. This time, they found that the beaks of this finch were on the large side. How can you explain that on the return trip to the island, larger beaks were found on more of the finches?</td>
<td>“Because they needed bigger beaks to eat the stuff. Maybe the smaller beaks could not eat it, maybe they were not strong enough. They were adapted to that kind of seed so maybe only the big ones, maybe they could adapt to them, so maybe only they could survive. Say like they were smaller and bigger ones. Maybe those smaller ones could evolve into the bigger ones eventually, as time goes on they could adapt to new stuff. Maybe they could adapt to eat the bigger seeds.” (Leslie, age 14) CODE: mixed informed/novice naturalistic reasoning</td>
</tr>
<tr>
<td>Scientists think the humans and chimps shared a common ancestor as recently as 5 million years ago. Describe how you think that both a chimp and a human could arise from the same kind of ancestor?</td>
<td>“Um, I think they are very alike, but I think they, does not like the apes, we were apes and I think we kind of separated. I do not know how, but we definitely separated and they kind of lived in a different environment to us. Separated from us.” (Victor, age 14) CODE: novice naturalistic reasoning</td>
</tr>
<tr>
<td>There are many types of algae in Yellowstone Lake; however, scientists have found a kind of algae in this lake that is not found anywhere else. These algae first appeared 14,000 years ago, at that time, the climate was warming. Describe how you think that this new kind of algae came to be in Yellowstone Lake?</td>
<td>“Bacteria could have evolved. I do not really know.” (Rebecca, age 14) CODE: novice naturalistic reasoning</td>
</tr>
</tbody>
</table>
was a specific type of algae only found in one place [Yellowstone Lake]?, how do you think the
chimps and humans arose from the same kind of ancestor?), and natural selection (why were
the finches beaks larger on scientists’ return trip to the island?). One further question asked
students to predict the consequences for biological entities if global warming continues (if the
sea temperature continues to rise, what will happen to the seals?). The questions about the
finches, algae, and chimps have been used in many past studies on evolution (Diamond &
Evans, 2007; Evans et al., 2010; Spiegel et al., 2012). We created the questions about the
seals and the Tasmanian tigers to gauge students’ understanding of species-environment fit. The
seals question also tapped into students’ knowledge about extinction. After the interview,
the researcher thanked the participants for their time and asked if they had any questions for the
researcher. None of the students had any further questions or comments following
the interview. Each participating school received either a £100 donation or a talk from the
researcher about studying psychology in higher education.

Interviews were transcribed and coded into three main reasoning patterns. The coding system
is described below.

Coding Scheme. The coding system was adapted from Evans et al. (2010) with a change
described in more detail at the beginning of the scoring section. The coding scheme was divided into
three categories: informed naturalistic reasoning (INR), novice naturalistic reasoning (NNR), and
denial of evolutionary reasoning (DR). These codes were not mutually exclusive. Students could use
more than one type of reasoning simultaneously (e.g., INR/NNR, NNR/DR, etc.), which allowed for a
total of eight different reasoning pattern combinations: INR, NNR, DR, INR/NNR, INR/DR, INR/
NNR/DR, NNR/DR, and “don’t know.” The code “don’t know” was used only when participants
specifically stated “don’t know” or provided no alternative answers. Table 2 provides examples.

To test for reliability, two coders independently coded a subset (20%) of the transcripts, which is
recommended by Bakeman and Gottman (1997) in developmental psychology, Thorndike and
Neuendorf (2002) argues that at least 50 units or 10% should be coded. Twenty-percent of transcripts
were chosen at random to be coded. The logic behind attaining inter-rater reliability on 20% is similar
to the logic behind inferential statistics in that the kappa coefficient generalizes to the remaining
transcripts. Reliability was calculated using kappa coefficients. Kappa coefficients between 0.60
and 0.75 were considered good and over 0.75 was excellent (Fleiss, 1981). A kappa coefficient of
0.81 was achieved for all codes in these transcripts. The first author coded the remainder of
the transcripts. Next, the second author read each transcript in its entirety and examined the first
author’s coding to her coding. The second author agreed with the vast majority of the remaining
coding (98%). All disagreements were resolved through discussion. The codes were as follows:

Informed Naturalistic Reasoning (INR)
The scientific model of evolution is based on variation, inheritance, selection, and time (Evans
et al., 2010). For students’ responses to be coded in this category, they needed to have alluded to
the idea that each living entity has evolved because of naturally occurring variations within a
population of species. These variations arise through genetic mutation or through genetic sexual
recombination; variations can be beneficial, harmful, or neutral to the survival of the individual.
Individuals that possess beneficial traits are more likely to survive to reproduction age and produce
offspring, whereas those that do not possess the traits are less likely to survive. Over time the
species population will have more individuals who possess the beneficial traits. Species no longer
suited to the environment will become extinct. Concepts included in this category include
extinction, inheritance, evolution, and common ancestor.

Journal of Research in Science Teaching
Table 2

<table>
<thead>
<tr>
<th>Coding scheme definitions and examples</th>
<th>Operational Definition</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Informed naturalistic reasoning</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extinction or death</td>
<td>Reference to animals not being able to adapt; or the specific mention of extinction.</td>
<td>“The animal cannot adapt fast enough so they die out.”</td>
</tr>
<tr>
<td>Inheritance</td>
<td>Reference to traits or characteristics being inherited. It is not sufficient for the students to suggest that the species will reproduce, but he/she will also need to specify that a certain trait has been passed on to the next generation.</td>
<td>“… the big beaked birds will probably reproduce, so they all have the genetic characteristics of a bigger beaked…”</td>
</tr>
<tr>
<td>Evolution</td>
<td>Reference to the underlying mechanisms of evolution. If students mention evolution without an explanation, it will not be coded in this category.</td>
<td>“Because all the small beaked birds probably died out because they were unable to get food, so the big beaked birds will probably reproduce, so they all have the genetic characteristics of a bigger beaked…”</td>
</tr>
<tr>
<td>Common ancestor</td>
<td>Mention that there were ancestors in common and explain this.</td>
<td>“There was a third species that was common to both the chimp and the human”</td>
</tr>
<tr>
<td><strong>Novice naturalistic reasoning</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static adaptation</td>
<td>References to the organism-environment fit.</td>
<td>“It is only adapted to living in the conditions that are in Tasmania”</td>
</tr>
<tr>
<td>Intention</td>
<td>The use of mental states to discuss change.</td>
<td>“Because the seeds were tougher, they need to develop bigger beaks to eat them”</td>
</tr>
<tr>
<td>Similarity</td>
<td>References to the similarity between organisms.</td>
<td>“They look the same, like chimps have five fingers, we have five fingers, they can stand on two feet, we can too”</td>
</tr>
<tr>
<td>Reaction or mutated</td>
<td>References to reactions to external matter.</td>
<td>“Maybe because it reacted with the chemicals in the lake, that is why they (algae) are there”</td>
</tr>
<tr>
<td>Movement</td>
<td>Animal moved somewhere either through another organism, by its own actions, or moved with the land.</td>
<td>“Because the Tasmanian tigers were on that bit of land when the Pangaea separated, and they could not swim, so that is why they were there”</td>
</tr>
<tr>
<td>Evolutionary term</td>
<td>Use the words evolve, adapted, adapted, adaptation, evolution without providing further explanation.</td>
<td>“The seals would adapt and evolve to live in the environment”</td>
</tr>
<tr>
<td>Teleological</td>
<td>Suggests that change occurs due to an end-point.</td>
<td>“Because they needed to eat the seeds, that is why their beaks were longer the second time the scientist came back”</td>
</tr>
<tr>
<td>Essentialist</td>
<td>References to the species having always been there, references to species stability.</td>
<td>“The algae has always been there, the scientists just did not find it before”</td>
</tr>
<tr>
<td>Hybrid</td>
<td>References to the interbreeding of two unrelated species.</td>
<td>“Maybe a wolf and a tiger mated and that is why you have the Tasmanian tiger”</td>
</tr>
<tr>
<td>Creationist reasoning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denial</td>
<td>Participant ejects information provided by researcher.</td>
<td>“I do not think humans and chimps shared a common ancestor”</td>
</tr>
<tr>
<td>Religious</td>
<td>Where participant makes reference to God or a supernatural being.</td>
<td>“Because God put them [Tasmanian tigers] there”</td>
</tr>
</tbody>
</table>

Journal of Research in Science Teaching
**Novice Naturalistic Reasoning (NNR)**

This category captures reasoning patterns derived from intuitive evolutionary concepts. In this conceptual framework, participants view individual animals as intentional agents who evolve as and when needed for survival, or that individual animals change to suit the environment. This reasoning pattern also includes responses where participants allude to essentialist (i.e., essentialism, movement) and teleological ideas (i.e., “humans won’t evolve anymore because they have already reached the highest”).

**Denial of Evolutionary Reasoning (DR)**

This final category captures reasoning patterns whereby students make reference to a creator or a supernatural being. There are two concepts in this reasoning category: rejection and religious. The code rejection is used when participants rejected a piece of information provided by the researcher (e.g., about the common ancestry of chimps and humans, “I don’t think that is true.”). The code religious was used when students referenced God as the instigator of evolution (e.g., “Because God made it that way”). We did not label these codes as creationist reasoning because few students invoked explicit creationism in their reasoning.

**Scoring.** One change was made to the original coding scheme. Evans et al. (2010) coded the mere mention of the words adapt, evolution, evolve, etc. as informed naturalistic reasoning. However, upon reading the transcripts in full, we saw that students frequently used these words while referencing concepts from novice naturalistic reasoning. For example, one 16-year-old girl used the term adapt, but then goes on to explain a teleological concept, “the drought caused the bodies of the finches who were living during the drought to adapt so they can have bigger beaks to store more food in cases of drought perhaps”. Thus, we did not credit her with having informed naturalistic reasoning. Instead, we coded students’ mere mention of such words as evolutionary term, and categorized them as belonging to the novice naturalistic reasoning category. In another interview, a student (male, age 14) answering the same question explained that the Galapagos finches adapted to eat the seeds. However, when the interviewer asked the student what he meant by “adapt,” the student endorsed a teleological stance:

Child: Because the seeds became bigger and harder, so then the beaks of the finches adapted to it, so they could eat the seeds.
Interviewer: What do you mean by adapt?
Child: Get used to the new food, so it can eat it.

If the above excerpt was coded following Evan’s coding scheme, the student would have been coded as using both informed naturalistic reasoning (for using an evolutionary term) and novice naturalistic reasoning (for endorsing teleological reasoning). However, with our new and more stringent coding scheme, we only coded this student as using novice naturalistic reasoning. This is because that we felt confident that the student did not have a normative understanding of the term “adapt” following the researcher’s probing question. It is important to note, however, that the youngest participants in Evans et al. (2010) were 18 years rather than 12 and would have been more likely than our participants to use the vocabulary in a scientifically accurate manner. Thus, our scheme is adapted for a younger age group.

When coding the interview transcripts, each of the reasoning categories was coded for presence (1) or absence (0). Thus, for each coding category, participants were scored as either 0 or 1. First, to determine students’ use of multiple epistemologies, students’ endorsement of each of the eight reasoning patterns was summed across the five questions ranging from a minimum of 0 (students did not use this type of reasoning in at all in answering the questions) to a maximum of...
5 (students used this type of reasoning consistently in all of the questions). Subsequently, to determine the types of reasoning students endorsed for each question, the original scoring for reasoning pattern (ranging from 0 to 1) was used. Finally, to contextualize our quantitative findings, we will provide representative examples of responses from interview conversations.

Results

Before testing our hypotheses statistically, we conducted a five (Question) × eight (Type of Reason) ANCOVA with age group and school as a between-subjects variables to rule out differences between schools. There was no main effect of school nor were there interaction effects between school and any of the variables. Thus, we can rule out school effects that could hinder our ability to generalize beyond these schools. To increase statistical power and test our hypotheses, we next conducted a five (Question) × eight (Type of Reason) ANOVA with age group as a between-subject variable. We used mixed design ANOVA models because we wanted to be able to examine the use of reasoning types across the five questions for each participant. Entering three factors into the mixed-design ANOVA also protected the alpha level. Analysis of variance (ANOVA) models were used to analyze dichotomous data. These procedures are preferable to log-linear analytical procedures when analyzing dichotomous and repeated measures designs (see Wainryb, Shaw, Laupa, & Smith, 2001). Moreover, ANOVA can be used with dichotomous data when the degrees of freedom for the error terms are greater than 40 (Lunney, 1970). Bonferroni post hoc tests were used to explore significant effects. Where assumptions of sphericity were not met because of a significant Mauchly’s test, Greenhouse-Geisser corrected results are reported. Finally, only significant main effects and interactions are reported, partial eta squared \( \eta^2 \) was used to calculate the effect size for ANOVA effects, and Cohen’s d \( d \) was used to calculate effect sizes for all pairwise post-hoc comparisons. Cohen’s \( d \) values between 0.20 and 0.50 indicate a small effect size, Cohen’s \( d \) values between 0.50 and 0.80 indicate a medium effect, and \( d \) values greater than 0.80 indicate a large effect (Cohen, 1988). Figure 1 shows the means of the different types of reasons used by each age group.

Multiple Epistemologies

To determine whether participants used target-dependent epistemologies in their explanations as would be expected by the third hypothesis, a five (Question) × eight (Type of Reason) × three (Age) repeated-measures ANOVA was conducted. Findings indicate a significant effect of Type of Reason, \( F(7, 707) = 331.57, p < 0.001, \eta^2 = 0.77 \), an interaction of Question × Type of Reason, \( F(28, 2828) = 20.46, p < 0.001, \eta^2 = 0.17 \), and a three-way interaction of Question × Type of Reason × Age, \( F(56, 2828) = 1.54, p = 0.009, \eta^2 = 0.03 \).

To follow up on the main effect of Type of Reason, Bonferroni post-hoc tests were conducted and revealed that students of all ages were more likely to endorse NNR (\( M = 3.38, SD = 1.10 \)), than all other reasoning patterns (INR/NNR, \( M = 0.63, SD = 0.70, t(103) = 18.32, p = 0.0001, d = 2.98 \); DK, \( M = 0.39, SD = 0.72, t(103) = 18.64, p = 0.0001, d = 3.21 \); INR, \( M = 0.38, SD = 0.58, t(103) = 22.52, p = 0.0001, d = 3.40 \); DR, \( M = 0.12, SD = 0.38, t(103) = 25.58, p = 0.0001, d = 3.95 \); NNR/DR, \( M = 0.11, SD = 0.34, t(103) = 27.05, p = 0.0001, d = 4.00 \); INR/DR and INR/NNR/DR, \( M = 0.00, SD = 0.00, t(103) = 31.32, p = 0.0001, d = 4.33 \)). Thus, NNR was by far the most typical reasoning pattern for all ages, which supports the first hypothesis that all age groups would rely on novice naturalistic reasoning. Second, students were more likely to rely on combined INR/NNR in their answers than DK, \( t(103) = 2.27, p = 0.03, d = 0.33 \), INR, \( t(103) = 2.36, p = 0.02, d = 0.38 \), DR, \( t(103) = 6.49, p = 0.0001, d = 0.91 \), NNR/DR, \( t(103) = 6.83, p = 0.0001, d = 0.95 \), INR/NNR/DR, \( t(103) = 9.11, p = 0.0001, d = 1.27 \), or
INR/DR, \( t(103) = 9.11, p = 0.0001, d = 1.27 \). Third, students gave “don’t know” responses more than DR, \( t(103) = 3.83, p = 0.0001, d = 0.47 \), NNR/DR, \( t(103) = 3.64, p = 0.0001, d = 0.51 \), INR/NNR/DR, \( t(103) = 5.61, p = 0.0001, d = 0.77 \), or INR/DR, \( t(103) = 5.61, p = 0.0001, d = 0.77 \). The use of “don’t know” and INR did not differ, \( t(103) = 0.10, p = 0.92 \). Fourth, students used INR more than DR, \( t(103) = 3.85, p = 0.0001, d = 0.55 \), NNR/DR, \( t(103) = 4.21, p = 0.0001, d = 0.58 \), INR/NNR/DR, \( t(103) = 6.77, p = 0.0001, d = 0.94 \), or INR/DR, \( t(103) = 6.77, p = 0.0001, d = 0.94 \). Finally, DR was used more often than, INR/NNR/DR, \( t(103) = 3.12, p = 0.002, d = 0.43 \), or INR/DR, \( t(103) = 3.12, p = 0.002, d = 0.43 \). Students’ endorsement of DR did not differ from their use of NNR/DR, \( t(103) = 0.19, p = 0.85 \). INR/NNR/DR and INR/DR were never used. See Figure 1 for use of the different reasons across age groups.

Next we followed up on the two- and three-way interactions related to our hypotheses. We examined each type of reasoning separately, and report age effects where significant to understand better the Question × Type of Reason × Age and Question × Type of Reason interaction effects. The finding from these analyses demonstrated the ways in which participants varied in their reasoning based on the question they were answering. Findings are reported in order of most to least endorsed reasoning type.

*Figure 1. Mean of students’ endorsement of each type of reasoning by age. [Color figure can be viewed at wileyonlinelibrary.com]*

*Journal of Research in Science Teaching*
Novice Naturalistic Reasoning

To examine students’ endorsement of novice naturalistic reasoning, a five(Question) × three (Age) repeated-measures ANOVA was conducted for NNR. Results indicated that there was a main effect of Question in students’ endorsement of NNR, $F(4, 404) = 25.56, p < 0.001, \eta^2_p = 0.20$. A Bonferroni post hoc test indicated that the questions about the tigers and algae were more likely to evoke NNR than the question about the seals (tiger-seal comparison, $p < 0.001, d = 1.19$, algae-seal comparison, $p < 0.001, d = 1.12$) and the finches (tiger-finch comparison, $p < 0.001, d = 0.63$; seal-finch comparison, $p < 0.001, d = 0.58$). Students were least likely to use NNR when discussing the seals (seal-human comparison, $p < 0.001, d = 0.47$). Students’ endorsement of NNR for tiger and algae did not differ from each other ($p = 1.00, d = 0.05$). Once again, there was no main effect of age in students’ endorsement of NNR, $F(2, 101) = 0.48, p = 0.62, \eta^2_p = 0.01$, indicating that participants in all age groups were equally likely to endorse novice naturalistic reasoning. Nor was there a significant interaction of Question × Age, $F(8, 404) = 0.88, p = 0.53, \eta^2_p = 0.02$.

Informed/Novice Naturalistic Reasoning

To examine students’ endorsement of INR/NNR, a five(Question) × three (Age) repeated-measures ANOVA was conducted and found a main effect of Question, $F(2.34, 235.97) = 37.59, p < 0.001, \eta^2_p = 0.27$. A Bonferroni post hoc test indicated that students were most likely to endorse INR/NNR when discussing the seals than all other entities (seal-finch comparison, $p < 0.001, d = 0.25$, seal-tiger comparison, $p < 0.001, d = 1.00$, seal-chimp/human comparison, $p < 0.001, d = 1.09$, and seal-algae comparison, $p < 0.001, d = 1.11$). Participants were also more likely to endorse this type of reasoning when discussing the finches than the chimps/humans ($p = 0.01, d = 0.43$), and the algae ($p = 0.01, d = 0.43$). Students’ endorsement of INR/NNR was not significantly different from each other for the tigers, the chimps/humans, and the algae (see Table 3). The interaction between Question × Age was not significant, $F(4.67, 235.97) = 1.12, p = 0.352, \eta^2_p = 0.02$. Nor was there a main effect of age, $F(2, 101) = 2.83, p = 0.06, \eta^2_p = 0.05$, thus indicating that students of all ages are equally likely to use INR/NNR in their answers.

Informed Naturalistic Reasoning

To determine whether or not student’s endorsement of INR varied by question, a five(Question) × three (Age) repeated-measures ANOVA was conducted for INR. The main effect for age was not significant, $F(2, 101) = 1.36, p = 0.26, \eta^2_p = 0.03$. Findings, however, indicate that there was a significant effect of Question, $F(2.36, 238.27) = 11.69, p < 0.001, \eta^2_p = 0.10$, and a significant interaction of Question × Age, $F(4.72, 238.27) = 2.75, p = 0.022, \eta^2_p = 0.05$, suggesting that students’ endorsement of informed naturalistic reasoning varied both by question and by age of the participant. A Bonferroni post hoc test suggested that students were more likely to endorse INR when answering the question about the seals than the question about the tigers ($p < 0.001, d = 0.62$), the chimps/humans ($p = 0.002, d = 0.52$), and the algae ($p < 0.001, d = 0.62$). They were also more likely to endorse INR when discussing the finches than the tigers ($p = 0.002, d = 0.51$), the chimps/humans ($p = 0.02, d = 0.40$), and the algae ($p = 0.002, d = 0.51$). Participants’ endorsement of INR for the questions about the finches and the seals were not significantly different from each other, nor were participants’ endorsement of INR for tigers, chimps/humans, and algae (see Table 3). To follow up on the significant Question × Age interaction, five one-way ANOVAs were calculated comparing the three age groups across each question separately, with a protected $p$-value of 0.01 (0.05 divided by 5). Only the question about the finches reached significance, $F(2, 105) = 5.28, p = 0.007, \eta^2_p = 0.09$. As predicted by
Table 3

Mean proportion by age

<table>
<thead>
<tr>
<th>Entity</th>
<th>INR Only</th>
<th>NNR Only</th>
<th>DR Only</th>
<th>INRNNR</th>
<th>INRNNRDR</th>
<th>INRDR</th>
<th>NNRDR</th>
<th>Don’t Know</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age 12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tiger</td>
<td>0.00 (0.00)</td>
<td>0.92 (0.27)</td>
<td>0.03 (0.16)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.05 (0.22)</td>
</tr>
<tr>
<td>Seals</td>
<td>0.26 (0.44)</td>
<td>0.38 (0.49)</td>
<td>0.00 (0.00)</td>
<td>0.33 (0.48)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.03 (0.16)</td>
</tr>
<tr>
<td>Finches</td>
<td>0.03 (0.16)</td>
<td>0.67 (0.48)</td>
<td>0.00 (0.00)</td>
<td>0.05 (0.22)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.26 (0.44)</td>
</tr>
<tr>
<td>Humans</td>
<td>0.05 (0.22)</td>
<td>0.67 (0.48)</td>
<td>0.10 (0.31)</td>
<td>0.05 (0.22)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.08 (0.27)</td>
</tr>
<tr>
<td>Algae</td>
<td>0.03 (0.16)</td>
<td>0.84 (0.37)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.13 (0.34)</td>
</tr>
<tr>
<td>Age 14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tiger</td>
<td>0.00 (0.00)</td>
<td>0.77 (0.43)</td>
<td>0.00 (0.00)</td>
<td>0.10 (0.31)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.10 (0.31)</td>
<td>0.03 (0.18)</td>
</tr>
<tr>
<td>Seals</td>
<td>0.13 (0.34)</td>
<td>0.32 (0.48)</td>
<td>0.00 (0.00)</td>
<td>0.52 (0.51)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.03 (0.18)</td>
</tr>
<tr>
<td>Finches</td>
<td>0.13 (0.34)</td>
<td>0.65 (0.49)</td>
<td>0.00 (0.00)</td>
<td>0.16 (0.37)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.06 (0.25)</td>
</tr>
<tr>
<td>Humans</td>
<td>0.00 (0.00)</td>
<td>0.81 (0.40)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.13 (0.34)</td>
<td>0.06 (0.25)</td>
</tr>
<tr>
<td>Algae</td>
<td>0.00 (0.00)</td>
<td>0.84 (0.37)</td>
<td>0.03 (0.18)</td>
<td>0.03 (0.18)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.03 (0.18)</td>
<td>0.06 (0.25)</td>
</tr>
<tr>
<td>Age 16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tiger</td>
<td>0.03 (0.03)</td>
<td>0.86 (0.35)</td>
<td>0.00 (0.00)</td>
<td>0.03 (0.17)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.08 (0.28)</td>
</tr>
<tr>
<td>Seals</td>
<td>0.17 (0.38)</td>
<td>0.33 (0.48)</td>
<td>0.00 (0.00)</td>
<td>0.47 (0.51)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.03 (0.17)</td>
</tr>
<tr>
<td>Finches</td>
<td>0.28 (0.45)</td>
<td>0.47 (0.51)</td>
<td>0.03 (0.17)</td>
<td>0.17 (0.38)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.03 (0.17)</td>
<td>0.03 (0.17)</td>
</tr>
<tr>
<td>Humans</td>
<td>0.03 (0.17)</td>
<td>0.75 (0.44)</td>
<td>0.14 (0.35)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.08 (0.28)</td>
</tr>
<tr>
<td>Algae</td>
<td>0.00 (0.00)</td>
<td>0.83 (0.38)</td>
<td>0.03 (0.18)</td>
<td>0.03 (0.17)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.14 (0.35)</td>
</tr>
<tr>
<td>Combined</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tiger</td>
<td>0.01 (0.10)</td>
<td>0.86 (0.35)</td>
<td>0.01 (0.10)</td>
<td>0.04 (0.19)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0.03 (0.17)</td>
<td>0.06 (0.23)</td>
</tr>
<tr>
<td>Seals</td>
<td>0.19 (0.40)</td>
<td>0.36 (0.48)</td>
<td>0 (0)</td>
<td>0.42 (0.50)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0.03 (0.17)</td>
</tr>
<tr>
<td>Finches</td>
<td>0.14 (0.35)</td>
<td>0.59 (0.49)</td>
<td>0.01 (0.10)</td>
<td>0.13 (0.33)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0.01 (0.10)</td>
<td>0.13 (0.33)</td>
</tr>
<tr>
<td>Humans</td>
<td>0.03 (0.17)</td>
<td>0.74 (0.44)</td>
<td>0.09 (0.28)</td>
<td>0.02 (0.14)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0.06 (0.23)</td>
<td>0.07 (0.25)</td>
</tr>
<tr>
<td>Algae</td>
<td>0.01 (0.10)</td>
<td>0.84 (0.37)</td>
<td>0.01 (0.10)</td>
<td>0.02 (0.14)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0.01 (0.10)</td>
<td>0.12 (0.32)</td>
</tr>
</tbody>
</table>

Note: Standard deviations in parentheses. This table displays the mean number of students endorsing each of the eight reasoning patterns for the five entities discussed. For each entity, the means of students’ endorsement of each reasoning pattern sum up to 100%. For column values marked with subscripts for the combined group, under each type of reasoning, means that do not share subscripts differ by p < 0.05 according to Bonferroni post hoc tests.

Indicate significant age differences (p ≤ .01).
hypothesis 2, findings indicate that 16-year-old students were more likely than 12-year-old students to endorse INR, \( t(42.97) = 3.15, p = 0.003, d = 1.61 \).

**Denial of Evolutionary Reasoning**

Similarly, students’ endorsement of denial of evolutionary reasoning was analyzed using a five(Question) × three (Age) repeated-measures ANOVA. The analysis revealed an effect of Question in students’ endorsement of DR, \( F(1.75, 176.77) = 5.85, p = 0.005, \eta_p^2 = 0.06 \). A Bonferroni post hoc test indicated that students were more likely to endorse DR when discussing the chimps/humans than the seals (\( p = 0.04, d = 0.45 \)) questions. Participants’ endorsement of DR did not differ for the question about the tigers, seals, finch, or algae (all \( ps = 1.00 \)). There was no Question × Age effect of students’ endorsement of DR, \( F(3.50, 176.77) = 2.13, p = 0.089, \eta_p^2 = 0.04 \) nor was there a significant main effect of Age, \( F(2, 101) = 1.08, p = 0.34, \eta_p^2 = 0.02 \).

**Mixed Novice Naturalistic/Denial of Evolutionary Reasoning and Other Mixed Reasoning Patterns**

Finally, to examine students’ endorsement of mixed NNR/DR, a five(Question) × three (Age) repeated-measures ANOVA was conducted. The ANOVA model revealed a main effect of Question, \( F(2.26, 228.50) = 3.29, p = 0.03, \eta_p^2 = 0.03 \). However, the Bonferroni post hoc test did not indicate any significant differences in students’ endorsement of this type of reasoning pattern for the different questions. Other mixed reasoning patterns (INR/NNR/DR, INR/DR) were not explored as none were endorsed by any of the participants.

“Don’t Know”. Since “don’t know” responses made up a significant proportion (39.4%) of student responses, we will also explore how these are distributed across questions and age groups. Once again, a repeated-measures ANOVA for DK was conducted for the five questions, and revealed a significant interaction of Question × Age, \( F(7.04, 356.06) = 2.17, p = 0.04, \eta_p^2 = 0.04 \). To follow up on this interaction, five one-way ANOVAs were conducted comparing the three age groups across each question separately, with a protected \( p \)-value of 0.01 (0.5 divided by 5). Only the question on finches reached significance, \( F(2, 106) = 5.64, p = 0.005, d = 0.10 \). A Bonferroni post hoc test indicated that 12-year olds were more likely than 14- \( p = 0.04, d = 0.53 \) and 16-year olds \( p = 0.007, d = 0.69 \) to respond with “I don’t know” for the finch question.

**Comparison Between Evan’s et al. (2010) Coding Scheme and the New One**

To examine whether there were differences based on the changes we made to the coding scheme, we conducted analyses using the original Evans’s et al. (2010) scheme. We first compared our codes to Evans’s more explicitly. As expected, INR (\( M = 0.84, SD = 0.92 \)) and INR/NNR (\( M = 1.45, SD = 1.06 \)) were coded more frequently with Evans’s coding scheme than with the new coding scheme, \( t(101) = 6.28, p = 0.0001, d = 0.60 \) and \( t(101) = 7.89, p = 0.0001, d = 0.91 \), respectively. In contrast, NNR (\( M = 1.83, SD = 1.16 \)) was coded less frequently with the new coding scheme than with the old coding scheme, \( t(101) = 12.52, p = 0.0001, d = 1.37 \). Thus, in adopting more stringent criteria in the new coding scheme, more students were classified as novice naturalistic reasoners than informed naturalistic reasoners than if we had used the old coding scheme (Evan’s et al., 2010).

Next, we conducted the main analyses of a three (Age) × five (Question) × eight (Type of Reasoning) ANOVA using the original scheme. First, with the coding based on Evans, there continued to be a difference in Type of Reason, \( F(3, 265) = 331.57, p = 0.001, \eta_p^2 = 0.77 \), as there was in the new coding scheme. Whereas NNR was the most frequently used reason with the new scheme, NNR (\( M = 1.83, SD = 1.16 \)) and INR/NNR (\( M = 1.46, SD = 1.06 \)) were used with equal
frequency. INR/NNR continued to be endorsed more often than the other codes. INR was the next most frequently used code ($M = 0.85, SD = 0.92$).

Second, we examined age effects. There was a significant Type of Reasoning $\times$ Age effect, $F(6, 300) = 3.23, p = 0.004, \eta^2 = 0.06$. To examine this further, we conducted three separate one-way ANOVAs with age as an independent variable on NNR, INR, and INR/NNR (we would not have expected differences on DK and DR because the coding schemes were identical for these codes and codes combined with DR were so rare that we would not expect effects). In terms of age effects, remember that there were no main effects for age with the new coding scheme. However, with the Evans’ scheme, there was a main effect of age on INR, $F(2, 101) = 3.59, p = 0.03, d = 0.59$, with 16 year olds ($M = 1.14, SD = 1.13$) using INR more frequently than 12 year olds ($M = 0.58, SD = 0.72$). In contrast, 12 year olds ($M = 2.26, SD = 1.13$) used NNR more frequently than 16-year olds ($M = 1.52, SD = 1.23$), $F(2, 101) = 4.59, p = 0.01, d = 0.62$. Fourteen year-old students did not differ from the other two age groups in their use of INR ($M = 0.83, SD = 0.79$) or NNR ($M = 1.63, SD = 0.96$). Finally, using Evans’s coding scheme, there was no significant age effect of INR/NNR, $F(2, 102) < 1$.³

Summary. In sum, fewer students were classified as having used informed naturalistic reasoning when using the new coding scheme than when using Evans’s coding scheme. In contrast, more students were classified as having used novice naturalistic reasoning when using the new coding scheme than Evans’s coding scheme. The key difference between the two coding schemes is that whereas Evans’s considers the mere mention of the terms evolve/evolution/adapt as evidence of having an informed naturalistic understanding, we did not. Students were considered to have an informed naturalistic understanding only if they were able to explain the relevant terms in that context. Because two different coding schemes were used to analyze the same dataset, differences as a result of the separate analyses must therefore be because of differences in the coding schemes. That fewer students were classified as having used informed naturalistic reasoning when using the new coding scheme than when using Evans’s coding scheme suggests that Evans and colleagues may have overestimated the proportion of people using informed naturalistic reasoning in their study.

Qualitative Changes in Students’ Reasoning About Evolution by Age

After conducting our quantitative analysis, we went back through our transcripts to confirm that the patterns we found quantitatively were apparent in our transcripts. Again, we found clear qualitative differences in the way students from the three age groups reasoned about evolution. Whereas students in the youngest age cohort focused on surface features to help them reason about evolution (e.g., the relationship between beak size and size of the seeds), older students’ responses incorporated more evolutionary terms (e.g., evolve, adapt, mutated; see Table 4), suggesting an attempt to reason about the mechanism of change. A further investigation into the pattern of change revealed that it is at age 14 that students begin to incorporate relevant terminology in their responses. That younger students focused on more surface features when answering the questions is in line with the research by Nehm and Ridgway (2011) who found that novices were more likely than experts to focus on surface features when solving evolutionary problems. Furthermore, because 14-year olds have had more instruction about evolution in school science than 12-year olds, we can attribute this change in part to students progressing through the National Curriculum. However, a careful reading of 14-year olds’ responses reveals that while they have appropriated the relevant terminology in their responses, they have yet to grasp the conceptual meaning conveyed by the terminology. Students’ misinterpretation of relevant terminology may have contributed to older students’ continued biases about evolutionary processes.

Journal of Research in Science Teaching
Only a subset of 16-year olds developed a scientific understanding about biological evolution (Table 4). The more advanced reasoning exhibited by a minority of 16-year olds can be attributed to the more elaborate content being taught in classrooms. For example, students are taught that variation can be traced back to genes, heredity is a biological process, and some species share a common ancestor (Ryan, 2014). As students learn about evolution in greater depth, we also find an increase in students’ use of these concepts (e.g., survival of the fittest, species environment fit, and species change over time). However, similar to the pattern we have seen among the 14-year olds, many 16-year olds also failed to grasp the scientific definition of the relevant terms introduced in the science classroom (see Anirudh in Table 4). It may be that students’ earlier misinterpretation of relevant terminology contributed to their continued misunderstandings about evolutionary processes. Rather than thinking about how environmental changes contribute to species changes within a population, some students assume that evolutionary changes occur at the individual level where individuals have direct control over gene mutation. For example, Harshil (age 16, Table 4)
proposed that the finches adapted to the harder seeds by mutating to have larger beaks. He also proposed that this mutation happens at the gene level and would be passed on to the next generation. Here, Harshil draws on cultural knowledge learned in the science classroom from the teaching of evolutionary theory, which includes relevant terminology, the understanding that mutation occurs at the gene level, and that genes are inherited from one generation to the next. He combines information from his newly represented epistemology with his prior naïve theory of biology (teleology) in a coexistence model.

In another example, Anirudh (age 16, Table 4) again constructed his response based on multiple epistemologies, however, his reasoning is less well integrated than Harshil’s. First, Anirudh acknowledged that the finches with smaller beaks would not survive, and that to survive, finches would need bigger beaks. When the interviewer repeated the original question how the finches on the Galapagos Islands came to have bigger beaks, Anirudh responded by incorporating relevant terminology (e.g., natural selection). However, it was clear from his response that he did not have a causal mechanistic explanation for the change.

In sum, our qualitative analysis suggests that as students learn more about evolution as part of the mandatory National Curriculum, they incorporate their new understandings about evolution into their existing reasoning patterns. However, students’ misinterpretation of key words (e.g., adapt, change, evolve, etc.) meant that their reasoning about evolution was inherently flawed allowing them to retain their representations of their naïve theories. Early misinterpretation of relevant terminology at age 14 may have contributed students’ erroneous reasoning about evolution. The resulting conceptual structure at age 16 is an amalgamation of students’ prior knowledge and discrete pieces of information recalled from school science.

Discussion

The purpose of this study was to understand developmental changes in secondary school students’ reasoning about evolution. We followed a procedure similar to the one outlined in Evans et al. (2010), with one notable change. Whereas Evans et al. (2010) coded participants’ mere mention of the terms evolve or adapt as evidence of informed naturalistic reasoning, participants in our study also needed to be able to explain these terms in context to be considered using informed naturalistic reasoning. Following this coding procedure, despite older students being more able than younger students to apply scientific concepts to their evolutionary reasoning of some entities (partial support for hypothesis 2), the majority of participants at all age groups continued to endorse novice naturalistic reasoning more than any other reason (hypothesis 1). Furthermore, there was evidence that students used different types of reasoning when answering the five questions (hypothesis 3). Specifically, students were more likely to endorse scientific concepts when reasoning about the seals and finches than about humans, algae, and tigers. Finally, students were more likely to endorse denial of evolutionary reasoning for the question about the chimps/humans than the seals; no other effect of students’ endorsement of denial of evolutionary reasoning was found (hypothesis 4).

Findings in this study echo that of past research in that secondary school students’ understandings about evolution are largely underdeveloped and guided by their intuitive reasoning (Bishop & Anderson, 1990; Clough & Wood-Robinson, 1985; Evans et al., 2010). Our findings are in line with research by Beggrow and Nehm (2012) that increased exposure to more advanced content is insufficient in helping students build more expert-like understandings about evolution. Perhaps the most important contribution of this research is the development of a more stringent coding scheme. By using a coding scheme that coded for students’ conceptual understandings rather than their use of evolutionary terms, this study was able to describe age differences in secondary school students’ conceptual understandings of evolutionary processes. Furthermore,
because our coding scheme allowed for more than one type of reasoning pattern to be coded, there was evidence to suggest that students’ preference for certain types of reasoning when thinking about biological evolution was already present by age 12. Let us further explore the role of relevant terminology in students’ reasoning about evolution and also patterns in students’ use of multiple epistemologies.

**Learning Relevant Terminology Toward Greater Evolutionary Understanding**

As detailed above, students frequently attach teleological or intentional definitions to evolutionary terms such as *adapt* and *evolve*. Within our sample, students also tended to use *change* interchangeable with *adapt* and *evolve*. Students’ misuse of scientific terminology is commonplace, especially for novice learners (Nehm, Rector, & Ha, 2010; Rector, Nehm, & Pearl, 2013; Ryan, 1985). Acquiring expert knowledge of relevant terms, concepts, and processes is a gradual process (Rector et al., 2013). Using appropriate language can help learners structure their thoughts so that they are more able to reason about problems in a scientifically appropriate way.

Figures 2 and 3 are examples of students’ explanations before and after they learn the relevant terminology. In Figure 2, we see that students recognize that there is a relationship between the size of the finches’ beaks and the size of the seeds that they eat, however, their explanations of how the change occurred relied mainly on intuitive explanations (e.g., moving away, digging deeper into the ground, nutrients causing the beaks to grow). In contrast, 14- and 16-year olds used relevant language to structure their explanation. For example, Auhamud (Figure 3) started his answer by stating “adaptation”. But when the researcher asked what he meant by adaptation, he cited his source in biology and extended what he thought was the correct interpretation to the new problem. Also from this response, it is clear that Auhamud’s misunderstanding did not arise during the transfer phase. Instead, Auhamud’s misconception was a result of his endorsement of a colloquial rather than a scientific understanding of the term *adaptation*. Nevertheless, comparing Auhamud’s response and those typical of 12-year olds, students who have learned the relevant terminology focused on the process of change, either at the individual level or at the population level:

- "Well, larger finches probably migrated to the island and the smaller finches were slightly driven off yeah.” (Alex, 12)
- "Maybe they can eat more food and they can get bigger things into their mouths.” (Elena, 12)
- "Um, maybe what the plants that they have been eating has helped their beaks to grow?” (Jessica, 12)
- "Because they had eaten more of these seeds so they would have to search further down [...] The seeds may be underground, so the more they eat, the further they will have to search for them." (Andrew, 12)

*Figure 2.* Evolutionary reasoning with relevant language.

*Journal of Research in Science Teaching*
level. Figures 2 and 3 show examples of reasoning patterns before and after students learned the relevant terminology.

Being able to use relevant language in the way it is intended is central to science learning (Ryan, 1985). Learning the relevant language has the potential to help students establish and organize their conceptual understanding of the topic (Lemke, 1990). Indeed, confusing scientific terms with colloquial ones and vice versa is common in science learning (Lemke, 1990). When students first learn about a scientific concept, they will grasp the semantic and conceptual relationships first before learning the relevant words (Lemke, 1990). However, in our study, students’ colloquial interpretations of relevant terms seemed to have hindered students’ representation of the intended scientific interpretation. Considering that many students at all levels have difficulty using evolutionary terms in an appropriate manner, a coding scheme that did not consider students’ understandings of relevant terminology are inherently flawed. Instead, students should only be considered to have an informed naturalistic reasoning (i.e., a normative understanding) about evolution when they are able to explain the causal mechanisms of evolutionary change.

Although many students held misconception about relevant terminology, a subset of students continued on to develop scientifically appropriate reasoning strategies in evolution. It is not clear why some students eventually developed a scientific understanding about evolution while others did not. One reason may be that even though all students followed the same mandatory National Curriculum, each student interpreted information through their previous epistemological filter (Gee, 2008). As such, what is being taught in classrooms (e.g., input) is not the same as what is learned (e.g., uptake; Gee, 2008). That not all 16-year olds eventually developed a normative understanding about evolution is in line with recent research on people’s evolution understanding. Abraham-Silver and Kisiel (2008) investigated museum visitors’ understanding and acceptance of evolution in the US, UK, Canada, and Australia. They found that although British museum visitors were less likely to reject evolution than their US counterparts, the British were as likely as those from other English speaking countries to hold misconceptions about evolution. Furthermore, in intervention studies where significant changes in students’ understandings about evolution have been documented there remained a proportion of students whose intuitive ideas were resistant to change (e.g., Banet & Ayuso, 2003; Geraedts & Boersma, 2006). Thus, we cannot expect all students to develop a normative understanding about evolution as a result of having progressed through the National Curriculum alone.

Journal of Research in Science Teaching
One implication from this research is that in teaching evolution, teachers must not assume that students will learn the scientific concepts behind the terminology as a result of working through the National Curriculum; teachers must also address and challenge students’ misconceptions in the teaching and learning process. In addition, students’ use of coexistence models suggests that teachers must also consider what students already think they know about evolution when teaching the topic. This will allow the teachers to address alternative conceptions if necessary.

**Endorsement of Multiple Epistemologies**

Next, let us examine factors influencing students’ use of multiple epistemologies. As predicted in hypothesis 3, participants endorsed different epistemologies across the five interview questions. Specifically, we found that the questions about the finch and seals elicited more informed naturalistic reasoning than the questions about tigers, humans, and algae, and that the questions about the tigers and the algae evoked more novice naturalistic reasoning than the question about finches. This pattern of endorsement is in contrast to findings by Evans et al. (2010) where adult museum visitors were more likely to endorse informed naturalistic reasoning for mammals and birds while endorsing novice naturalistic reasoning for microscopic species and invertebrates. Also in contrast with findings by Evans et al. (2010) is that students in this study were not more likely to endorse creationist reasoning for the chimps/humans questions than all other entities. Instead, participants were more likely to endorse denial of evolution reasoning for chimps/humans only in comparison to the question about the seals. Previous research has indicated that people are less likely to extend evolution to humans than other entities because of negative ramifications for the purpose of humans’ existence, such as not having a purpose or free will (Brem, Ranney, & Schindel, 2003). We do not believe this explanation applies to our sample. Instead, we argue that the lack of support for denial of evolutionary reasoning is the result of British students being exposed to a set of explanations (at home and in formal education settings) that less frequently include creationist explanations that in the southern and midwestern US. Although the UK and the US are similar in terms of industrialization, the southern and midwestern US, where Evans et al. (2010) collected her data tends to be more religious than the UK (Micklethwait & Wooldridge, 2005). Furthermore, as Kelemen (2003) has previously argued, differences in the majority culture’s willingness to openly endorse creationist reasoning may result in children growing up in the two countries being exposed to slightly different types of explanations in parent-child conversations (Tenenbaum & Hohenstein, 2016). Thus, our study provides further support that minute cultural differences may have a significant impact on how people learn to reason about biological evolution.

One limitation of our study, nonetheless, is that it is cross-sectional. As a result, we cannot know how students continue to build an understanding about evolution as a result of their experiences in formal and informal learning settings. Thus, we were not able to explain why some students developed a normative understanding about evolutionary theory while others did not. Future research needs to incorporate longitudinal and even micro-genetic methods to further our understandings of the underlying processes in the development of students’ reasoning about evolution. In addition, this study highlights the need for future research to explore more exhaustively the role of culture in students’ reasoning about biological evolution. We have demonstrated that secondary students in England tend not to endorse creationist explanations about evolution, even when they fail to provide a plausible alternative. However, because our study was not a comparison study, further evidence is needed to determine the role of culture in people’s reason about biological evolution.

*Journal of Research in Science Teaching*
Conclusion

In summary, this research has found a clear developmental trend in students’ emerging understandings about evolution where 16-year olds are better able than their younger counterparts to incorporate scientific concepts in their explanations about evolution, but only within certain contexts. First, students were able to learn relevant vocabulary without associating this language with more scientific understandings. Relevant language in turn helped students structure their reasoning about species change. Whereas age effects on informed naturalistic reasoning were found using Evans’s et al. (2010) coding scheme across items, using our coding scheme, age effects on informed naturalistic reasoning were only found for the finch question. We believe that these effects demonstrate that adolescents have learned the vocabulary but not the concepts necessary for generalising concepts about evolution across a range of domains. Second, from a theoretical perspective, this study has contributed to the literature by generating a more valid coding scheme to be used with secondary school students to better understand their reasoning about evolution.

Considering current research findings within a teaching and learning context, when teaching about biological evolution, there is a genuine need for teachers to allow students multiple opportunities to practice using scientific theories in varying contexts. Doing so will ensure that students recognize that the underlying mechanism for evolution is the same across different type of entities and across different contexts. Furthermore, in the assessment of students’ understanding about biological evolution, in addition to being able to use relevant terminology, teachers need to encourage students to explain what they mean by adapt and evolve, such that misconceptions can be modified. Given that effective teaching strategies are those that take into consideration students’ prior knowledge and specific understandings (Bang & Medin, 2010), educators must tailor classroom instruction to dispel misconceptions students hold that can inhibit their reasoning about new concepts. Through such methods, students can begin to develop a richer understanding of evolutionary process that they can apply across different organisms.

Notes

1 Students in this study followed this curriculum. The recent change in curriculum introduced in 2014 to include evolution in the primary science curriculum did not affect the students in this study.

2 GCSE stands for the general certificate of secondary education. This is an academic qualification award in a specified subject taken in a variety of subjects by students aged 14 and 16. Schools in England, Wales, and Northern Ireland require these awards. GCSE performance is usually a good indicator of how well students do in advanced level studies (Gardner, 2014).

3 Third, there was a significant interaction effect of Question × Response, \( F (28, 997) = 12.96, p = 0.0001, \eta_p^2 = 0.12 \). We followed up this interaction by conducting separate ANOVAs for each type of response. For INR, there was a main effect of question, \( F (3.29, 338.81) = 10.42, p = 0.0001, \eta_p^2 = 0.09 \). Students used INR more for the seals (\( M = 0.32, SE = 0.47 \)) and finches (\( M = 0.26, SE = 0.44 \)). Students’ endorsement of INR did not differ from one another than for the questions on chimps/humans (\( M = 0.09, SE = 0.28 \)), algae (\( M = 0.09, SE = 0.28 \)), or tiger (\( M = 0.10, SE = 0.30 \)). For NNR, there was a main effect of question, \( F (4, 412) = 23.96, p = 0.0001, \eta_p^2 = 0.19 \). Students used NNR most for the algae (\( M = 0.63, SE = 0.49 \)) and humans (\( M = 0.46, SE = 0.50 \)). Next, NNR was used most for the tiger question (\( M = 0.40, SE = 0.49 \)), this did not differ significantly from the human question (\( M = 0.46, SE = 0.50 \)). NNR was used the least for finches (\( M = 0.18, SE = 0.39 \)) and seals (\( M = 0.15, SE = 0.36 \)) questions. Finally, for INR/NNR, there was a main effect of question,
$F(4, 400) = 11.61, p = 0.0001, \eta_p^2 = 0.10$. Students used INR/NNR most for the seals ($M = 0.50, SE = 0.05$) and finches ($M = 0.42, SE = 0.05$), which did not differ from one another. The use of NNR/INR did not differ between finches and tigers ($M = 0.38, SE = 0.05$). It was used the least for the human question ($M = 0.22, SE = 0.04$) and or algae ($M = 0.15, SE = 0.04$) questions.

References


Journal of Research in Science Teaching