# Bridging a Conceptual Divide: How Peer Collaboration Facilitates Science Learning

Proceedings of the Annual Meeting of the Cognitive Science Society, 39(0)

Shtulman, Andrew
Young, Andrew

2017

# Bridging a Conceptual Divide: How Peer Collaboration Facilitates Science Learning

**Andrew Shtulman (shtulman@oxy.edu)**
**Andrew Young (ayoung2@oxy.edu)**
Department of Psychology, Occidental College
1600 Campus Road, Los Angeles, CA 90041

## Abstract

Collaboration is generally an effective means of learning new information, but is collaboration productive in domains where collaborators may hold qualitatively different conceptions of the domain's causal structure? We explored this question in the domain of evolutionary biology, where previous research has shown that most individuals construe evolution as the uniform transformation of an entire population (akin to metamorphosis) rather than the selective survival and reproduction of a subset of the population. College undergraduates (*n* = 44) completed an assessment of their evolutionary reasoning by themselves (pretest), with a partner (dyad test), and several weeks later (posttest). Collaboration proved ineffective for the higher-scoring partner in each dyad, as their scores generally remained unchanged from pretest to dyad test to posttest, but it proved effective for the lower-scoring partner. Not only did lower-scoring partners increase their score from pretest to dyad test, but they maintained higher scores at posttest as well. Follow-up analyses revealed that participants' posttest scores were predicted by their partners' pretest scores but only for lower-scoring partners, and the relation was negative: the smaller the difference between pretest score, the greater the gain from pretest to posttest for lower-scoring partners. These findings indicate that collaboration in domains characterized by conceptual change is possible, but that learning from such collaboration is asymmetric (i.e., individuals with low levels of understanding benefit more than their partners do) and unequal (i.e., individuals with low levels of understanding benefit more if their partner's understanding is only moderately higher). Thus, bridging the gap between a novice's view of a conceptually complex domain and an expert's view appears to require instruction more aligned with the former than the latter.

**Keywords:** collaboration, conceptual development, science learning, intuitive theories, evolutionary reasoning

## Introduction

Some ideas are more difficult to learn than others. Ideas that can be encoded in terms of preexisting concepts, like the name of an unfamiliar animal or the function of an unfamiliar artifact, are much easier to learn than ideas that require new concepts for their encoding, like the reason the seasons change or the reason projectiles fall to the ground in a parabolic path. Learning the latter requires *conceptual change*, or knowledge restructuring at the level of individual concepts (Carey, 2009; Chi, 1992). Conceptual change is an intrinsic part of science learning. Most domains of science entail entities, properties, and mechanisms that defy our intuitive knowledge of how the world works and can only be represented if that knowledge is reorganized and restructured (Nersessian, 1998; Vosniadou, 1994).

Conceptual change is empirically distinguishable from other forms of knowledge acquisition in that it results in systematic failures of teaching and learning. In domains requiring conceptual change, individuals who have yet to undergo that change exhibit misconceptions about the domain's content that are internally coherent and developmentally widespread. These misconceptions are robust in the face of counterevidence or counterinstruction, and they create impasses in communication between those who have achieved conceptual change and those who have not (for reviews, see Carey, 2009; Shtulman, 2017).

Consider the domain of evolutionary biology—the domain of choice in the present study. Evolution results from differential survival and differential reproduction within a population; the traits possessed by the most reproductively successful individuals spread through the population over time. Most people, however, view evolution as the uniform transformation of an entire population, where every organism is guaranteed to have offspring more adapted to the environment than it was at birth (Bishop & Anderson, 1990; Shtulman, 2006). This view is grounded in the commonsense assumption that all members of a species share the same inner nature, or essence, which determines their outward appearance and behavior (Gelman, 2003; Shtulman & Schulz, 2008). Evolution is thus seen as a kind of cross-generational metamorphosis; selection plays no role in the process. This essentialist view of evolution has been documented in people of varying ages (Berti, Toneatti, & Rosati, 2010; Shtulman, Neal, & Lindquist, 2016) and educational backgrounds (Coley & Tanner, 2015; Gregory & Ellis, 2009), and it characterizes how a person reasons about several aspects of evolution, including variation, inheritance, adaptation, domestication, speciation, and extinction (Shtulman & Calabi, 2013).

The focus of the current study is a particular hallmark of conceptual change: impasses in communication between those who have achieved conceptual change and those who have not. Such impasses have been observed in conversations between children and adults (e.g., Carey, 1985), in conversations between science students and science teachers (e.g., Wiser & Amin, 2001), and in conversations between scientists working within different theoretical paradigms (e.g., Kuhn, 1977). Such impasses are often encountered in the context of learning—e.g., a child learning about the properties of living things from a parent or a student learning about the properties of thermal systems from a teacher—but it is unclear how they affect learning. Achieving conceptual change requires overcoming the conceptual gap responsible for the impasse, but how and with whom?

Conceptual impasses in communication are particularly important to study in light of the finding that learning is often facilitated through collaboration. For many types of inductive problems, individuals who collaborate on those problems are more likely to solve them—and learn from them—than individuals who work alone (Gauvain & Rogoff, 1989; Laughlin, Vanderstoep, & Hollingshead, 1991; Leman, Skipper, Watling, & Rutland, 2016). Collaboration is effective for several reasons. It opens partners' eyes to ideas they would not have generated on their own, highlighting alternative approaches to the same problem (Schwarz, Neuman, & Biezuner, 2000; Young, Alibali, & Kalish, 2012) or alternative explanations for the same phenomenon (Ames & Murray, 1982; Howe, 2009). It forces collaborating partners to articulate their reasons for endorsing a particular hypothesis or favoring a particular solution strategy and defend those reasons with evidence (Okada & Simon, 1997; Teasley, 1995). And it introduces social incentives for completing the task at hand, increasing partners' motivation to persist in the face of unexpected obstacles (Butler & Walton, 2013).

Given the pedagogical benefits of collaboration, we sought to determine whether collaboration is useful—or even possible—in domains requiring conceptual change. The answer to this question has both practical and theoretical implications. From a practical point of view, educators who instruct students on topics requiring conceptual change (e.g., evolution, microbiology, mechanics, thermodynamics, fractions) would benefit from knowing whether collaboration is an effective instructional strategy or a dead end. From a theoretical point of view, models of conceptual change would be further informed by research clarifying which kinds of input foster conceptual change and which do not. Parent-child conversation, for instance, is a form of collaboration that may help foster conceptual change (Gunderson & Levine, 2011; Jipson & Callanan, 2003), but it is unclear how beneficial this activity is relative to other domain-specific activities (e.g., refutation-based instruction, inquiry-based instruction, informal exploration).

Previous studies have found that collaboration in conceptually complex domains can be successful. For instance, Asterhan and Schwarz (2007) found that undergraduates who collaborated on devising evolutionary explanations for two instances of biological adaptation (mosquitos developing resistance to an insecticide, cheetahs acquiring the ability to run faster than any other mammal) provided more sophisticated (selection-based) explanations for biological adaptation from pre-collaboration to post-collaboration. Likewise, Loyens, Jones, Mikkers, and van Gog (2015) found that undergraduates who collaborated on determining the paths traced by three projectiles (a child jumping from a swing, an object falling on someone's head, a coyote falling from a cliff) drew more accurate motion paths from pre-collaboration to post-collaboration. In both the domain of evolution and the domain of motion, learning a correct, scientific view of the domain is difficult to achieve; direct (lecture-based) instruction on these topics has typically proven unsuccessful (see Shtulman, 2017, for a review).

These studies demonstrate that collaboration can facilitate learning in domains characterized by conceptual change, but they are limited in that they explored only one aspect of those domains—explanations for adaptation in the study by Asterhan and Schwarz (2007) and trajectories of projectile motion in the study by Loyens et al. (2015). The present study explored whether collaboration is effective for learning several facets of a conceptually complex domain—namely, the phenomena of variation, inheritance, adaptation, domestication, speciation, and extinction within the domain of evolutionary biology. We chose this domain because its content is notoriously difficult to understand and because individuals who are asked to reason about domain-relevant problems would most likely hold different levels of understanding. Collaboration under these circumstances thus provides a stringent test of whether, and how, collaboration can facilitate conceptual change.

# Method

## Participants

The participants were 44 college undergraduates, recruited from introductory psychology and cognitive science courses and compensated either with extra credit in those courses or with a small stipend. They were taken from a larger dataset of 174 participants, assigned to one of 87 dyads. All participants in the larger dataset were invited to complete a posttest (for a $12 Amazon gift card) but only 44 did. Those 44 came from 36 different dyads. Approximately half were the higher-scoring partner in their dyad ($n = 25$) and half were the lower-scoring partner ($n = 19$). In other words, approximately half collaborated with a partner who demonstrated a higher level of understanding prior to the collaboration, and half collaborated with a partner who demonstrated a lower level of understanding.

The 44 participants who completed a posttest earned similar pretest scores to those who did not complete a posttest. That is, the 19 low scorers who completed a posttest scored similarly to the 68 who did not ($M = 9.4$ vs. $M = 9.7$, $t(85) = 0.22$, $p = 0.83$), and the 25 high scorers who completed a posttest scored similarly to the 62 who did not ($M = 17.1$ vs. $M = 15.9$, $t(85) = 1.13$, $p = 0.26$). While there may be motivational differences between those who opted to complete a posttest and those who did not, there were no reliable knowledge differences between the groups (relative to participants' classification as a low scorer or a high scorer).

## Materials

Participants were assessed on their understanding of evolution using an instrument developed by Shtulman (2006). The assessment consisted of six sections, each devoted to a different biological phenomenon (inheritance, variation, adaptation, domestication, speciation, and extinction). Participants' understanding of the phenomenon was assessed with five questions or tasks designed to elicit either an essentialist interpretation or a selection-based interpretation.

With respect to inheritance, for instance, participants were asked to make predictions about parent-offspring resemblance with questions like the following: "Imagine that biologists discover a new species of woodpecker that lives in isolation on a secluded island. These woodpeckers have, on average, a one-inch beak and their only food source is a tree-dwelling insect that lives, on average, one-and-a-half inches under the tree bark. Compared to its parents, the offspring of any two woodpeckers should develop: (a) a longer beak, (b) a shorter beak, or (c) either a longer beak or a shorter beak; neither is more likely." The correct response is (c), because offspring vary randomly from their parents, but most people select (a), reasoning that offspring will inherit whatever traits will help them survive—traits conferred by an underlying essence that adaptively changes in response to the species' current needs.

As another illustration, consider this task designed to probe participants' understanding of within-species variation: "During the 19th century, England's native moth species, *Biston betularia*, evolved darker coloration in response to the pollution produced by the Industrial Revolution. Imagine that biologists gathered a random sample of *Biston betularia* once every 25 years from 1800 to 1900. What range of coloration would you expect to find at each point in time?" Participants were given a five-by-five matrix of moth outlines and instructed to shade the moths to reflect what the moths might look like at 1800, 1825, 1850, 1875, and 1900. The two most common response patterns are depicted in Figure 1. The pattern on the left depicts a mutation for darker coloration spreading through the population over time and is consistent with a selection-based view of evolution. The pattern on the right depicts the uniform transformation of the population, such that variation occurs between generations but not within generations, and is consistent with an essentialist view.

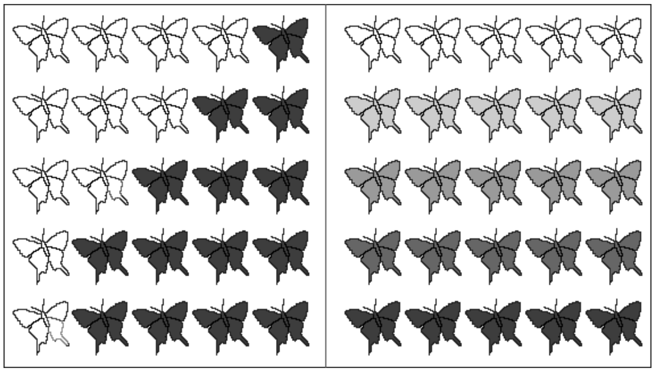

Figure 1: A selection-based response pattern (left) and an essentialist response pattern (right) on the moth-shading task of the evolution comprehension assessment.

The full battery of questions can be found in the Appendix of Shtulman (2006), along with criteria for scoring each question or task. Participants were assigned 1 point for every correct. selection-based response and 0 points for every incorrect, essentialist response. Responses too vague to be counted as selection-based, were also assigned 0 points. Participants' scores thus ranged from 0 to 5 per section and from 0 to 30 for the assessment as a whole.

## Procedure

The evolution comprehension assessment was administered on a computer. It took between 30 and 45 minutes to complete, and participants completed it twice by themselves (pretest and posttest) and once with a partner (dyad test). Participants were tested in pairs in a room in the Psychology Department. They completed the pretest by themselves, and they completed the dyad test together immediately following the pretest. Participants typically did not know their dyad partner, and they were given no instruction on how to coordinate their responses. They were simply asked to complete the survey as a pair, on a single computer. Their conversations were recorded and transcribed at a later date. (Data from the conversations are not reported here, for lack of space).

The posttest was administered one semester (i.e., half a year) after the dyad test. The average delay between dyad test and posttest was 7.6 months, and the delay for the high scorers was equivalent to the delay for the low scorers ($M$ = 6.8 vs. $M$ = 8.8, $t(42) = 1.13$, $p = .138$). Effects of collaboration detectable after half a year arguably represent long-term changes in understanding, as participants' memory for the episodic details of the collaboration session would likely have faded.

# Results

For the analyses below, we used linear mixed models (LMMs) with random-effects structures specified according to the procedure recommended by Bates and colleagues (Bates et al., 2015). We used likelihood ratio-test (LRT) comparisons and 95% confidence intervals (CI) for inference.

## Evolution Scores at Pretest, Dyad Test, and Posttest

Overall assessment scores at pretest, dyad test, and posttest are shown in Figure 2 (left panel). Dyads generated higher scores than individuals at pretest, $\beta = 2.16$, 95% CI [.27, 4.05]. However, individuals' posttest scores were similar to their pretest scores, $\beta = .91$, 95% CI [-.98, 2.80].

## Low- versus High-Scoring Partners

An individual's ability to profit from collaboration likely depends on both their own prior knowledge and their partner's prior knowledge. For example, individuals paired with partners that demonstrated greater conceptual knowledge of evolution at pretest might have a greater opportunity to learn from collaboration than those paired with partners with less conceptual knowledge. To explore this possibility, we categorized participants in terms of whether they were the lower or higher scoring partner in their respective dyads at pretest. Figure 2 shows pretest, dyad test, and posttest scores for lower-scoring partners (middle panel) and higher-scoring partners (right panel). There was an interaction between scoring status and test, LRT $\chi^2(2) = 20.13$, $p < .001$. For high-scoring partners, pretest, dyad test, and posttest scores were similar. However, for low-scoring partners, dyad tests were greater than pretests, $\beta = 6.53$, 95%

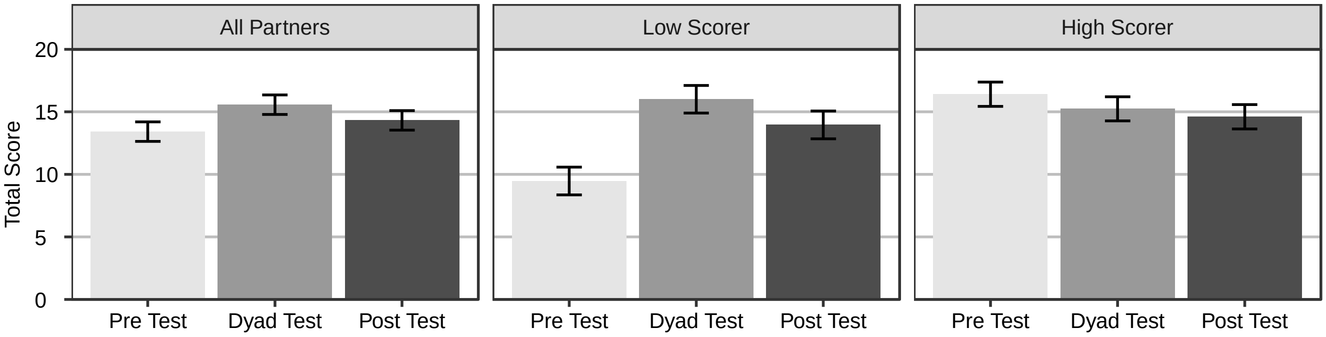

Figure 2. Mean evolution score at pretest, dyad test, and posttest for all participants, low scoring participants, and high scoring participants. Error bars represent ± SE.

CI [3.92, 9.13], and posttests were greater than pretests, β = 3.37, 95% CI [1.87, 7.08].

### Low- versus High-Scoring Partners by Section

Only 5 of the 44 participants demonstrated consistently greater or poorer pretest performance than their partner across all six sections of the assessment. Categorizing participants as low- or high-scoring by section potentially provides a more nuanced view of performance than overall pretest scores. Figure 3 shows lower- and higher-scoring partners' pretest, dyad test, and posttest scores for each of the six sections. A three way interaction between scoring status, test, and section suggested variation in low and high scorers' learning across section, LRT $\chi^2(10) = 19.10$, $p = .0317$.

Scores for Inheritance and Speciation demonstrated little pretest to posttest change for both low and high scorers. Most consistent with the overall assessment, lower scorers demonstrated pretest to posttest gains for Domestication, β = 1.50, 95% CI [.54, 2.46], and Extinction, β = .76, 95% CI [-.02, 1.55], whereas high scorers had similar pretest and posttest scores. Low scorers again demonstrated pretest to posttest gains for the Adaptation, β = 1.71, 95% CI [.84, 2.57], and Variation, β = 1.43, 95% CI [.47, 2.38]. However, high scorers surprisingly demonstrated pretest to posttest losses for Adaptation, β = -1.50, 95% CI [-2.26, -.74], and Variation, β = -1.21, 95% CI [-2.03, -.39].

### Predicting Posttest Scores

For low scorers, collaborating with a more advanced partner yielded pretest to posttest improvement in 4 out of 6 sections. For high scorers, collaborating with a less advanced partner yielded pretest to posttest decline in 2 out of 6 sections. These results suggest that participants' posttest performance was influenced both by their own understanding of the domain and by their partner's understanding.

To explore this possibility further, we fit an LMM on posttest scores (by section) with participant pretest scores, partner pretest scores, scoring status (high vs. low within a participant's respective dyad), and their interactions as fixed effects. Participant pretest score was a positive predictor of posttest score, β = .38, 95% CI [.08, .67], and did not interact with scoring status. In contrast, partner pretest score interacted with scoring status, LRT $\chi^2(1) = 3.86$, $p = .049$. Partner pretest scores were not predictive of posttest scores for high scorers, β = .08, 95% CI [-.15, .31]. However, partner pretest scores were negatively related to posttest scores for low scorers, β = -.22, 95% CI [-.43, .00]. Thus, it appears that low scorers learned more from partners with slightly greater knowledge than themselves at pretest compared to partners with much greater knowledge.

## Discussion

Collaboration is an effective and efficient means of devising new hypotheses (Okada & Simon, 1997) and learning new problem-solving strategies (Schwarz et al., 2000), but is collaboration possible in domains where individuals are known to hold vastly discrepant views of the domain's causal structure? The answer appears to be yes. Individuals who collaborated on tasks within the domain of evolutionary biology—a domain characterized by qualitatively different theories of what evolution is and how evolution works (Shtulman, 2006)—demonstrated a higher level of understanding together than they did individually. This finding was far from guaranteed given the content of the task. When partners disagreed, their disagreements typically reflected fundamental differences in their understanding of the task domain. Resolving those disagreements entailed more than just recognizing who knew the answer. It entailed recognizing which of two answers—a selection-based answer and an essentialist answer—was more plausible or justifiable.

Collaboration not only facilitated more accurate responding, it also facilitated learning, though the effects were nuanced. Individuals who entered the collaboration with lower levels of understanding demonstrated increased understanding at posttest (several months later), whereas individuals who entered the collaboration with higher levels of understanding demonstrated no gains at posttest.

The learning exhibited by less-knowledgeable partners was generally robust across different sections of the assessment, as was the stasis exhibited by more-knowledgeable partners. That said, there were some sections on which the less-knowledgeable partners exhibited no gains from pretest to

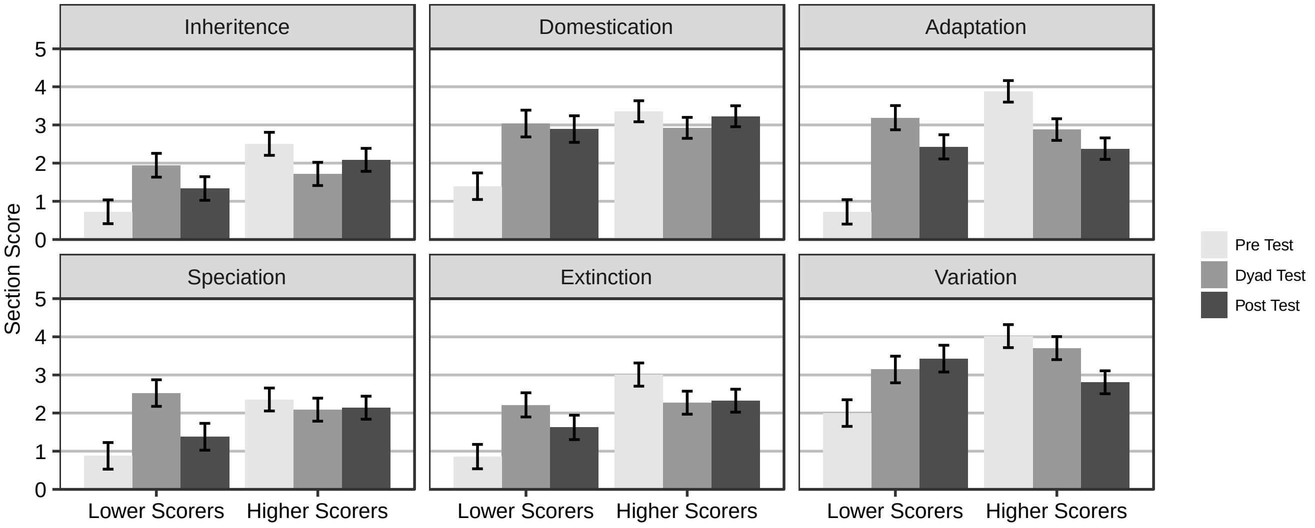

Figure 3. Mean pretest, dyad test, and posttest scores for low scoring and high scoring partners by section. Error bars represent ± SE.

posttest and some sections on which the more-knowledgeable partners exhibited losses from pretest to posttest. Collaboration in conceptually complex domains may thus hinder learning for some individuals in some contexts. Still, the net benefits of collaboration were positive, which is a surprising finding given that (a) collaborators often had to communicate across a conceptual divide and (b) the collaboration itself consisted solely of discussion. There were no opportunities to generate evidence or test hypotheses, which suggests that such activities may not be necessary for learning in cases where the primary intellectual challenge is just interpreting what one's partner is saying.

Perhaps the most provocative finding was that, among participants who learned from collaboration (the low scorers), those who learned the most collaborated with partners who had moderately higher levels of understanding. Individuals who collaborated with partners with substantially higher levels of understanding benefited less, at least by posttest. This finding, though tentative, may have resulted from differential impasses in communication; the greater the discrepancy between partners' understanding of the domain, the more likely they encountered impasses in communication and the more strained their collaboration may have become. Consider, for instance, the following conversation between a participant who earned a pretest score of 19 (P1) and one who earned a pretest score of 2 (P2) about the woodpecker question presented above:

P1: Alright, for the first one I put either a shorter or longer beak because it says compared to its parents, and compared to its parents it pretty much has the same beak because it has the same genes.

P2: Okay. Hmm. I put longer beak ... because, yeah, they have to eventually evolve into the thing, but I can see what you are saying about, like, it wouldn't take one generation.

P1: Well ... the next generation would end up with a longer beak, but this one particular woodpecker would have the same [beak] as its parents, if you understand what I'm saying. The generations would get longer beaks because the ones with the shorter beaks will be killed off. [But] no matter what, the offspring are gonna have beaks pretty much the same as its parents.

P2: Okay, I see what you're saying. Yeah, I guess I just assumed that they would interbreed or they would have a woodpecker from a different... Okay, I see what you're saying.

P2 claims to understand what P1 is saying, but P2's attempts to resolve the discrepancy—by acknowledging P2's answer as correct on the assumption that "it wouldn't take one generation" or that birds with different beak lengths did not "interbreed"—do not actually address P2's point that evolutionary change occurs at the population level, not the individual level. This type of impasse may be more common in conversations between partners with discrepant levels of understanding than in conversations between partners with similar levels of understanding, though confirmation of this pattern awaits further analysis of the conversational data.

The finding that participants benefited most from collaborating with individuals who were only moderately more knowledgeable about the domain helps answer the question of how individuals are able to communicate across a gap in conceptual understanding. Communication may be possible only if the gap is not too wide; wider gaps may lead to irreconcilable differences in how partners perceive or analyze the problems at hand. We plan to test this idea directly by analyzing the dynamics of participants' conversations in relation to their score differences from pretest to dyad test and from pretest to posttest. Previous research on how domain experts converse with domain novices suggests that the experts supply novices with specialized knowledge, in the moment, by adjusting how they label or how they describe objects of shared attention (Clark & Schaefer, 1989; Isaacs & Clark, 1987). However, such studies have involved domains in which the difference between a novice's knowledge and an expert's knowledge is quantitative rather than qualitative (e.g., knowledge of New York City landmarks). It remains an open question how domain novices and domain experts are able to bridge differences in knowledge, through discourse patterns, when that knowledge entails conceptual change.

## References

Ames, G. J., & Murray, F. B. (1982). When two wrongs make a right. *Developmental Psychology, 18*, 894-897.

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

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1-48.

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. *The 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.

Butler, L. P., & Walton, G. M. (2013). The opportunity to collaborate increases preschoolers' motivation for challenging tasks. *Journal of Experimental Child Psychology, 116*, 953-961.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

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

Chi, M. (1992). Conceptual change within and across ontological categories. In R. Giere (Ed.), *Cognitive models of science* (129-186). Minneapolis: U of Minnesota Press.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science, 13*, 259-294.

Coley, J. D., & Tanner, K. (2015). Relations between intuitive biological thinking and biological misconceptions in biology majors and nonmajors. *CBE: Life Sciences Education, 14*, 1-19.

Gauvain, M., & Rogoff, B. (1989). Collaborative problem solving and children's planning skills. *Developmental Psychology, 25*, 139-151.

Gelman, S. A. (2003). *The essential child*. Oxford, UK: Oxford University Press.

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

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.

Howe, C. (2009). Collaborative group work in middle childhood. *Human Development, 52*, 215-239.

Isaacs, E. A., & Clark, H. H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology , 116*, 26-37.

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

Kuhn, T. S. (1977). *The essential tension: Selected studies in scientific tradition and change*. Chicago: University of Chicago Press.

Laughlin, P. R., Vanderstoep, S. W., & Hollingshead, A. B. (1991). Collective versus individual induction. *Journal of Personality and Social Psychology, 61*, 50-67.

Leman, P. J., Skipper, Y., Watling, D., & Rutland, A. (2016). Conceptual change in science is facilitated through peer collaboration for boys but not for girls. *Child Development, 87*, 176-183.

Loyens, S. M., Jones, S. H., Mikkers, J., & van Gog, T. (2015). Problem-based learning as a facilitator of conceptual change. *Learning and Instruction, 38*, 34-42.

Nersessian, N. J. (1989). Conceptual change in science and in science education. *Synthese, 80*, 163-183.

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

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

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

Shtulman, A. (2017). *Scienceblind: Why our intuitive theories about the world are so often wrong*. New York: Basic Books.

Shtulman, A., & Calabi, P. (2013). Tuition vs. intuition: Effects of instruction on naïve theories of evolution. *Merrill-Palmer Quarterly, 59*, 141-167.

Shtulman, A., Neal, C., & Lindquist, G. (2016). Children's ability to learn evolutionary explanations for biological adaptation. *Early Education and Development, 27*, 1222-1236.

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

Teasley, S. D. (1995). The role of talk in children's peer collaborations. *Developmental Psychology, 31*, 207-220.

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

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

Young, A. G., Alibali, M. W., & Kalish, C. W. (2012). Disagreement and causal learning: Others' hypotheses affect children's evaluations of evidence. *Developmental Psychology, 48*, 1242-1253.