The Role of Thought Experiments in Science and Science Learning

A. Lynn Stephens and John J. Clement

University of Massachusetts—Amherst

Correspondence:

Lynn Stephens

428 Lederle GRT

710 N. Pleasant St.

Amherst, MA 01003-9305

Ph: 413-230-4634

Fax: 413-545-4884

Email: lstephens@educ.umass.edu

John Clement

428 Lederle GRT

710 N. Pleasant St.

Amherst, MA 01003-9305

Ph: 413-545-0133

Fax: 413-545-4884

Email: clement@educ.umass.edu

The original publication is available at www.springerlink.com

Abstract. This chapter will (1) briefly review selected studies examining the nature of thought experiments in science; (2) review previous studies on the role that thought experiments can play in science instruction; (3) give case study examples of thought experiments (TEs) proposed by both teachers and students and the ensuing classroom discussions. We discuss several definitions for the term thought experiment and examine methods that have the potential to illuminate issues such as the following: students can generate their own TEs as well as discuss ones proposed by the teacher; students give evidence of using imagery during TEs as indicated by certain imagery indicators; one can track how a TE spreads "contagiously" between students in a discussion and how it is modified and improved in the process. We will conclude that student TEs can be similar to expert TEs in many ways and raise possible factors that make teacher generated TEs foster student discussion and sense making.

Keywords: informal learning, inquiry, learning, nature of science, research methods

In this chapter we review selected studies of thought experiments used by both experts and students and attempt to develop some useful definitions and conceptual distinctions. We then apply these in an analysis of a classroom episode as an example of the roles thought experiments can play in productive whole class discussions. We are interested in this area because thought experiments are one example of the kinds of creative reasoning of which experts and students appear to be capable under the right conditions.

REVIEW OF SELECTED STUDIES ON THOUGHT EXPERIMENTS IN SCIENCE EXPERTS

Certain writers in philosophy of science have been intrigued with thought experiments (TEs) for some time because if effective, they seem to contradict the spirit of empiricism that dominated philosophy of science for much of the 20th century. The idea of obtaining new knowledge from internal mental manipulations alone does not sit comfortably within an empiricist framework.

Authors such as David Brown (1991) and Roy Sorensen (1992) have compiled collections of TEs that were important in the history of science. By now it is widely recognized that at least some TEs in the history of science have been noticeably, if not spectacularly, germane to a scientist's investigation. Famous examples include those used in the Einstein–Bohr debates on quantum mechanics. Nancy Nersessian (1992) has analyzed historical records of Maxwell's breakthroughs in electromagnetic field theory, finding that a series of thought experiments involving gears and then fluid vortices played a role in his theory formulation.

TEs also have been considered somewhat enigmatic and exotic. The reason for this is captured in what John Clement (2002, p.32) called the "Fundamental Paradox of Thought Experiments," namely, "How can findings that carry conviction result from a new experiment conducted entirely within the head?" The idea of an experiment (involving observation) being conducted in the head (without observation) appears self-contradictory.

Purposes for Thought Experiments

One line of investigation is to examine the purpose served by thought experiments. Thomas Kuhn (1977) argued that the purpose of a TE is to disconfirm a theory by disclosing a conflict between ones existing concepts and nature. Undoubtedly, TEs are probably most impressive when they act to disconfirm an established theory in science; then they actually seem to be doing something as powerful as a critical experiment or anomaly can do.

On the other hand, Brown (1991) identified several purposes for TEs including constructive as well as destructive (conflict-generating) purposes. He also theorized that a few special TEs could serve both functions. Similarly, Nersessian's (1992) analysis of Maxwell's work hypothesized that a TE could expose conflicts in an existing theory but also point to new constraints that help guide positive modifications of the theory, thus playing both a destructive and constructive role. Interestingly, Athanasios Velentzas, Krystallia Halkia and Constantine Skordoulis (2005) found that textbooks in relativity and quantum mechanics use constructive but not destructive TEs; they feel the inclusion of destructive TEs could increase student interest.

Clinical Studies

Evidence in historical and philosophical studies has been indirect because these studies have not been able to examine real time evidence for purposes and mechanisms of TEs as they are being used. Clement (2008, in press) attempted to collect such evidence by interviewing experts thinking aloud about unfamiliar explanation problems. Think-aloud transcripts are not perfect or complete records of thinking but they do provide considerably more detail than historical papers. He found cyclical sequences of model construction and evaluation, and different TEs being used for model generation (constructive) and model evaluation purposes. He also found that within the evaluation category, TEs could be either disconfirmatory or confirmatory. These studies also confirmed that TEs could be used as a part of the actual thinking process, not just pedagogically.

One problem used was the 'Spring Problem', which asks whether a first spring would stretch more than a second spring that is identical except with coils twice as wide in diameter. In the simplest possible example of a TE, one subject simply tried to imagine which spring would be harder to pull, saying:

> Episode 1: I'm **going to try to visualize it** to imagine what would happen— my guess would be that it [wider spring] would stretch more— this is a kind of **kinesthetic sense that somehow a bigger spring is looser**… .

This is certainly a more primitive experiment than the famous TEs in history of science, and yet it has the basic qualities of imagining the results of an experiment in the head. (The bold type in these episodes denotes imagery indicators, to be discussed later.)

A more creative experiment was generated when this subject engaged the question of whether the deformation in the spring wire is due primarily to bending or to twisting of the wire as the spring stretches. He generated the case of a spring made of a vertically oriented band of material, depicted in Figure 1. The reader might imagine the thin metal strip unwound from a coffee can, reshaped to make a spring 8 cm or so in diameter.

> Episode 2: How about a spring made of something that can't bend. And if you showed that it still behaved like a spring you would be showing that the bend isn't the most important part— **How could I imagine** such a structure?— I'm thinking of something that's made of a band— **we're trying to imagine configurations** that wouldn't bend. Since its cross section is like that [see Figure 1]— it can't bend in the up-down **[indicates up/down directions with hands]** direction like that because it's too tall. But it can easily twist **[gestures as if twisting an object]**.

He infers that the spring can still stretch even though it cannot bend, arguing against the theory of bending as necessary for stretching. Here there is more clearly a design process leading to a contradiction.

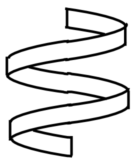


Fig. 1 Band spring

Definitions

A problem in the literature is that there is no consensus on a definition of a TE. Sorensen (1992, p. 205) defines a thought experiment as "(A)n experiment that purports to achieve its aim without the benefit of execution". However, this shifts much of the burden to the term "experiment". Experiment is defined as "a procedure for answering or raising a question about the relationship between variables by varying one (or more) of them and tracking any response by the other" (p. 186). But as we shall see, some TEs appear to be less formal than a procedure and some appear to envision a single event without systematic variation; alternative definitions may be worth exploring.

The range of TEs in the above episodes—from simple to complex—motivated our formulation of a broad definition and a narrow definition for TEs (Clement 2008, p. 278), as follows:

> *Broad definition:* Performing an (untested) thought experiment (in the broad sense) is the act of considering an untested, concrete system (the "experiment" or case) and attempting to predict aspects of its behavior. Those aspects of behavior must be new and untested in the sense that the subject has not observed them before nor been informed about them.

The word "untested" is used to rule out cases where the subject simply replays a previously observed event. Still, the above definition is intentionally quite broad and encompasses cases as simple as in the first episode above.

> *Narrow definition:* Performing an evaluative Gedanken experiment is the act of considering an untested, concrete system designed to help evaluate a scientific concept, model, or theory-- and attempting to predict aspects of the system's behavior.

The second band spring episode above had these characteristics since it was designed to test the theory that bending is the source of stretching in springs. In the first episode, the subject was trying to make a prediction only for the specific system and not to test a broader theory.

Possible advantages of these definitions are that they are more inclusive by not depending crucially on the subject having proposed a formal experiment; they are somewhat more operational (possible to agree on recognizing) in emphasizing a process rather than a product; and the first one fits the paradox better by being somewhat broader than the set of carefully designed scientific Gedanken experiments.

Mechanisms: What Processes do Scientists Use to Run TEs?

It is difficult to analyze the mental processes that allow a scientist to generate and run a TE during an investigation by using historical data because the original thought process can easily be buried under many changes and refinements authors carry out before publishing a thought experiment. Also, for many TEs it is hard to know whether they were originally part of a discovery process or created after the investigation to convince others. Nevertheless, working from the thought experiments themselves, a number of authors have hypothesized at least a rough description of processes that may have been involved. Debates have emerged among disparate theories ranging from those defending an empiricist view to those proposing a rationalist alternative.

Several intermediate positions have been postulated. Miriam Reiner and John Gilbert (2000) ask what is the source of conviction in TEs. They point out, for instance, that Poisson conducted a TE that led him to make a professionally high-risk claim—without having performed the experiment. They theorize that the intellectual power of a TE is in the integration of conceptuo-logical beliefs, mental visual imagery and bodily knowledge, and suggest that the last two bring tacit knowledge to bear on the problem. Nersessian (1992) hypothesized that TEs utilize simulative mental models and that, "The constructed situation inherits empirical force by being abstracted from both our experiences and activities in, and our knowledge, conceptualizations, and assumptions of, the world" (p. 297). Likewise, Reiner (1998) posited that one necessary component for thought experimentation is construction of mental imagery in order to build the hypothetical world of a TE.

Clement (1994) attempted to speak to mechanism questions on the basis of real time data by looking for imagery indicators in videotapes of experts. The bold type in the two episodes above denotes several instances of imagery indicators. In order of appearance, they are: Episode 1—announces intent to form image, kinesthetic imagery report; Episode 2—announces intent to form image, imagery report, and depictive gestures.

Such imagery indicators accompanied many TEs in these videotapes, leading to the proposal that a process of imagistic simulation underlay those TEs. In this process, a perceptual motor schema generates dynamic imagery, complimented by nonformal, rationalistic contributions from general spatial reasoning operations and the ability to combine two such schemas in new combinations. Evidence from these studies suggests that premises can be in the form of implicit physical intuitions apprehended in imagistic simulations, rather than being explicit linguistic propositions or axioms, and that reasoning with these can involve spatial reasoning or constructed compound simulations that are less formal than rule-based arguments. These mechanisms provided a way to speak to the TE paradox, showing how a TE could feel empirical but actually involve a considerable amount of reasoning inside the head (Clement 2008). Much of the prior work on this topic has involved the analysis of TE cases from the history of science; only recently has data been collected on the *process* of producing and running TEs.

Analytical schemes for TEs

Several investigators have suggested analytical schemes for TEs. For instance, Reiner (1998) identified a five-part structure of TEs: hypothetical world, hypothesis, experiment, results and conclusion. She hypothesized that the conclusion of a TE is based on logical derivations, although in a later paper (Reiner & Gilbert 2000) she stressed that TEs have a non-propositional aspect. The extent of the role of logical derivation has also been examined by Clement (2008). This analysis of spontaneous expert TEs indicates that TEs are often run in a non-formal, imagistic, or intuitive manner.

How TEs can go wrong

Miriam Reiner and Lior Burko (2003) analyze five TEs from history of science according to Reiner's five stages (1998), and identify stages at which errors occurred. In the TEs studied, errors were usually made in the first two stages: constructing the hypothetical world and formulating the hypothesis. Reiner and Burko draw implications for the use of TEs in education; this will be discussed further below.

REVIEW OF PREVIOUS STUDIES ON ROLES THOUGHT EXPERIMENTS CAN PLAY IN SCIENCE INSTRUCTION

TEs can be used by students

Early work by Hugh Helm, John Gilbert and Michael Watts (1985) describes students spontaneously generating their own TEs. Since then, a number of studies have documented the fact that TEs can be used by students in educational contexts. In most of these studies, Sorensen's definition is used or the concept of TE is left undefined.

Reiner (1998) found that episodes containing at least three parts from her five-part structure of TEs (described in the expert section above) were prevalent in the transcripts of 12 grade-eleven students working in small groups at computers with interactive schematic representations. In this study, it was assumed that interactive graphical dynamic representations generated by computer served as "basic tools for learning processes that require(d) imagery" (p. 1046). Therefore, the imagery of the students was scaffolded by a display jointly viewed by several students. It might not seem surprising that, in Reiner's view, these students appeared to share mental animations that yielded similar results. However, Reiner also documents instances where students reasoned about variations of the system that had not yet been shown on the screen and agreed on predictions for these absent configurations. Especially in these instances, she argues, the students appeared to be relying on mental imagery. Working with older students, Reiner and Gilbert (2000) observed senior undergraduate physics majors and physics education majors as they solved problems designed to elicit TEs. They found that thought experimentation was a frequently used strategy.

In another instructional approach, Gilbert and Reiner (2004) found that 12 and 13 year old students working in small groups constructed and ran thought experiments intertwined within the processes of conducting physical experiments. Transcripts showed students making progress toward scientific ideas by alternating between these imaginary and physical models. The students also used gestures and drawings to communicate ideas when trying to model how a physical system worked. This study suggests that the interplay between experiments, drawings, and thought experiments can be very rich.

Maria Nunez-Oviedo, John Clement and Mary Anne Rea-Ramirez (2008) investigated the role of TEs with a similar age group. In middle school classrooms, the teacher was observed inviting students to run TEs both to support modification of ideas and to disconfirm ideas. Nunez-Oviedo et al. report that students were able to reason with the scenarios to arrive at scientifically accepted ideas. They argue that TEs can be used and are plausibly important at the middle school level.

Thought experiments—even Gedanken experiments—spontaneously generated and run by high school students need not be jointly constructed, though they may be inspired by the comments of other students. Lynn Stephens and John Clement (2006) found that students independently could generate novel scenarios, make predictions from those scenarios, and evaluate those predictions on their own during class discussion. David Hammer (1995) considered thought experiments in high school physics class discussions as one of several kinds of process skills that were exhibited by students when the teacher in his case study took care to foster an open attitude toward contributing ideas.

Importance of TEs in teaching and learning

Gilbert and Reiner's (2004) work suggests that TEs can play an important role in physical (real) experimentation, suggesting modifications to physical experiments and alternating with them to lead to a convergence on accepted scientific concepts (in this case, of unusual sophistication for middle school level, as the students themselves spontaneously generated the beginnings of a concept of magnetic field). Helm et al. (1985) speculate that TEs can play an important role in conceptual change because they have the ability to arouse dissatisfaction with existing conceptions. There are several questions they believe need to be answered, including: Is the classic structure of TEs drawn from physics the ideal structure of TEs to be used in pedagogical contexts? How far does TE overlap with analogy? What can be done to support students in their spontaneous generation of TEs?

Some recent studies have begun to address these and similar questions. For instance, what gives a model the ability to generate dynamic imagery, which then can be used to generate predictions during a TE? Clement (2008) hypothesized that some primitive physical intuitions have this kind of 'runnability' built into them in the form of perceptual motor schemas (such as a schema embodying ideas about pressure). When these are used as components in an explanatory model, the model can inherit this capability for generating dynamic imagery. This transfer of runnability is used to explain the ability of some analogies to serve as seed material for developing an explanatory model. So, for example, a student can develop a model of electric circuits based on a metaphor of electric pressure, with pressure spreading equally throughout equipotential (connected) areas of a circuit and pressure differences driving flow through resistors. When such a model is used to make a prediction for the first time, or used flexibly on a transfer problem involving a circuit with a type of geometry the student has never seen before, this is an instance of a thought experiment in the broad sense of the term used here—they are making an as yet untested prediction. In this case, it is being run via an imagistic simulation.

This hypothesis of transfer of runnability was supported by case study evidence (John Clement and Melvin Steinberg 2002). A subject's spontaneous use of depictive gestures over drawings while she processed an air pressure analog case, and her use of similar gestures during later instructional circuit episodes, indicated that she was using a similar type of imagistic simulation in the two cases. Furthermore, the subject's spontaneous use of similar depictive gestures during a later posttest provided evidence that the instruction fostered development of a dynamic mental model of fluid-like flows of current caused by differences in electric pressure, a model that could generate new imagistic simulations for understanding relatively difficult transfer problems.

Thus, in addition to the use of Gedanken experiments, students making a prediction for an unfamiliar analogy or running a new model for the first time, or applying a model to an unfamiliar transfer problem, are doing an untested thought experiment. There is case study evidence from both experts and students that all of these operations can involve imagistic simulation (Clement 2008). This suggests that this kind of rationalistic, hypothetical, imagistic thinking via TEs can be important in many more learning situations than we might initially imagine, and that it is an extremely important complement to empirical and algorithmic work. A related theme was developed by Hammer (1995), who identified a number of rationalistic process goals being fulfilled in whole class discussion that are quite different from the classic, more empirically oriented process goals in science originally identified by Michael Padilla (1991). This points to the importance of understanding student use of TE processes in both the broad and narrow senses.

A CASE STUDY

In the interest of aiding further research on TEs in instruction, we will illustrate a method using the two-tiered definition of thought experiments from Clement (2002) to identify transcript evidence that students can generate TEs at both tiers. We will also illustrate how a set of imagery indicators from Clement (2008) can be used to show that there is evidence for the involvement of mental imagery as students ran the TEs.

These recent analysis methods (Stephens and Clement submitted) are aimed at questions such as the following:

- Can we identify evidence that students use TEs?
- Can we identify evidence that students can generate and run their own TEs?
- Are the appearances of TEs isolated or do they have impact on classroom discussion?
- Can students evaluate TEs? Can they modify or improve them?
- Can we associate student use of imagery with the running of TEs?
- If so, can we identify evidence for particular kinds of imagery; i.e., visual or kinesthetic?

The 2-tiered definition applied to transcript analysis

We have examined a number of transcripts of classroom activity to see whether evidence for student-generated TEs could be identified (Stephens and Clement 2006). In most of this classroom activity, guided inquiry methods of teaching and learning were being employed. We developed coding criteria based on the two-tiered definition for TEs, and we selected for more detailed analysis portions of transcripts where creative student reasoning appeared to be occurring. We were able to identify what seemed to us a surprising number of instances that met our criteria for student-generated thought experiments including several evaluative Gedanken experiments.

For coding purposes, the definition for the broad category of untested TEs (above) was broken into two requirements, which were coded for separately:

1. Subject attempts to predict behavior of concrete system;
2. Subject has not observed the experiment before, nor been informed about its behavior.

Example. A physics class is discussing possible causes for gravity including the rotation of the Earth (a common misconception). A student refers to a chalkboard drawing of the Earth with a stick figure of a man standing on it (Figure 2).

Line 40, S5: Well, I just think that gravity has nothing to do with rotation, but maybe with rotation, like, that guy is trying to get thrown off the Earth. So he's getting pulled at the same rate but he's also getting pushed away.

S5 attempts to predict the behavior of a concrete system, a rotating Earth with a man standing on it. He has never observed the Earth from this vantage point and certainly has not experienced it spinning rapidly enough to feel the effects of being thrown off. This meets our criteria for a TE in the broad sense.

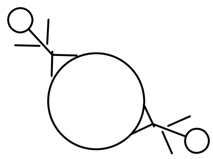


Fig. 2 US/Australia Case

For all episodes that had been coded as having evidence for TEs in the broad sense, we applied more restrictive coding criteria to establish whether each episode also met our definition for TEs in the narrow sense, evaluative Gedanken experiments. In addition to 1 and 2 above, we required that

3. The case appears to have been designed or selected by the subject in order to help evaluate a scientific concept, model, or theory.

The TE of Line 40 above appeared to have been selected by the subject in order to help evaluate the theory that rotation is a cause of gravity and so met the additional criterion of a Gedanken experiment.

All cases that met Criteria for TEs in either the broad or narrow senses were also analyzed for the following factors:

- whether the TE was generated by the teacher or the student,
- whether the TE was run by the teacher or the student.

The distinction between generating a TE and running it is an interesting one. What we have termed a pedagogical TE[1] may be generated in order to ask ones audience to make a prediction about a system where the results are unknown to the audience but known to the generator. Often, the pedagogical TEs in the transcripts we analyzed were generated by the teacher and run by the students; however, there are several incidences where we believe a student generated a TE, the outcome of which he or she was already certain, in order to convince fellow students of a point. At other times, a student generated and ran an untested TE and another student refined and reran it as a Gedanken with differing or refined results, or a student proposed a concrete case as an exemplar of some idea and another student used the case to generate an untested prediction, thus running it as an untested TE. Because of this network-like aspect of suggested test cases, untested TEs run on those cases, and Gedanken experiments (which might incorporate multiple earlier TEs from either tier), it was difficult to count the TEs in an unambiguous way until we considered the generation of TEs separately from their running.

Evidence of Spontaneous TEs From a Classroom Transcript

In Stephens and Clement (2006), the transcript under analysis was of a whole class discussion that comprised 42 minutes over the span of two days in a senior level high school physics class. The transcript began when the teacher first introduced the topic of gravity.

We organized our data by "case" (denoted Case 1, Case 2, and so on), "variation of a case" (denoted 1a, 3f, and so on), and "episode" ("S5 reruns Case 2d as a Gedanken"). A case is a concrete example of a system, such as the case of one person standing in the US and another standing in Australia, each person experiencing gravitational forces. A variation of a case involved the same concrete example of the system but with some variable changed in a significant way (such as being taken to extreme beyond the normal range for the system) or an additional variable highlighted. For instance, when a student introduced the rotation of the Earth into the discussion about Case 1, we counted this as Case 1a. An episode involved a single student either generating or running a case or variation.

We identified six separate cases that were topics of discussion in this transcript. These included: Case 1, a spherical mass such as a planet with one or more people upon it experiencing gravitation; Case 2, two small objects not touching and not experiencing noticeable gravitational forces due to each other; Case 3, gravity inside a bell jar; Case 4, a spinning fair ride and the forces due to spinning felt by the riders; Case 5, a catapult and the forces experienced by a projectile; and Case 6, a space ship rapidly orbiting the sun. The teacher introduced Cases 1 and 3 as part of the planned lesson; Case 1 then gave rise to numerous variations by students. The other four cases were introduced spontaneously by students.

The discussion begins with the teacher asking the students to consider a drawing on the board (Figure 2).

The teacher explains that the upper stick figure is standing in the US and the lower in Australia and asks the students to "vote" on a "ballot" they have been given.

> Line I-5, T: Now. Vote Number 1 … (A)h, compared to the United States, gravity in Australia is: a little less, equal, a little bit more.

Students have differences of opinion on this, leading to a very active discussion. This is Case 1 in the chart in Figure 4 below.

Soon after the teacher presents this case, S4 responds that he thinks "somehow the fact that [the Earth] spins causes a lot of the main force of gravity". This is the Spinning Earth variation, Case 1a. The student has introduced spinning as an important variable, indicating that his model of gravity includes spinning. This was not coded as a TE because the student did not make a prediction about the behavior of the system; the outcome (that spinning causes the main force of gravity) was assumed beforehand.

Several students attempt to address this student's misconception, including S5, who reruns the Spinning Earth case as a TE (Line 40, described above). In fact, S5's prediction, that spinning will throw "that guy" off the Earth, becomes a hot topic of debate in the class. Note that he speaks of "that guy" as though it were the drawing on the board along with its stick figure that is doing the rotating. The student appears to use the case to help evaluate the effect of spinning in his mental model of gravity so, even though the student did not generate the case, we have classified the episode as the running of a TE, and in the narrow sense (that is, as Gedanken experimentation).

In spite of the attempts of several students to counter the idea, S4 and S6 continue to defend rotation as a cause of gravity. This leads to an incident where a student appears to adopt the case another student invented, convert it into an extreme case, and then run it as an evaluative Gedanken experiment. In Line 49, S7, who had been quiet until this point, suggests the following.

> Line 49, 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 imagine a ball floating in space you tape your feet to. And you start spinning the ball around, you're gonna feel like you're gonna be 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 feel the forces being spun around in a centrifugal force.

The massive earth has shrunk to a small ball and the spinning has increased from one revolution a day to many times a minute judging from his gestures on the videotape.

The transcript of the first day provides sufficient evidence to code five episodes of untested TEs being generated, two of them by students. Both of the latter were also coded as Gedankens. In addition, there is evidence that two students ran TEs generated by the teacher. At other points, students appear to be generating predictions but in each of those instances there is not enough information to determine whether the system in question was untested for those students (Lines 88 and 89). Coding in this conservative manner yielded four episodes in less than 20 minutes of tape where there was evidence for students generating and/or running TEs.

In Day 2, there was a new round of discussion in which, over 25 minutes, there is evidence for the generation of six new untested TEs, the first three by the teacher and the last three by students. Again, all three of the student-generated TEs were judged to be Gedankens. In addition, there were instances where students appeared to run TEs generated by other students or by the teacher.

The methodology used here resulted in the identification of evidence in 42 minutes of videotape for 11 episodes of TE generation, 5 of them Gedanken experiments generated by students. In addition, there was evidence for 7 episodes of students running TEs formulated by others, including two where they were run as Gedankens. Figure 3 gives a breakdown of coding results.

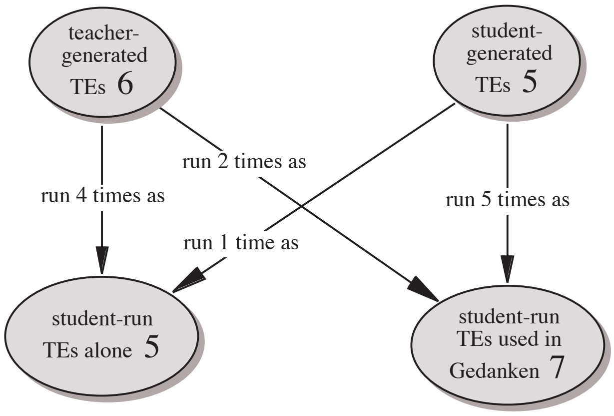


Fig. 3 Breakdown of TEs: TEs were run multiple times and in various combinations, so the number of TEs generated (top row) does not match the number of times TEs were run (bottom row). If the same TE was run twice by the same student, it was not double-counted.

Evidence of Imagery Use

Whether TEs are considered in the broad or the narrower sense, there is some evidence that they can involve imagery-rich mental simulation and that this dynamic imagery can enable the user to access implicit knowledge, rendering it more explicit (Clement 1994, in press). Identification of imagery-use indicators (Clement et al. 2005; Clement 2008) has allowed us to address further the question of whether classroom TEs can involve dynamic imagery.

We regard depictive gestures, which appear to depict an imaginary image "in the air" near the speaker, as providing some evidence for the involvement of mental imagery. In particular, we are interested in evidence for the use of animated or runnable mental imagery, which we obtain from gestures that appear to depict an imaginary motion or force. Identifying these types of gestures gives us a potential foothold on distinguishing between static and animated mental imagery.

For the Gedanken experiment of Line 40 discussed above, here is the same passage with gestures described.

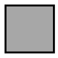
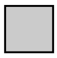
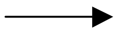
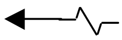
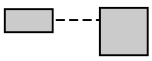
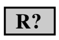
> Line 40, S5: Well, I just think that gravity has nothing to do with rotation, but maybe with [right forefinger rotates quickly, inscribing tiny circles in the air] rotation like [points to chalkboard] that guy is trying to get [emphatic, sweeping movement with his right hand and arm, moving across the front of his body from right to left] **thrown** off the Earth. So he's getting [repeats sweeping movement] **pulled** at the same rate but he's also getting [reverses previous movement, pulling his right hand and arm back to the right] **pushed** away.

With the exception of the pointing gesture, which refers to a real object rather than an imaginary image, the rest of these gestures were coded as depictive. With video sound off, the first depictive gesture was classified as motion indicating[2] and the last three as force indicating. The written transcript was then coded for force-indicating terms. Examining the results, our classification of the last three gestures as force indicating was confirmed by the fact that force-indicating terms (in bold) co-occurred with them. In fact, the co-occurring gestures appear to depict the terms—throwing, pulling, pushing. Throughout this videotape, depictive gestures were observed in abundance.

Coding Results

After reaching agreement on the coding for the gestures, the verbal imagery indicators, TEs in the broad sense, and Gedankens, we compared the results to see how often imagery indicators coincided with evidence for TEs. Figure 4 is a chart of the results. The discussion is represented chronologically from left to right and top to bottom; the numbers across the top of each row are transcript line numbers. Table 1 shows the key to Figure 4.

Table 1 Key to the chart in Figure 4 below

Symbol	Indicates
	Imagery indicators are present.
	Both gestures and verbal imagery indicators are present.
	There is evidence for a TE in the broad sense, an untested TE.
	There is evidence for a TE in the narrow sense, a Gedanken Experiment.
	The teacher is introducing a new case or explicitly proposing a TE.
	The later case is a variation of the earlier case or incorporates it.
	The later case appears designed to dispute the results of the earlier one.
	7 depictive gestures (for ex.) were coded for this line of dialog.
	There is evidence that a Gedanken was Generated and Run.
	An evolving case was described by a single speaker through multiple transcript lines interspersed with transcript lines spoken by others.
	Though a TE appears to have been run, there is not sufficient evidence to determine whether the system was untested by the student.

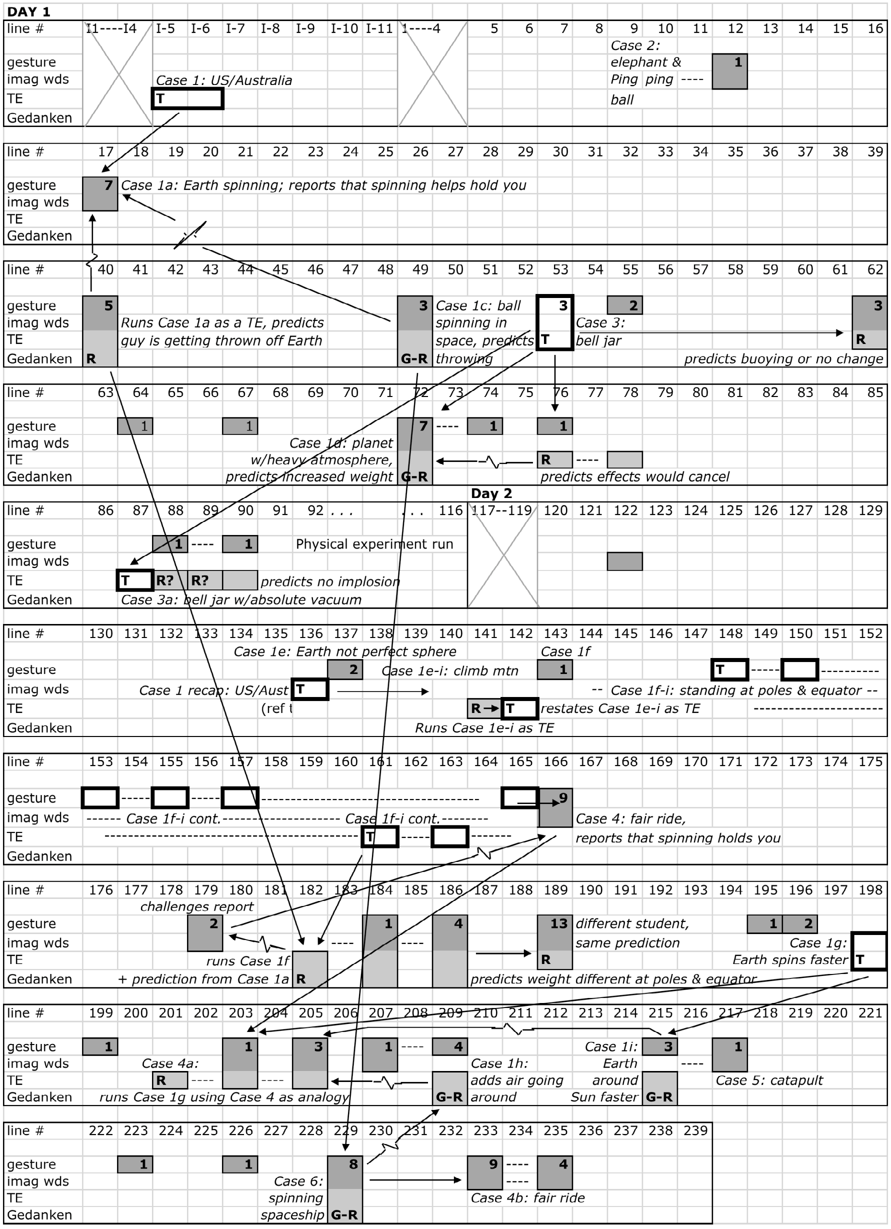


Fig. 4 Gravity class TEs and imagery use, Days 1 and 2

A sampling of features that can be seen in the kind of chart in Figure 4:

- There are large blocks of transcript with no teacher-generated cases as in Lines 1-52 and Lines 199-239. Here, the students were generating the cases and maintaining the discussion.
- We can see at a glance whether a TE was confirmatory or disconfirmatory of the idea it sought to address by whether the line connecting it to a previous case under discussion is straight or jagged.
- The individual TEs appear reactive to other TEs and to other ideas.
- We can easily see which TEs were associated with evidence for imagery by whether light gray blocks on the bottom two rows are paired with dark gray blocks directly above them.

Potential of the methodology: Sample of findings

This analysis, using the conceptual categories and methodology developed, demonstrates that evidence can be collected for the following (see also Stephens and Clement 2006):

1. *Thought experiments in the broad sense*. In the transcript discussed above, we found evidence for 6 teacher-generated and 5 student-generated untested TEs. There was explicit evidence from 12 transcript statements for the TEs being run by students.
2. *The involvement of imagery during the running of the TEs*. There were 14 episodes where evidence for generation or running of TEs was paired with evidence for the use of imagery. Eleven of these episodes had evidence for imagery from both gesture and verbal data.
3. *Kinesthetic imagery*. The most frequent form of evidence for imagery use in these transcripts was the use of force terms coupled with gestures that appeared to depict what the force terms were describing.
4. *Evaluative Gedanken experiments*. Students designed cases and used them to evaluate explanatory models. A few of these were discussed, but, as a look at Figure 4 will reveal, there were many other instances coded.
5. Students can make sense of and discuss TEs proposed by the teacher; same for TEs proposed by other students;
6. TEs can spread "contagiously" between students in a discussion, become modified and improved; this is an indication of the coherence of discussion.

CONCLUSIONS

Definitions. A problem in the literature is that there is no consensus on a definition of TE. In much of the literature, Sorensen's definition (Sorensen 1992) is used or the concept of TE is left undefined. An issue with Sorensen's definition is that it shifts much of the burden to the term experiment. TEs pose a paradox (Clement 2002, p. 32), namely, "How can findings that carry conviction result from a new experiment conducted entirely within the head?" Motivated by the paradox, a two-tiered definition is proposed; it is more inclusive by not depending crucially on the subject having proposed a formal experiment, slightly more operational in emphasizing a process rather than a product, and the broader tier fits the paradox better than the narrower set of carefully designed scientific Gedanken experiments.

Reiner (1998) has proposed a five-part structure of TEs: hypothetical world, hypothesis, experiment, results, conclusion. This provides potentially useful fine structure; however, expert studies indicate that TEs can also be run in a nonformal or intuitive manner. A less fine-grained but perhaps more easily codeable breakdown is that of Stephens and Clement (submitted) between generating and running a TE.

Existence in classrooms. There is some initial evidence that middle and high school students can run teacher-generated TEs and Gedankens and generate and run TEs of their own. However, given the broader definition for TE that has been proposed, it is possible that additional middle or elementary school student utterances will be reinterpreted as evidence for this kind of TE in the future. As for student-generated Gedankens, this may be an advanced skill. There is evidence from case studies that, on occasion, some students in physics classes have done this. An interesting question for future research is whether this skill can be taught.

Overall, this suggests that rationalistic, hypothetical thinking via TEs can be important in many more learning situations than we might initially imagine. A related theme was developed by Hammer (1995), who identified a number of rationalistic process goals being fulfilled in whole class discussion that are quite different from the classic, more empirically oriented process goals in science originally identified by Padilla (1991).

Purpose. Different kinds of TEs can be used to construct or evaluate (disconfirm or confirm) a model. Clement (2008) has identified a number of thinking processes that can incorporate and utilize TEs (defined in the broad sense), including the use of analogies, extreme cases, and runnable mental models.

TE Mechanisms. There is case study evidence from gestures and other indicators from both experts and students that TEs used for all of the above purposes can involve imagistic simulation. This suggests that imagistic thinking via TEs can also be important in many more learning situations than we might initially imagine.

Ongoing work on mechanisms in expert TEs points to the involvement in many TEs of perceptual motor schemas that drive imagistic simulations with the help of spatial reasoning processes. This is providing some initial explanations for the thought experiment paradox concerning the origins of conviction in TEs.

Instructional Implications

Effectiveness. In the gravity transcripts described earlier, we saw examples of creative co-construction of explanatory models for phenomena and argumentation about their validity (see also Clement and Rea-Ramirez, 2008). These are valuable higher order process goals for science instruction. The generation of TEs in favor of the scientific model indicates the potential of student TEs to contribute also to content goals.

Gilbert and Reiner (2004) found that the process of alternating between experimenting empirically and experimenting in thought can lead towards a convergence on scientifically acceptable concepts. However, to date, findings on effectiveness come exclusively from case studies (e.g. Reiner and Gilbert 2000; Stephens and Clement 2006).

We end by hypothesizing a possible general framework for viewing the role of imagery and TEs in instruction. Firstly, TEs require somewhat risky, hypothetical reasoning that is different from the security of deduction or induction by enumeration. But because they usually involve "stretching" a concept or schema to use it in a new domain, they may be a very important learning tool. The idea of extending a schema to be used for a problem outside of its normal domain of application is one way to promote sense making by building on what is known and extending or modifying it.

Secondly, imagistic simulation may be a very important sense making process. If imagistic simulation is a major mechanism for sensemaking, then we need to find ways to foster it, as it is a very different mode of thinking from recalling memorized facts or executing algorithms. TEs in the broad sense may be a way of promoting imagistic simulation as a key element of sense making.

NOTES

[1] This is a broader category than Gilbert and Reiner's (2000) *teaching* TE in that a pedagogical TE need not be related to any existing consensus TE.

[2] With sound off, classifying a gesture as motion indicating was considered more conservative than classifying it as force indicating. The fact that rotation implies a force to the physicist was not deemed sufficient here.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grants REC-0231808 and DRL-0723709, John J. Clement, PI. 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.

REFERENCES

Brown, J. R. (1991). *The laboratory of the mind: Thought experiments in the natural sciences*. London: Routledge.

Clement, J. (1994). Use of physical intuition and imagistic simulation in expert problem solving. In D. Tirosh (Ed.), *Implicit and explicit knowledge* (pp. 204-244). Norwood, NJ: Ablex Publishing Corp.

Clement J. (2002). Protocol evidence on thought experiments used by experts. In W. Gray & C. Schunn (Eds.), *Proceedings of the twenty-fourth annual conference of the Cognitive Science Society* (p. 32). Mahwah, NJ: Erlbaum.

Clement, J. (2008). *Creative model construction in scientists and students: The role of imagery, analogy, and mental simulation*. Dordrecht, The Netherlands: Springer.

Clement, J. (in press). The role of imagistic simulation in scientific thought experiments. *TOPICS in Cognitive Science*.

Clement, J., & Rea-Ramirez, M. A. (2008). *Model based learning and instruction in science*. Dordrecht, The Netherlands: Springer.

Clement, J., & Steinberg, M. (2002). Step-wise evolution of models of electric circuits: A "learning-aloud" case study. *Journal of the Learning Sciences, 11*, 389-452.

Clement, J., Zietsman, A., & Monaghan, J. (2005). Imagery in science learning in students and experts. In J. Gilbert (Ed.), *Visualization in science education* (pp. 169-184). Dordrecht, The Netherlands: Springer.

Gilbert, J., & Reiner, M. (2000). Thought experiments in science education: Potential and current realization. *International Journal of Science Education, 22*(3), 265-283.

Gilbert, J., & Reiner, M. (2004). The symbiotic roles of empirical experimentation and thought experimentation in the learning of physics. *International Journal of Science Education, 26* (15), 1819-1834.

Hammer, D. (1995). Student inquiry in a physics class discussion. *Cognition and Instruction*, *13*(3), 401–430.

Helm, H., Gilbert, J., & Watts, D. M. (1985). Thought experiments and physics education, Part II. *Physics Education, 20*, 211-217.

Kuhn, T. (1977). *The essential tension*. Chicago: University of Chicago Press.

Nersessian, N. (1992). In the theoretician's laboratory: Thought experimenting as mental modeling. In Hull, D., Forbes, M., & Okruhlick, K. (Eds.), *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association,* 1992, Volume 2: Symposia and Invited Papers, 291-301. East Lansing, MI: Philosophy of Science Association.

Nunez-Oviedo, M. C., Clement, J., & Rea-Ramirez, M. A. (2008). Developing complex mental models in biology through model evolution. In J. Clement & M. A. Rea-Ramirez (Eds.), *Model based learning and instruction in science* (pp. 173-194). Dordrecht, The Netherlands: Springer.

Padilla, M. J. (1991). Science activities, process skills, and thinking. In S. Glynn, R. Yeany, & B. Britton (Eds.), *The psychology of learning science* (pp. 205-217). Hillsdale, NJ: Erlbaum.

Reiner, M. (1998). Thought experiments and collaborative learning in physics. *International Journal of Science Education*, *20*(9), 1043–1058.

Reiner, M., & Gilbert, J. (2000). Epistemological resources for thought experimentation in science learning. *International Journal of Science Education, 22*(5), 489-506.

Sorensen, R. (1992). *Thought experiments*. Oxford, UK: Oxford University Press.

Stephens, L., & Clement, J. (2006, April). *Designing classroom thought experiments: What we can learn from imagery indicators and expert protocols*. Paper presented at the annual conference for the National Association for Research in Science Teaching, San Francisco, CA.

Stephens, L., & Clement, J. (submitted). Documenting the use of expert scientific reasoning processes by high school physics students.

Velentzas, A., Halkia, K., & Skordoulis, C. (2005). Thought experiments in the theory of relativity and in quantum mechanics: Their presence in textbooks and in popular science books. *Proceedings of the International History, Philosophy, Sociology & Science Teaching Conference*. Retrieved 2 March, 2009, from http://www.ihpst2005.leeds.ac.uk/papers.htm.

Lynn Stephens is an advanced doctoral student in the School of Education at the University of Massachusetts in the Mathematics, Science, and Learning Technologies program, currently employed with John Clement in the Scientific Reasoning Research Institute. Her interests are in the design of online simulations and animations for physics and other areas of science and mathematics education and how these can be made to support mental modeling for a broader cross section of learning styles, cognitive abilities, and motivational issues. Correspondence: University of Massachusetts—Amherst, 428 Lederle, 710 N. Pleasant St., Amherst, MA 01003-9305. Email: lstephens@educ.umass.edu.

John Clement is a Professor in the School of Education and the Scientific Reasoning Research Institute at the University of Massachusetts. His current research is focused on methods for helping students form and use visualizable models in science and studies of mental model construction by expert scientists. He has recently authored the book "Creative Model Construction in Scientists and Students: Imagery, Analogy, and Mental Simulation" and edited another with M. Ramirez, "Model-Based Learning in Science". Correspondence: University of Massachusetts—Amherst, School of Education, Furcolo Hall, Amherst, MA 01003-9305. Email: clement@educ.umass.edu.