

# Documenting the use of expert scientific reasoning processes by high school physics students

A. Lynn Stephens\* and John J. Clement†

*School of Education and Scientific Reasoning Research Institute, University of Massachusetts, Amherst, Massachusetts 01003, USA*

(Received 13 January 2010; published 24 November 2010)

We describe a methodology for identifying evidence for the use of three types of scientific reasoning. In two case studies of high school physics classes, we used this methodology to identify multiple instances of students using analogies, extreme cases, and Gedanken experiments. Previous case studies of expert scientists have indicated that these processes can be central during scientific model construction; here we code for their spontaneous use by students. We document evidence for numerous instances of these forms of reasoning in these classes. Most of these instances were associated with motion- and force-indicating depictive gestures, which we take as one kind of evidence for the use of animated mental imagery. Altogether, this methodology shows promise for use in highlighting the role of nonformal reasoning in student learning and for investigating the possible association of animated mental imagery with scientific reasoning processes.

DOI: [10.1103/PhysRevSTPER.6.020122](https://doi.org/10.1103/PhysRevSTPER.6.020122)

PACS number(s): 01.40.Fk, 01.40.Ha, 01.40.ek

## I. INTRODUCTION

In this paper, we describe a method for studying student reasoning processes and mental imagery within the methodologically noisy environment of a classroom. We illustrate elements of the method as we examine two high school physics class discussions in order to document student engagement in three types of reasoning processes. We also attempt to identify evidence for students' use of mental imagery in association with the processes. We will find that we can identify evidence for students' spontaneous use of analogies, extreme cases, and Gedanken experiments even when these three processes are used in combination. We will also be able to identify evidence that students can use mental imagery in connection with the processes, specifically animated mental imagery. This set of tools allows us to make a number of "existence demonstrations" of imagery-based nonformal reasoning, which, taken together, suggest that the role of imagery in physics learning should be taken seriously as a topic for future research.

### Purpose

Although there has been limited research on students' use of active reasoning processes such as the invention of analogies, extreme cases, and thought experiments while they are considering questions in the science classroom, to our knowledge this has not been demonstrated in any other way than by narrative description. Harrison and de Jong [1], and Cosgrove [2], have identified the student generation of analogies in the classroom and Hammer [3], Schultz and Clement [4], and Harrison and Treagust [5] have identified multiple types of reasoning in classroom discussions. But these pioneering efforts undertaken at the level of narrative description point out the need for more precise observational definitions for the reasoning processes identified. Even in the clinical laboratory, the study of such processes as the use of

analogy has proven difficult [6]. Consequently, it has not been possible to agree on the importance of these processes in student thinking, and it is not surprising that researchers and educators alike differ in their opinions on the importance of fostering the use of these processes in the classroom.

We asked ourselves whether we could clearly identify the use of analogies, extreme cases, Gedanken experiments, and other active reasoning processes as they appear in videotapes of whole class discussion. This has required the honing of clear and precise definitions, the development from these definitions of lists of observables that can be coded in a transcript, and, finally, the creation of new methods of coding. We want to progress beyond the stage of open coding (the process of identifying themes in the data without using any prior assumptions about what might be found [7,8]) so that: theoretical concepts are separated from observations, evidence is triangulated where possible, coders can reach consensus in joint coding, and observations are coded over complete transcripts according to fixed definitions and criteria. (See Supplemental Appendix A for a description of a sequence of methodological stages that a field may pass through, starting from open coding.) The objective is to develop a clearer set of defined concepts so that we may better understand these nonformal reasoning processes.

As regards imagery, the relative importance of its use in the scientific thinking of experts and students is a matter of debate. Some educators prefer to stress the importance of propositional thinking while others believe that imagistic processes (those that employ mental imagery) are just as important as propositional thinking, if not more so. If it is challenging to make the case that a reasoning process has occurred, it is even more challenging to make a case that a subject has used mental imagery, though Hegarty [9] has done pioneering work in this area by giving subjects problems in individual interviews that would appear to require mental animation to solve. To create a plausible argument that mental imagery is being used by students while in the methodologically noisy environment of the classroom is considerably more difficult—some might say nearly impossible. Here we attempt this using data from gestures and other transcript indicators.

Our purposes in the present paper:

\*lstephens@educ.umass.edu

†clement@educ.umass.edu; clement@srri.umass.edu

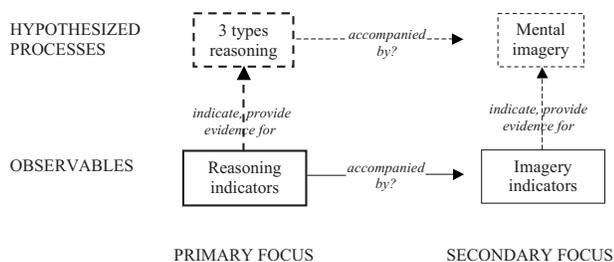


FIG. 1. Relationships among observables and hypotheses in the methodology.

(1) Propose criteria for identifying when a reasoning process has occurred and give transcript examples that illustrate the criteria;

(2) Report the results of two case studies in which the two authors jointly coded entire transcripts for the presence of the reasoning processes;

(3) Describe *depictive gestures* as one kind of indicator for the presence of mental imagery and give examples;

(4) Report evidence obtained via exploratory analysis of several types of depictive gesture.

The logic behind the methodology is illustrated in Fig. 1, which shows a diagram of relationships among the observables and the hypotheses they support.

We will argue that these methods are capable of revealing a richness and density of creative reasoning processes seldom documented in students. We also show that, in the transcripts considered here, most instances of evidence for reasoning processes were accompanied by evidence for the use of mental imagery, and in particular, for the use of animated mental imagery.

What we mean by ‘evidence’ in the paragraph above and by ‘indicator’ in Fig. 1 is discussed in the Methodology section.

## II. PREVIOUS RESEARCH AND THEORETICAL FRAMEWORK

This review of previous literature is organized around three points: in addition to student processes for model *evaluation*, student processes for model *construction* may also be important for learning; animated mental imagery may be an important form of mental representation used in these processes; depictive gesture can provide the researcher with a window onto the mental imagery of his or her subjects. These points will be used to frame the purposes of the study.

### A. Reasoning processes used in scientific model construction

The ability to engage in scientific model construction appears to be a crucial aspect of science [10] and also of student thinking [11,12]. In fact, it is argued that science textbooks are organized around such models [13]. Research continues to indicate the importance of mental modeling in both experts [14] and students [15–17], but Driver’s work [18] indicates that students often need to be helped to assimilate their prior experience [19] into scientifically accepted

models. We will ask whether certain nonformal reasoning processes such as the design and use of *analogies*, *extreme cases*, and *Gedanken experiments* can play an important role as students engage in scientific model construction. Each of these can involve the creative generation of a new concrete case (that is, a concrete example of some system) as the first step in the reasoning process.

Previous work on argumentation (Osborne, Erduran, and Simon [20], McNeill and Krajcik [21], Clark and Jorde [22], Duschl and Osborne [23] and others) has done much to document certain student argumentation processes, primarily centered around the question of whether students can support or discount a claim, e.g., by observations or analogies. The starting point for this vein of work is Toulmin’s [24] analysis of arguments to evaluate claims. In the tradition of Toulmin, there is a tendency for argumentation studies to place a primary focus on the process of evaluating claims as opposed to generative processes for creating new thought experiments, extreme cases, or analogies. It is true that since Walton’s [25] work, several argumentation studies have described student arguments by analogy to support a claim, but they have not said much about whether students can generate new analogies that employ novel cases expressly generated for the purpose at hand.

Clement [26–28] has conducted think aloud protocol studies that provide evidence that the generation and use of analogies, extreme cases, and *Gedanken experiments* [29] can be central during scientific model construction by expert scientists. These three processes have also been documented in history of science studies by Nersessian [30] and Darden [10]. This raises the question of whether these creative, generative processes might be important during student model construction as well, and this is a central question of this paper.

The work of Gentner and Gentner [12] and others [5,29,31] suggests that people can use *analogies* to help construct mental models and that carefully constructed analogies can be used to address students’ preconceptions in physics [11]. An important kind of analogy works by grounding instruction on students’ “anchoring” intuitions or familiar prior knowledge [32]. Recent work by Podolefsky and Finkelstein [33] indicates that analogies can enable students to generate useful inferences. In addition to teacher-constructed analogies, student-generated analogies can also be used as a tool for understanding [34], although reports on the aptness of student generated analogies have been mixed [35]. However, people appear to use analogies differently in the laboratory than they do in nonlaboratory contexts [6], making it a challenge to study the effectiveness of this type of reasoning. This suggests the importance for developing and honing methodologies for identifying and studying analogies as they occur spontaneously in nonlaboratory contexts such as classrooms.

*Extreme case* reasoning is another nonformal reasoning process used by experts [29] that can play a role in instruction [36]. We believe this process warrants study as well.

*Gedanken experimentation* (and more broadly, thought experimentation [27]) appears to be a powerful way to evaluate a mental model [37,38]. Previous philosophical analyses include theories of the structure of thought experiments [39]

and their function in scientific thinking [40,41]. However, there is no consensus in the literature on a definition for the terms “thought experiment” or “Gedanken experiment.” Indeed these terms could be applied quite broadly to any mental simulation or quite narrowly to formal imagined experiments. More recently, Clement [27,29,42] has investigated Gedanken experiments that were spontaneously generated and used by experts during problem solving, but Gedanken experiments also appear capable of playing an important role in teaching and learning [43,44]. Using a four-part definition of thought experiment, Reiner and Gilbert [45,46] found that some students can and will use thought experiments to find solutions to problems when the problems are formulated in a way to encourage this, especially in small-group collaborative settings. Very few studies, however, have investigated the role of thought experiments in large class discussion. Exceptions are Hammer [3], who identified thought experiments in use in large class discussions in high school physics, and Nunez-Oviedo, Rea-Ramirez, and Clement [47], who identified them in middle school physical science classes. However, we need a methodology that can be used to code consistently for the presence of thought experiments and Gedanken experiments over entire transcripts if we are to deepen the study of their use.

### B. Previous work on imagery

Recent findings from cognitive science [48] reinforce the notion of many physicists, e.g., Miller [49] and Hestenes [50], that imagery is an important form of mental representation in science. Ronald Fink [51] has defined imagery as the mental invention or recreation of an experience that in at least some respects resembles the experience of actually perceiving an object or an event (but see Hestenes [50] for a slightly different definition).

It appears to be important to be able to animate this imagery. Frederiksen, White, and Gutwill [52] found that high school students who were given only initial and final snapshots of an aggregate model of electric circuit behavior underperformed those who were also provided transient snapshots. Hegarty [9] hypothesizes that her subjects used mental animation as a mechanism to run their mental models as they evaluated them. Hegarty and others have investigated the use of mental animation in problem solving by experts [53] and students [9,54]. Some of the mental imagery involved appears to be kinesthetic in nature, as when expert physicists imagine manually exerting a push or a pull [38,53]. Kinesthetic imagery appears to be associated with physical intuition [26,55] and has been used in instruction [43,56]. Kinesthetic thinking appears to have an effect in problem solving in domains other than the physical sciences such as in geometry [57], which suggests that the role of this form of thinking may be more fundamental than previously thought.

In this study, we ask whether we can identify some of the points at which students are using imagery and whether the use of imagery is involved in student use of the three scientific reasoning processes considered here. We regard depictive gestures, which appear to depict an imaginary object or action “in the air” near the speaker, as providing some evi-

dence for the involvement of mental imagery. In particular, we will discuss evidence for the use of animated or *runnable* mental imagery, which we obtain from motion- and force-indicating gestures, described below. Identifying these types of depictive gesture gives us a potential foothold on distinguishing between static and animated mental imagery. Because gesture is the primary form of non-verbal evidence discussed in the present paper, we will briefly review what the literature says about the connection between gesture and mental representation.

### C. Gesture as a window onto mental imagery

We agree with Reiner and Gilbert [45] that only a small portion of the kind of knowledge accessed by nonformal reasoning processes can be articulated verbally; we seek non-verbal forms of evidence for this knowledge. Most of the recent research on gesture has focused on *representational* gesture, a broad category that includes any gesture that conveys semantic content, as by using shape, placement, or motion of the hands [58,59]. Representational gesture excludes gestures used merely for rhythmic emphasis. *Depictive* gesture, which is our focus, is a subset of representational gesture and depicts an object, force, or event; it excludes stylistic representational gestures such as the “thumbs up” sign [26,54,60].

We are not focused here on whether gesture helps the gesturer. Rather, our point is that gesture can help the *researcher* (c.f. Ochs [14] and Scherr [60]); we believe gesture provides at least a partial window onto subjects’ mental imagery. We can summarize our argument for this as a sequence of findings from the literature, expanded upon in Supplemental Appendix B.

(i) Type and amount of gesture appear to be closely associated with the nature of the subject’s internal representation (Lozano and Tversky [61], Iverson and Goldin-Meadow [62,63]).

(ii) Representational gesture production, in particular, appears to be associated with visuospatial and other imagistic processes (Iverson and Goldin-Meadow [62], Krauss [64], Hostetter and Alibali [65], Alibali [58], Feyereisen and Harvard [66]).

(iii) Evidence suggests that representational gesture is not merely a translation of subjects’ verbal meanings, but can reveal unspoken thought. An alternative hypothesis is that gesture is a physical translation of words, but a number of studies have cast doubt on the plausibility of this hypothesis (Goldin-Meadow [67,68], Roth [69]).

(iv) Depictive gesture appears to be a natural way of expressing the results of mental animation and conveys information about the animation not revealed in subjects’ words (McNeill [59], Hegarty, Mayer, Kriz, and Keehner [70]).

These selected results from the literature lend strong support to the idea that subjects’ gestures can provide information about their mental imagery, as described further in Supplemental Appendix B.

Depictive gestures are included in a list of *imagery indicators* developed by Monaghan and Clement [71], a set of observables that are hypothesized to indicate the presence of

mental imagery in the thinking of a subject. We hypothesize that some of these observables indicate the presence of *animated* imagery in particular. In prior research [26,27,29], evidence from imagery indicators, including depictive gesture, has lent weight to the idea that at least some expert physicists generate animated mental imagery (with kinesthetic as well as visual components) as they use nonformal reasoning processes to solve problems. We will apply our method in two classroom case studies in an attempt to identify evidence that students can make spontaneous use of these forms of scientific reasoning in the classroom. We will also identify evidence that can address whether such reasoning can involve animated imagery.

### III. METHODOLOGY

#### A. Unit of analysis

We organized our data by *case*, *variation* of a case, and *reasoning episode*. A *case* is a concrete example of a system. A case introduced by a teacher during a discussion about the causes of gravity, the US-Australia case, comprised the Earth, two people standing on it, and the gravitational forces between the Earth and the people. A *variation* of a case involves the same concrete example of a system but with some variable changed in a significant way (such as to create an extreme case) or with an additional variable highlighted. For instance, when a student introduced the rotation of the Earth into the discussion as a possible factor causing gravity in the US-Australia case, we counted this as a variation of that case. A *reasoning episode* involves a single student (or teacher) reasoning about a case or variation or drawing a comparison between cases or variations. Our unit of analysis in the present study is the *reasoning episode*.

#### B. Identifying evidence for reasoning processes

The processes we wish to identify and study have been defined somewhat inconsistently in physics education research (PER). Part of our methodology was to define them in a way that could allow us to code transcripts for their spontaneous occurrence, even in instances where students were not yet expert in their use or were not as articulate as we might wish in expressing them. In this section, we propose definitions and clarify transcript indicators for the processes.

When we have used the term *indicator* in this paper, we do not mean something as objective as many of the variables measured in physics. We are investigating an area of student reasoning, and no direct measures of such mental processes are possible. Thus, we are in a situation similar to other developing areas of science, where the best we can do is to make hypotheses about hidden processes and then to support or disconfirm them with the best data we have. Thus, “indicator” refers to an observation pattern (e.g., “depictive gestures”) that provides evidential support for the presence of a hidden cognitive process (e.g., “imagery”). Theory development in the field of the psychology of science learning is at a much more embryonic state than it is in most areas of classical physics. Case studies are an appropriate method for obtaining viable initial process hypotheses in such a field and

working out relationships between observations and theory.

Analogical reasoning, extreme case reasoning, and evaluative Gedanken experimentation are among reasoning processes previously identified by Clement [29] as part of a network of reasoning processes that allow experts to generate ideas divergently and then to evaluate them convergently. We define them here as follows.

*Analogical Reasoning.* This occurs when (1) a subject, in thinking about a target situation A, refers to another situation B where one or more features ordinarily assumed fixed in the original problem situation A are different; that is, the analogous case B violates a “fixed feature” of A (to be defined below); (2) the subject indicates explicitly or implicitly that certain structural or functional relationships (as opposed to surface attributes alone) may be equivalent in A and B; (3) the related case B is described at approximately the same level of abstraction as A; and (4) there is an explicit or implicit intent to aid in reasoning about (infer findings in) A from consideration of B. As used here, ‘fixed features’ are those features of the problem situation that would commonly be assumed givens not subject to change, as opposed to problem variables, or features assumed to be changeable or manipulable.

*Extreme Case Reasoning.* This occurs when a subject, in thinking about a target situation A, shifts to consider a situation E (the extreme case) where some feature of interest from situation A has been taken to an unusually high or low value; (2) there is an explicit or implicit intent to aid in reasoning about (infer findings in) A from consideration of E.

*Evaluative Gedanken Experimentation.*<sup>1</sup> This occurs when a subject considers an untested, observable system designed to help evaluate a scientific concept, model, or theory and attempts to predict aspects of its behavior [27]. In these experiments, an element of a theory is evaluated as it is applied to the untested system. By untested, we mean that the subject has not observed that aspect of the system before nor been informed about its behavior [29].

Making these definitions more precise is not the same as committing to disjoint categories. As defined here, these three processes are not intended as mutually exclusive disjoint categories along a single dimension; in some circumstances more than one can apply to a single instance of reasoning. For example, the categories of extreme case reasoning and evaluative Gedanken experimentation can sometimes apply to the same reasoning episode because they describe different dimensions of a single instance of student reasoning about a case.

The definitions were developed from work with expert protocols during which subjects with a scientific background were reasoning about problems unfamiliar to them. In the expert situations, it was clear that subjects, while reasoning aloud, were inventing their own analogous cases and thought experiments and frequently drawing conclusions about those cases. If they ran an evaluative Gedanken experiment, evi-

<sup>1</sup>Clement [29] has identified both broad and narrow categories of thought experiments. The Evaluative Gedanken Experiment, the most restrictive category, is the only category that will be considered in the present work.

dence often could be found in the transcript that the experiment employed cases invented on the spot for the purpose. When coding classroom transcripts, however, we confronted situations in which reasoning was more distributed. One student might suggest a case as analogous, whereupon other students might draw multiple (and sometimes contradictory) conclusions from the analogy or modify it slightly in order to question the conclusions of their fellow students. Our impression certainly was that we were seeing forms of reasoning reminiscent of the expert processes, but how to code the student utterances, often interrupted and stated with varying degrees of articulateness, proved a challenge. We eventually began to make use of the distinction between *generating* a reasoning process and *running* the process. We made this distinction by analyzing the case(s) used in the process. Thus, if a student suggested that a case was analogous to the target case and this analog case had not yet been mentioned in the discussion, we coded the reasoning episode as involving a *spontaneously generated analogy*. If another student, without prompting, suggested conclusions for the target citing the previously suggested analog case as a basis, then we coded the episode as involving a *spontaneously run analogy*.

At times the same student could both generate and run an analogy. Another question arose: if we counted numbers of instances of generated analogies and numbers of instances of run analogies, how many instances of analogical reasoning could we say we had witnessed? If one student suggests an analogy and another student reasons with it, the result could be considered a single jointly constructed analogy. However, it was common for several students to reason about a single case, sometimes repeating each other. In addition, it was our impression that it was rare for all steps of a student's thinking to find their way into the class discussion.

We found that coding for all student utterances that met criteria (described below) for being generated *or* run allowed us a rich description of distributed reasoning; these descriptions are discussed in the first half of the paper and help to give a sense of the quality of discussion and the student-to-student transmission of ideas. In the second half of the paper, we address a different purpose: we wish to make a conservative estimate of the amount of spontaneous reasoning of the above types taking place in these two class discussions. For this purpose, we restrict ourselves to the most conservative tally afforded by our data: we tally the number of times students generated the reasoning processes; that is, the number of times they made new suggestions for cases to be used in analogies, as extreme cases, or in Gedanken experiments. Frequently, such cases appeared to be novel ones designed by the students for the purpose.

The coding criteria below were an outcome of an iterative process of coding classroom transcripts using, critiquing, and refining applicable portions of the Clement definitions above.

*Generating an Analogy.* (The subject spontaneously suggests the case.)

(1) Is the subject attempting to facilitate reasoning about a target situation A by suggesting or implying that findings from a situation B (the base) be applied to A, where B is at about the same level of generality as A and differs in some significant way from A?

(2) Is this the first time in this discussion that the situation B has been mentioned in connection with A?

*Running an Analogy.* (The subject attempts to draw conclusions by using a case suggested by himself or herself or another.)

(3) For an analogy generated as above, does the subject draw a prediction or implication from B or attempt to apply findings from B to A?

*Generating an Extreme Case.* (The subject spontaneously suggests the case.)

(1) Is the subject attempting to facilitate reasoning about a target situation A by suggesting a situation E (the extreme case) where some variable from A has been taken to an unusually high or low value?

(2) Is this the first time in this discussion that the situation E has been mentioned as an extreme case?

*Running an Extreme Case.* (The subject attempts to draw conclusions by using a case suggested by himself or herself or by another.)

(3) For an Extreme case generated as above, does the subject make a prediction or implication from E or attempt to apply findings from E to A?

*Generating an Evaluative Gedanken Experiment.* (The subject spontaneously suggests the case or cases.)

(1) Does the subject spontaneously introduce a system (or variation on a system) for which it is likely she or he has never observed nor heard of the results? Or, if the subject is proposing the experiment to others, is it likely that the others have never observed nor heard of the results?

(2) Does the subject propose an activity that, if it could be conducted, could yield empirical observations?

(3) Does the subject make an implicit or explicit suggestion that a prediction be inferred for an aspect of the behavior of the system?

(4) Was the activity designed or selected to help evaluate a scientific theory, a scientific or mathematical concept, or an explanatory or mathematical model?

(5) Is this the first time in this discussion that the activity has been mentioned to help evaluate this scientific theory, concept, or model?

*Running an Evaluative Gedanken Experiment.* (The subject attempts to draw conclusions from a case suggested by himself or herself or by another.)

(6) For an evaluative Gedanken experiment generated as above, does the subject attempt to infer a prediction for an aspect of the behavior of the system?

Note that in an analogy, case B must differ "in some significant way" from A, while for an extreme case, it is the value of a problem variable that differs. The criteria above hint at the fact that evaluative Gedanken experiments are more complex than the other two forms of reasoning defined here. Examples of our coding will be given in a later section.

### C. Identifying evidence for imagery

After coding the videotape transcripts using the above criteria, we coded the videotapes themselves for the presence of *depictive gestures*. These gestures appear to depict an imaginary object, action, or location, and are taken as an indication

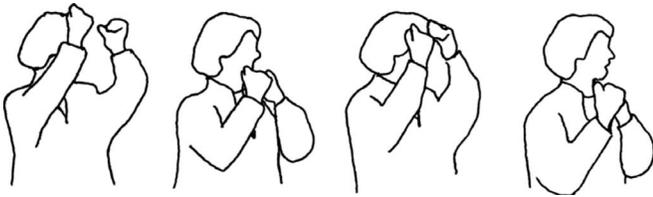


FIG. 2. Force-indicating depictive gestures. “The Earth has more pull on you.”

that mental imagery is being used [29]. We did not include stylized gestures (such as the thumbs up gesture) or beat gestures (used for rhythmic emphasis). Visual inspection alone was sufficient to identify the presence of depictive gesture. In a later step using both visual and verbal information, the depictive gestures were assigned to one of three subcategories. *Shape-indicating* gestures [G-S] appear to depict a shape and are taken as evidence for the presence of mental imagery. *Motion-indicating* gestures [G-M] appear to indicate the motion of an object (it may be a point-object) and are taken as evidence for animated mental imagery. *Force-indicating* gestures [G-F] appear to indicate the action of a force; these can be quite emphatic. These gestures are taken as evidence for the presence of animated imagery that contains kinesthetic components; an example is shown in Fig. 2. At times, an educated guess can be made from the appearance of the gesture alone as to whether it is intended to convey a motion or a force; however, we rely on the subject’s use of *force terms* such as “pulling” or “throwing” as additional evidence for our choice between these two categories. We call both motion- and force-indicating gestures *action gestures*. (For a larger list of imagery indicators, see Clement [29].)

After completing the coding for the reasoning episodes and the presence of depictive gesturing, the results of the two coding procedures were compared to determine which reasoning episodes were accompanied by depictive gestures.<sup>2</sup> Again, our unit of analysis was the episode; our intent was to establish, for each episode in which a student was reasoning about an analogous case, extreme case, and/or Gedanken experiment, whether depictive gesturing was also occurring. (This bears similarity to a practice followed by other gesture researchers who have used transcript utterances as the unit of analysis [72], though our emphasis is different.) We do not attempt here to establish an exact number of depictive gestures during each reasoning episode, but only whether depictive gestures were present during that episode. After this was

<sup>2</sup>Note that written utterances were used for coding for presence of the processes while visual inspection of the videotape was used for coding for the presence of depictive gestures. It is the presence of these two kinds of indicators that was compared. Force terms from the utterances were used later to aid in sub-categorizing the depictive action gestures as either force- or motion-indicating, not to establish their presence.

established, a final determination was made for each episode as to which subcategories its associated gestural sequences belonged.<sup>3</sup>

We have identified additional categories of nonformal reasoning discussed elsewhere [73,74] and we continue to refine our list. Here we consider only three types of nonformal reasoning from that list; therefore, we cannot compare gesturing during all episodes of nonformal reasoning vs gesturing at other times. In addition, this study is not yet about establishing typicality of processes but about developing viable ways of recognizing nonformal reasoning processes when they do occur. As we are using coding as a way of developing stable categories, in all cases coding was jointly agreed upon by the two authors and disputes were used as a mechanism for refining and clarifying the coding criteria.

#### D. Data sources

We use two case studies to illustrate the kinds of distinctions that can be made with this methodology. Lengthy discussions were triggered when a physics teacher presented target cases designed to elicit student misconceptions so that they could be addressed. The two discussions from which we draw examples occurred in different class sections in a middle class suburban high school in the northeastern United States. The teacher was using an innovative curriculum [56]. The classes, which were videotaped, were both at the college preparatory level but were on different topics in physics though gravity was a factor in each. These two transcripts were selected for analysis because they appeared to contain the phenomena for which we wished to refine categories; this means that the frequency of student-generated nonformal reasoning processes may have been higher than is typical, which is appropriate for a study of this kind. In social science research, this is referred to as purposive sampling [75,76]. (This is similar in practice to the study of understudied phenomena in other domains; e.g., if very little work has been done on the structure of optic nerves, one may purposefully start with a study of the largest available—that of the giant squid. If no previous dissections have been done at this level of detail, a thorough dissection of even one or two animals can be enormously informative as a starting point.)

Discussions were animated and, although the students were arguing about basic concepts, they appeared to be using some interesting scientific reasoning. Each of the discussions lasted about 45 min.

#### IV. EPISODES OF REASONING IDENTIFIED FROM TRANSCRIPTS

In each transcript, we identified many reasoning episodes involving each of the three scientific reasoning processes; these included student-generated, student-run, and teacher-

<sup>3</sup>To establish whether a flowing sequence of nearly identical gestures was depictive and whether it was shape-, motion-, or force-indicating did not require us to establish the exact number of gestures within the sequence. The repetitive sequence in Fig. 2 is an example.

generated examples. In the present section, we give examples of each of these and attempt to give the reader a feel for the flow of the class discussions. In a later section, we will discuss student-generated examples in particular.

### A. “Book on Table” class transcript

In this lesson, the teacher wanted students to consider whether a table exerts an upward force on objects resting on its surface. A common conception prior to instruction is that static inanimate objects cannot exert forces. The target model for the lesson was one in which objects exert normal forces that are equal and opposite to the weight of objects resting on them. The whole lesson was structured around a series of analogies (see the curriculum [56], also Clement [11]), and the teacher repeatedly mentioned to the class that he was using analogies.

#### 1. Analogy between base and target (teacher-generated)

At the beginning of the lesson, the teacher placed a book on his desk and called students’ attention to it, then drew two figures on the chalkboard. One was a simple line drawing of a book on a table (which we identify as the *target*), and another of a hand pressing downward on a spring (identified as the *base* of the analogy). He asked the students to compare the two cases (to engage them in analogical reasoning) and to vote on whether they thought the table could exert a force on the book. The teacher has reported that he hoped all of the students would believe that the spring pushed up on the hand and that he could use this as an anchoring base case for the lesson. It had become clear in previous years that, although many of his students had believed the spring would exert a force on the hand, a large number had not believed the table would exert a force on the book. Therefore, the teacher planned to introduce a series of analogies designed to help transfer intuitions from the hand on spring case to the book on table case.

#### 2. Series of bridging analogies (student-generated)

Before the teacher could introduce the planned series of analogies, his students preempted him, producing their own series of cases. They spontaneously invented a number of novel scenarios to support their positions and they evaluated and modified the scenarios of others. For example, a series of student-generated modifications began with S 15’s suggestion to imagine “building a table out of something else, like um, uh, a balloon...” Later the following exchange took place, with multiple (though not always identifiable) student voices.

- S15: Wouldn’t it, it make more sense if we build the table out of something pliable—  
 S3: Like plywood.  
 T: Suppose we build the table out of something really cheap, I think I hear—  
 S (off camera): Yeah.  
 T: Really thin plywood, or—  
 S15: A piece of cardboard.  
 S15: A piece of cardboard.

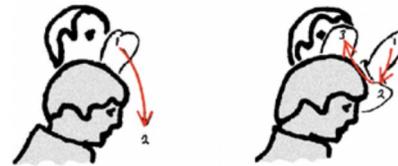


FIG. 3. (Color) “(I)t [G-F] punctures the thing; it [G-S] warps it too much.”

T: —or a piece of cardboard—

S15: Or a piece of paper.

S (off camera): Bounty (a brand of paper towel, heavily advertised as “strong”).

We view this joint construction as a series of bridging analogies that appear to span the conceptual gap between a sturdy laboratory table and the original “hand on spring” analogy. Note that the analogies were constructed mostly by the students rather than by the teacher. One way to code the above passage would be as three analogies: plywood, cardboard, paper. In addition, we could code at least one extreme case: the extremely flexible table made of paper towels (as the variable of flexibility has been taken well beyond the range normal for a table). However, to remain conservative, in the Results section of the present study we will count this entire episode as a single episode of student-generated analogical reasoning.

#### 3. Extreme case reasoning (student-generated)

Somewhat later, S5 returned to the case of a thin warpable table to argue that, unlike the spring, the table cannot exert a normal force; the table does not have enough power to “exceed” the weight of an object to move it in the other direction, “and as soon as (*the weight*) gets too great then the table collapses.” S15 then recast S5’s statement as an extreme case in order to argue that even though the table is ultimately breakable, elastic warping could be present up to that point (and therefore, presumably, a normal force could be present):

S15: (S5’s) idea is compatible with the warped table theory. The idea is that the [G-S] elephant sitting on the table is too much [G-S] for the material that the table is made out of, and it [G-F] *punctures* the thing; it [G-S] warps it too much.

“Punctures” is in italics to denote that it is a force term. [G-S] and [G-F] refer to shape- and force-indicating gestures, respectively, and are placed at the point in the transcript where the student began the gesture. Figure 3 is a tracing from the videotape of the final two gestures; the gesturer is the student in the rear. The shape depicted by the final gesture was a deep curve, concave from above, much deeper than a table could normally form without breaking. By pushing the warped table to an extreme, the student had transformed the warped table into the broken table (moving from extreme warping into the new regime of breaking). He argued that the possibility of breakage was not evidence against the prior presence of elastic warping. The numerous gestures give evidence of the presence of visual imagery, and



FIG. 4. (Color) “(T)he book would just [G-M] fly off into space.”

the action gesture accompanied by the force term *punctures* suggests the presence of kinesthetic imagery as the student appeared to embody the act of puncturing.

#### 4. Analogical reasoning (student-generated)

Later in the class, S14 drew an analogy between the book situation and a situation the class had studied earlier, that of a boat powering upstream just hard enough to counteract a current and stay stationary. If the current were to stop suddenly while the engine continued running, the boat would move upriver; likewise, if the table were suddenly not there pushing against the book, the book would fall down. We counted this as a student-generated analogy because the student was attempting to facilitate reasoning about the book-on-table situation by reintroducing a different case from a prior class (about canceling velocities) that had not been mentioned in the context of the present discussion (about forces). In this instance, the student also ran the analogy: he drew an implication from the boat and river situation (the boat would move upstream) and applied it to the book and table situation (the book would fall down).

#### 5. Evaluative Gedanken experiment (student-generated) involving an analogy (student-run)

S15 replied to S14 by using the same analogy between the book situation and the boat situation. However, rather than imagining the current stopping, he imagined the force of the boat engine disappearing and predicted what would happen to the boat due to the current. In doing so, he refined the description of the analogical relationships. While S14 had focused on a comparison between the *movements* of objects in the two scenarios, S15 appeared to specify a relationship between the *force* of the engine and the *force* of gravity, and then predicted what would happen to the book if the force of gravity disappeared:

S15: But by the same analogy, then, if gravity disappeared, right, the *force* of the [G-F: *sudden thrust downward*] engine on the book, even the book would just [G-M: *flings arms upward and outward*] fly off into space.

Here we have included descriptions of the gestures to convey their energetic quality (see Fig. 4). We take these to be indications of the student’s use of animated mental imagery.

The student appears to be saying that if the engine disappeared, the current would move the boat, and by analogy, if

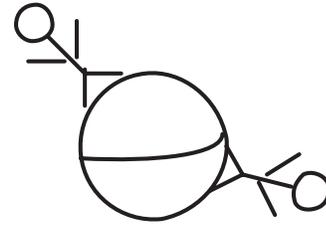


FIG. 5. US-Australia Case

gravity disappeared, the normal force would send the book off into space. (The table would suddenly unwarp—a correct inference, although the effect would be extremely small.) We consider this to be an evaluative Gedanken experiment. The case of gravity disappearing is an untested system and the student attempted to predict an aspect of its behavior—what would happen to a book on a table in such a situation. Also, the activity of imagining gravity disappear to see what would happen to the book had not been mentioned earlier in the discussion; this activity appears to have been selected by this student to evaluate an aspect of the theory of the existence of normal forces.

Later in the class period, students were presented with a model of solid matter as being made of atoms with springlike bonds between them. This was followed by a classroom demonstration that optically magnified the effect of warping in an apparently solid table and there was further discussion. Next, students voted again on whether they thought the table could exert a force on the book and most were convinced that, if the table could warp, it could push back against objects resting upon it.

See Supplemental Appendix C for more instances of student-generated reasoning processes identified in this discussion, many of them accompanied by depictive gestures.

#### B. “Gravity” class transcript

The second transcript was of a class that had finished a unit on density and was just beginning a unit on gravity. Common conceptions of students prior to instruction are that causes of gravity include the rotation of the Earth and/or the “downward” pressure of the atmosphere. The target model of the lesson was one in which every particle of matter pulls on every other particle. The teacher planned to introduce three cases during the course of the lesson; however, his students pre-empted him and came up with the third case on their own before the teacher could introduce it.

The first case was designed to elicit misconceptions such as those just mentioned and to stimulate discussion. The teacher drew a figure on the board (reproduced in Fig. 5) and asked the class to vote on the following: “Compared to the United States, gravity in Australia is: a little less, equal, a little bit more.”

After the students had recorded their votes on voting sheets, the teacher opened the discussion by asking, “Just what is it that causes gravity, anyway?” What followed was a very lively discussion in which the teacher played a role that was almost neutral, restating student positions, asking for



FIG. 6. (Color) Shape-indicating gesture. “[G-S] Imagine a ball floating in space....”

clarification, and occasionally recasting a student utterance in a slightly altered form.

### 1. Evaluative Gedanken experiment involving extreme case reasoning and analogical reasoning (student-generated)

Early in the class, some students suggested that the rotation of the Earth either causes gravity or contributes to it. Although several students countered this idea, the proponents of the rotation model of gravity appeared not to be convinced. Another student suggested the following Gedanken experiment (Figs. 6 and 7).

S7: Well, in reference to rotation and gravitational force, I think of them as being two opposite forces because if you stand on—let’s just [G-S] imagine a ball floating in space you tape your feet to. And you start spinning the ball around, you’re gonna [G-M] *feel like you’re gonna be* [G-F] *thrown off*. But if it’s a small ball, then the attraction between you and that little small mass is negligible so that you’re just gonna [G-F] *feel the forces being spun around* in a centrifugal force.



FIG. 7. (Color) Force-indicating gesture, “You’re just gonna [G-F] *feel the forces being spun around*....” (Student rapidly moves finger tip in horizontal circles.)

This is an imaginary case that appears designed to evaluate the gravity-from-spinning theory by pitting it against a strong conflicting intuition. When weighing oneself, the spinning of the Earth does, in fact, reduce the reading on the scale slightly everywhere except at the poles, but many students have trouble imagining and understanding this effect, and instead guess that spinning may be one of the causes of gravity. The ball with a person’s feet taped to it was introduced by S7 as analogous to the Earth. (It *differed significantly* from the Earth-person system both by initial lack of spin and by addition of the force of the tape as a substitute for most of the force of gravity; therefore, the first sentence meets our coding criteria for a generated analogy. In our experience, students commonly believe there are many differences between the fixed features of a ball and a planet—an earthly object and a celestial object—and believe the two objects would behave quite differently.) Then he rendered it equivalent to the Earth by spinning it, though his rapid gesturing indicated a degree of spinning that would be well out of the normal range for a planet. He pointed out an additional equivalence: the ball is “a little small mass.” Therefore, two variables, mass and rotation, have been *taken to unusually low and high values*, respectively, for a planet-human system and the second two sentences meet our coding criteria for extreme case.

S7 generated a prediction from this untested situation: “you’re gonna feel like you’re gonna be thrown off.” The prediction of an effect opposite to the effect of gravity is a result (observable, at least in principle) that would tend strongly to discount spinning as a causal factor in the pull of gravity. This reasoning episode therefore meets our definition of an evaluative Gedanken experiment: the student *considered an untested, observable system that appears to have been designed to help evaluate a theory about the cause of gravity and predicted an aspect of the behavior of this system*.

We hypothesize that this episode can be viewed as a student’s effort to design a case that maximized the potential of the rotating-globe scenario to evoke comprehension via kinesthetic imagery. It appears designed to help him and his classmates convincingly distinguish between the (felt) effects of rotation and the (felt) effects of the downward pull of gravity. His depictive gestures provide evidence for his own use of animated imagery, including some with kinesthetic components, throughout this reasoning episode.

### 2. Extreme case (teacher-generated, student-run, student-run with modification)

An episode that illustrates more interchange and co-construction between the teacher and students is the following. In response to a question about whether gravity would change if one climbed a mountain, a student replied,

S4: I think how far you are from the poles has more to do with it.

The teacher responded by paraphrasing S4’s statement:

T: Now the other issue that you’re bringing up ... was that gravity has to do with the Earth spinning, also is

TABLE I. Evidence for student-generation of three nonformal scientific reasoning processes in Book on Table discussion (approx. 45 min): Numbers of episodes.

A	B	C
Combinations of processes	No. of episodes in which these combinations were identified	No. of episodes in Col. B accompanied by depictive gestures
Analogy alone	3	2
Extreme case alone	3	3
Gedanken alone	1	1
Extreme case and Gedanken	1	1
<b>Total episodes identified</b>	<b>8</b>	<b>7</b>

another issue that was mentioned. If that's the case, let's give a little bit of thought about what (S4) is saying. If I were to stand at the North Pole, say the pole is here and I hold a hand on the pole, how long does it take me to spin around that pole?

Once the class reached agreement that it would take one day and that the speed of movement around the pole would be very small, the teacher continued,

T: Let me point out, if I stand on the equator, however—

and a student from off-camera replied,

S: You're going real fast.

The teacher's responses had converted S4's vague phrase "how far you are from the poles" into a comparison between people standing at the longitudinal extremes of the North Pole and the equator, and the students promptly began to reason about this teacher-generated variation. In fact, this extreme case comparison continued as the topic of discussion over the next several minutes. Later, there was evidence that S4 was able to run the extreme case as he used it to generate predictions concerning the net effects of rotation at the pole vs at the equator. He predicted that if it were true that rotation "throws you," the effect of rotation at the equator would be to throw one away from the Earth, whereas the effect of rotation near the pole would be to throw one sideways. (We coded S4's prediction and the reasoning that led to it as an evaluative Gedanken experiment; see the second table in Supplemental Appendix C, Table II, transcript line 182). It is doubtful that S4 would have been able to reason in this way with his own original statement.

Another student reran this same extreme case with a slight modification that increased the accuracy of his results:

S9: What were we arguing about? Well I'm basically taking (S4's) position in that [G] when the Earth spins, it seems logical to me—although (*another student*) says it's wrong—but it seems logical to me that there would be a [G] force—say you're on the equator and you're going around, there's this greater force pushing you off the Earth than if you were on the pole and you're doing this little circle. It's just much less of a force throwing you that way. But if gravity is the same [G: indicates top of an invisible object in front of him,

*presumably a globe about the size of a basketball*] here, and gravity is the same [G: indicates side of the "globe," *presumably at the equator*] here, it seems that you would weigh less [G: side of "globe"] here because you're being thrown off more [G: indicates motion laterally away from the "equator"] that way. Although you'd still stick to the Earth. You could still—I think you would weigh less.

S9's refinement specified the forces involved (though one was a pseudoforce), enabling him to state not only that one's weight would register slightly less at the equator than at the poles, but why that would be so. (Both students appear to have been equating the reading that would result on the scale with the quantity of weight; if one equates "show a smaller reading on the spring scale" with "weigh less," S9's prediction is accurate.) Such episodes suggest to us that it is very important for each student to go through the mental activity of running a simulation of such a case if it is going to be understood.

Additional student-generated examples and another evaluative Gedanken experiment from this discussion, together with accompanying depictive gestures, are included in Supplemental Appendix C.

## V. NUMBER OF IDENTIFIED STUDENT-GENERATED REASONING EPISODES

Above, we looked at episodes of reasoning that were student-generated, student-run, and teacher-generated. We wish to obtain an estimate of students' *spontaneous* use of these scientific reasoning processes in these classes. One way to gain a conservative estimate of this is to consider only reasoning episodes that were *student-generated*. In order to do this, for each episode in which one or more of the three reasoning processes had been identified, we determined whether the activity or situation that the student was reasoning about had been used previously in this reasoning process in this classroom discussion (e.g., whether a student was reasoning about an analogy suggested by another, even if this student drew a different conclusion). If the activity or situation had not been used previously for this reasoning process, the episode was coded as having evidence for a *student-generated reasoning process*. (See the coding criteria above.) In Table I, we summarize the results of coding the Book on

TABLE II. Evidence for student-generation of three nonformal scientific reasoning processes in Gravity discussion (42 min): Numbers of episodes.

A	B	C
Combinations of processes	No. of episodes in which these combinations were identified	No. of episodes in Col. B accompanied by depictive gestures
Analogy alone	4	4
Extreme case alone	2	1
Gedanken alone	2	2
Analogy and Gedanken	1	1
Extreme case and Gedanken	1	1
Extreme case and Analogy and Gedanken	1	1
<b>Total episodes identified</b>	<b>11</b>	<b>10</b>

Table discussion for student-generation of the three processes considered here; again, our methods are designed to produce a very conservative estimate.

The numbers in Column C are the numbers of reasoning episodes during which gesturing occurred (not numbers of gestures).<sup>4</sup> Again, the reasoning episode is our unit of analysis in the present study; this is reminiscent of the practices of others [72]. Note that:

Seven of the eight episodes involving the spontaneous generation of one or more of the three reasoning processes were accompanied by depictive gestures that were visible on the videotape—and all seven of these included action gestures (indicating force or motion). The use of action gestures provides evidence for the use of animated imagery in conjunction with these processes.

Our definitions for evaluative Gedanken experiment generation, analogy generation, and extreme case generation have allowed us to identify when each process was being generated even when they were used in combination. This has allowed us, for example, to describe how a student's incorporation of an extreme case helped strengthen the design of his evaluative Gedanken experiment, as in our analysis of the Gravity transcript. This is reflected in Table II. (Clement [29] has also documented cases where experts use analogies or Gedanken experiments that are also extreme cases.)

This class was at least as rich in evidence for imagery as the preceding one<sup>5</sup>:

Ten of the 11 episodes involving the spontaneous generation of one or more of these expert reasoning processes were accompanied by depictive gestures that were visible on the videotape—and all ten of these included action gestures (indicating force or motion).

As before, we consider this to be evidence for student use of animated imagery in conjunction with the generation of these reasoning processes.

<sup>4</sup>However, one of the authors did use McNeill's methods to identify 53 separate gestures in 45 min of the Book on Table videotaped discussion, with multiple gestures identified during each of the episodes listed above. Because many students in this large classroom were partially obscured to the camera, this number is conservative.

<sup>5</sup>One of the authors identified 105 gestures in 42 min of the Gravity videotaped discussion.

The analysis reported in this section was restricted to numbers of *episodes* of student reasoning rather than to numbers of individual processes within those episodes; furthermore, it was restricted to spontaneous student generation of the processes; and to three processes from a longer list of nonformal reasoning processes. Although our decision to restrict our analysis in these ways may, perhaps, render our numbers less impressive, we believe this exhibits the potential of the method to build a relatively firm “existence proof” or “existence demonstration” for the presence of these processes—even when we do not have the luxury of examining students in a controlled environment.

## VI. FINDINGS

We have attempted to propose a set of viable, definable constructs for nonformal reasoning processes and to show how such processes can be teased apart and connected to imagery indicators. Specifically:

(1) The conceptual distinctions and definitions we have developed allow us to use transcripts from classroom videotapes to identify student use of several categories of nonformal reasoning processes that are also used by experts: generating and/or running *analogies*, *extreme cases*, and *evaluative Gedanken experiments*. It has been possible to identify these three processes even when the processes are used in combination.

(2) Some of the cases were generated and run cooperatively between multiple students or between the teacher and a student.

(3) All instances of the processes listed in Tables I and II, however, were spontaneously generated by the students during discussions about important conceptual issues.

(a) For instructors interested in science process goals, this constitutes an “existence demonstration” that students can engage in these creative scientific reasoning processes in classroom discussions.

(4) It is possible to gather evidence from classroom videotapes that indicates students can use mental imagery when engaged in these three types of expert reasoning.

(5) We identified three types of gesture—shape indicating, motion indicating, and force indicating. This distinction allows us to identify evidence for students' use of, respec-

tively, mental imagery, *animated* mental imagery, and animated mental imagery with *kinesthetic* components. The techniques used in this study provide us with a set of tools for analyzing nonformal reasoning processes in classroom discussions.

## VII. LIMITATIONS

This paper is qualitative, descriptive, and exploratory in character; its intent is to work toward more stable and precise concepts concerning scientific reasoning processes exhibited by students. This can be done with a small sample size that allows more intensive analysis but this means that we make no claims about the typicality of the frequency of these processes. The aim, rather, is to produce stable definitions that can be used more broadly in the future. The conclusions we draw are in the form of “existence demonstrations” of categories of processes that can suggest fruitful avenues for future research.

## VIII. SOME WIDER IMPLICATIONS OF THE METHODOLOGY

In trials with individual subjects, Catrambone and Holyoak [77] found that many subjects required direct prompting in order to engage in analogical problem-solving, even when applicable source cases had been provided earlier. In contrast, the present study provides an “existence demonstration” that high school students can engage in analogy and other nonformal scientific reasoning processes even when direct prompts of source cases are not provided. Even where we are preaching to the converted (e.g., those teachers who have seen this in their classes), it is a different matter to provide empirical evidence that this has occurred. Although some prior evidence exists, we did not find definitions sufficiently precise to allow for the kind of coding that we wished to do. Therefore, we believe that the present formulation of definitions in terms of observables can aid further research.

The definitions give us a tool for documenting students’ spontaneous use of such reasoning in classrooms and so open the door for further investigation. This tool should allow us in the future to investigate whether such reasoning processes are central or peripheral when students are constructing and revising their mental models. It should also allow us to investigate the role of mental imagery in these processes in a way that has not before been possible. To our knowledge, there are only a few previous researchers who have made an evidence-based argument for the involvement of imagery in analogical reasoning in science contexts [29,43,78], and in extreme cases or Gedanken experiments [29,79].

We asked whether we could develop a set of observables that could provide evidence for the use of mental imagery while students are learning. Our method was influenced by the prior work of one of us [29] which indicates that many of the 11 experts he studied made primary and integral use of imagistic, nonformal reasoning methods when problem solving in physics. However, we believe that many science teachers do not recognize the importance of these processes in scientific thinking. Clement developed definitions that relied

on observables for several of the expert processes so that expert transcripts could be coded for the occurrence of these. However, in the present study, when trying to apply these definitions and imagery indicators to student videotapes, we ran into a number of challenges. In addition to the interruptions and general difficulties in interpreting student statements in a classroom discussion, there were several problems in particular that faced us methodologically:

(1) How should we deal with an episode of reasoning that is split between two or more students?

(2) How should we deal with a reasoning process that is initiated by the teacher but developed by the students (or vice versa)?

(3) Should comments on or modifications of a previous case be counted as a new case?

(4) How should we deal with episodes in which two or more kinds of reasoning are used on a single case?

Assembling a way to detect reasoning and imagery indicators under these conditions was challenging and involved several cycles of refinement in order to attain stable definitions. An important decision for us was to split each reasoning type into the *generation* of a case and the *running* of the case. This made it easier to assign part of the reasoning about a case to different students; or part to the teacher and part to a student. Here, we were fairly conservative when students modified a case, most often not counting it as a new case. However, we believe some flexibility is warranted depending on the purpose of a study, and other authors who wish to focus on student modifications could choose to count such revisions.

Our long-term objective is to develop definitions that can capture student reasoning that is scientific even though it is nonformal. That is, *we want to be able to code for thinking that goes beyond the parroting of facts or the solving of formulaic problems*, that grapples with the essence of scientific phenomena in a way that reflects the practice of science. In this study we have taken the initial steps in developing such procedures.

### Instructional implications

In this section we go beyond our data based findings to speculate on possible educational implications. Several of the teachers in our study have reported they are now noticing when students are using extreme cases and analogies, apparently lending a new dimension to these teachers’ awareness of the qualities of student discussion. In our experience, the tendency is for some teachers to see such spontaneous reasoning as a distraction away from their lesson plans. We believe that becoming acquainted with clear definitions cast in terms of observables and being exposed to examples of their appearance in classroom discussion could help teachers recognize these reasoning processes and, more importantly, recognize fruitful discussion when it occurs.

It may come as a surprise to some (as it did to us) that students in the gravity lesson did not all immediately see that the Earth’s rotation would tend to make scale readings smaller rather than larger. The extended discussion and efforts of students to improve upon their Gedanken experi-

ments and analogies, as they struggled to convince others of this point, testifies to its import as a nontrivial topic for these students. After studying the entire tape, our impression is that the type of imagery involved in grasping this was doable but challenging for most of the students in the class. Yet this would appear to be at the heart of an understanding based on sense making for why gravity cannot be caused by rotation.

In these discussions, students' plausible reasoning arguments were made both for and against the scientific point of view, but this appeared to add fuel to the discussion. For instance, the teacher needed to make a judgment in the gravity lesson as to whether to interrupt with a discussion of the nature of centrifugal force as a pseudoforce and decided not to. Since this lesson came early in the curriculum, he felt that the students were not ready for the subtler issues involved, and that these could distract students from the more basic issues at hand. In both discussions, over a period of about 45 min, the student arguments did converge on reasons in favor of the accepted scientific views of gravity and of normal forces. This suggests that these student-generated processes can in some cases also contribute to content goals. (In each class, the teacher also made sure that the students understood his position on the major issues by the end of class.)

More directly, the two analyses provide an initial "existence demonstration" that high school students can engage spontaneously in nonformal scientific reasoning processes that speak to process (scientific thinking) goals. We hypothesize that the teacher facilitated this by encouraging open discussions that were guided to stay on topic but that were open to a variety of student ideas both for and against the canonical view. The three processes were all exemplar-based; we hypothesize that encouraging student generation and modification of concrete exemplars, along with the use of gestures and drawings to communicate these, may help facilitate such processing.

We presented some examples where the teacher generated a case and one or more students ran the case. In these examples, the process of running an imagistic simulation on a concrete case appeared to be an intuitive process for these students. It is important to know that if the right case is generated by the teacher, the students may be able to invest in it by running it themselves if time is allowed for this. Spontaneous student generation of a strategic case for mental simulation may be a more difficult task, but students in these classes were also observed to do just that.

## IX. CONCLUSION

One of the purposes of exploratory case studies is to raise issues for future research. Though our sample was small, we believe that the presence of imagery indicators during most episodes of student generation of the three scientific reasoning processes we documented here suggests that the role of imagery in physics learning should be taken seriously as a topic for future research.

If further study shows imagery to be generally central in this kind of reasoning, it would strongly suggest that teachers pay increased attention to strategies that support imagery, such as having students make drawings, encouraging the use of gestures and drawings for visual communication, and emphasizing the importance of mentally animating diagrams and drawings.

The present case studies demonstrate that it was possible for these students to engage in creative scientific reasoning during classroom discussions. In the present study there was evidence that imagistic simulation occurs in conjunction with these processes. The most challenging tasks for the researchers in this study were those of defining, criticizing, redefining, and refining the central concepts behind, and observable indicators for, these three types of reasoning, and delineating types of gestures as imagery indicators. We hope this will contribute to further research on the nature of student reasoning in classrooms.

## ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. REC-0231808. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

## SUPPLEMENTARY AUXILIARY MATERIAL

See separate auxiliary material supplement for Appendixes A, B, and C.

- 
- [1] A. Harrison and O. de Jong, Exploring the use of multiple analogical models when teaching and learning chemical equilibrium, *J. Res. Sci. Teach.* **42**, 1135 (2005).
- [2] M. Cosgrove, A study of science-in-the-making as students generate an analogy for electricity, *Int. J. Sci. Educ.* **17**, 295 (1995).
- [3] D. Hammer, Student inquiry in a physics class discussion, *Cogn. Instruct.* **13**, 401 (1995).
- [4] K. Schultz and J. Clement, in *Preconceptions in Mechanics: Lessons Dealing with Conceptual Difficulties*, 1st ed., edited by C. Camp *et al.* (Kendall Hunt, Dubuque, Iowa, 1994), Appendix, p.1.
- [5] A. G. Harrison and D. F. Treagust, A typology of school science models, *Int. J. Sci. Educ.* **22**, 1011 (2000).
- [6] K. Dunbar, in *The Analogical Mind: Perspectives from Cognitive Science*, edited by D. Gentner, K. J. Holyoak, and B. N. Kokinov (MIT Press, Cambridge, MA, 2001), p. 313.
- [7] B. Glaser, *Emergence vs. forcing: Basics of Grounded Theory analysis* (Sociology Press, Mill Valley, CA, 1992).
- [8] A. Strauss and J. Corbin, in *Handbook of Qualitative Research*, edited by N. Denzin and Y. Lincoln (Sage Publications, Thousand Oaks, CA, 1994), p. 273.

- [9] M. Hegarty, Mental animation: Inferring motion from static diagrams of mechanical systems, *J. Exp. Psychol. Learn. Mem. Cogn.* **18**, 1084 (1992).
- [10] L. Darden, *Theory Change in Science: Strategies from Mendelian Genetics* (Oxford, New York, 1991).
- [11] J. Clement, Using bridging analogies and anchoring intuitions to deal with students' preconceptions in physics, *J. Res. Sci. Teach.* **30**, 1241 (1993).
- [12] D. Gentner and D. R. Gentner, in *Mental Models*, edited by D. Gentner and A. L. Stevens (Erlbaum, Hillsdale, NJ, 1983), p. 99.
- [13] R. N. Giere, *Explaining Science: A Cognitive Approach* (Chicago University Press, Chicago, 1988).
- [14] E. Ochs, P. Gonzales, and S. Jacoby, in *Interaction and Grammar*, edited by E. Ochs, E. A. Schegloff, and S. A. Thompson (Cambridge University Press, Amsterdam, 1996), pp. 328–369.
- [15] D. Gentner, in *International Encyclopedia of the Social and Behavioral Sciences*, edited by N. J. Smelser and P. B. Bates (Elsevier Science, Amsterdam, 2002), p. 9683.
- [16] N. Nersessian, Should physicists preach what they practice? Constructive modeling in doing and learning physics, *Sci. Educ.* **4**, 203 (1995).
- [17] M. C. Nunez-Oviedo, Doctoral dissertation, University of Massachusetts, 2003.
- [18] R. Driver, *The Pupil as a Scientist?* (Open University Press, Milton Keynes, 1983).
- [19] J. Smith, A. diSessa, and J. Roschelle, Misconceptions reconceived: A constructivist analysis of knowledge in transition, *J. Learn. Sci.* **3**, 115 (1994).
- [20] J. F. Osborne, S. Erduran, and S. Simon, Enhancing the quality of argumentation in school science, *J. Res. Sci. Teach.* **41**, 994 (2004).
- [21] K. L. McNeill and J. Krajcik, Scientific explanations: Characterizing and evaluating the effects of teachers' instructional practices on student learning, *J. Res. Sci. Teach.* **45**, 53 (2008).
- [22] D. B. Clark and D. Jorde, Helping students revise disruptive experientially-supported ideas about thermodynamics: Computer visualizations and tactile models, *J. Res. Sci. Teach.* **41**, 1 (2004).
- [23] R. A. Duschl and J. Osborne, Supporting and promoting argumentation discourse in science education, *Stud. Sci. Educ.* **38**, 39 (2002).
- [24] S. Toulmin, *The Uses of Argument* (Cambridge University Press, Cambridge, England, 1958).
- [25] D. Walton, *Argumentation Schemes for Presumptive Reasoning* (Erlbaum, Mahwah, NJ, 1996).
- [26] J. Clement, in *Implicit and Explicit Knowledge: An Educational Approach*, edited by D. Tirosh (Ablex Publishing Corp., Norwood, NJ, 1994).
- [27] J. Clement, The role of imagistic simulation in scientific thought experiments, *Topics in Cognitive Science* **1**, 686 (2009).
- [28] J. Clement, in *Model-Based Reasoning in Science and Engineering*, edited by L. Magnani (King's College Publications, London, 2006).
- [29] J. Clement, *Creative Model Construction in Scientists and Students: The Role of Imagery, Analogy, and Mental Simulation* (Springer, New York, 2008).
- [30] N. Nersessian, in *Philosophy of Science, Cognitive Psychology and Educational Theory and Practice*, edited by R. Duschl and R. Hamilton (State University of New York Press, Albany, NY, 1992), p. 48.
- [31] A. Collins and D. Gentner, in *Cultural Models in Thought and Language*, edited by D. Holland and N. Quinn (Cambridge University Press, Cambridge, England, 1987), p. 243.
- [32] J. Clement, D. Brown, and A. Zietsman, Not all preconceptions are misconceptions: Finding 'anchoring conceptions' for grounding instruction on students' intuitions, *Int. J. Sci. Educ.* **11**, 554 (1989).
- [33] N. S. Podolefsky and N. D. Finkelstein, Use of analogy in learning physics: The role of representations, *Phys. Rev. ST Phys. Educ. Res.* **2**, 020101 (2006).
- [34] E. D. Wong, Self-generated analogies as a tool for constructing and evaluating explanations of scientific phenomena, *J. Res. Sci. Teach.* **30**, 367 (1993).
- [35] M. J. Else, J. Clement, and M. A. Ramirez, in *Model Based Learning and Instruction in Science*, edited by J. Clement and M. A. Ramirez (Springer Netherlands, Dordrecht, 2007), Chap. 12, pp. 215–231.
- [36] A. Zietsman and J. Clement, The role of extreme case reasoning in instruction for conceptual change, *J. Learn. Sci.* **6**, 61 (1997).
- [37] N. Nersessian, *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, edited by D. Hull, M. Forbes, and K. Okruhlick (Philosophy of Science Association, East Lansing, MI, 1992), Vol. 2: Symposia and Invited Papers, p. 291–301.
- [38] D. Gooding, *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, edited by D. Hull, M. Forbes, and K. Okruhlick (Philosophy of Science Association, East Lansing, MI, 1992), Vol. 2: Symposia and Invited Papers, p. 280–290.
- [39] J. R. Brown, Thought experiments since the scientific revolution, *International Studies in the Philosophy of Science* **1**, p.1 (1986).
- [40] T. S. Kuhn, in *The Essential Tension*, edited by T. Kuhn (University of Chicago Press, Chicago, 1977) pp. 240–265.
- [41] R. A. Sorensen, *Thought Experiments* (Oxford University Press, Oxford, 1992).
- [42] J. Clement, *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society*, 2002, edited by W. Gray and C. Schunn (Erlbaum, Mahwah, NJ, 2002), Vol. 22, p. 32.
- [43] J. Clement and M. Steinberg, Step-wise evolution of mental models of electric circuits: A "learning-aloud" case study, *J. Learn. Sci.* **11**, 389 (2002).
- [44] J. K. Gilbert and M. Reiner, Thought experiments in science education: Potential and current realization, *Int. J. Sci. Educ.* **22**, 265 (2000).
- [45] M. Reiner and J. Gilbert, Epistemological resources for thought experimentation in science learning, *Int. J. Sci. Educ.* **22**, 489 (2000).
- [46] M. Reiner, Thought experiments and collaborative learning in physics, *Int. J. Sci. Educ.* **20**, 1043 (1998).
- [47] M. C. Nunez-Oviedo, J. Clement, and M. A. Ramirez, *Model Based Learning and Instruction in Science*, Developing Complex Mental Models in Biology Through Model Evolution, edited by J. Clement and M. A. Ramirez, Models and modeling in Science Education, edited by J. K. Gilbert, Vol. no. 2, (Springer, Dordrecht, The Netherlands, 2008), Chap. 10, pp.

- 173–193.
- [48] M. Kozhevnikov, M. Hegarty, and R. Mayer, *Annual Meeting of the American Educational Research Association*, Montreal (unpublished).
- [49] A. Miller, *Imagery in Scientific Thought* (MIT Press, Cambridge, MA, 1986).
- [50] D. Hestenes, Secrets of genius: Review of Imagery in Scientific Thought, *New Ideas Psychol.* **8**, 231 (1990).
- [51] R. Finke, *Principles of Mental Imagery* (MIT Press, Cambridge, MA, 1989), p. 2.
- [52] J. R. Frederiksen, B. Y. White, and J. Gutwill, Dynamic mental models in learning science: The importance of constructing derivational linkages among models, *J. Res. Sci. Teach.* **36**, 806 (1999).
- [53] J. Clement, *Proceedings of the NARST 2006 Annual Meeting*, San Francisco, CA, 2006 (The National Association for Research in Science Teaching, Reston, VA, (2006).
- [54] J. Clement, A. Zietsman, and J. Monaghan, in *Visualization in Science Education*, edited by J. Gilbert (Springer, Dordrecht, The Netherlands, 2005).
- [55] D. Gooding, Creative rationality: Towards an abductive model of scientific change, *Philosophica* **58**, 73 (1996).
- [56] C. Camp, J. Clement, D. Brown, K. Gonzalez, J. Kudukey, J. Minstrell, K. Schultz, M. Steinberg, V. Veneman, and A. Zietsman, *Preconceptions in Mechanics: Lessons Dealing with Conceptual Difficulties* (Kendall Hunt, Dubuque, Iowa, 1994).
- [57] J. A. Sellarès and G. Toussaint, On the role of kinesthetic thinking in computational geometry, *Int. J. Math. Educ. Sci. Technol.* **34**, 219 (2003).
- [58] M. W. Alibali, Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information, *Spat. Cogn. Comput.* **5**, 307 (2005).
- [59] D. McNeill, *Hand and Mind: What Gestures Reveal About Thought* (University of Chicago Press, Chicago, IL, 1992).
- [60] R. E. Scherr, Gesture analysis for physics education researchers, *Phys. Rev. ST Phys. Educ. Res.* **4**, 010101 (2008).
- [61] S. Lozano and B. Tversky, *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society*, edited by B. Bara and M. Bucciarelli, Stresa, Italy, (Erlbaum, Mahwah, NJ, 2005).
- [62] J. M. Iverson and S. Goldin-Meadow, What's communication got to do with it? Gesture in children blind from birth, *Dev. Psychol.* **33**, 453 (1997).
- [63] J. M. Iverson and S. Goldin-Meadow, Why people gesture when they speak, *Nature (London)* **396**, 228 (1998).
- [64] R. M. Krauss, Why do we gesture when we speak? *Curr. Dir. Psychol. Sci.* **7**, 54 (1998).
- [65] A. Hostetter and M. Alibali, *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society*, edited by K. Forbus, S. Gentner, and T. Regier (Erlbaum, Chicago, IL, 2004), p. 589.
- [66] P. Feyereisen and I. Havard, Mental imagery and production of hand gestures while speaking in younger and older adults, *J. Nonverbal Behav.* **23**, 153 (1999).
- [67] S. Goldin-Meadow, The role of gesture in communication and thinking, *Trends Cogn. Sci.* **3**, 419 (1999).
- [68] S. Goldin-Meadow, *Hearing Gesture: How Our Hands Help Us Think* (Harvard University Press, Cambridge, London, 2003).
- [69] W.-M. Roth, From gesture to scientific language, *J. Pragmat.* **32**, 1683 (2000).
- [70] M. Hegarty, S. Mayer, S. Kriz, and M. Keehner, The role of gestures in mental animation, *Spat. Cogn. Comput.* **5**, 333 (2005).
- [71] J. M. Monaghan and J. Clement, Use of a computer simulation to develop mental simulations for understanding relative motion concepts, *Int. J. Sci. Educ.* **21**, 921 (1999).
- [72] M. Alibali and M. Nathan, in *Video Research in the Learning Sciences*, edited by R. Goldman, R. Pea, B. Barron, and S. Derry (Lawrence Erlbaum Associates, Mahwah, NJ, 2007).
- [73] A. L. Stephens and J. Clement, *2006 Physics Education Research Conference*, edited by L. McCullough, L. Hsu, and P. Heron (American Institute of Physics, Melville, NY, 2006), Vol. 883, p. 89.
- [74] A. L. Stephens and J. Clement, *Annual Meeting of the American Educational Research Association*, San Francisco, 2006 (unpublished).
- [75] M. Q. Patton, *Qualitative Evaluation and Research Methods*, 2nd ed. (Sage Publications, Newbury Park, CA, 1990).
- [76] M. Miles and A. M. Huberman, *Qualitative Data Analysis*, 2nd ed. (Sage, London, 1994).
- [77] R. Catrambone and K. J. Holyoak, Overcoming contextual limitations on problem-solving transfer, *J. Exp. Psychol. Learn. Mem. Cogn.* **15**, 1147 (1989).
- [78] N. Nersessian, *Creating Scientific Concepts* (MIT Press, Cambridge, MA, 2008).
- [79] M. Reiner and J. Gilbert, The symbiotic roles of empirical experimentation and thought experimentation in the learning of physics, *Int. J. Sci. Educ.* **26**, 1819 (2004).