Unpacking Dimensions of Evidentiary Knowledge and Reasoning in the Teaching and Learning of Science

Ala Samarapungavan (co-chair), Kari Clase, Nancy Pelaez, Stephanie Gardner, and Chandrani Misra
ala@purdue.edu; klclase@purdue.edu; npelaez@purdue.edu; sgardne@purdue.edu; mishrac@purdue.edu
Purdue University

Ravit Golan Duncan (co-chair), Rutgers University, ravit.duncan@gse.rutgers.edu
Clark Chinn (co-chair), Rutgers University, clark.chinn@gse.rutgers.edu
Sarit Barzilai, University of Haifa, sarit.barzilai@edtech.haifa.ac.il
Leema Berland, University of Wisconsin-Madison, lberland@wisc.edu,
Katherine McNeill, Boston College, kmcneill@bc.edu
Eve Manz, Boston University, eimanz@bu.edu
Alison Wylie (discussant), University of British Columbia, alison.wylie@ubc.ca
William Sandoval (discussant), University of California Los Angeles, sandoval@gseis.ucla.edu

Abstract: Although the study of evidentiary reasoning has a long history in psychology and science education, much of this scholarship has focused on how people coordinate evidence with knowledge claims. Less attention has been paid to our notions of evidence itself and how these develop, especially in the context of schooling. In this symposium, presenters draw from scholarship in science studies and the philosophy of science, cognitive work on epistemic reasoning, and research in science education to unpack dimensions of evidentiary reasoning. Our collective focus is on identifying aspects of evidentiary knowledge and reasoning that prevail in scientific practice but are typically absent from classroom implementations of inquiry in science education. Further, the symposium will address sources of challenge for teachers and students as they engage with evidence in the science classroom and discuss ways in which educators can scaffold the development of more sophisticated reasoning with and about evidence.

The study of evidentiary reasoning has long been a focus of psychologists and science educators. Psychologists interested in scientific reasoning have focused on how people develop the ability to coordinate theories with evidence. Science educators have examined how the design of learning environments can influence students' understanding and use of evidence in contexts of argumentation and inquiry. Despite this rich history of interest, there is surprisingly little discussion in the literature about our notions of evidence itself. Science educators have typically posited that phenomena become "evidence" when connected to a knowledge claim by argument. Although this definition is a helpful starting point, it needs much unpacking to be useful in the design of learning environments. The scholarship in contemporary science studies suggests that scientific evidence is not a simple unitary construct but rather a rich, multidimensional construct. For example, reasoning with evidence involves considerations of its relevance, significance or weight, quality, and concordance with other lines of evidence, in relation to some circumscribed set of hypotheses or models under consideration. Scientific disciplines typically evolve internal methodological norms, standards, and procedures for gathering and evaluating evidence. For example, scientists working in a discipline share knowledge of relevant variables and plausible mechanisms. As part of their education, they learn how variables are typically operationalized in investigative designs, norms and standards for the precision and accuracy of instrumentation, experimental procedures, and measurement, bandwidths and density of sampling, models for aggregating and analyzing data, and conventions for communicating results. Both the psychological literature on scientific reasoning and the science education literature have neglected many of these aspects of evidentiary knowledge and reasoning. In this symposium, we aim to advance research by attempting to address three core questions:

1. What are some dimensions of reasoning with evidence that are prevalent in scientific practice but are mostly missing from classroom implementations of inquiry?
2. What aspects of reasoning with evidence are most challenging for students and teachers to engage with in the science classroom?
3. What are some ways we can scaffold and support students' engagement with more sophisticated ways of reasoning with and about evidence?

This symposium will be conducted in an interactive format. The co-chairs will briefly introduce the rationale and core questions that the symposium will address (3-5 minutes). Each individual presentation will take approximately 12 minutes. Presenters will draw from their theoretical and empirical work to offer a lens for considering what it means to think with and about evidence in the context of scientific work. The empirical work that will inform the questions addressed by the symposium draws from several grade bands ranging from primary (Manz) and middle school (Berland & McNeill; Duncan, Chinn, & Barzilai) to secondary and post-secondary settings (Samarapungavan, Clase, Pelaez, Gardner, & Misra). Additionally, the presenters examine reasoning in an array of task settings for evidentiary reasoning, including discourse about evidence in contexts of explanation and argumentation, reasoning with digital data such as simulations, from personal every day experiences of phenomena, and from data collected during laboratory investigations. Our two discussants (Wylie and Sandoval) bring different perspectives to the symposium. Wylie, a philosopher of science, has most recently engaged in a program of research (with her collaborator Robert Chapman) that examines evidentiary reasoning in anthropology and how the methodological norms and standards employed in the discipline allow anthropologists to reach consensus about evidence. Sandoval, a learning scientist, has been at the forefront of work on the design of science learning environments to support students' evidentiary reasoning. During the last segment of the symposium, our two discussants (Wylie and Sandoval) will lead a critical consideration of the presentations, engage the audience in questions and discussion of the issues, and synthesize crosscutting theoretical and practical themes and directions for future work.

"In real nature, does the wind blow three times?" Making the representational nature of evidence visible in classroom investigations

Eve Manz, Boston University

This paper addresses one important aspect of scientific reasoning that is typically left out of students' experience constructing and critiquing claims and evidence—namely, the relationship between an investigation and the phenomenon it is meant to represent. I use a second grade landforms experiment as a context for analyzing student reasoning about evidence. I show how opportunities to consider how the experiment represented (or failed to represent) the focal phenomenon of wind and water shaping land both supported episodes of rich reasoning and provided challenges for teachers and students.

While empirical investigations have long been a focus of research, we know little about how students reason about the transitions represented in Figure 1 and how this reasoning might be supported in classroom learning environments. Whether in contexts of inquiry, or explanation and argumentation (e.g., Kuhn & Pease, 2008; Masnick & Klahr, 2003; Toth, Suthers, & Lesgold, 2002), prior studies focus on the relationship between experiment and evidence, overlooking experiments' function as a way to get a grip on aspects of the world that are difficult to isolate and test in situ (for two reviews, see Cavagnetto, 2010; Manz, 2015a). I conceptualize the relationships between evidence, explanations, and empirical work using a framework drawn from literature in Science Studies (e.g. Gooding, 1990; Latour, 1987) and Science Education (Lehrer & Schauble, 2006; Manz, 2012; 2015b). Figure 1 represents these relationships using an elementary school experiment which has been redesigned to more accurately represent scientific work: here, the *empirical investigation* (placing plants in different light conditions to study their success) is generated to understand a complex *phenomenon* (a backyard characterized by patterns of shade and light and a corresponding distribution of plants). *Observations and evidence* are determined in light of an understanding of the phenomenon, and must be made sense of both to draw a conclusion about the investigation and to develop an *explanation or explanatory model* (here that, different plants are successful in different light conditions).

The analysis presented here focuses on a second grade landforms investigation co-designed with five second grade teachers by adapting lessons from a commercial science kit. The original kit did not provide direction for teachers to support students to think about the transitions represented in Figure 1. In the re-designed lessons, students first examined photographs and discussed how wind and water might shape land; designed investigations using straws and spray bottles to test their ideas; developed, presented and critiqued claims and evidence about how wind and water move earth materials; and then discussed the phenomenon again based on their investigations. Data collected and analyzed for all lessons included videotape of classroom instruction, classroom artifacts, field notes, student work, and an individual semi-structured interview with teachers. Analysis showed that (1) how the experiment represented the focal phenomenon was a relevant question for students and (2) that it influenced how they generated and evaluated evidence. First, in designing experiments, differences in students' strategies for representing the phenomenon, supported the generation of different forms of evidence. For example, some groups squirted the spray bottles at the earth materials and used measures of distance the materials traveled as evidence, while others decided to first put water into petri dishes, then move the water; producing forms of evidence such as floating or absorption of water. Second, students bounded their conclusions based on the levels of variables represented in their experiments: that is, many were unwilling to claim that the experiment showed that wind *cannot* move rocks, as there were rocks in the world bigger than those tested and more extreme forces of wind and water were likely to move rocks. Third, students questioned whether the design of experiments represented "real-world" processes. One student argued against his teacher's attempt to ratify the choice to blow three times on each material by stating "Because in real life, in real nature, does the wind blow three times and wait for ten minutes, and then blow three times again?"

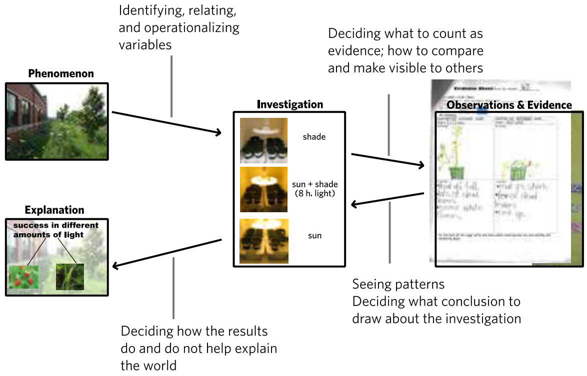

Figure 1. Conceptual Framework for Empirical Work.

Opening the experiment up to these choices, tensions, and disagreements provided important opportunities for students to move past objectifying evidence (Sandoval & Çam, 2011) However, these openings also provided challenges to teachers, particularly when students sensibly argued against canonical aspects of evidence production that are reified in school practices. I end with a conundrum that I will explore further in the symposium: are some classroom experiments more useful for reasoning about evidence if we focus on the opportunities that emerge from their problematic aspects, rather than their function in producing "evidence" to support desired content understandings? What might this mean for how we design science learning environments and support teachers to orchestrate them?

How can personal experiences be leveraged as "scientific evidence" In K-12 classrooms?

Leema Berland, University of Wisconsin-Madison, and Katherine McNeill, Boston College

There is a shared understanding throughout the education world that we learn by connecting prior knowledge/experiences with new knowledge (NRC, 2015). In science education, this means that students should be enabled—nay encouraged—to bring their prior experiences into the class' sense making discussions. However, this call, while easy to make and almost universally supported, has deep underlying complexities. In science, scientists dialogically build knowledge about natural phenomena (Ford, 2012) by manipulating representations of that phenomenon (Duschl, 1990). This suggests a particular definition of scientific evidence in which the information is both phenomena based and transformable (McNeill & Berland, 2017). Consequently, a tension can arise between the everyday experiences students bring to the classroom and a particular view of what counts as scientific evidence.

For instance, Table 1 includes an example from two middle school students during a life science unit. The task asked students to analyze data from an online simulation and to choose which of two provided claims (Desiree's or Abde's) is better supported by evidence. In this conversation, we see a tension in terms of what the two students are using to justify their claims. Ignacio is focused on the simulation and talks about "what we've seen" and "his energy." Ignacio is using the scientific evidence from the simulation, which consists of data. In contrast, Julie is focused on her own experiences with running and eating. Her language focuses on "you eat a lot" and "you run faster."

Table 1: An example from two middle school students during a life science unit

Speaker	Quote
Ignacio	Well, I-I think she thinks that Desiree, uh, Desiree's claim is the smarter one.
Julie	Desiree's is not the smarter one. Abde's is.
Ignacio	How?
Julie	Well, if <u>you</u> eat a lot before you run you just, you know, <u>you</u> run faster you-
Ignacio	Yeah, but what we <u>what we've seen</u> is after the three minute mark <u>his energy</u> starts to drop insanely fast.
Julie	Yeah but…if <u>you</u> eat a lot then more then maybe more than Abde did, then… <u>you'll</u> make it to the end. And he got past halfway.
Ignacio	Past halfway, yeah, but <u>Desiree still went the whole way</u> keeping <u>his energy up</u> around 90%.

This discussion illustrates a tension that often occurs in classrooms. We want students to be collaboratively connecting to their everyday resources to make sense of what is happening in the classroom. But, how can we dialogically build knowledge based on phenomena that we have not all experienced? How can they use these resources in their sense making if they do not question and interpret them? How can they fold these resources into their sense making so they work in concert with observations and experiences they make in their classroom? In short: Is it possible for Julie's experience to be a productive resource for sense making in this discussion?

In this paper, we argue that it *is possible* to leverage everyday resources in ways that allow the class to use them as a piece of evidence as they interpret their more formal observations (what we might call scientific evidence). This is possible when teachers and students work with these resources in ways that are consistent with three design heuristics for identifying and using evidence: phenomena-based, transformable and used dialogically. Table 2 uses the design heuristics to show both how Julie and Ignacio *did* position her everyday resources (we show these as italicized quotes in the table) in the conversation and how Julie, Ignacio, or a teacher, *could have* positioned them (we show this as non-italicized text). The presentation will explore this further, exemplifying the various ways everyday resources might be used in class discussions of evidence and how these discussions can be refined in ways that allow the everyday resources to be leveraged as evidence.

<u>Table 2: Example using design heuristics</u>

	Low	*High*
Phenomena-based	Information students have been told, or information that is not directly and obviously connected to experience (e.g., the hypothetical: "*if <u>you</u> eat a lot before you run you just, you know, <u>you</u> run faster you")*	Information that is based in observed experiences (e.g., I run a lot and I run differently depending on when and what food I eat)
Transformable	Asking students to describe what they have seen, or not engaging with it (i.e., Have you ever gone for a long run?) (We note this question, while not encouraging transformation, may set it up by shifting the conversation to a specific phenomenon.)	Asking students questions that challenge their interpretations of the experience (i.e., When do you eat large meals for running – close to the run, the week before, etc.?)
Used Dialogically	Experiences or information that are not shared and not easily related to other experiences (i.e., Everyone's body is different, that is what works for me)	Experiences that are common enable students to question and challenge one another (i.e., emphasizing the commonality: "*if <u>you</u> eat a lot before you run you just, you know, <u>you</u> run faster*)

Problematizing and expanding our conceptualization of evidence in science instruction

Ravit Golan Duncan, Clark A. Chinn, Rutgers University, and Sarit Barzilai, University of Haifa

Research in science education has investigated students' reasoning with and about evidence in the context of evidence use in arguments to support or refute claims (e.g. Berland & McNeill, 2010), as well as use of evidence in constructing and evaluating models (e.g. Passmore & Svoboda, 2012). Yet we argue that despite these efforts, the construct of evidence remains relatively undifferentiated in the science education community and in science instruction. One consequence of this uniform view of evidence is that classrooms often feature evidence that is epistemically simpler than evidence in science. Whereas evidence in science varies noticeably in amount, scope, comprehensiveness, methodological quality, robustness, technical complexity, and types of inferential connections to explanations, evidence in science classrooms is often simplistically used to determine whether or not evidence supports or contradicts a claim, and stating a reason why.

We thus argue that there is still a need to problematize and unpack the nature and development of evidential reasoning. To make progress, we must complexify the construct of evidence and explicitly deal with a wider range of dimensions of reasoning with and about more authentic forms of evidence. We therefore propose a theoretical framework for reasoning with and about evidence that expands current conceptualizations of evidence. The main objective of this framework, which we call *grasp of evidence,* is to complexify the concept of evidence in ways that will facilitate introducing more authentic forms of evidence and more sophisticated ways of engaging with evidence in science classrooms. Our work builds on recent insightful analyses by McNeill & Berland (2017) and by Samarapungavan (in press).

We focus on developing a *grasp of evidence*, which draws on Ford's (2008) construct of *grasp of practice*. For Ford, a grasp of practice involves internalization of two interrelated roles critical for scientific knowledge building: *constructing* and *critiquing* claims (Ford, 2008). "Grasp" implies that the knowledge at hand is not purely declarative, but also includes knowledge of how to engage in critique, as well as epistemic justifications about why such critiques are necessary and which are relevant. Such a grasp is socially constructed and negotiated within a community of scientists (or learners). From a lay perspective, a grasp of evidence affords becoming a competent outsider, and making informed decisions about the credibility of scientific claims and evidence even in the absence of deep domain knowledge (Feinstein, Allen, & Jenkins, 2013). The grasp-of-evidence framework consists of two "axes." The first axis theorizes four *dimensions* that comprise grasp of evidence:

Analysis: To reason with evidence, one must first identify and comprehend its components (e.g., goals, methods, results, conclusion) and their interrelations. Science studies have shown the prominence of reading and comprehending in the work of scientists (e.g., Tenopir, King, Boyce, Grayson, & Paulson, 2005). This involves *analysis* of studies into its components, and understanding how these components fit together.

Evaluation: The second cluster of practices involves *evaluating* evidence. These are the familiar processes of evaluating the full range of methods used in a particular study—whether this means critiquing someone else's study or thinking through how to construct studies that withstand critical evaluation. Evidence evaluation is a central evidentiary practice in science (e.g., Staley, 2004).

Interpretation: The third cluster of practices involves *interpreting evidence*. These practices are also at the grain size of the individual study, as scientists work out how to interpret or reinterpret the results of a study in terms of one or more models, explanations, or theories under consideration. Understanding the nature and strength of the relationships between the evidence and competing claims and models is thus a core aspect of working with evidence (Chapman & Wylie, 2016; Galison, 1997).

Integration. The fourth cluster involves *integrating evidence*. In science this involves a variety of processes for identifying bodies of relevant evidence, considering how types of research fit together to support one model over another, and weighing evidence in various ways (e.g., Solomon, 2015).

The second axis of our theoretical framework derives from the AIR model of epistemic cognition (Chinn et al., 2014), which posits that epistemic cognition includes three central components: (A) *Aims and value* are the goals that individuals and communities set (aims), such as knowledge, and the importance of that knowledge (value). (B) *Epistemic Ideals* are the criteria used to evaluate whether epistemic aims have been achieved and the quality of resulting scientific products such as evidence or models. (C) *Reliable epistemic processes* are the diverse processes used to achieve epistemic aims, such as protocols for carrying out observations or conducting experiments, approaches to conducting meta-analyses, and so on. The framework involves applying the AIR model to unpack and specify the four dimensions of grasp of evidence. As an example, consider the *evidence interpretation* dimension. The core epistemic aim we associate with evidence interpretation is determining model validity using strong evidence. By strong evidence we mean evidence that is more tightly interconnected to one model and thus supports that model differentially over others. Several ideals can be used to judge evidence strength including its relevancy to the model in question, its ability to provide support (or to refute) core aspects of the model (as opposed to peripheral ones), and its diagnosticity (differentially supporting one model while refuting another). To meet these ideals students can engage in reliable processes such as careful consideration of which part of a model evidence supports, designing experiments that can provide diagnostic evidence, and so on.

We argue that a framework for grasp of evidence can help educators and education researchers in at least three ways: (a) it can help decide how to engage students in reasoning with and about evidence; (b) it can provide the basis for better assessments of reasoning with and about evidence; and (c) it can suggest instructional approaches that can help students develop a grasp-of-evidence.

Deconstructing evidence: Contextualizing students' understanding of methods for gathering and interpreting evidence in biology

Ala Samarapungavan, Kari Clase, Nancy Pelaez, Stephanie Gardner, and Chandrani Misra, Purdue University

In this presentation, we draw from a conceptual framework for contextualizing students' evidentiary reasoning in disciplinary knowledge and practices in biology (Samarapungavan, in press) and from our recently initiated empirical work as part of the Exploring Biological Evidence (EBE) project (funded by the National Science Foundation) to address the core questions posed by this symposium. Scholarship in science studies has emphasized the role of shared disciplinary norms and standards for inquiry and the generation and valuation of evidence as a basis for scientific consensus (Chapman & Wylie, 2016; Giere, 2010). For example, Galison (1997) has examined how the theoretical commitments of particle physicists shape their design of experiments, strategies for data reduction and decisions about whether the data represent something "real" in the world. Psychologists and educators have long recognized that science learners may interpret phenomena differently from scientists because they draw on different funds of knowledge. Indeed, research in psychology and science education has placed considerable emphasis on this aspect of evidentiary reasoning (Lehrer, & Schauble, 2006). The kinds of scaffolds that have been used to support students evidentiary reasoning tend to be generic in nature. For example, technology prompts in WISE (a digital inquiry environment) urge students to evaluate evidence for "usefulness" and "relevance" as they generate scientific explanations (Kali & Linn, 2008). Educators have paid much less attention to the rich methodological knowledge embedded in disciplinary practice that shapes and constrains fruitful evidentiary reasoning for scientists. Yet it is precisely this kind of knowledge that becomes critical to more advanced science learning in the secondary and post-secondary years, a period in which US students, for example, show sharp declines in interest for and achievement in science. The EBE project attempts to address this gap by designing and evaluating the impact of varied types of disciplinary scaffolds to support students' considerations of methodology in evidentiary reasoning. We will present preliminary data from our first round of implementation to illustrate how the contextualization of school inquiry practices in theoretical and methodological aspects of relevant disciplinary knowledge, can be used to enhance students' evidentiary reasoning. Based on a synthesis of research from science studies, the psychology of scientific reasoning, and science education, Samarapungavan's (in press) *Conceptual Analysis of Disciplinary Evidence (CADE)* framework highlights four broad, reciprocally related, categories of evidentiary relationships in scientific practice that are shaped by disciplinary knowledge. Because of space constraints, we focus here on the three of the four CADE categories to illustrate our approach:

1. Theory to Evidence relationships involve the framing and articulation of potentially testable models. Disciplinary knowledge circumscribes and problematizes focal phenomena that scientific models are designed to represent and explain. These relationships come to define what counts as evidence, where we should look for it, and how we will collect, interpret and use it. Recent science education research has grappled with ways of connecting the theoretical and empirical in student reasoning and sense making during inquiry (Berland & McNeill, 2017; Manz, 2012; Sandoval, 2005, 2014). For example, our own prior work as well that of others shows that in the teaching and learning of evolutionary biology in secondary school, discussions of natural selection often focus on species features and behaviors that confer survival benefits, such as success at finding food or evading predators (Samarapungavan, 2011, Sandoval & Reiser, 2004). In contrast, little attention is paid to reproductive success which biologists consider to be the mechanism by which population changes occur over generationsy. Therefore, to support evidentiary reasoning about evolution, disciplinary scaffolds might include reminders to consider the specific factors needed for the preservation/transmission (or lack thereof) of traits and behaviors across generations of species (i.e, reproductive success).

2. Evidence to Data relationships involve models for designing and executing investigations including the set up and use of instrumentation for data gathering, as well as models for aggregating and analyzing data. The work presented by Manz suggests that even young science learners (second graders) can begin to consider the extent to which their designs for gathering evidence make sense given what they know in a particular domain. However, these aspects of evidentiary reasoning remain problematic at more advanced levels of science learning. For instance, disciplinary research traditions develop contextualized internal norms and standards for sampling, which include knowledge of appropriate sample sizes but also such things as what intervals or range of values to sample. Students often have not learned (or have learned but do not remember to contextually employ) such disciplinary norms as they engage in inquiry. While students thinking about natural selection might know that they need to look at survival data over a time span rather than a single point in time, they often pick a time span that is too short to observe evolutionary adaptations because they do not consider the reproductive cycles of a particular species and how long it will take for several successive generations of that species to reproduce. To support a more effective methodological framing of student inquiry from evolutionary data bases such as the Galapagos finch simulation (Howard Hughes Medical Institute, 2015), disciplinary scaffolds might explicitly prompt students to consider the time it takes for a species to produce a new generation of members, and to consider how many generations they would need to observe in order to draw conclusions about evolutionary change.

3. *Evidence to Theory* relationships include disciplinary contextualizations that constrain the interpretation and evaluation of evidence gathered from a particular set of investigations. They involve evaluations of the evidence along such dimensions as the consistency of evidence across related experiments, strength of effects, boundary conditions, relationships to a previously established body of evidence in the field, or relationships to some set of disciplinary models. In the practice of science, evaluations of evidence are highly contextualized in disciplinary knowledge. For example, a member of our research team (Pelaez) asked undergraduate biology students to design an aqueous media for diluting, preserving, and observing red blood cells from specific animals (rabbit, cow, goose, chicken, etc.). Preliminary analyses of student work suggest they had trouble integrating pH and osmolyte concentration variables in their treatment of blood cells from different animals (Pelaez & Liu, in preparation). Later, given a table of normal blood serum parameter ranges including pH and osmolyte concentration measures (extracted from published research studies), they had trouble selecting and integrating all the relevant blood serum parameter evidence for clustering species on an evolutionary tree. In constructing a tree for a set of species including the birds and mammals, some compared absolute differences in the numerical upper value for some osmolytes with overlapping ranges, which is less meaningful than small differences in other osmolytes with non-overlapping ranges). Furthermore, even those who successfully clustered species based on the blood serum evidence had trouble integrating all of the evidence. Many drew an evolutionary tree with guinea pigs and rabbits correctly clustered on a branch with a more recent ancestor than with a cow, but they ignored evidence for putting the chicken and goose together on another branch that shares an even more distant common ancestor with the rabbit, guinea pig, and cow (Pelaez & Liu, in preparation). Although the students were given generic prompts to consider "variability" in their data set, had learned about the specific osmolytes under consideration in prior coursework, and had studied the chronology of evolutionary processes, they did not spontaneously use disciplinary knowledge to contextualize their interpretations of the evidence in the lab. Disciplinary scaffolds to support students' evidentiary reasoning in this instance would include prompts to first identify and explain how and why the range and distribution of serum pH and osmolyte values for each animal differs, second to consider what magnitude of differences in value would be considered significant by biologists, and third to explain their findings in terms of the chronology of evolutionary processes that produced diversification (Kong et al., 2017) resulting in the different mammal and bird species. Making such considerations explicit should help students interpret the evidence for different evolutionary relationships in ways that are more consistent with disciplinary norms of biology.

Although, educators have made important strides in trying to understand and support the development of evidentiary reasoning in science learning, we propose that in order to support sophisticated epistemic reasoning in the teaching and learning of science, educators must unpack the notion of evidence itself and reconnect it to its disciplinary contexts.

References

Berland, L. K., & McNeill, K. L. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. *Science Education, 94*, 765-793.

Cavagnetto, A. (2010). Argument to foster scientific literacy. *Review of Educational Research, 80*, 336-371.

Chapman, R. & Wylie, A. (2016). *Evidentiary reasoning in archaeology*. London, UK: Bloomsbury.

Chinn, C. A., Rinehart, R. W., & Buckland, L. A. (2014). Epistemic cognition and evaluating information: Applying the AIR model of epistemic cognition. In D. Rapp and J. Braasch (Eds.), *Processing inaccurate information* (pp. 425-453). Cambridge, MA: MIT Press.

Duschl, R. A. (1990). *Restructuring science education*. New York: Teachers College Press.

Feinstein, N., Allen, S., & Jenkins, E. (2013). Outside the pipeline: Reimagining science education for nonscientists. *Science*, *340*, 314–317.

Ford, M. (2008). "Grasp of practice" as a reasoning resource for inquiry and nature of science understanding. *Science & Education, 17*, 147-177.

Ford, M. J. (2012). A dialogic account of sense-making in scientific argumentation and reasoning. *Cognition and Instruction*, *30*, 207–245.

Galison, P. (1997). *Image & logic: A material culture of microphysics*. Chicago, IL: University of Chicago Press.

Giere, R. N. (2010). An agent-based conception of models and scientific representation. *Synthese*, *172*, 269-281.
Gooding, D. (1990). *Experiment and the making of meaning*. Dordrecht, the Netherlands: Kluwer.
Howard Hughes Medical Institute (2015). Evolution in action: Data analysis. *Biointeractive Resources.* http://www.hhmi.org/biointeractive/evolution-action-data-analysis
Kali, Y., & Linn, M. C. (2008). Technology-enhanced support strategies for inquiry learning. In D. Jonassen, M. J. Specter, M. Driscoll, M. D. Merrill, & J. van Merrienboer (Eds.), *Handbook of research on educational communications and technology* (pp. 145-161). New York: Taylor & Francis.
Kong, Y., Thawani, A., Anderson, T.R., & Pelaez, N. (2017). A model of the use of evolutionary trees (MUET) to inform K-14 biology education. *The American Biology Teacher*, *79*, 79-88.
Koslowski, B., Marasia, J., Chelenza, M., & Dublin, R. (2008). Information becomes evidence when an explanation can incorporate it into a causal framework. *Cognitive Development, 23,* 472-487.
Latour, B. (1987). *Science in action*. Cambridge, MA: Harvard University Press.
Lehrer, R., & Schauble, L. (2006). *Cultivating model-based reasoning in science education*. Cambridge: Cambridge University Press.
Manz, E. (2012). Understanding the codevelopment of modeling practice and ecological knowledge. *Science Education*, *96*, 1071-1105.
Manz, E. (2015a). Representing student argumentation as functionally emergent from scientific activity. *Review of Educational Research*, *85*, 553-590.
Manz, E. (2015b). Resistance and the Development of Scientific Practice: Designing the mangle into science instruction. *Cognition and Instruction, 33*, 89-124.
Masnick, A., & Klahr, D. (2003). Error matters: An initial exploration of elementary school children's understanding of experimental error. *Journal of Cognition and Development, 4*, 67-98.
Mayo, D. G., & Spanos, A. (2010). *Error and inference*. Cambridge University Press.
McNeill, K. L., & Berland, L. (2017). What is (or should be) scientific evidence use in k-12 classrooms? *Journal of Research in Science Teaching, 54*(5), 672-689.
National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas.* Washington, DC: National Academies Press.
National Research Council. (2015). *Guide to implementing the Next Generation Science Standards*. Washington, D.C.: National Academies Press.
Passmore C., & Svoboda J. (2012). Exploring opportunities for argumentation in modeling classrooms. *International Journal of Science Education*, *34*, 1535–1554.
Pelaez, N. J., & Liu, C. (in preparation). *Opportunities and problems with a model of the use of evolutionary trees (MUET) for tree thinking in the laboratory classroom*. Unpublished manuscript.
Samarapungavan, A. (2011). Ontological assumptions about species and their influence on children's understanding of evolutionary biology. In R. Taylor, & M. Ferrari (Eds.), *Epistemology and science education: Understanding the evolution vs. intelligent design controversy.* New York: Routledge.
Samarapungavan, A. (in press). Construing scientific evidence: The role of disciplinary knowledge in reasoning with and about evidence in scientific practice. In K. Engelmann, F. Fischer, J. Osborne, & C. A. Chinn (Eds.), *Scientific reasoning and argumentation: The roles of domain-specific and domain-general knowledge.* New York: Routledge.
Sandoval, W. (2014). Science education's need for a theory of epistemological development. *Science Education*, *98*, 383-387.
Sandoval, W. A., & Çam, A. (2011). Elementary children's judgments of the epistemic status of sources of justification. *Science Education*, *95*, 383-408.
Sandoval, W. A., & Reiser, B. J. (2004). Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education*, *88*, 345-372.
Solomon, M. (2015). *Making medical knowledge*. Oxford: Oxford University Press.
Staley, K. W. (2004). *The evidence for the top quark*. Cambridge, UK: Cambridge University Press.
Tenopir, C., King, D. W., Boyce, P., Grayson, M., & Paulson, K. L. (2005). Relying on electronic journals: Reading patterns of astronomers. *Journal of the Association for Information Science and Technology*, *56*, 786-802.
Toth, E., Suthers, D., & Lesgold, A. (2002). "Mapping to know": The effects of representational guidance and reflective assessment on scientific inquiry. *Science Education, 86*, 264-286.

Acknowledgements

The EBE project is supported by a grant (# 1661124) from the National Science Foundation. The opinions expressed are those of the authors and do not represent views of the National Science Foundation.