IEJEE

Parent-child conversations about evolution in the context of an interactive museum display

Andrew SHTULMAN[*]
Occidental College, Los Angeles, USA

Isabel CHECA
Occidental College, Los Angeles, USA

Abstract

The theory of evolution by natural selection has revolutionized the biological sciences yet remains confusing and controversial to the public at large. This study explored how a particular segment of the public – visitors to a natural history museum – reason about evolution in the context of an interactive cladogram, or evolutionary tree. The participants were 49 children aged four to twelve and one accompanying parent. Together, they completed five activities using a touch-screen display of the phylogenetic relations among the 19 orders of mammals. Across activities, participants revealed similar misconceptions to those revealed by college undergraduates in previous studies. However, the frequency of those misconceptions was attenuated by the level of parental engagement, particularly the frequency of turn-taking between parents and children. Overall, these findings suggest that evolutionary reasoning may be improved by the kinds of collaborative discussions fostered by interactive museum displays, so long as the affordances of those displays encourage multi-user interactions.

Keywords: Conceptual Development, Evolution Understanding, Parent-Child Conversation, Informal Learning Environments, Science Education.

Introduction

In 1996, the U.S. National Academy of Sciences identified evolution as one of five "unifying concepts and processes" that should be taught in all grades, from K through 12 (National Research Council, 1996). The rationale behind this recommendation was that "evolution is the central organizing principle that biologists use to understand the world. To teach biology without explaining evolution deprives students of a powerful concept that brings great order and coherence to our understanding of life" (p. 3, National Research Council, 1998). Despite the force of this recommendation, many schools continue to fail to teach evolution in *any* grade (Griffith & Brem, 2004), and many Americans continue to deny the very fact of evolution, particularly human evolution (Miller, Scott, & Okamoto, 2006; Newport, 2010).

[*] ✉ Andrew Shtulman, Department of Psychology, Occidental College, Los Angeles, CA 90041. Phone: 323-259-2633. E-mail: shtulmanoxy.edu

One of the problems underlying, if not exacerbating, public denial of evolution is that most people fail to understand what evolution actually is and how evolution actually occurs (Shtulman & Calabi, 2012). Recent research has found that individuals of all ages and educational backgrounds tend to construe evolution as a kind of holistic transformation, by which organisms are predisposed to produce offspring more adapted to the environment than they were at birth (Shtulman, 2006; Shtulman & Calabi, in press). On this erroneous theory, the sole mechanism behind species adaptation is *need*: if a species *needs* to adapt, then it *will* adapt (Bishop & Anderson, 1990; Brumby, 1984; Southerland, Abrams, Cummins, & Anzelmo, 2001). Selection plays no causal role on this view, which, as a result, bears more resemblance to pre-Darwinian theories of evolution than post-Darwinian ones (Mayr, 1982). Developmentally, this view appears to be an outgrowth of an early emerging "essentialist" construal of biological kinds, in which an organism's outward appearance and behavior is determined by some kind of internal force, or "essence," conferred from parent to child at birth (Gelman, 2003; Hatano & Inagaki, 1994; Solomon & Zaitchik, 2012). While essentialism may be useful for reasoning about the properties of individual organisms (e.g., Gelman & Coley, 1990; Waxman, Medin, & Ross, 2007), it has been shown to be detrimental for reasoning about the properties of entire species, as it leads students to overvalue variation *between* species and undervalue variation *within* a species (Nettle, 2010; Shtulman & Schulz, 2008). As a consequence, students have difficulty understanding mechanisms of change that operate specifically over the variation within a species, namely, natural selection.

To date, numerous studies have documented the nature of students' essentialist, need-based views of evolution (for reviews, see Gregory, 2009; Sinatra, Brem, & Evans, 2008). Less attention, however, has been paid to the ways in which alternative views of evolution manifest themselves in everyday discourse and everyday interactions. The present study attempted to explore this issue in the context of parent-child conversations at a natural history museum. In particular, we sought to elicit conversations about the phenomena represented by one of the most canonical depictions of evolutionary change: the *cladogram*. More commonly referred to as an "evolutionary tree," cladograms are branching diagrams that depict patterns of common ancestry among three or more groups of organisms, or *taxa*. A sample cladogram, depicting the evolutionary relations among primates, can be seen in Figure 1.

Even though cladograms are ubiquitous in biology textbooks (Catley & Novick, 2008) and natural history museums (Torrens & Barahona, 2012), they are notoriously difficult to interpret, partly because they contain unfamiliar notational conventions (Novick & Catley, 2007) and partly because they are amenable to inaccurate, essentialist interpretations of evolutionary change (Shtulman, 2006). Drawing on recent empirical investigations of "tree thinking" in introductory biology students, Gregory (2008) outlined 10 such misconceptions:

   1. Interpreting taxa on one side of a cladogram as "higher" or "lower" than those on the other side.

   2. Interpreting the longest line in a cladogram as the "main line" from which other taxa have deviated or side-tracked.

   3. Inferring information about relatedness from the ordering of a cladogram's terminal nodes rather than from its branches.

   4. Interpreting cladograms as representations of morphological similarity rather than common ancestry.

   5. Interpreting some taxa in a cladogram as the ancestors of other taxa rather than interpreting all taxa as "siblings" or "cousins."

6. Interpreting the length of the branches in a cladogram as measures of evolutionary change (or lack thereof).

7. Inferring that the taxa on one side of a cladogram appeared, in their current form, earlier than those on the other side of the cladogram.

8. Interpreting the length of the longest branch of the cladogram as a measure of time.

9. Interpreting the number of intervening nodes between two taxa as a direct measure of their relatedness.

10. Interpreting internal nodes in a cladogram as representing precise moments of speciation, with little to no change occurring before or after that point in time.

Many of the misconceptions are overlapping (e.g., 1 and 7) and some are mutually exclusive (e.g., 6 and 10), but all represent illegitimate inferences from the information at hand.

As an illustration, consider the cladogram depicted in Figure 1. The only information this diagram provides is information about common ancestry – namely, that humans share a common ancestor with chimpanzees more recently than with any of the other primates, that humans and chimpanzees share a common ancestor with gorillas more recently than with any of the other primates, and so forth. Nevertheless, most people are prone to infer that:

1. Humans are more highly evolved than other primates.

2. Human evolution represents the "main line" of evolution, whereas the evolution of other primates represents sidetracks from this main line.

3. Chimpanzees are related to gorillas more closely than humans are related to gorillas (because the nodes of the former pair are adjacent but the nodes of the latter pair are not).

4. Humans are most similar to chimpanzees (the closest node to humans) and least similar to new world monkeys (the farthest node from humans).

5. Each primate is the descendent of the primate on its left and the ancestor of the primate on its right.

6. Humans have undergone the more evolutionary change than other primates (because their connection to the root node is longest).

7. Each primate appeared, in its current form, earlier than the primate to its right.

8. Each primate is older than the primate to its right.

9. Humans are related to orangutans less closely than chimpanzees are related to orangutans (because of differences in the number of intervening nodes).

10. Chimpanzees came into being instantaneously at the point denoted by the rightmost node.

These inferences are not just logically unwarranted; they are also empirically incorrect. Indeed, inferences like 4 and 8 are not even meaningful on a scientific understanding of speciation, let alone correct or incorrect.

Misconceptions of this nature have been documented both in the classroom (Baum, Smith, & Donovan, 2005; Meir, Perry, Herron, & Kingsolver, 2007) and in carefully controlled
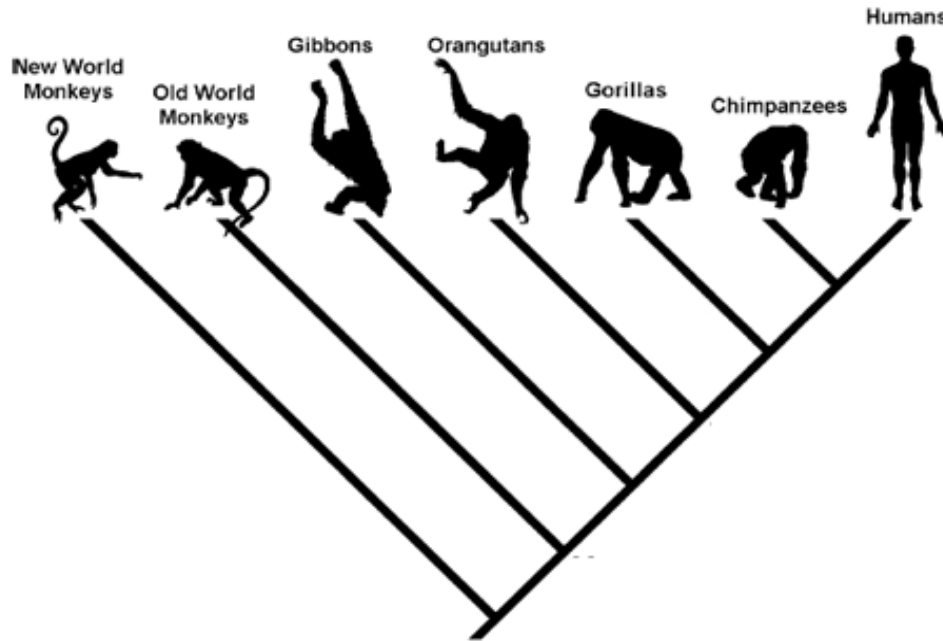
Figure 1. Cladogram depicting the phylogenetic relations among seven primates (adapted from Gregory, 2008).

laboratory studies (Catley, Novick, & Shade, 2010; Novick & Catley, 2007; Novick, Shade, & Catley, 2010). However, all such studies have involved college undergraduates, and no studies, to our knowledge, have explored the prevalence of macroevolutionary misconceptions in a non-college population (though see Berti, Toneatti, & Rosati, 2010, and Samarapungavan & Wiers, 1997, for research on children's misconceptions about other aspects of evolution). To address this gap in the literature, we explored the prevalence of macroevolutionary misconceptions among preschool- and elementary-school-aged children visiting a natural history museum with their parents. This sample served to broaden the scope of inquiry not only in terms of age but also in terms of context, as each child was interviewed as part of a dyad with his or her parent. In other words, children's reasoning about evolution was explored within the naturalistic context of a parent-child conversation.

Previous research on parent-child conversation suggests that conversations of this type can be a double-edged sword, with some aspects of the conversation scaffolding learning and other aspects obscuring, or even obstructing, learning. For instance, Jipson and Callanan (2003) found that parents of preschool-aged children typically use the word "grow" in a literal sense, to refer to biological changes in size (e.g., "the mushroom grew taller"), but also occasionally use the word in a metaphorical sense, to refer to non-biological changes in size (e.g., "the rock grew bigger"), yielding a linguistic signal that is reliable yet noisy nonetheless. In a similar vein, Rigney and Callanan (2011) found that parents at a marine science center ascribed biological properties to typical animals (e.g., sharks) no more often than they ascribed biological properties to atypical animals (e.g., anemones), potentially reinforcing the inclusion of atypical animals in the category of *living things*, a notoriously difficult concept to acquire (see, e.g., Anggoro, Waxman, & Medin, 2008). However, these same parents ascribed intentional states, like beliefs and desires, to both typical and atypical animals significantly more often than their children did, thus modeling a scientifically inappropriate form of reasoning. In short, parental input can serve as a source of accurate reasoning but by no means guarantees accurate reasoning (see also Gleason & Schauble, 2000; Gunderson & Levine, 2011).

In the present study, we explored how parents converse with their children about evolution, which we anticipated would be a difficult topic for both parties. We elicited these conversations by recruiting parent-child dyads from the floor of the Los Angeles Natural History Museum to complete a series of activities centered around an interactive cladogram. Two questions were of primary interest. First, how well do parent-child dyads interpret the information contained in cladograms, given that they are perhaps the most prevalent representation of evolutionary change in modern culture (Torrens & Barahona, 2012) yet are largely misunderstood by most biology students (Catley et al., 2010)? Research by Evans et al. (2010) and Spiegel et al. (2012) suggests that museum visitors hold a variety of preconceptions about *micro*-evolutionary change, some consistent with the principle of natural selection (e.g., need-based reasoning) and some inconsistent with it (e.g., creationist reasoning). Still, it remains an open question as to how museum visitors interpret displays representing *macro*-evolutionary phenomena, such as speciation, extinction, and common descent.

Second, what factors influence the accuracy of dyads' reasoning? Three candidate factors were identified from prior research on shared scientific thinking: (a) the child's gender, (b) the child's age, and (c) the dyad's overall style of interaction. In terms of gender, we predicted that dyads with male children would outperform dyads with female children, owing to the finding that parents are more likely to explain scientific phenomena to their sons than to their daughters (Crowley, Callanan, Tenenbaum, & Allen, 2001; Diamond, 1994) and might thus devote more attention to their sons in the activities at hand. In terms of age, we predicted that dyads with older children would outperform dyads with younger children, owing to the finding that older children are generally more familiar with evolutionary ideas than younger children (Berti et al., 2010; Legare, Lane, & Evans, in press; Spiegel et al., 2012) and might thus comprehend the purpose of the activities more thoroughly. Finally, in terms of interaction style, we predicted that dyads exhibiting higher levels of collaboration would outperform dyads exhibiting lower levels of collaboration, owing to the finding that parents at a science museum tend to hone their children's exploration of the exhibits in conceptually constructive ways (Crowley, Callanan, Jipson, Galco, Topping, & Shrager, 2001; Tare, French, Frazier, Diamond, & Evans, 2011), even if parents do occasionally provide conflicting or confusing input.

To preview our results, we found that parent-child dyads espoused the same kinds of misconceptions documented among college-level biology students. However, the frequency of such misconceptions varied by dyad, with dyads exhibiting low levels of collaboration espousing more misconceptions than those exhibiting higher levels of collaboration. This effect of dyad interaction was larger and more consistent than any of the other effects documented and thus has potentially important implications for both evolution education and informal science learning.

Method

Participants

The participants were 49 parent-child dyads recruited from the "Age of Mammals" exhibit at the Los Angeles Natural History Museum (see Figure 2A). We chose the Age of Mammals exhibit because it is thoroughly grounded in evolutionary findings and evolutionary principles and thus served as an ideal venue for eliciting conversations about evolution. All parents accompanying children between the ages of four and twelve were approached by
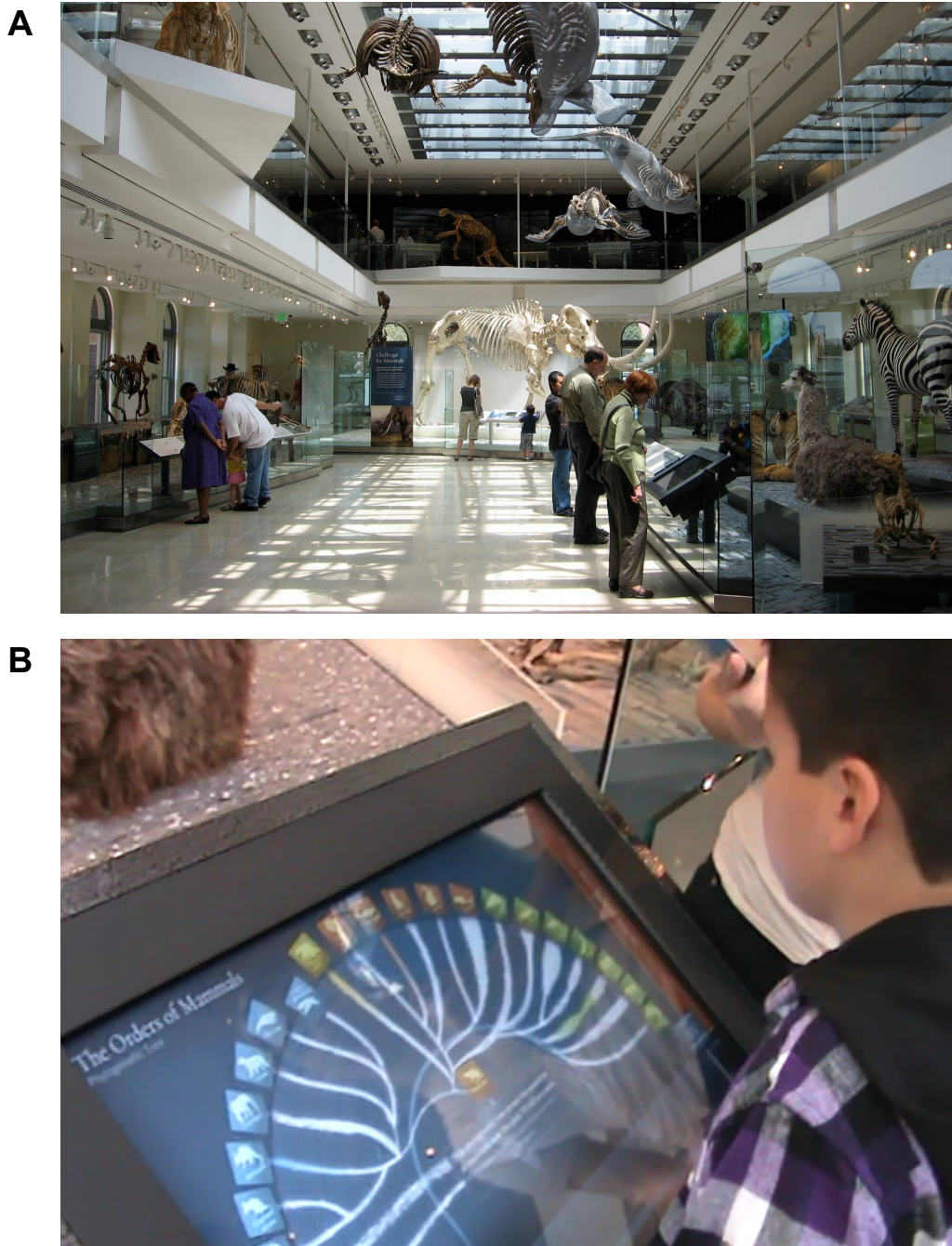
Figure 2. (A) The "Age of Mammals" exhibit at the Los Angeles Natural History Museum; (B) the interactive cladogram used to elicit participants' evolutionary reasoning.

the research team as they entered the exhibit and were invited to participate in a study. They were informed that the study consisted of how parents and children communicate about complex biological concepts, like evolution and common descent. Fifty-five parent-child dyads consented to participate, but only 49 completed the entire study. The ages of the participating children were evenly distributed across the age range sampled; half were between the ages of four and eight ($M = 6.8$, $SD = 1.3$, $n = 24$) and half were between the ages of nine and twelve ($M = 10.3$, $SD = 1.0$, $n = 25$). As for gender, 46% of the children were female, and 46% of the accompanying parents were female. Visitors who took part in the study did not receive any monetary compensation for their participation.

Procedure

The Age of Mammals exhibit features a total of 240 biological specimens, including both articulated skeletons of extinct species and taxidermied specimens of extant species. The specimens are organized into collections intended to illustrate three key principles: "Continents move. Climates change. Mammals evolve." These principles are also illustrated in the form of various touch-screen, computerized displays. One such display, an interactive cladogram, comprised the focus of the present study (see Figure 2B). All parent-child dyads who consented to participate were taken directly to the interactive cladogram, briefly familiarized with its features and functions, and then led through a series of five activities exploring the information contained within.

Technically, the display was not a "cladogram," in the biologist's sense of the word, because it attempted to represent more than just patterns of common ancestry; it also represented the times that taxa diverged and whether the taxa are extant or extinct. Nevertheless, we will refer to the display as a cladogram because its *primary* function was to depict patterns of common ancestry. Five display-based activities were designed to probe participants' understanding of macroevolution and the ways in which macroevolutionary relations are represented in cladograms. These activities, which were derived from previous research on students' macroevolutionary misconceptions (e.g., Catley et al., 2010; Gregory, 2008), are described below in relation to participants' actual responses. All responses were video recorded and transcribed at a later date. At the completion of the five-activity interview, participants were encouraged to explore other areas of the exhibit, but their conversations at those other areas were not recorded.

Coding

Measures of engagement. Each interview was transcribed by two independent coders, the first producing a written record of all utterances and actions and the second editing or embellishing that record to account for details the first coder appeared to have missed. The coders then summed the number of distinct utterances and actions for each dyad member and each activity, resulting in 490 engagement scores (5 scores per participant for 98 participants). Any self-contained thought, question, or response was coded as a distinct utterance, even if that thought did not take the form of a complete sentence (e.g., "yes" or "OK" were counted as distinct utterances). Likewise, any attempt to deliberately manipulate the display (via tapping or scrolling), or manipulate a partner's view of the display (via pointing or waving), was coded as a distinct action. Extended sequences of dialogue or activity were broken into distinct utterances and actions, with the exception of experimenter requested actions (e.g., "point to the monkey, the tree shrew, and the flying lemur"), which were counted simply as one action so as not to inflate engagement scores for some activities relative to others. Coders agreed on their tabulation of distinct utterances and actions 93% of the time, and all disagreements were resolved through discussion.

Measures of accuracy. Participants' responses to each activity were assigned a score that ranged from -1 to +2. Responses that revealed a positive misconception about the material at hand received a score of -1. Responses that were vague, ambiguous, or equivocal (including "don't know" responses) received a score of 0. Responses that revealed a partial understanding of the material at hand received a score of +1. And responses that revealed a full understanding of the material at hand received a score of +2. What constituted a correct response or an incorrect response is discussed below in relation to the corresponding activity. It should be noted that each dyad received a single score per activity, rather than separate scores for each dyad member, as almost all dyads offered a single response by the

conclusion of almost all activities. Two coders independently assigned scores to all 245 responses (5 responses per dyad for 49 dyads). Overall agreement was 82% (Cohen's κ = .76), and all disagreements were resolved through discussion.

Results

How well do parent-child dyads reason about evolutionary phenomena?

Below we describe each activity used to elicit participants' evolutionary reasoning and the nature of their responses, followed by an analysis of participants' engagement with each activity and how that engagement was related to response accuracy per activity (summed across dyads). We address the question of whether, and how, engagement influenced response accuracy per dyad (summed across activities) in the following section.

Activity 1: Ordering. The introductory screen of the interactive cladogram featured all 19 orders of mammals, arranged in a semicircle with primates in the center (see Figure 2B). The first activity was designed to elicit participants' beliefs about the necessity, and potential flexibility, of this particular ordering. As mentioned previously, the ordering of the taxa in a cladogram is, to a large extent, arbitrary. While taxa that share a most recent common ancestor must be adjacent (e.g., chimpanzees and humans), their ordering relative to one another is arbitrary (i.e., chimpanzees can be on the left and humans on the right or humans can be on the left and chimpanzees on the right). The representation of nested taxa is governed by the same constraint, meaning that entire groups of taxa can be swapped with one another so long as the underlying branching relations are preserved. Thus, any one taxon could appear at any point in the row of terminal nodes, and any cluster of taxa could be reordered in many different ways.

We attempted to elicit participants' understanding of the ordinal relations among taxa by asking them to locate three particular taxa – the monkey, the tree shrew, and the flying lemur – and reflect on the ordering of those taxa. Specifically, we asked, "Does it matter that the flying lemur is on the left, the tree shrew is in the middle, and the monkey is on the right? Or could they be reordered so that the monkey is on the left, the flying lemur is in the middle, and the tree shrew is on the right? Why or why not?"

Because lemurs and shrews share a common ancestor with each other more recently than either shares with monkeys (as depicted by the relevant branching relations), the only constraint on ordering was that the lemur and the shrew had to be adjacent, which was true of the hypothetical ordering we asked participants to consider. Nevertheless, 17 dyads claimed that the taxa could not be reordered, justifying their judgment with an affirmation that ordering matters (scored -1). Most dyads (*n* = 23) were unsure whether or not the taxa could be reordered (scored 0), and only a few dyads (*n* = 9) claimed the taxa could be reordered but were unable to provide an adequate justification for their judgment (scored +1). None of the dyads provided both a correct judgment (that the taxa could be reordered) and a correct justification (that only the branching relations matter), and thus none received a score of +2. On the contrary, dyads that received a score of 0 or 1 seemed genuinely unsure of whether, and how, the ordinal properties of the display reflected information about the species' evolutionary origins.

Activity 2: Branching. Participants' understanding of the branching relations in a cladogram was elicited more directly in the second activity. Participants were asked to locate two non-placental mammals – the kangaroo and the platypus – and to read about their features in a pop-up window that appeared upon touching each. We chose non-placental mammals as our target taxa for this activity because their divergence from the other mammals occurred earliest (around 170 million years ago for egg-laying mammals and 130 million years ago for

marsupials) and was thus highly salient. After participants had read about the features of kangaroos and platypuses, they were asked, "How are these two mammals different from the other mammals in the tree? Is this difference reflected in the tree itself somehow?"

Most dyads were able to identify a genuine morphological or geographic difference between the non-placental mammals and the other mammals (e.g., "only kangaroos have pouches and only platypuses lay eggs," "they're the only ones that live in Australia"), but very few were able to identify how that difference was reflected in the tree. Six dyads claimed the difference was *not* reflected in the tree (scored -1); 15 dyads claimed they were unsure whether or not the difference was reflected in the tree (scored 0); and 23 dyads claimed the difference was reflected in the tree but did not refer to the branching relations, e.g., "they [the labels] are different colors" or "they [the animals] just look different" (scored +1). Only 5 dyads correctly identified the branching relations as the relevant form of representation, e.g., "this one's branching out, completely separate from these" (scored +2).

Activity 3: Speciation. One of the unique features of the interactive cladogram was a slider at the bottom of the screen for manipulating the timeline, allowing users to scroll between the beginning of the divergence of the 19 orders of mammals (65 million years ago) and the present day. Moving back in time "shrank" the cladogram such that branching events that occurred after that time no longer appeared on the screen. We used this feature of the display to elicit participants' beliefs about the origin of species. Specifically, we asked participants to move the slider to 40 million years, which caused the taxon representing an extinct, hippo-like creature – the paleoparadoxiid – to disappear from the screen. We then asked, "Did you see that the paleoparadoxiid disappeared from the tree? Why do you think that happened? What might have occurred between 40 million years ago and 30 million years ago that led to the appearance of paleoparadoxiid?"

Of interest was whether participants could identify a biologically plausible cause of divergence – i.e., geographic isolation, reproductive isolation, or unique selection pressures. Five dyads did, in fact, cite such a factor, e.g., "maybe the climate changed" (+2). Fourteen dyads noted that the paleoparadoxiid must have evolved during the time period of interest but were unclear on what factors may have driven its divergence from its closest relative (scored +1). The majority of dyads (*n* = 21) were unsure of what the appearance and disappearance of the paleoparadoxiid icon was supposed to represent biologically (scored 0), and the remaining dyads (*n* = 9) acknowledged that the appearance and disappearance of the paleoparadoxiid icon was intended to represent a speciation event but cited biological implausible means of speciation, e.g., "the sea cow and the hyrax had a baby" or "it came from the ground" (scored -2).

Activity 4: Relatedness. Another unique feature of the interactive cladogram was that users could "launch the story" of a particular taxon, which opened a window containing detailed information about the taxon's habitat, diet, and morphology. Also contained in these windows were additional cladograms, depicting the relations among other species within the taxa not explicitly depicted in the main display. We used this feature as a vehicle for eliciting participants' understanding of the relatedness of species whose morphological features seemingly belie their evolutionary origins: camels, pigs, and whales. Participants were asked to "launch the story" for the camels, which then brought up a cladogram depicting pigs on the left, camels in the middle, and whales on the right. The branching relations among these taxa indicated that camels are more closely related to whales than to pigs; the outward behavior and morphology of these animals, however, would suggest that camels are more closely related to pigs. Faced with this tension, participants were asked to determine whether camels were more closely related to pigs or to whales and to explain how

they were able to discern that fact from the tree. The story itself, we should note, did not explicitly reference this tension, nor did it give participants any clues as to which of the three species were most closely related.

The vast majority of dyads (*n* = 39) claimed that the tree indicated that camels are more closely related to pigs than to whales (scored -1). Three dyads claimed not to be able to discern what the tree indicated about relatedness (scored 0); four claimed that the tree indicated that camels are more closely related to whales but were unable to explain how it indicated that (scored +1); and three claimed the tree indicated that camels are more closely related to whales and were also able to justify that inference by reference to the branching relations (scored +2).

Activity 5: Extinction. The interactive cladogram contained another feature not represented in standard cladograms: it allowed participants to select images of the specimens on view next to the display and learn about those specimens in the context of the phylogenic relations and historical timeline contained within. We used this feature to elicit participants' reasoning about the relation between extinct species and extant species. Participants were asked to navigate to the "Mammals on Display" tab, select the "entelodont" (an extinct, pig-like creature), and decide (a) whether it still exists and (b) where it might fit in the cladogram in the main display. The entelodont's status as an extinct species could be discerned from the sliding timeline at the bottom of the screen, which made the entelodont non-selectable if moved to a point past its estimated date of extinction. It could also be discerned from the information contained in the pop-up window describing the species. Thus, the question of whether or not the entelodont is extinct was of less interest than the question of where it might fit in the cladogram, which depicted 19 extant species but only one extinct species (the paleoparadoxiid).

As expected, virtually all dyads were able to discern that the entelodont is extinct. Only 19 dyads, however, were able to discern that it would be located among the ungulates (scored +2). Twenty-two dyads indicated that the entelodont would be located in some other part of the display, typically the bottom, near the cladogram's root (scored +1); 6 dyads were unsure of where the entelodont would be located (scored 0); and 2 dyads claimed that the entelodont would *not* fit into the cladogram because it was extinct and extinct species cannot be represented in cladograms (scored -1).

Patterns of engagement. Participants' engagement with each activity was operationalized as the number of distinct utterances and actions produced during that activity. Engagement scores were calculated separately for each dyad member and each activity. To determine whether engagement scores varied by activity or by dyad composition, we ran a repeated-measures analysis of variance (ANOVA) of the engagement scores in which activity (ordering, branching, speciation, relatedness, extinction) and dyad member (parent, child) were treated as within-participants variables and parent's gender (male, female), child's gender (male, female), and child's age (4-8, 9-12) were treated as between-participants variables. This analysis revealed significant effects of activity ($F(4,164) = 5.08$, $p < .01$) and dyad member ($F(1,41) = 30.74$, $p < .001$), but no significant effects of parent's gender, child's gender, or child's age and no significant interactions either.

The effects of activity and dyad member on engagement scores are illustrated in Figure 3A. While dyads produced more utterances and actions for some activities (e.g., branching) than for others (e.g., extinction), this effect was small in comparison to the effect of dyad member. Across activities, the mean engagement score for children was 4.6 (*SD* = 1.4, range = 1.2 to 8.8), whereas the mean engagement score for parents was only 1.8 (*SD* = 2.5, range = 0 to 11.6). In fact, 27 of the 49 parents produced fewer than one utterance or action per activity,

despite the fact that the interview was framed, in both verbal and written communications, as "a study of how parents and children converse about complex biological topics, like evolution and common descent." The remaining 22 parents, however, tended to produce as many utterances and actions as their children, if not more. This variance, while unexpected, ultimately proved fruitful in determining how engagement scores related to response accuracy at the level of the dyad (discussed subsequently).
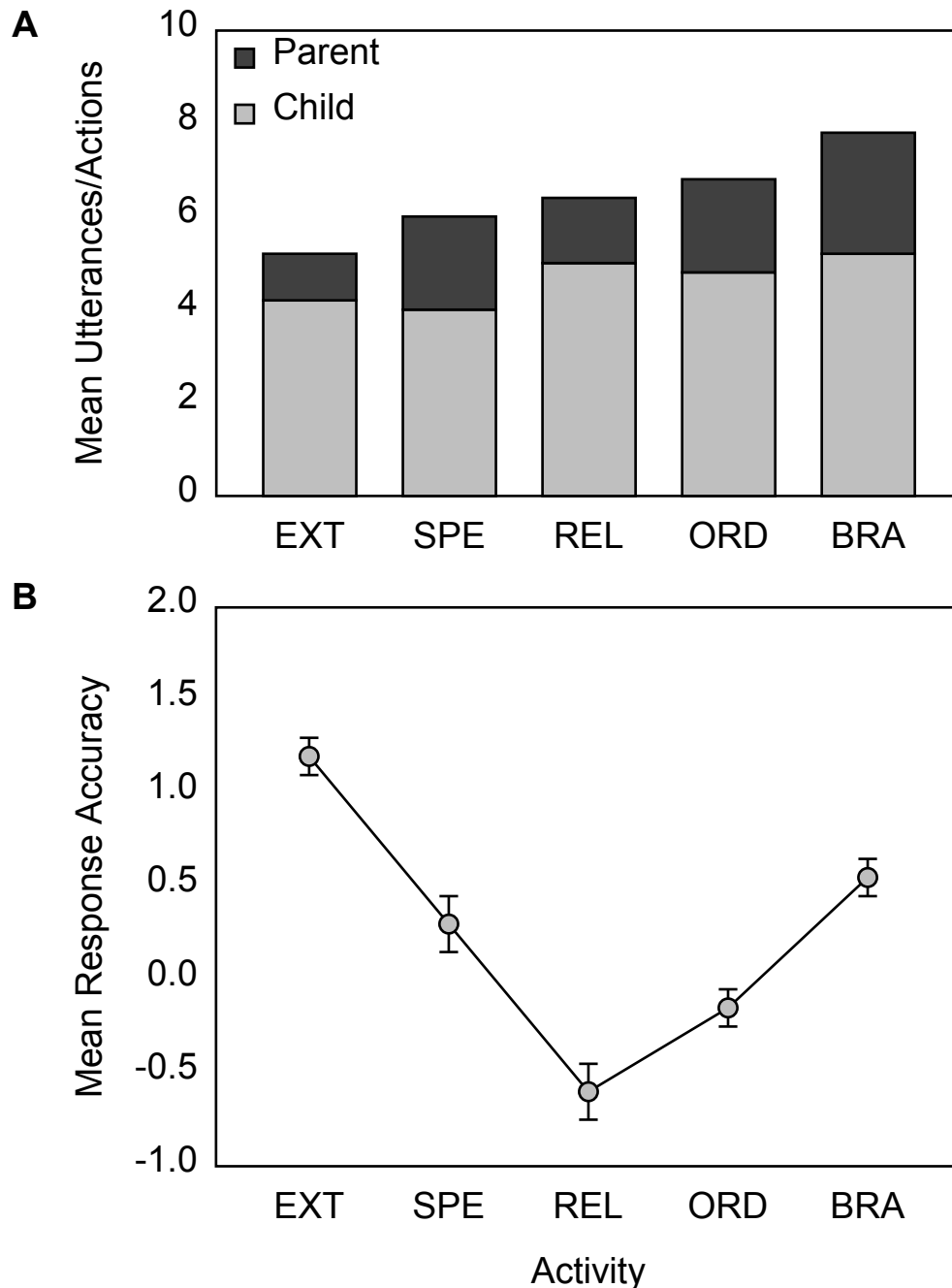

Figure 3. (A) Mean number of utterances/actions per activity per dyad member; (B) Mean accuracy of dyad responses per activity (+ *SE*). EXT = Extinction, SPE = Speciation, REL = Relatedness, ORD = Ordering, BRA = Branching.

Patterns of reasoning. Participants' mean accuracy on each activity is displayed in Figure 3B, ordered by the mean level of engagement with those activities. A repeated-measures

ANOVA revealed not only that accuracy scores differed by activity ($F(4,192) = 34.43$, $p < .01$) but also that accuracy scores were quadratically related to mean engagement scores ($F(1,48) = 96.41$, $p < .001$, for the quadratic contrast). The source of this quadratic relation is not entirely clear, though one possibility is that the association between accuracy scores and engagement scores may have been moderated by a third variable – task difficulty – such that engagement was negatively correlated with response accuracy for easy tasks but positively correlated with response accuracy for harder tasks. Put differently, increased engagement may have been productive for the more difficult tasks but counterproductive for the easier ones.

We did not, however, collect any independent measures of task difficulty, so this speculation requires further verification. Nevertheless, the two activities that differed most in engagement – activity 5 (extinction) and activity 2 (branching) – yielded opposite correlations between accuracy scores and engagement scores at the level of the dyad. Dyads who exhibited more engagement with the branching activity tended to produce more accurate responses ($r = .27$), whereas dyads who exhibited more engagement with the extinction activity tended to produce less accurate responses ($r = -.20$). These correlations were not significantly different from zero (given the small sample size), but they were still significantly different from one another ($z = 3.25$, $p < .01$), suggesting that the relation between engagement and accuracy may differ depending on the conceptual demands of the task.

What factors influenced accuracy of reasoning?

Dyads differed substantially in the accuracy of their reasoning, from providing responses scored -1 on four of the five activities to providing responses scored +2 on four of the five activities. To determine which factors, if any, were associated with accuracy, we regressed each dyad's accuracy score against five dyad-specific measures: the child's age (in years), the child's gender (0 = female, 1 = male), the parent's gender (0 = female, 1 = male), the parent's mean engagement score, and the child's mean engagement score. Only two measures emerged as significant predictors of response accuracy: the child's age ($\beta = .32$, $t = 2.39$, $p < .05$) and the parent's mean engagement score ($\beta = .36$, $t = 2.61$, $p < .05$). These effects are displayed in Figure 4, with child's age dichotomized as "younger" (4-8) and "older" (9-12) and parent's engagement scores dichotomized as "low" ($M < 1$ utterance or action per activity) and "high" ($M > 1$ utterance or action per activity). Dyads with older children produced more accurate responses than dyads with younger children, and dyads with high parental engagement scores produced more accurate responses than dyads with low parental engagement scores. A univariate ANOVA revealed no interaction between child's age and parental engagement ($F(1,45) < 1$, ns), though it did reveal that the effect of parental engagement ($\eta^2 = .13$) was nearly three times as large as the effect of child's age ($\eta^2 = .05$).

To explore the effect of parental engagement further, we computed an additional measure of engagement: the number of alternations between the child's contribution to the conversation (utterance or action) and the parent's contribution. While this measure was strongly correlated with our initial measure of parental engagement ($r = .74$), it was not entirely redundant with that measure. Figure 5 illustrates how the two measures diverged. The transcript shown in 5A contains multiple alternations between parent and child, with parent A actively eliciting information from the child, either from the display or from memory. The transcript shown in 5B, on the other hand, contains far fewer alternations, with parent B generally dominating the conversation. While parents in both dyads produced a
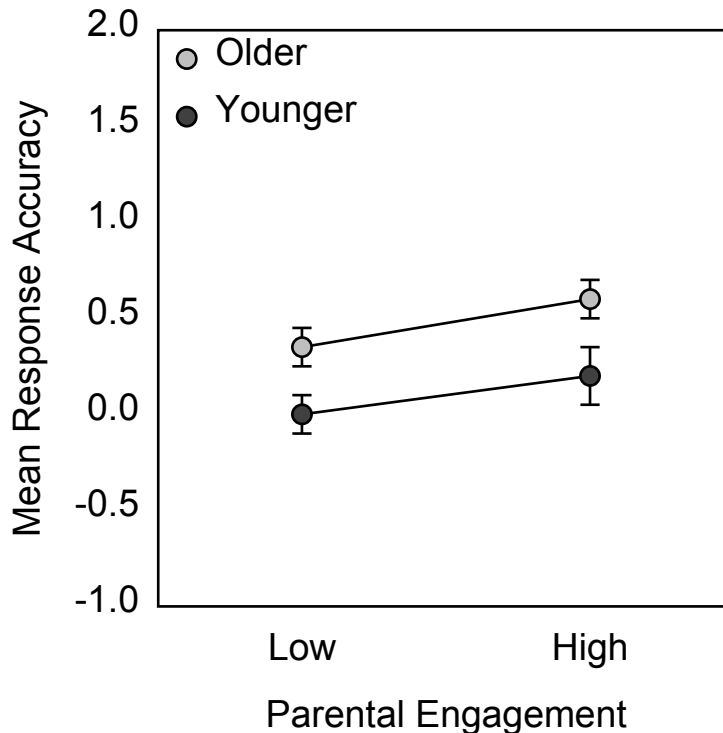
Figure 4. Mean accuracy of dyad responses (+ *SE*) as a function of parental engagement (low vs. high) and child's age (younger vs. older).

similar number of utterances or actions per activity (*M* for parent A = 5.0, *M* for parent B = 4.8), the two dyads received very different alteration scores (*M* for dyad A = 4.8, *M* for dyad B = 1.8). It should be noted that dyads A and B were two of only 33 dyads for which alternation scores could be computed at all. The remaining 16 dyads included too little parental engagement, typically because the parents in these dyads resigned themselves to simply watching their children complete the activities. It should also be noted that the overall paucity of parental involvement rendered the use of more detailed, content-specific coding schemes (like those used by Crowley et al., 2001, or Evans et al., 2010) impractical.

Among the 33 dyads for which alternations could be computed, the mean number of alternations was 14.2 (*SD* = 10.7, range = 2 to 44), and alternations were strongly correlated with response accuracy (*r* = .55). Alternations actually proved to be a stronger predictor of response accuracy than parental engagement in general (*r* = .35). But were alternations a *unique* predictor of response accuracy? We addressed this question by submitting dyads' mean accuracy scores to a hierarchical regression in which alternation scores were entered subsequent to parental engagement scores. Whereas parental engagement scores explained 14% of the variance in response accuracy in the initial model (*F*(1,31) = 5.00, *p* < .05), alternation scores explained an additional 17% of variance in the subsequent model (*F-change*(2,30) = 6.69, *p* < .01). In fact, the partial correlation between alternation scores and response accuracy, controlling for parental engagement scores, was nearly as large as the zero-order correlation between these two variables (*r* = .44, *p* < .01), whereas the partial correlation between parental engagement scores and response accuracy, controlling for alternation scores, was no longer significant (*r* = .06, *ns*). The *nature* of parental engagement thus appeared to be more important than the mere act of engagement.

A

E: Select the kangaroo and the platypus and read about their features.

P: Can you find the kangaroo?
C: [Hits the kangaroo]
P: [Reads it to her] OK, now the platypus. You keep looking over it.
C: [Hits the kangaroo]
P: [Reads it to her]

E: How are these two mammals different from the other mammals in the tree?

P: Where do their babies grow?
C: In their pouch.
P: Right, what about the platypus?
C: I don't know.
P: Remember it lays eggs, now where do they come from?
C: Australia and New Guinea.

E: Is this difference reflected in the tree itself somehow?
P: By the colors.

B

E: Select the kangaroo and the platypus and read about their features.

P: Kangaroo and platypus.
C: [Leans on display looks around]
P: OK, here's the kangaroo and the platypus.
C: [Reads them]
P: OK, now this one, most other mammals don't lay eggs, but the platypus is a mammal that lays eggs. That's a big deal, will you remember that?

E: How are these two mammals different from the other mammals in the tree?

P: Well, the platypus lays eggs and live in the water, and the kangaroo has a pouch and keeps their young in the pouch.

E: Is this difference reflected in the tree itself somehow?

P: Well, they are all mammals, but [points to tree, sweeping motion]. Hmm. It's obvious about the platypus...

Figure 5: Examples of high parental engagement involving (a) frequent alternations and (b) infrequent alternations between parent and child contributions. E = Experimenter, P = Parent, and C = Child. Actions are denoted in brackets.

Conclusions

Evolutionary concepts like *speciation*, *extinction*, and *common descent* are notoriously difficult to understand (Gregory, 2008; Shtulman, 2006), and canonical representations of those concepts are notoriously difficult to interpret (Catley et al., 2010; Novick et al., 2010). The

present study sought to determine how parents and children discuss such concepts and interpret such representations in the context of an interactive museum display. We found that parents and children, like the college-aged students tested in previous research, exhibited significant difficulty in interpreting the core feature of a cladogram – namely, the branching relations among its taxa. Parents and children also revealed the same kinds of evolutionary misconceptions documented in previous research, including (a) that the ordering of a species in a cladogram carries biologically relevant information, (b) that the morphological overlap between species is a reliable indicator of shared ancestry, and (c) that speciation occurs through a kind of instantaneous transformation or emergence. Nevertheless, the frequency of those misconceptions was negatively correlated with the degree to which parents were involved in generating a response, particularly the degree of turn-taking between parents and children. The more often the two collaborated (by this measure), the more often they generated accurate interpretations and explanations of the phenomena at hand.

These findings have implications for both the study of evolutionary reasoning and the design of informal learning environments. With respect to evolutionary reasoning, they suggest that collaboration may be an effective means of reducing, or even eliminating, evolutionary misconceptions. Previous research by Asterhan and Schwartz (2007, 2008) found that undergraduates who answered evolution-based problems on their own learned less about evolution, in both the short-term and the long-term, than those who worked in pairs. They also found that the style of a dyad's interaction influenced learning such that dyads who engaged in argumentation exhibited greater learning gains than those who merely shared information. Assuming that turn-taking between parents and children served as a proxy for argumentation, the present study not only replicates Asterhan and Schwartz's findings but also extends those findings to (a) populations of different ages and (b) conversations elicited in more naturalistic contexts.

That said, the studies by Asterhan and Schwartz employed controls absent from the present study, including the provision of instruction on how to conduct a critical discussion and the administration of pre- and post-intervention measures of evolution understanding. Controls like these are difficult to implement in a museum setting, where participants' time and attention are limited resources. It may thus be beneficial to test parent-child dyads in a more controlled (laboratory) setting, particularly for the purpose of assessing *learning*. While we did, in fact, document a strong relation between the level of interaction between dyad members and the accuracy of their responses, it is unclear whether greater dyadic interaction was a cause, or merely a symptom, of more accurate responding. Dyads who entered the museum with better evolution understanding may have engaged in more discussion simply because they were more familiar, or more comfortable, with the topics at hand.

At least two considerations militate against this interpretation, however. First, not all forms of engagement were associated with higher response accuracy. Children's engagement scores, for instance, were uncorrelated with response accuracy, as were parental engagement scores after controlling for parent-child alternations. Second, engagement was positively correlated with response accuracy for some activities but negatively correlated with others. Still, future research should explore the effects of parent-child interaction on learning more directly by including both pretests and posttests and by administering them (separately) to both parents and children.

With respect to informal learning environments, our findings suggest that environments that encourage dyadic interaction – particularly parent-child interaction – may be more effective

at fostering learning than those that encourage more monadic forms of exploration. These findings echo many other findings in the science education literature. Van Schijndel, Franse, and Raijmakers (2010), for instance, found that children who were actively coached by their parents during a museum visit engaged more effectively with the exhibits than children who were not coached in this way. Tenenbaum, Prior, Dowling, and Frost's (2010) found that parent-child dyads who explored a museum with a booklet of activities spent more time exploring and discussing the exhibits than dyads who explored the museum in a less structured way. And Tare et al. (2011) found that parents' explanatory behavior at a museum exhibit was positively correlated with their children's explanatory behavior. Two aspects of our findings, however, stand in contrast to these earlier findings.

First, the relation between dyads' level of engagement and accuracy of reasoning was not entirely straightforward. Dyads reasoned most accurately for activities in which they were either *most* engaged or *least* engaged; intermediate levels of engagement were, on the other hand, associated with the poorest performance (see Figure 3). Whatever the cause of this effect may be, the effect itself implies that increased engagement is not always productive or beneficial and that learners may need additional guidance in how to allocate their attention to the displays at hand (see, e.g., Allen & Gutwill, 2004).

Second, we did not explicitly instruct parents on how to engage with their children during the activities, leaving open the option that parents might chose not to engage at all, and, to our surprise, approximately half adopted that option. Anecdotal evidence suggests that what led to such high levels of parental attrition was not confusion about the purpose of the study but the nature of the touch-screen display, which was clearly designed for a single user. If two users touched the screen at the same time, then either the cursor on the display would toggle between the two points of contact or the display would reset itself altogether, closing all pop-up windows and restoring all default settings (e.g., restoring the sliding timescale to "present day"). Dyads quickly became aware of this contingency and delegated the task of manipulating the display to a single individual, typically the child. As a result, many parents seemed to become increasingly disengaged over the course of the interview, an observation backed by a significant drop in parental engagement scores from the first activity to the last (M = 1.9, 2.6, 2.0, 1.4, 1.0; linear contrast: $F(1,48)$ = 12.33, $p < .01$). This unexpected finding highlights a paradox in the use of touch-screen displays increasingly populating the halls of science museums today: such displays are typically designed for a single user, yet, in our study, dyads in which a single member was actively engaged with the display profited less from the display than those in which both members were actively engaged.

The conceptual benefits of collaborative activity over solitary activity have been documented in a variety of studies (e.g., Craig, Chi, & VanLehn, 2009; Okada & Simon, 1997, Schwarz, Neuman, & Biezuner, 2000). The mechanism behind this effect appears to be the addition of a "social impetus" to explain and justify one's reasoning. Museum displays that can capture or create this kind of impetus would thus seem to be more efficacious than those that cannot. That said, the design of such displays is constrained by additional, pragmatic factors that might effectively limit interactivity, including how long visitors can be expected to use the display, how well visitors can discern the purpose of the display's affordances, and how constructively visitors can intervene on the phenomena of interest (Allen & Gutwill, 2004). Which considerations to privilege over others is a question that likely merits a different answer for every display. Still, our research suggests that, at a minimum, touch-screen displays should be tolerant of multiple points of contact so that parents and children may jointly interact with the display without accidentally erasing one another's paths of exploration.

In conclusion, visitors to a science museum are no more immune to evolutionary misconceptions than other populations. Science museums, however, provide unique opportunities for collaboratively discussing evolutionary phenomena, and such discussions appear to help attenuate evolutionary misconceptions. While further research needs to be done on how parent-child conversations foster accurate evolutionary reasoning and whether such conversations lead to long-term learning, the current findings point to a promising new method for increasing evolution understanding among the general public.

● ● ●

Andrew SHTULMAN. Received a Ph.D. in Psychology from Harvard University in 2006. He has been a member of the Department of Psychology at Occidental College since 2007. His research focuses on conceptual development and conceptual change, particularly as they relate to science education.

Isabel CHECA. Is an undergraduate at Occidental College, anticipated to graduate in 2013. She has been a member of Dr. Shtulman's lab since 2010 and is interested in early childhood education.

References

Allen, S., & Gutwill, J. P. (2004). Designing with multiple interactives: Five common pitfalls. Curator, 47, 199-212.

Anggoro, F. K., Waxman, S. R., & Medin, D. L. (2008). Naming practices and the acquisition of key biological concepts: Evidence from English and Indonesian. Psychological Science 19, 314-319.

Asterhan, C. S. C., & Schwartz, B. B. (2007). The effects of monological and dialogical argumentation on concept learning in evolutionary theory. Journal of Educational Psychology, 99, 626-639.

Asterhan, C. S. C., & Schwartz, B. B. (2009). Argumentation and explanation in conceptual change: Indications from protocol analyses of peer-to-peer dialog. Cognitive Science, 33, 374-400.

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

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.

Catley, K. M., & Novick, L. R. (2008). Seeing the wood for the trees: An analysis of evolutionary diagrams in biology textbooks. BioScience, 58, 976-987.

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

Craig, S. D., Chi, M. T. H., & VanLehn, K. (2009). Improving classroom learning by collaboratively observing human tutoring videos while problem solving. Journal of Educational Psychology, 101, 779-789.

Crowley, K., Callanan, M., Jipson, J., Galco, J., Topping, K., & Shrager, J. (2001). Shared scientific thinking in everyday parent-child activity. Science Education, 85, 712-732.

Crowley, K., Callanan, M.A., Tenenbaum, H. R., & Allen, E. (2001). Parents explain more often to boys than to girls during shared scientific thinking. Psychological Science, 12, 258-261.

Diamond, J. (1994). Sex differences in science museums: A review. Curator, 37, 17-24.

Jipson, J. L., & Callanan, M. (2003). Mother-child conversation and children's understanding of biological and non-biological changes in size. Child Development, 74, 629-644.

Evans, E. M., Spiegel, A., Gram, W., Frazier, B. F., 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. Oxford, UK: 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.

Gleason, M. E., & Schauble, L. (2000). Parents' assistance of their children's scientific reasoning. Cognition and Instruction, 17, 343-378.

Gunderson, E. A., & Levine, S. C. (2011). Some types of parent number talk count more than others: Relations between parents' input and children's cardinal-number knowledge. Developmental Science, 14, 1021-1032.

Gregory, T. R. (2008). Understanding evolutionary trees. Evolution: Education and Outreach, 1, 121-137.

Gregory, T. R. (2009). Understanding natural selection: Essential concepts and common misconceptions. Evolution: Education and Outreach, 2, 156-175.

Griffith, J. A., & Brem, S. K. (2004). Teaching evolutionary biology: pressures, stress, and coping. Journal of Research in Science Teaching, 41, 791-809.

Hatano, G., & Inagaki, K. (1994). Young children's naïve theory of biology. Cognition, 50, 171-188.

Legare, C. H., Lane, J., & Evans, E. M. (in press). Anthropomorphizing science: How does it affect the development of evolutionary concepts? Merrill-Palmer Quarterly.

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

Meir, E., Perry, J, Herron, J. C., & Kingsolver, J. (2007). College students' misconceptions about evolutionary trees. American Biology Teacher, 69, 71-76.

Miller, J. D., Scott, E. C., & Okamoto, S. (2006). Public acceptance of evolution. Science, 313, 765-766.

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

National Research Council (1998). Teaching about evolution and the nature of science. Washington, DC: National Academy Press.

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

Newport, F. (2010). Four in 10 Americans believe in strict creationism. Gallup Organization.

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

Novick, L. R., Shade, C. K., & Catley, K. M. (2010). Linear versus branching depictions of evolutionary history: Implications for design. Topics in Cognitive Science, 3, 536-559.

Okada, T., & Simon, H. (1997). Collaborative discovery in a scientific domain. Cognitive Science, 21, 109-146.

Rigney, J. C., & Callanan, M. A. (2011). Patterns in parent-child conversations about animals at a marine science center. Cognitive Development, 26, 155-171.

Samarapungavan, A. & Wiers, R. W. (1997). Children's thoughts on the origin of species: A study of explanatory coherence. Cognitive Science, 21, 147-177.

Schwarz, B. B., Neuman, Y., & Biezuner, S. (2000). Two wrongs may make a right...if they argue together! Cognition & Instruction, 18, 461-494.

Shtulman, A. (2006). Qualitative differences between naive 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, UK: Oxford University Press.

Shtulman, A., & Calabi, P. (in press). Tuition vs. intuition: Effects of instruction on naive theories of evolution. Merrill-Palmer Quarterly.

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

Sinatra, G. M., Brem, S. K., & Evans, E. M. (2008). Changing minds? Implications of conceptual change for teaching and learning about biological evolution. Evolution: Education and Outreach, 1, 189-195.

Solomon, G. E. A., & Zaitchik, D. (2012). Folkbiology. WIREs Cognitive Science, 3, 105-115.

Southerland, S. A., Abrams, E., Cummins, C. L., & Anzelmo, J. (2001). Understanding students' explanations of biological phenomena: Conceptual frameworks or p-prims? Science Education, 85, 328-348.

Spiegel, A., Evans, E. M., Frazier, B. F., Hazel, A., Tare, M., Gram, W., & Diamond, J. (2012). Changing museum visitors' concepts of evolution. Evolution: Education and Outreach, 5, 43-61.

Tare, M., French, J., Frazier, B., Diamond, J., & Evans, E. M. (2011). Explanatory parent-child conversation predominates at an evolution exhibit. Science Education, 95, 720-744.

Tenenbaum, H. R., Prior, J., Dowling, C. L., & Frost, R. E. (2010) Supporting parent-child conversations in a history museum. British Journal of Educational Psychology, 80, 241-254.

Torrens, E., & Barahona, A. (2012). Why are some evolutionary trees in natural history museums prone to being misinterpreted? Evolution: Education and Outreach, 5, 76-100.

van Schijndel, T. J. P., Franse, R. K., & Raijmakers, M. E. J. (2010). The exploratory behavior scale: assessing young visitors hands-on behavior in science museums. Science Education, 94, 794-809.

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.

www.iejee.com

This page is intentionally left blank