

Intertwining evidence- and model-based reasoning in physics sensemaking: An example from electrostatics

Rosemary S. Russ and Tor Ole B. Odden

*Department of Curriculum and Instruction, University of Wisconsin, Madison,
225 North Mills Street, Madison, Wisconsin 53706, USA*

(Received 22 March 2017; published 9 August 2017)

Our field has long valued the goal of teaching students not just the facts of physics, but also the thinking and reasoning skills of professional physicists. The complexity inherent in scientific reasoning demands that we think carefully about how we conceptualize for ourselves, enact in our classes, and encourage in our students the relationship between the multifaceted practices of professional science. The current study draws on existing research in the philosophy of science and psychology to advocate for intertwining two important aspects of scientific reasoning: *using evidence from experimentation* and *modeling*. We present a case from an undergraduate physics course to illustrate how these aspects can be intertwined productively and describe specific ways in which these aspects of reasoning can mutually reinforce one another in student learning. We end by discussing implications for this work for instruction in introductory physics courses and for research on scientific reasoning at the undergraduate level.

DOI: [10.1103/PhysRevPhysEducRes.13.020105](https://doi.org/10.1103/PhysRevPhysEducRes.13.020105)

I. INTRODUCTION

Reformers have repeatedly called for science instruction to move away from teaching students an inert body of facts and toward engaging them in the process of scientific knowledge construction. In the physics classroom, this means that we, as instructors, are increasingly tending away from teaching students disconnected equations, definitions, and rote problem-solving techniques, and towards helping them think “like physicists”.

Educators have advanced a variety of constructs to operationalize this call. In the early 1900s, Dewey used the notion “science as method” to argue that students need to understand how science generates reliable knowledge [1], and later Schwab introduced “science as enquiry” to highlight the varied and theoretically laden activities of scientists’ knowledge generation [2]. At the K-12 level, policy makers and standards documents continue to push for science learning grounded in “scientific practices” [3] that engage students in the work of professional science.

Undergraduate science, technology, engineering, and mathematics (STEM) programs also endeavor to teach more than just science content by emphasizing the “authentic practice of science and thinking scientifically” [4] (p. 020111-1). Physics education research (PER), in particular, has devoted substantial attention to the ways in which physics instructors can help students cultivate

“scientific reasoning,” “scientific thinking,” “reasoning skills,” and “scientific abilities” [4–11]. These reasoning skills, researchers argue, may help students to make decisions, interpret evidence, or solve complex problems even long after the details of their physics content knowledge have faded from memory [6].

Together this suite of constructs embodies a systematic effort to help students understand and engage in the knowledge construction work of professional scientists [12,13]. This goal is often described as a shift in students’ epistemic beliefs or personal epistemologies—that is, students should develop specific knowledge, beliefs, and attitudes about scientific knowledge and science learning [14]. For example, it is widely agreed that students should, over the course of taking a science class, come to understand science learning as an active process of meaning making rather than a passive process of receiving information [14,15]. Within undergraduate physics courses, scholars are dedicated to pursuing this epistemological goal [15–17] and have developed a variety of instruments to examine the success of instructional interventions in achieving it [18–21].

While there are many important differences among these efforts to promote productive epistemological views of science, all of them either tacitly or explicitly enumerate the multifaceted nature of the intellectual work of scientists. There is no one singular “scientific practice” [22], nor is there one all-encompassing “scientific reasoning skill” [23]. Thus, the fundamental approaches and assumptions that a physics student will learn are often radically different from those learned by students in biology, chemistry, or mathematics. Furthermore, even within a particular field, successful students will need to grasp and take up numerous different scientific practices and ways of thinking, often without explicit instruction.

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.

Since science is inherently a complex enterprise, the work of scientists must be embodied in classrooms as such. However, this multiplicity of practices leaves researchers and educators with an important theoretical and practical question: What is the relationship between the different facets of scientists' work? Are they linear? Iterative? Composite? Are they even independent and/or separable? If our ultimate aim is to help students develop productive epistemological understandings of science, then we need to be clear about the relationship between the various aspects of scientific reasoning. If we are not clear—both for ourselves and for our students—we run the risk of tacitly adopting overly reductionist models of these relationships which would dramatically misrepresent the science. In doing so, we would ultimately undermine our own efforts to support students in learning how professional scientists construct knowledge about the physical world.

In light of this need for clarity, the current study elaborates a model for the relationship between two important aspects of scientific reasoning: *using evidence from experimentation* and *modeling*. In the following section, we examine the literature on these two aspects separately, and theoretically argue for the need to intertwine them. Sections III and IV present a case from an undergraduate physics course that illustrates how these aspects can be intertwined productively, and describes specific ways in which these aspects of reasoning can mutually reinforce one another in this context. In Sec. V, we discuss implications for this work for instruction in introductory physics courses and for research on scientific reasoning at the undergraduate level.

II. THEORETICAL PERSPECTIVES

In their research documenting the knowledge-building practices of professional scientists, Klahr and Dunbar [24] examine the science studies literature and note that

The successful scientist, like the successful explorer, must master two related skills: knowing where to look and understanding what is seen. The first skill... involves the design of experimental and observational procedures. The second skill... involves the formation and evaluation of theory.

Although researchers study other aspects of scientific reasoning [25], the importance of these two skills for practicing scientists has resulted in researchers devoting substantial study to their applications for K-16 science learning. The first skill is generally described in terms of evidence-based reasoning, and the second in terms of modeling.

A. Evidence-based reasoning as scientific reasoning

Traditionally, the scientific reasoning literature has heavily emphasized the importance of *evidence-based*

reasoning. Specifically, this body of research has focused on how students conduct investigations in order to collect and analyze evidence [26–30]. This focus is reflected in Lawson's Classroom Test of Scientific Reasoning [7], which is commonly used in studies of undergraduate student reasoning. This multiple-choice diagnostic includes 11 questions (out of 24) that probe students' abilities to design or explain evidence from valid experiments. Student scores on this test have been used extensively either as dependent variables to demonstrate the efficacy of instructional interventions or as independent variables to explore the relationship between multiple aspects of learning [11,23,31].

Within PER, Etkina and colleagues [6,8,9] have also highlighted the importance of evidence in scientific reasoning. They have developed several instructional tools and rubrics to assess students' development in critical skill areas for scientists, which include categories such as “devise and test physical relationships” and “collect and analyze data” [8] (p. 020103-1). Of the seven scientific abilities these researchers identify, four directly involve experimentation and/or using empirical results to evaluate predictions. Furthermore, these researchers use students' lab reports from their experiments to evaluate reasoning abilities [6]. Underlying this work is an assumption that students should be able to collect and interpret evidence to make claims because that is a critical part of professional scientific practice.

B. Modeling as scientific reasoning

In addition to exploring how students make sense of data and experiments, researchers have also emphasized the importance of *modeling* in scientific reasoning [32–34]. In physics, Halloun and Hestenes [35–38] were among the first to highlight the centrality of modeling for physics learning, and that pedagogical approach has since become increasingly popular in undergraduate physics courses and laboratories [13,39–46]. Within K-12 science education, the Next Generation Science Standards contend that modeling should be a central activity of students in science classrooms, and so researchers have designed multiple instructional supports to support that activity [34,47–49].

Models are representations—either physical or mental—of the key features, relationships, and processes at work within a system that account for how the system “runs” [50,51]. Within the discipline of physics, these models are often *mechanistic*, meaning that they describe the underlying process for how a cause brings about an effect [52]. Scientific models are grounded in physical or natural phenomena in the real world, built from prior knowledge, and change dynamically as our knowledge of those phenomena changes [39]. This dynamic nature of modeling is central to using it to support learning; a crucial part of the modeling process involves identifying the different components of models—the key features, relationships, or objects involved in the

mechanistic model that produce the phenomenon of study [50]. However, as students learn about, observe, or remember additional components they must add these into their models, which can significantly change the model, in turn making room for new components. So, from a modeling perspective, students learning science should iteratively select or construct, critique, and revise models (and model components) either individually or in groups [34,35,47].

C. Separation of evidence-based reasoning and modeling

It is important to note that there has been little overlap between these two bodies of literature; in other words, educators and researchers often explore one aspect of scientific reasoning—either evidence-based reasoning or modeling—at a time. This separation may arise out of a need to restrict the space of scientific practices in order to make the task of studying (or teaching) the complexities of scientific reasoning tractable. However, this separation may also be rooted in the historical evolution of the field of science education, evidenced by the fact that most (though not all) of the literature on evidence-based reasoning was published prior to the literature on modeling.

For many years, research and policy documents shared the assumption, either tacitly or explicitly, that the most valuable facet of scientific reasoning, which distinguished it from other types of reasoning, was formal empirical investigation. In the mid-1990s and early 2000s, the standard for scientific inquiry was based on students' abilities to interpret evidence from formal experimentation by asking testable questions and negotiating among dependent, independent, and controlling variables during data collection [3,28,53,54]. Educational research explicitly studied and promoted those skills with new assessments and curricula [29,55,56]. Furthermore, researchers' assumptions of the primacy of evidence-based reasoning in science were evident from the contexts in which they investigated student reasoning: these studies almost always took place during experimentation.

However, in the mid-to-late 2000s there was a shift in researcher attention with more focus placed on student abilities to engage in theoretical modeling [34]. Although Halloun and Hestenes [36] had long since developed their Modeling Instruction paradigm in physics education, very little research attention outside of their work was focused on the ways in which students engaged in and learned from modeling. In fact, it was not until the work leading up to the Next Generation Science Standards [57] and the identification of modeling as an important scientific practice that research began theorizing about and examining modeling in science classrooms.

Whatever the reason for the separation—whether for ease of study or the historical development of the field—physics education has largely treated evidence-based reasoning and modeling as independent aspects of scientific

practice. Why should we be concerned about such a separation? First, separating the practices undermines student abilities to construct robust understandings of physical phenomena. Research in psychology and philosophy of science demonstrate how models and evidence are both essential components of understanding for laymen and scientists, respectively [52]. As such, separation inhibits conceptual learning—one of our main goals for undergraduate physics instruction.

Second, separating these practices conveys inaccurate epistemological messages to students about what constitutes knowing and learning in physics. As a physics faculty member at a colloquium recently told the first author, “the thing that sets physics apart from nonsciences is that we use evidence to construct models of the world.” If students do not understand the value of intertwining these activities, we have failed to convey to them the substance of our discipline and have presented a “grossly misleading view of the relationship between theory and evidence” [58] (p. 26). As a result, separating them undermines our field's goals for student epistemological learning about professional science practice [12,13].

Finally, if we separate modeling- and evidence-based reasoning, we may fail to tap into how students (and laypeople generally) naturally reason about the world around them. People commonly and spontaneously use both model-based and evidence-based reasoning to make sense of causal situations [30,59,60]. For example, people treat both covariational and mechanistic information as essential to their reasoning both in everyday interpersonal situations (e.g., choosing whether to eat a particular food) [59] and in everyday physical situations (e.g., understanding variation in gas mileage) [60]. Given that evidence- and model-based reasoning are used together in life outside of school, separating them in the classroom portrays physics as epistemologically discontinuous with students' everyday reasoning. As a result, students may fail to tap into those reasoning abilities during physics class.

D. Intertwining evidence-based reasoning and modeling

Although historically research has treated these two types of student reasoning as independent, a few researchers examining undergraduate physics courses have recently tackled this misleading separation. They have begun to ask, if evidence-based reasoning and modeling are both important elements of physics learning, how are they related to one another? Initial studies have found that the two are not separable: we do not start by collecting evidence and then stop our evidence collection to select a final model based on that evidence, nor do we first select a model and then use it to conduct experiments without refinement. Instead, modeling- and evidence-based reasoning occur together as an iterative, “nonlinear, recursive process” [41] (p. 1).

Zwickl and Hu [39] systematize this iterative process in their Experimental Modeling Framework. They describe how students move fluidly between constructing and revising models of the physical system and the experimental or measurement system. Specifically, their case study shows how, over the course of 35 minutes of an upper-division laboratory, a student transitions 9 times between what we have called modeling- and evidence-based reasoning. Dounas-Frazer and colleagues identify similar patterns in students' joint use of these practices [41]. In other words, they provide evidence that these two practices are intertwined in the work of both scientists and students [32,33,52,61].

However, knowing *that* students transition fluidly between these two practices does little to help us understand precisely *how* these practices inform one another in student reasoning. What role does evidence play in modeling? And how does modeling affect evidence collection or interpretation?

Research from psychology and philosophy of science indicate at least three ways in which intertwining evidence-based reasoning and modeling may support scientists' and students' reasoning. First, we use plausible models to determine which evidence to collect and which variables are likely to yield causal information. As Ahn and Kalish [62] point out, "To determine causal candidates for covariational analyses, one must start out with some understanding of causal mechanisms [models] because, otherwise, one soon runs into a computational explosion" (p. 216). It would take too much time to perform the "computational explosion" of all the possible controlled experiments on all possible causal variables for a given phenomenon. Developing models allows those engaged in scientific reasoning to narrow down the space of evidence desired and required.

Second, we combine models with evidence to make claims about causality. Even once we have data from an experiment, the presence of correlation does not necessarily indicate causality, nor does its absence necessarily imply a lack thereof [52,63]. However, when one knows the underlying model, one's judgment about whether two events are causally related is more likely to be accurate than if one were relying on covariation data alone [59,64] or model data alone. Koslowski *et al.* [65] make an even stronger claim: models do not just support data, they make data evidential. We cannot recognize data as evidence until we have an underlying model in place. Zwickl and his colleagues [39] in physics education call this type of connection *interpretation*. The "interdependence of explanation or theory [models] and evidence" [65] (p. 472), thus allows those engaged in scientific reasoning to be more confident in the appropriateness and sufficiency of their causal claims.

Third, we supplement evidence with mechanism (in the form of causal models) to make sense of new phenomena. Evidence is narrow in that it only gives insight into a particular case. In contrast, information about underlying

models is generalizable. Ahn *et al.* [59] claim that this property of generalizability is why mechanistic or model-based explanations are preferable to those based on covariation alone:

[M]echanism-based explanations are projective or generative in the sense that we can make predictions about novel situations... Suppose John had a traffic accident and the mechanism-based explanation for the event is that John is drunk. What if a particular factor were different? For example, what if John was wearing a hat? Would the effect still have occurred? The explainer already knows the preconditions for accidents involving drunk driving and he/she knows that the effect would occur by necessity as long as the conditions are satisfied. Therefore, the projection into this new situation can be easily carried out. (pp. 340, 341)

The example of John changing his hat may seem silly, but this silliness arises precisely because we have no mechanism relating hat wearing to car accidents. Zwickl and his colleagues refer to this type of connection between modeling the physical system and the measurement system as *prediction* [39]. Thus, causal modeling supports us in reasoning about future predictions from covariation data.

III. RESEARCH QUESTION

Despite the fact that research on scientists' and laymen's reasoning suggests dynamically shifting between evidence-based reasoning and modeling is essential for reasoning [52,63], we know little about precisely how students do so in their physics courses and what it affords them when they do. In this work, we ask, How do students shift between evidence-based reasoning and modeling as they learn about physical phenomena?

Rather than focusing on student difficulties with coordinating these two practices [39,66], the case we present here provides examples of ways students productively connect these two reasoning practices. Identifying such productive strategies allows us as a field to explicitly consider whether and how we want to make them an intentional part of the epistemology we convey to students.

Our analysis here is similar to work by Dounas-Frazer and his colleagues [40,41] who examined how students in an undergraduate physics lab engaged in model-based reasoning while troubleshooting an experimental apparatus and design. Like those researchers, we are interested in understanding how students' use of evidence contributes to their modeling of physical systems and vice versa. However, in their work students were explicitly engaged in a researcher-generated task of trying to make sense of data in a formal experiment (i.e., troubleshooting). In the example we explore, students are engaged in the task of answering a conceptual question by modeling the physical system using observational and everyday evidence. Although we might

expect some of the students' strategies to be similar, the different contexts, authorities, and goals for the activities across the two studies may result in different student engagement patterns.

IV. METHODS

A. Study context and data collection

As our context for observing the relationship between evidence and modeling, we explore student reasoning in a large-enrollment, algebra-based, reformed introductory physics course taught by a physics education researcher at a large public university on the East Coast. This particular iteration of the course was part of a broader, multiyear course redesign focused on highlighting and supporting the epistemological aspects of learning science.

The course was structured into three 50-minute lectures each week with approximately 200 students, most of whom were junior and senior health and life science majors. In addition to the large lecture, each week there was a required 2-hour laboratory and a 1-hour tutorial session that served 20–24 students. The tutorials were conceptual worksheets in which students worked in small groups with limited guidance from instructors [67]. However, these particular tutorials were modified to explicitly help students reconcile their everyday intuitions with the formal physics concepts. Lab and tutorial sessions occurred back to back each week, with students working in the same small groups with the same instructors for the entire 3-hour block of time. Instructors were trained both to guide small-group work in lab and tutorial and also to facilitate large-group discussion between the various groups. All class activities—lecture, lab, and tutorial—were video recorded, and these recordings serve as the data source for this study.

B. Participants and case study selection

To identify how students might productively intertwine evidence and modeling, we conducted a case study analysis of a conversation between four students during tutorial. Given that only a few studies have yet examined the specific strategies of students using modeling and experimentation together, case study methodology is an appropriate early step. Case studies provide researchers with an understanding of the particulars and complexity of individual cases [68,69] and are not intended to be generalized. It is precisely these particulars and complexities of *how* students reason we wish to unpack in this work.

The specific case for this study was selected via intensity sampling [70]. We selected an example from the larger-project data that was rich with the phenomenon of interest but was not extreme in any sense. Specifically, we looked for candidate cases in which students were drawing on both evidence and models to make sense of a physical situation. Tutorials, in general and for this student group in particular, were known to provoke thoughtful discussions where the

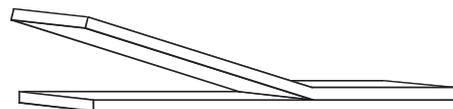


FIG. 1. Two pieces of tape are stuck together halfway.

students attempted to make sense of unfamiliar phenomena using mechanistic modeling and a variety of evidence. An initial viewing of the data confirmed that they were, at least intuitively, rich with students engaging in these practices.

The conversation analyzed here comes from a single group of four students working on one tutorial during the third week of the second semester of the course. This particular tutorial focuses on helping students develop a model for static electricity from their own experiences. Other research has explored student understanding of this concept more broadly [71,72]. During this session, the students discussed ideas consistent with that research both independently and with the instructor, but the conversation highlighted here involves them working alone.

The following is a synopsis of their discussion. The students begin the tutorial with some basic experimentation: they observe whether or not static electricity is created when various objects (two foam plates, a foam plate and a cloth, a foam plate and a sweater, or two hands) are rubbed together. After making these observations and recalling their everyday experiences with static electricity, the students spontaneously discuss *why* some pairs of objects charge and others do not. They come up with several possible models, none of which seem to account for all of their experiential data. In particular, they have trouble reconciling an observation involving two pieces of scotch tape that are placed on a table, one on top of the other, so that the sticky side of one is touching the slick side of the other (Fig. 1).

During lecture and again in tutorial, the students observe that these two pieces of tape attract one another when they are pulled apart. The students return to this phenomenon repeatedly (in the excerpts below they refer to the phenomenon as “the tape”). After discussion with the instructor, the students decide that the two objects must be made from different materials (“provide two different environments for the electrons”) in order to produce charge. In addition, they describe that electrons jumping from one object to the other create sparks and a resultant charge.

C. Data analysis

We transcribed the video recording of the tutorial conversation to include both audible speech and iconic gestures students used to convey meaning to their peers. We conducted fine-grained discourse analysis on each student utterance in the transcript. For the purposes of this analysis, we define an utterance as a complete conversational turn at talk, regardless of its length or any pauses

TABLE I. Mechanism coding scheme.

Code	Description	Example
Describing target phenomenon for the model	Students state or demonstrate the particular phenomenon or result they are trying to explain	“So the two pieces of tape attract.”
Describing set-up conditions for the model	Students identify particular enabling conditions of the environment that must be met at the beginning in order for the mechanism to run	“This is like after its [the tape] already been stuck together and you’re pulling it apart.”
Identifying entities in the model	Students articulate objects that affect the outcome of the phenomenon	“The actual electrons go to the other one.”
Identifying activities in the model	Students’ actions and interactions that occur among the entities, the things that entities “do”	“The electrons jump from one piece of tape to the other one.”
Identifying properties of entities in the model	Students articulate general features of entities that are necessary for this particular mechanism to run	“One side of the tape is more prone than the other to accepting or donating electrons to the surface.”
Identifying organization of entities in the model	Students describe or model where entities are located spatially and how they are organized	“Only the inside surfaces that touch are different, not the whole piece of tape.”
Chaining: Backward or forward in the model	Students use knowledge about the causal structure of the world to make claims about what must have happened previously to bring about the current state of things (backward) or what will happen next given that certain entities or activities are present now (forward).	“The electrons leave one piece and go to the other one so that this whole piece has more positives, and this whole piece has more negatives. So it cannot be just the inside surfaces that are different.”

during its completion. A change in speaker marks the end of an utterance.

After segmenting the transcript into utterances, the first author and a second researcher coded the data along two different dimensions. For the first dimension, we identified utterances in which students drew on empirical evidence; for the second, we identified moments in which students used mechanistic reasoning to describe model components and relationships.

1. Coding for use of evidence

We counted as evidence any time a student referenced a specific observable phenomenon regardless of whether they recalled it from their everyday experiences outside class or derived it from the experimental equipment during the tutorial. This definition of evidence is different from that used in other literature [39,40,66] because in this context the students are not engaged in formal experimentation. Instead, they are marshaling their own evidence—instances of specific phenomena in the world—to build their understanding of static electricity.

From the physicist’s perspective, combining these two types of evidence in our analysis here may seem problematic. As physicists, we treat evidence gathered anecdotally from everyday life very differently from evidence gathered in a controlled experiment; one has more reliability than the other and as such holds more empirical strength in our reasoning about phenomena. However, it is not clear that students assign empirical strength in the same way. In fact, it is even possible that because their everyday experiences

are authentic (e.g., “I have been shocked plenty of times outside of a ‘controlled’ experiment”) and repeated (e.g., “I have been shocked many times in many different places”), students may treat them as more “real” and thus more reliable than evidence gathered in the classroom.

A preliminary analysis of this group’s conversation suggests they treat these two types of evidence similarly. In their talk, students draw on both kinds of evidence, often using one type of evidence to make sense of the other. For example, when making sense of the “tape experiment” from class, Jennifer and Emily reference their everyday experience, saying, “Well we didn’t rub the tape together we just stuck it together and pulled it apart. [But] like Saran Wrap, that already has like static in it... from rubbing against itself.” At times, they even refer to both types of evidence in the same utterance. For example, Emily compares her body as a conductor (everyday evidence) to a Styrofoam plate (classroom experiment) saying, “Your whole body is [only one thing]. [But] Like this is a plate and you can charge this [Styrofoam plate].” Because of this continuity in how these students treat anecdotal and experimental evidence, in this work we group them into a single category and analyze both together.

2. Coding for components of causal modeling

To code for instances of student modeling, we used a discourse analysis coding scheme developed from the philosophy of science literature [73]. Elsewhere, the first author has demonstrated the reliability and utility of this coding scheme for identifying evidence of student

reasoning about causal mechanisms for physical phenomena [50,74]. In this work, the coding scheme allows us to identify moments when students describe mechanistic model components and connections between those components. As such, it allows us to unpack the *process* of student reasoning about modeling, not just the final model they produce. Table I lists the mechanism (modeling) codes, a description of those codes, and hypothetical examples of the codes using student ideas about the tape-tape interaction question (Fig. 1).

After coding for both evidence use and identifying model components and relationships, we placed the codes for these two dimensions side by side to clearly highlight moments when students simultaneously engaged in evidence-based reasoning and mechanistic modeling. As such, the coding provided a starting point for closer, qualitative interpretive analysis of how the two scientific practices inform one another. We thus used the codes from the mechanism coding scheme to ground the interpretive analysis of the specific classroom episode and gain insights presented in Sec. V [75].

V. RESULTS

There are three main results in the sections that follow. First, we demonstrate that this group of students, when trying to make sense of a confusing phenomenon, spontaneously intertwines modeling and using evidence. Second, we confirm findings by others that students move quickly fluidly back and forth between modeling and using evidence over the span of their short conversation. Finally, we identify specific strategies these students use when intertwining modeling and using evidence, and the ways in which those strategies are productive for helping them in their sensemaking.

A. Spontaneous intertwining of evidence and modeling

The data include many instances of students using both evidence and causal modeling in their reasoning about static electricity. Across the approximately 15 minutes of the discussion analyzed here, there were 53 instances of students using evidence and 127 instances of them modeling. In addition, nearly half of the time students used evidence, they also—in the same utterance—used model-based reasoning. While this finding is not surprising given the case was selected to demonstrate these practices, it is nonetheless an important finding to highlight, given the focus in the literature on students having difficulties with this practice [66] and needing substantial scaffolding to do it well [9]. Not only were students able to reason about evidence and modeling together, they did so spontaneously without the prompting of either their instructor or the instructional materials. And there is no reason to suspect that these students are extraordinary in any way; they were

not known to their instructors as particularly thoughtful or invested students, nor did they regularly outperform their peers on class assessments.

B. Dynamic intertwining of evidence and modeling

Throughout the entirety of their discussion about electrostatics, these students reason about both physical models and evidence based on their formal and informal experiments. Additionally, Fig. 2 shows they quickly move back and forth between reasoning about evidence and reasoning about modeling.

This representation is consistent with timelines depicting student reasoning during formal experimental troubleshooting presented by both Zwickl and colleagues and Dounas-Frazer and colleagues [39–41]. However, they use horizontal blocks of color to represent spans of time when students are engaged in different kinds of reasoning. That representation is consistent with their analytic approach of cutting the transcript into equal segments of time and coding those blocks of time rather than individual utterances. In contrast, we coded individual utterances and represent that coding using individual points on the timeline.

Our choice of analytic method and representational form is purposeful. When using points on the graph and utterance-level coding—instead of blocks of time and time-segmented coding—it becomes clear that it is nearly impossible to pinpoint precise moments when students transition from modeling to evidence and back. Instead, students reason about models and mechanisms in one utterance and then draw on evidence *within that same*

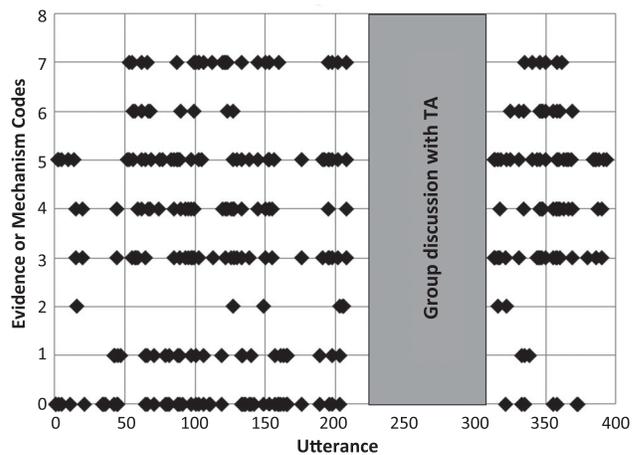


FIG. 2. Graph of student use of evidence and modeling over time in the conversation about electrostatics. Code (0) was for evidence and (1)–(7) are the mechanistic modeling codes: (1) describing target phenomenon, (2) set-up conditions, (3) identifying entities, (4) identifying activities, (5) identifying properties of entities, (6) identifying organization of entities, and (7) chaining. Utterances 230–310 were not coded because the teaching assistant (TA) was present and guiding much of the conversation.

utterance. This fluid movement suggests it may be inappropriate to model student reasoning about the physical world as involving “transitions” between two different activities of modeling- and evidence-based reasoning.

This graph, however, while useful, still does not indicate *how* students use modeling to inform their evidence-based reasoning or vice versa. It is too coarse a tool to delineate the nature of the relationship between the two. So, next, we situate coded student talk into a larger narrative story to provide qualitative detail on what is happening in those moments when students use both evidence- and model-based reasoning. This fine-grained analysis provides further evidence that it may be not be epistemologically appropriate to think of these two types of reasoning as separable activities that occur consecutively. Instead, we may need to think of them as simultaneous activities that co-occur as part of a larger process of making sense of phenomena.

C. Strategic intertwining evidence and modeling

In what follows, we present several episodes all from this student group’s discussion of electrostatics to demonstrate four of the ways in which they use empirical data together with modeling to build on their understanding of electrostatics. Doing so allows us to flesh out the work these two practices do for one another when used together in student reasoning.

These four episodes, while not generalizable, show that the relationship between modeling and evidence is strongly recursive. Although they are not meant to span the space of possible evidence-modeling relationships, each of these examples shows a way in which modeling and evidence can mutually influence one another. In the first three examples, the students are using evidence to build or flesh out models, thereby proceeding from evidence to model building. In the last example, they use a model to determine what kind of evidence they should look for next, proceeding from a model to evidence gathering. However, while the first three examples show cases of students proceeding from evidence to additions or revisions of models, it is reasonable to assume that the students in question came into the interaction with some kind of model (partial or full) of the phenomenon already in place, so these relationships, too, are recursive.

1. Using evidence to motivate the need for a model

Early in the electrostatics tutorial, the students observe that rubbing a foam plate with a cloth creates static but rubbing two hands together does not. In the following segment, these students use this evidence to drive the need for a model of sparks:

Jennifer: So why, I don’t get this. What’s the rational explanation for why there’s no sparks between our hands? Is it ‘cause they’re moist?

Riku: Because—

Emily: Because of moisture.

Riku: —there’s so much moisture in the air and it’s a conductor so...

Emily: The air and in your skin. It’s mostly in your skin.

Emily: [To Jennifer] But why does, is it just moisture?

Jennifer’s observation of these two contrasting results prompts her to seek “the rational explanation” for the phenomenon. She both expects that an underlying model exists and wants to know it. Her next question about moisture gives a further indication as to the kind of explanation she wants. She is not looking for an explanation based on physical laws or rules (e.g., Newton’s laws or Kirchoff’s loop rule), such as “sparks always sum to zero” or “sparks move from high to low concentration.” Instead, she wants an explanation involving physical causes, and offers moisture as one possible relevant component in that model. Other members of the group suggest potentially relevant properties of moisture and conductivity, which prompts Emily to express her dissatisfaction and ask another question: how can moisture by itself constitute the model for static electricity? Jennifer and Emily are, in other words, explicitly requesting other entities or activities in the model to account for the data.

There are other cases of students asking for model components based on evidence throughout the conversation. For example, Riku later asks, “But why, why wouldn’t you get two charges on your hands if you can get two different charges on other things?” He wants to know why the model works for some situations but not others. In terms of the mechanism coding scheme, he implicitly asks what setup conditions, entities, and activities are required for the static electricity model to run. Second, when students observe that rubbing together two plates does not relate charge even though a plate rubbed with a cloth does, Emily asks, “Why if you rub the plates aren’t they excite, like, if you rub things together we say that electrons get excited and whatever a charge. But like, why wouldn’t it work with plates?” Emily suggests that the components of the model they have discussed, i.e., electrons (entities) getting excited (activity), are insufficient to account for this new result.

How is this relationship between evidence and modeling productive for students? The process of using anomalous evidence to question underlying models is an essential part of professional science practice, and as such can be productive for students in physics class. For example, Hammer [76] describes Bell’s discovery of pulsars where an unexpected observation of a regular radio signal prompted her to seek a causal explanation for the signal. In this example, students’ uneasiness or confusion around evidence—like Bell’s unexpected radio signal—spontaneously prompts them to search for model-based explanations for their results. This relationship supports student learning by providing organic motivation for students to generate

models *that are meaningful to them*. Other researchers have described similar instances when students use surprising or anomalous data to develop or change their theories and models for how the world works [30]. In these instances, students seek to satisfy their own curiosity, not the curiosity of the curriculum.

Although our example is spontaneous, other researchers have intentionally fostered this type of reasoning in undergraduate physics laboratories in the form of troubleshooting. This line of research argues that when students get a result that seems incongruous with their understanding, it is important for them to reason about the physical system to understand the origin of the result [39,41]. Specifically, these students' reasoning process may be related to what Dounas-Frazer and colleagues call *isolating subsystems as (mal)functioning* in formal experimentation.

2. Using evidence to hypothesize model components

Throughout the tutorial, the students in the group describe familiar instances of static electricity by observing and recalling evidence from their classes and everyday lives. In this excerpt, they recall the tape-tape experiment (Fig. 1) from lecture. They charge Styrofoam plates by rubbing them with a cloth but then also discuss and observe how rubbing your hands together or rubbing two plates together does not charge said objects. The students use this collection of evidence to try to construct a rule for when static electricity occurs, a key component of their model of static electricity.

Riku: Maybe it has to be two different objects. 'Cause its balloon to head, cloth to plate.
 Emily: But tape to tape remember?
 Riku: But it was on a desk.
 Emily: I'm pretty sure if you just took the tape [Laughter as she pulls her hand apart like she's holding tape.] The tape desk interaction. [Laughter.]
 Jennifer: The stupid tape.
 Riku: Could be.
 Jennifer: Well we didn't rub the tape together we just stuck it together and pulled it apart. 'Cause like Saran Wrap, that already has like static in it.
 Emily: From rubbing against itself when you pull it.
 Anika: So.
 Riku: Maybe foam's not a good conductor. [Putting finger quotes around conductor.] Or Styrofoam.
 Emily: No because why is, why is it a good conductor when you rub it with something else?
 Riku: Mmmm. [Shrugs his shoulders.]

In this episode, the students are using evidence to generate ideas for properties that may be relevant to their model. They cite the balloon-plate and cloth-plate interactions to support the entity-level property of different materials, tape-tape interaction and Saran Wrap against itself to support the

activity of rubbing, and plate-plate lack of interaction for the property of conductors. The students thus have ample experience with static electricity to draw on, and those different pieces of evidence suggest different model components that may be important for how the mechanism runs.

How is this relationship between evidence and modeling productive for students? There are times when a new physical situation is initially so disjointed from our previous experience that we may be uncomfortable even hazarding a guess about the underlying physical model. However, elements or properties that are consistent across a variety of different cases may indicate key parts of that model and provide an entry point for student reasoning. By first surveying informal evidence, these students are gaining insight into unfamiliar components of their model for static electricity. This strategy is similar to what Klahr and Dunbar [24] refer to as *inducing a hypothesis frame*, and what Dounas-Frazer [41] and his colleagues observe in their experimental context when students *brainstorm potential causes or discern functions of systems/components*.

3. Using evidence to elaborate further model components

Having hypothesized and discussed various model components, the students in this group now begin to use evidence to flesh out their models. Specifically, they focus on one piece of evidence everyone agrees on (when people touch their hands together they can and do shock one another) and explore the implications of that evidence for their model.

Jennifer: Well I can, if I take my hand and rub on her sweater then I can touch something spark. I don't think you can do two things that are made out of the same thing.
 Emily: [Simultaneously with Jennifer] Well your whole body is [only one thing]. Like this is a plate and you can charge this [Styrofoam plate]. But, but your whole body is kind-of like a conductor really.
 Emily: [To Riku] Yeah that's true. Because well it could go back to like what you [Jennifer] said before with like it travels through your body.
 Riku: But like, I mean, if you touch someone's hand, [it's fine], you shock each other.
 Jennifer: Yeah.
 Emily: You can create a charge maybe but it doesn't stay there?
 Riku: I think, I think [it distributes].
 Emily: Can you create a charge though?
 Riku: I don't know if I can get like two different charged hands.
 Jennifer: No. I don't think you can. I don't think just from your hands. [Rubbing her hands together] I think, I think what makes the charge is two different

substance things. ‘Cause if you figure you walk around and then shock somebody, the charge is—

Anika: It’s the floor. Yeah.

Jennifer: —probably created from the carpet on your feet and then you carried it through [your body to your hands].

In the previous example, when the students considered new evidence, they hypothesized new model components that were independent from others being considered. In this episode, however, Jennifer does more than simply add additional components to the model to account for Riku’s evidence. Instead, she elaborates on her original component (“different materials”) by describing new components that account for *how* the model runs, assuming that the “different materials” component is still relevant. Jennifer works with Emily to describe how charge (underlying entity) is created between the carpet and feet, carried (activity) through the body (organization of entities) to the hand, which then causes a shock with another hand. Although the evidence initially seems discontinuous with her model component (shocking another person involves two of the same material), she articulates that the charge is actually created between the carpet and feet—two different materials—thus lending plausibility to her original model component.

How is this relationship between evidence and modeling productive for students? When students have multiple possible components of a model, they must examine the plausibility of each because it is unlikely that every hypothesized component is part of the model. To do so, students articulate exactly how and when a particular component works to produce one piece of evidence, thereby showing how it might possibly account for the other evidence. Koslowski [52] describes this process as involving “specifying more precisely the situations in which a mechanism [model] will produce an expected covariation” (p. 96), and Schauble [77] observed students spontaneously constructing underlying models when they came across a surprising covariation. Etkina and colleagues [8] elicited this type of reasoning by engaging undergraduate physics students in “surprising data tasks” (p. 020103) that helped them modify and further develop their existing explanations. Across all of these cases, students compare potentially incongruous data to pare down the number of hypothetical variables by lending plausibility to specific model components that fit within a larger mechanism. Doing so allows students to further flesh out their mechanistic models to account for evidence.

4. Using models to guide the search for new evidence

After discussion with the teaching assistant, the student group articulates an electron transfer model of static electricity: two different materials can charge one another because electrons move from one to the other based on an

unspecified preference. More specifically, when two pieces of tape are stuck together and then pulled apart, electrons jump to either the sticky side of the top tape or the slick side of the bottom tape. This leaves one piece of tape positive and the other negative, causing them to attract. However, when the students encounter a specific question regarding the location of the charge in Fig. 1, they notice a gap in their model of static electricity.

Question: A bottom (B) piece of tape and a top (T) piece of tape are separated halfway as shown (Fig. 1). Use “+” and “−” symbols to indicate the parts of the tapes that are charged and the type of the charge on a diagram like the one below.

In response to this question, the students use their proposed model to determine what kinds of new evidence they should look for in order to elaborate on their model.

Emily: So basically one is plus, one is minus, but does it matter which—

Anika: Which piece? [?]

...

Jennifer: So does that just mean that the inside surfaces that touch are different? Or like the entire piece of tape is now [?]

...

Emily: [Pulls two pieces of tape off the table and away from each other, then brings both sides near each other]

No see, but both sides stick.

Jennifer: So you think it goes like, it goes through the—

Emily: Like slick to slick still touches, still is attracted [Anika nodding]. So it must charge the whole thing.

Jennifer recognizes the relevant entities in the model, which she calls the “pluses” and “minuses” (later she identifies these as electrons). However, she is not sure what exactly happens with those electrons and how they are organized on the tape. To fill in this gap in the model, Emily spontaneously collects new empirical evidence to test whether the charge is isolated on one side of the tape or distributed across both surfaces. This test allows the students to infer another component of the model (organization of entities) by chaining from her evidence—since both sides of the tapes attract, both sides of the tape must be charged. The students continue the conversation.

Anika: So like along this whole surface of this top one would be plus?

Emily: I think the whole piece.

Jennifer: Once you open it, right like, this part down here [Unpulled apart section of tape] isn’t charged yet right? ‘Cause you haven’t like taken it apart to make the—[Pulling her hands away from each other.]

- Emily: So, can, alright. Am I wrong in thinking of it like the actual electrons go to the other one?
- Anika: That's what I thought. They like jump from one piece of tape to the other one.
- Jennifer: I thought that when we like—
- Emily: To the other so that like the whole piece has more positives, and this whole piece has more negatives.
- Anika: Has more negatives, right.
- Jennifer: Okay.

Although Emily is satisfied by the empirical test, Anika and Jennifer remain puzzled. It is not until Emily explains why her conclusion about the entity organization must be true by chaining it from Anika and Jennifer's own model of jumping electrons that the gap in their mechanism is filled. In these segments, Jennifer's questions about the underlying mechanism (where are the electrons?) prompt Emily to construct a meaningful empirical test. However, the ensuing observations are not convincing for the students until she connects the findings of that test to other components of their model (the activity of electrons jumping).

How is this relationship between evidence and modeling productive for students? When students recognize gaps in their models, they may spontaneously search for evidence to elaborate their underlying models. Similarly, scientists often look to empirical results to "provide direct evidence for specific mechanism components" [78] (p. 35), or "fill in unspecified variables" [24] (p. 8) in their model. This relationship between mechanism and empirical results is different from those described in previous sections: it involves beginning with some sense of the underlying mechanism and seeking direct evidence for components as opposed to collecting indirect anecdotal evidence to form a model. In keeping with the cyclic nature of the relationship, students may then supplement their results by accounting for them with a complete mechanism.

Steinberg and Cormier [4] describe a similar type of student reasoning as the most sophisticated in their data. This reasoning occurs when a student "cites observations or experiments distinguishing between 2 models and supports choice with proper explanation" (p. 020111-3). Notice, however, that students in our example are not just citing observations, they are constructing their own observations to distinguish between two possible model components (organization of entities).

VI. DISCUSSION

In this work, we explore how undergraduate physics students intertwined evidence-based and model-based reasoning to construct explanations for static electricity. In doing so we verified the spontaneous existence of a phenomenon other researchers have observed in purposefully designed settings [40,41]. Additionally, we presented

four cases of *how* students use modeling and evidence to inform one another in their reasoning.

In the first three cases, the students used evidence as a resource for their model building. Specifically, they used evidence to *motivate* the need for a model, *hypothesize* which new components might be important or useful for that model, and *elaborate* on the model components they all agreed upon. From a mechanistic perspective, each of the three cases addresses a different way in which evidence can aid in the model-construction process. First, *motivating* the need for a model involves identifying both the phenomenon to be modeled and the starting conditions for a possible mechanism, as Jennifer did in the first selection when she pointed out that rubbed hands produce no static. Second, *hypothesizing* new components requires one to identify the entities and their associated actions, properties, and organization, as the group did in the second example when they listed the various objects that did or did not cause static when rubbed together. Finally, *elaborating* on components involves chaining backwards or forwards, as Jennifer did in the third example, where she described how a shock is caused by charges from the carpet being carried up to one's hands.

In contrast to the first three cases, the fourth case illustrates how the students were able to use their partially constructed model to guide the search for new physical and/or experiential evidence. Here, from a mechanistic perspective, students capitalized on their constructed mechanism—with its specific entities—to guide the next steps in their investigation so that they could identify the activities and entity organization.

The relationships we present here are asymmetrical in that we describe three ways students move from evidence to modeling but only one from modeling to evidence. In doing so we do not mean to imply there is an inherent asymmetry or that there are no other relationships to uncover. Rather, our explication of these relationships in terms of the mechanism framework highlights several particular ways we might expect students might move from models to evidence. For example, researchers [30,65] have argued that people use models to evaluate evidence, i.e., to determine the usefulness and/or accuracy of a particular observation or experience. We would hypothesize that students might also use a model to distinguish different types of evidence on the basis of the entities, actions, or properties involved, determine the conditions under which evidence should be collected (organization), or decide how much evidence they need. Future work on this topic may additionally uncover many more such relations.

VII. CONCLUSION

In this work, we have presented several cases in which the relationship between modeling and evidence is recursive; students use models to find evidence and evidence to

construct models. Furthermore, we have demonstrated the fluidity with which students move between these two types of reasoning.

The short time scale in which students move between these two types of reasoning, and the ease with which they do so, suggests that it may be inappropriate for us to think about evidence-based reasoning and model-based reasoning as two independent activities. Instead, it may be more appropriate to think of them as subcomponents of the larger activity of *sensemaking*. Researchers have described sense-making as a central practice in physics learning that involves adding elements and reconciling inconsistencies in one's knowledge framework through argumentation and explanation building [79–82]. Though the students in this case study were, as previously argued, intertwining evidence- and model-based reasoning, to focus just on these two elements would be to miss the larger-scale goal of their activity, which was to make sense of static electricity.

The shift to conceptualizing evidence-based and model-based reasoning as components of the larger practice of sensemaking, rather than separate activities, may seem semantic or arbitrary. However, we suggest that doing so gives researchers a more appropriate theoretical perspective on student reasoning, which has practical consequences: when we assume by default two things are separate, instances of separation are the norm, and instances of

them being used together are the exception. As such, we might tacitly assume the exception requires work to accomplish. If we assume that evidence-based reasoning and modeling are separate types of reasoning, then researchers may be predisposed to find student difficulties in doing the challenging work of using the practices together [5,66].

In contrast, if we assume by default we are looking at a sense-making practice with multiple components, then instances of that practice occurring (with all its components) are the norm and instances of separation are the exception. This type of lens leads us to assume that students will by default intertwine the practices and that failing to do so is the anomaly. In this way, researchers are primed to see strengths in sensemaking rather than weaknesses in connecting two disparate types of reasoning. Such a turn in theoretical stance affords us the opportunity to assume, and investigate the ways in which, students are capable of engaging in the epistemologically authentic work of reasoning about models and evidence in physics.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grants No. ESI-9986846 and No. REC-0440113.

-
- [1] J. Dewey, *How We Think* (D.C. Heath and Co. Publishers, Boston, 1910).
 - [2] J.J. Schwab, The teaching of science as enquiry, in *The Teaching of Science* (Harvard University Press, Cambridge, MA, 1962), pp. 3–103.
 - [3] National Research Council, *National Science Education Standards* (National Academy Press, Washington, DC, 1996).
 - [4] R. Steinberg and S. Cormier, Understanding and affecting science teacher candidates' scientific reasoning in introductory astrophysics, *Phys. Rev. ST Phys. Educ. Res.* **9**, 020111 (2013).
 - [5] N. G. Holmes and C. E. Wieman, Examining and contrasting the cognitive activities engaged in undergraduate research experiences and lab courses, *Phys. Rev. Phys. Educ. Res.* **12**, 020103 (2016).
 - [6] A. Karelina and E. Etkina, Acting like a physicist: Student approach study to experimental design, *Phys. Rev. ST Phys. Educ. Res.* **3**, 020106 (2007).
 - [7] A. E. Lawson, The development and validation of a classroom test of formal reasoning, *J. Res. Sci. Teach.* **15**, 11 (1978).
 - [8] E. Etkina, A. Van Heuvelen, S. White-Brahmia, D. T. Brookes, M. Gentile, S. Murthy, D. Rosengrant, and A. Warren, Scientific abilities and their assessment, *Phys. Rev. ST Phys. Educ. Res.* **2**, 020103 (2006).
 - [9] E. Etkina, A. Karelina, and M. Ruibal-Villasenor, How long does it take? A study of student acquisition of scientific abilities, *Phys. Rev. ST Phys. Educ. Res.* **4**, 020108 (2008).
 - [10] A. E. Lawson, The nature and development of scientific reasoning: A synthetic view, *Int. J. Sci. Math. Educ.* **2**, 307 (2004).
 - [11] J. C. Moore and L. J. Rubbo, Scientific reasoning abilities of nonscience majors in physics-based courses, *Phys. Rev. ST Phys. Educ. Res.* **8**, 010106 (2012).
 - [12] L. K. Berland, C. V. Schwarz, C. Krist, L. Kenyon, A. S. Lo, and B. J. Reiser, Epistemologies in practice: Making scientific practices meaningful for students, *J. Res. Sci. Teach.* **53**, 1082 (2015).
 - [13] E. Brewster, Modeling theory applied: Modeling Instruction in introductory physics, *Am. J. Phys.* **76**, 1155 (2008).
 - [14] B. K. Hofer, P. R. Pintrich, and R. Mayer, *Personal Epistemology: The Psychology of the Beliefs About Knowledge and Knowing* (Routledge, New York, 2004).
 - [15] C. S. Kalman, M. Sobhanzadeh, R. Thompson, A. Ibrahim, and X. Wang, Combination of interventions can change students' epistemological beliefs, *Phys. Rev. ST Phys. Educ. Res.* **11**, 020136 (2015).

- [16] S. Sharma, P. K. Ahluwalia, and S. K. Sharma, Students' epistemological beliefs, expectations, and learning physics: An international comparison, *Phys. Rev. ST Phys. Educ. Res.* **9**, 010117 (2013).
- [17] P. Zhang and L. Ding, Large-scale survey of Chinese precollege students' epistemological beliefs about physics: A progression or a regression?, *Phys. Rev. ST Phys. Educ. Res.* **9**, 010110 (2013).
- [18] C. W. Shubert and D. C. Meredith, Stimulated recall interviews for describing pragmatic epistemology, *Phys. Rev. ST Phys. Educ. Res.* **11**, 020138 (2015).
- [19] S. Yerdelen-Damar, A. Elby, and A. Eryilmaz, Applying beliefs and resources frameworks to the psychometric analyses of an epistemology survey, *Phys. Rev. ST Phys. Educ. Res.* **8**, 010104 (2012).
- [20] B. M. Zwickl, T. Hirokawa, N. Finkelstein, and H. J. Lewandowski, Epistemology and expectations survey about experimental physics: Development and initial results, *Phys. Rev. ST Phys. Educ. Res.* **10**, 010120 (2014).
- [21] B. R. Wilcox and H. J. Lewandowski, Research-based assessment of students' beliefs about experimental physics: When is gender a factor? *Phys. Rev. Phys. Educ. Res.* **12**, 020130 (2016).
- [22] National Research Council, *Next Generation Science Standards: A Framework for K-12 Science Education* (National Academies Press, Washington, DC, 2012).
- [23] V. P. Coletta and J. A. Phillips, Interpreting FCI scores: Normalized gain, preinstruction scores, and scientific reasoning ability, *Am. J. Phys.* **73**, 1172 (2005).
- [24] D. Klahr and K. Dunbar, Dual space search during scientific reasoning, *Cogn. Sci.* **12**, 1 (1988).
- [25] A. L. Stephens and J. J. Clement, Documenting the use of expert scientific reasoning processes by high school physics students, *Phys. Rev. ST Phys. Educ. Res.* **6**, 020122 (2010).
- [26] K. Dunbar, How scientists really reason: Scientific reasoning in real-world laboratories, in *The Nature of Insight*, edited by R. J. Sternberg and J. E. Davidson (MIT Press, Cambridge, MA, 1995), pp. 365–395.
- [27] D. Klahr and M. Nigam, The equivalence of learning paths in early science instruction, *Psychol. Sci.* **15**, 661 (2004).
- [28] D. Kuhn, Children and adults as intuitive scientists, *Psychol. Rev.* **96**, 674 (1989).
- [29] W.-M. Roth and A. Roychoudhury, The development of science process skills in authentic contexts, *J. Res. Sci. Teach.* **30**, 127 (1993).
- [30] C. A. Chinn and W. F. Brewer, An empirical test of a taxonomy of responses to anomalous data in science, *J. Res. Sci. Teach.* **35**, 623 (1998).
- [31] L. Ding, Verification of causal influences of reasoning skills and epistemology on physics learning, *Phys. Rev. ST Phys. Educ. Res.* **10**, 023101 (2014).
- [32] R. Lehrer, L. Schauble, and D. Lucas, Supporting development of the epistemology of inquiry, *Cognit. Dev.* **23**, 512 (2008).
- [33] C. Passmore, J. Stewart, and J. Cartier, Model-Based inquiry and school science: Creating connections, *School Sci. Math.* **109**, 394 (2009).
- [34] M. Windschitl, J. Thompson, and M. Braaten, Beyond the scientific method: Model-based inquiry as a new paradigm of preference for school science investigations, *Sci. Educ.* **92**, 941 (2008).
- [35] I. Halloun, Schematic modeling for meaningful learning of physics, *J. Res. Sci. Teach.* **33**, 1019 (1996).
- [36] I. Halloun and D. Hestenes, Modeling instruction in mechanics, *Am. J. Phys.* **55**, 455 (1987).
- [37] D. Hestenes, Modeling games in the Newtonian World, *Am. J. Phys.* **60**, 732 (1992).
- [38] D. Hestenes, Toward a modeling theory of physics instruction, *Am. J. Phys.* **55**, 440 (1987).
- [39] B. M. Zwickl and D. Hu, Model-based reasoning in the physics laboratory: Framework and initial results, *Phys. Rev. ST Phys. Educ. Res.* **11**, 020113 (2015).
- [40] D. R. Dounas-Frazer, K. L. Van De Bogart, M. R. Stetzer, and H. J. Lewandowski, The role of modeling in troubleshooting: An example from electronics, in *Proceedings of the Physics Education Research Conference* (American Association of Physics Teachers, College Park, MD, 2015), pp. 103–106.
- [41] D. R. Dounas-Frazer, K. L. Van De Bogart, M. R. Stetzer, and H. J. Lewandowski, Investigating the role of model-based reasoning while troubleshooting an electric circuit, *Phys. Rev. Phys. Educ. Res.* **12**, 010137 (2016).
- [42] E. Brewster, L. Kramer, and G. O'Brien, Modeling instruction: Positive attitudinal shifts in introductory physics measured with CLASS, *Phys. Rev. ST Phys. Educ. Res.* **5**, 013102 (2009).
- [43] E. Brewster, V. Sawtelle, L. H. Kramer, G. E. O'Brien, I. Rodriguez, and P. Pamelá, Toward equity through participation in Modeling Instruction in introductory university physics, *Phys. Rev. ST Phys. Educ. Res.* **6**, 010106 (2010).
- [44] J. De La Garza and H. Alarcon, Assessing students' attitudes in a college physics course in Mexico, *AIP Conf. Proc.* **1289**, 129 (2010).
- [45] C. Fazio, O. R. Battaglia, and B. Di Paola, Investigating the quality of mental models deployed by undergraduate engineering students in creating explanations: The case of thermally activated phenomena, *Phys. Rev. ST Phys. Educ. Res.* **9**, 020101 (2013).
- [46] A. Traxler and E. Brewster, Equity investigation of attitudinal shifts in introductory physics, *Phys. Rev. ST Phys. Educ. Res.* **11**, 020132 (2015).
- [47] C. V. Schwarz, B. J. Reiser, E. A. Davis, L. Kenyon, A. Achér, D. Fortus, Y. Shwartz, B. Hug, and J. Krajcik, Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners, *J. Res. Sci. Teach.* **46**, 632 (2009).
- [48] L. Kenyon, C. Schwarz, and B. Hug, The benefits of scientific modeling, *Sci. Child.* **46**, 40 (2008).
- [49] C. M. Passmore and J. Svodoba, Exploring opportunities for argumentation in modelling classrooms, *Int. J. Sci. Educ.* **34**, 1535 (2012).
- [50] R. S. Russ, R. E. Scherr, D. Hammer, and J. Mikeska, Recognizing mechanistic reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science, *Sci. Educ.* **92**, 499 (2008).
- [51] C. Passmore and J. Stewart, A modeling approach to teaching evolutionary biology in high schools, *J. Res. Sci. Teach.* **39**, 185 (2002).

- [52] B. Koslowski, *Theory and Evidence: The Development of Scientific Reasoning* (MIT Press, Cambridge, MA, 1996).
- [53] American Association for the Advancement of Science, *Science for All Americans* (Oxford University Press, New York, 1993).
- [54] National Research Council, *Inquiry and the National Science Education Standards: A Guide for Teaching and Learning* (National Academies Press, Washington, DC, 2000).
- [55] C. A. Chinn and B. A. Malhotra, Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks, *Sci. Educ.* **86**, 175 (2002).
- [56] B. Y. White and J. R. Frederiksen, Inquiry, modeling, and metacognition: making science accessible to all students, *Cognit. Instr.* **16**, 3 (1998).
- [57] NGSS Lead States, *Next Generation Science Standards: For States, By States* (The National Academies Press, Washington, DC, 2013).
- [58] D. Hodson, Toward a philosophically more valid science curriculum, *Sci. Educ.* **72**, 19 (1988).
- [59] W.-k. Ahn, C. W. Kalish, D. L. Medin, and S. A. Gelman, The role of covariation versus mechanism information in causal attribution, *Cognition* **54**, 299 (1995).
- [60] B. Koslowski, L. Okagaki, C. Lorenz, and D. Umbach, When covariation is not enough: The role of causal mechanism, sampling method, and sample size in causal reasoning, *Child Development* **60**, 1316 (1989).
- [61] E. D. Corpuz and N. S. Rebello, Investigating students' mental models and knowledge construction of microscopic friction. II. Implications for curriculum design and development, *Phys. Rev. ST Phys. Educ. Res.* **7**, 020103 (2011).
- [62] W. Ahn and C. W. Kalish, The role of mechanism beliefs in causal reasoning, in *Explanation and Cognition*, edited by F. C. Keil and R. A. Wilson (MIT Press, Cambridge, MA, 2000), pp. 199–226.
- [63] W. C. Salmon, Why ask, “Why?” An inquiry concerning scientific explanation, *Proc. Am. Philos. Assoc.* **51**, 683 (1978).
- [64] S. S. Glennan, Mechanisms and the nature of causation, *Erkenntnis* **44**, 49 (1996).
- [65] B. Koslowski, J. Marasia, M. Chelenza, and R. Dublin, Information becomes evidence when an explanation can incorporate it into a causal framework, *Cognit. Dev.* **23**, 472 (2008).
- [66] N. G. Holmes and C. E. Wieman, Assessing modeling in the lab: Uncertainty and measurement, in *Proceedings of the 2015 Conference on Laboratory Instruction Beyond the First Year* (American Association of Physics Teachers/Creative Commons Attribution, College Park, MD, 2015), pp. 44–47.
- [67] L. C. McDermott, Millikan Lecture 1990: What we teach and what is learned—Closing the gap, *Am. J. Phys.* **59**, 301 (1991).
- [68] M. Lichtman, *Understanding and Evaluating Qualitative Educational Research* (SAGE, Los Angeles, 2011).
- [69] R. K. Yin, *Case Study Research: Design and Methods* (SAGE Publications, Los Angeles, 2009).
- [70] J. W. Creswell, *Qualitative Inquiry and Research Design: Choosing among Five Approaches* (SAGE, Washington, DC, 2013).
- [71] M. Kohlmyer, M. Caballero, R. Catrambone, R. Chabay, L. Ding, M. Haugan, M. Marr, B. Sherwood, and M. Schatz, Tale of two curricula: The performance of 2000 students in introductory electromagnetism, *Phys. Rev. ST Phys. Educ. Res.* **5**, 020105 (2009).
- [72] S. Pollock, Longitudinal study of student conceptual understanding in electricity and magnetism, *Phys. Rev. ST Phys. Educ. Res.* **5**, 020110 (2009).
- [73] P. Machamer, L. Darden, and C. F. Craver, Thinking about Mechanisms, *Philos. Sci.* **67**, 1 (2000).
- [74] R. S. Russ, J. E. Coffey, D. Hammer, and P. Hutchison, Making classroom assessment more accountable to scientific reasoning: A case for attending to mechanistic thinking, *Sci. Educ.* **93**, 875 (2009).
- [75] R. E. Scherr, Video analysis for insight and coding: Examples from tutorials in introductory physics, *Phys. Rev. ST Phys. Educ. Res.* **5**, 020106 (2009).
- [76] D. Hammer, The variability of student reasoning, lectures 1-3, in *Proceedings from Research in Physics Education: Proceedings of the International School of Physics “Enrico Fermi” Course CLVI* (Italian Physical Society, Bologna, 2004), pp. 1–15.
- [77] L. Schauble, The development of scientific reasoning in knowledge-rich contexts, *Dev. Psychol.* **32**, 102 (1996).
- [78] L. Darden, Strategies for discovering mechanisms: Schema Instantiation, modular subassembly, forward/backward chaining, *Philos. Sci.* **69**, S354 (2002).
- [79] A. Gupta and A. Elby, Beyond Epistemological Deficits: Dynamic explanations of engineering students' difficulties with mathematical sense-making, *Int. J. Sci. Educ.* **33**, 2463 (2011).
- [80] D. Hammer, Two approaches to learning physics, *Phys. Teach.* **27**, 664 (1989).
- [81] S. Kapon, Unpacking Sensemaking, *Sci. Educ.* **101**, 165 (2016).
- [82] B. A. Danielak, A. Gupta, and A. Elby, Marginalized identities of sense-makers: Reframing engineering student retention, *J. Eng. Educ.* **103**, 8 (2014).