1

# Thought Experiments and Imagery in Expert Protocols

JOHN J. CLEMENT

ABSTRACT. This paper focuses on case studies from think-aloud protocols with expert scientists solving explanation problems in which they appear to make predictions for novel systems they have never seen by "imagining what will happen" or "running" the system in their heads. Nersessian [2002] has proposed, based on her reading of historical records of investigations in scientific work such as Maxwell's work on electromagnetic field theory, that thought experiments can play a role in scientific theory formulation. Such thought experiments are intriguing because (1) they appear to play a powerful role in science; and (2) the subject appears to gain something like empirical information without making any new observations. This raises what I call the *fundamental paradox of thought experiments*, expressed as: "How can findings that carry conviction result from a new experiment conducted entirely within the head?" Here I will analyze examples of thought experiments from think aloud protocols in order to examine some approaches to resolving the paradox and to begin to explore the breadth of circumstances in which thought experiments (TE's) of different types are used. Scholars have long been intrigued with the nature of thought experiments in science but the definition and scope of the term thought experiment has remained controversial, as have theories of the mechanisms by which they work. This motivates developing a taxonomy and theory of thought experiment processes based on observations from expert protocols. Findings from such studies may also give us a more systematic way to analyze the types of TE's used in the history of science and in instruction.

## Expert case study

The data base for this study comes from ten professors and advanced graduate students in scientific fields who were recorded while thinking aloud about the following "Spring Problem":

> A weight is hung on a spring. The original spring is replaced with a spring made of the same kind of wire, with the same number of coils, but with coils that are twice as wide in diameter. Will

Lorenzo Magnani, editor, *Model–Based Reasoning in Science and Engineering*, pp. 1–16

> the spring stretch from its natural length more, less, or the same amount under the same weight? (Assume the mass of the spring is negligible.) Why do you think so?

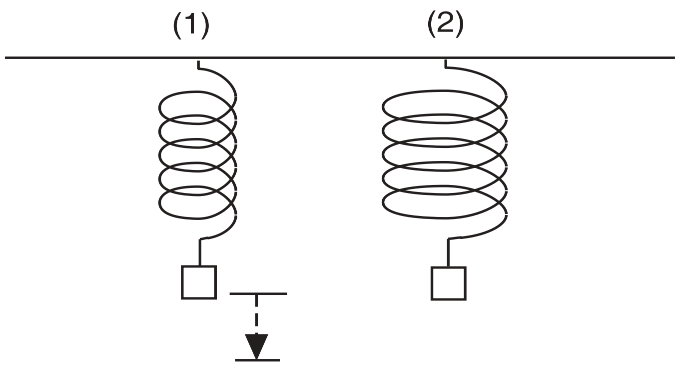


Figure 1.

This data base has yielded previous findings on analogies and creative model construction cycles [Clement, 1988; Clement, 1989; Clement, 2004]. For example, the 1988 paper coded frequencies for different types of analogies across all subjects in the data base, and the 1998 and 2004 papers extended this analysis to other tasks. The 1989 paper documented creative analogies, Aha! insights and cyclical model evaluation and revision processes. Griffith, Nersessian, and Goel [2000], developed a computational model for the generation and revision of analogies from these protocols. The present paper uses newly analyzed hand motions and other indicators in these video tapes to examine thought experiments.

Unfortunately, there is no consensus on a definition of ‘thought experiment’. I will present some examples of TE’s from one of these subjects and then give a definition. In the following episode subject S2 simply imagined what would happen to wide and narrow springs:

> **Protocol Section 1:** S: “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 kinesthetics sense* that somehow a bigger spring is looser – Umm, that’s high uncertainty.”

I will use italicized type to identify observations that have potential as evidence for imagery (both kinesthetic and visual) and simulation use. This appears to be a TE in the sense that he makes a prediction for a concrete situation, and it is one he has not previously observed. In this case his visualization, as he puts it, gives him the correct answer, but he is not very confident in the result, so it is a high–uncertainty TE.

S2 also generated an analogy in which he predicted that a long horizontal rod fixed at one end would bend more than a short one (with the same sized weight attached to the other end of each rod), inferring that segments of the wider spring would bend more and therefore stretch more. However, he was concerned about the appropriateness of this analogy at a deeper level because of the apparent lack of a match between (1) bending producing an increasing slope in the rod and (2) a lack of increasing slope in the wire in a stretched spring. One can visualize this discrepancy here by thinking of the increasing slope a bug would experience walking down a bending rod and the constant slope the bug would experience walking down the helix of a stretched spring. (The latter is my own descriptive analogy for purposes of clarity – not the subject's.) This discrepancy led him to question whether the bending rod was an adequate model for the spring.

> **Section 2**: "But then it occurs to me that there's something clearly wrong with that [bending rod] metaphor, because ...it would (raises hands together in front of face) droop (*moves r. hand to the right in a downward curve*) like that, its slope (*retraces curved path in air* with l. hand) would steadily increase, whereas in a [real] spring, the slope of the spiral is constant. . . "

Imagining the spring with an increasing slope appears to be a thought experiment in which he "runs" the idea of bending taking place in the spring as it stretches. (This implies that the coils would be farther apart at the bottom of the spring than the top.) This can be seen as a novel thought experiment in which he examines the consequences of running the "bending model" of the spring. This anomaly of a spring with an increasing slope produces a mismatch with his prior knowledge that spring coils should not become wider apart at one end (when the mass of the wire is negligible.) This anomaly appears to bother him considerably and drives further work on the problem, eventually discrediting the bending model. This TE also appears to be *generative* in that it generates a situation with new and novel properties, namely a spring with increasing slope. The passages in parentheses indicate *depictive hand motions* which suggest the use of imagery.

**Definition**

These examples motivate the following 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. I use the phrase "untested" thought experiment to emphasize that this does not include cases where the subject simply replays a previously observed event. Still, the above definition is intentionally very broad. Later I will give a second narrower definition for referring to a more specialized use of TE's. However, an advantage of the broad definition above is that it appears to designate cases that raise the fundamental paradox.

**Polygonal springs**

After spending nearly 30 minutes considering this and other analogies, he generates the polygonal coil cases in Figure 2a and 2b. While analyzing the hexagon in terms of bending effects below, it occurs to him in an Aha episode that there will also be *twisting* effects in the segments.

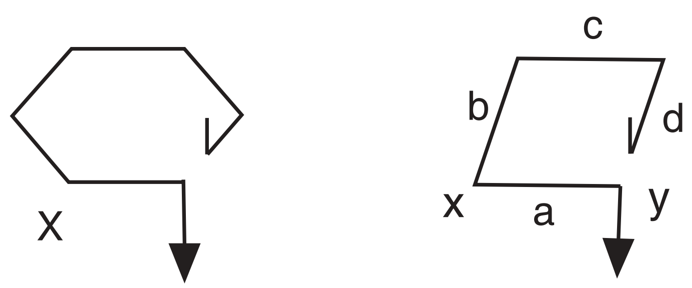


Figure 2. a and b.

> **Section 3** - *Running an analogy and activating a schema:*
>
> "Just looking at this it occurs to me that when force is applied here (at arrow in Figure 2a), you not only get a bend on this segment, but because there's a pivot here (point $x$), you get a torsion [strain produced by twisting movement of the wire] effect. Aha!! Maybe the behavior of the spring has something to do with twist (*makes twisting motion with right hand*) forces as well as bend forces. That's a real interesting idea."

Twisting of the wire and the resulting torsion is in fact a key element in the analysis of spring behavior as understood by engineers. Its discovery here represents a major insight in finding a new causal mechanism. The torsion discovery and Aha! phenomenon above is an interesting process in itself and is discussed in Clement [1989]. However in this paper I want to focus on examining the possibility that imagistic simulation plays a role in performing and making inferences for thought experiments rather than on how they are generated. The subject continues:

> **Section 4**: "Let me accentuate the torsion force by making a square where there's a right angle. I like that, a right angle. That unmixes the bend from the torsion. Now I have two forces introducing a stretch. I have the force that bends this segment a (Figure 2b) and in addition I have a torsion (*makes twisting motion with right hand*) force which twists [rod $b$) at vertex, um, $x$" [as if side $a$ were a wrench acting to twist side b.]

> **Section 5** "Now let's assume that torsion and bend (*makes bending motion with hands together*) don't interact... does this (points to square) gain in slope–toward the bottom? Indeed, we have a structure here which does not have this increasing slope as you get to the bottom. It's only if one looks at the fine structure; the rod between the $Y$ and the $X$, that one sees the flop (*moves left hand horizontally in a downward curve*) effect." "Now I feel I have a good model of sp- of a spring."

Because bending and twisting still allow the slope to "start over from zero" at each corner, the square coil is a new model in which the accumulating slope difficulty does not occur, suggesting a way to resolve his previous anomaly. He goes on to ask about the effect of coil width for the square coil model.

> **Section 6**: "Now making the sides longer certainly would make the [square] spring stretch more... The longer the segment, (*holds hands up in front as if holding something between them* ) the more (*makes bending motion with right hand*) the bendability."

> **Section 7**: "Now the same thing would happen to the torsion I think, because". "If I have a longer rod (moves hands apart), and *I put a twist* on it (*moves hands as if twisting a rod*), it seems to me–again, physical intuition–that it will twist more... I'm (raises hands in same position as before and holds them there continuously until the next motion below)... *imagining holding something that has a certain twistyness to it, a-and twisting it...*"

Figure 3. a, b, and c.

> **Section 8**: Now I'm confirming (*moves right hand slowly toward left hand*) that by using this method of limits. As (*moves right hand slowly toward left hand* until they almost touch at the word "closer") I bring my hand up closer and closer to the original place where I hold it, *I realize very clearly that it will get harder and harder to twist.* So that confirms my intuition so I'm quite confident of that. . . (see Figure 3).

(The reader may wish to try this thought experiment with images of coat hanger wire, bent to have "handles" at each end of the wire.) Later the subject distinguishes between confidence in the answer to the spring problem and confidence in his understanding of it, and indicates that the torsion analysis has increased his subjective feeling of understanding from "way, way down" to "like, 80%". At this point S2 appears to have a mental model of the spring as working like a square coil that contains elements that both bend and twist. S2 uses thought experiments to predict correctly that the wider spring will stretch more. These also suggest that the slope of the stretched spring will be constant throughout (also correct), resolving S2's previous anomaly about increasing slope.

Traditional descriptions of expertise focus on practiced skills and domain specific knowledge about systems that are very familiar, whereas in this setting, the subject exhibits something more like adaptive expertise wherein he tries to use more general heuristics to invent and evaluate new models of an unfamiliar system. Unlike practiced homework problems, the spring problem challenges S2 in this way, and appears to put him on the "frontier" of his own personal knowledge.

Under the definition proposed, each of the eight numbered sections above can be considered to contain an untested thought experiment. The epistemological status of thought experiments is controversial in terms of whether they can "officially" support or discount scientific theory in the realm of publicly tested theory. Real experiments are ordinarily preferred when available. Nevertheless, scientists such as Galileo, Newton, and Maxwell have included powerful thought experiments in their published works. Perhaps a more interesting question is whether they can play an important role in the generation and initial evaluation of theories by a scientist during the challenging process of theory construction. Nersessian [2002] discusses the way in which Maxwell realized through a TE that adjacent electromagnetic vortices in the ether would in effect "jam" like meshing gears trying to rotate in the same direction and the way that this TE led him to reject and improve upon this early model. Related processes that can be interpreted as utilizing thought experiments have been documented in the case of Michael Faraday by Ippolito and Tweney [1995] and in the case of astronomers by Trickett and Trafton [2002]. In the present case we appear to have evidence that (1) TE's have been part of the subject's generating and considering important new hypotheses (e.g. protocol sections 1 and 3); (2) a TE (section 2) has raised serious doubts about one hypothesis; (3) TE's have boosted his confidence in other hypotheses about certain aspects of system behavior (section 8) and have increased his feeling of understanding how the system works. Thus TE's appeared to play a central, not just peripheral, role within the thinking of this subject in helping to generate, cast serious doubt on, or support hypotheses.

## Use of imagistic simulation

In this section I will put forward an initial explanation of how TE's work by introducing the concept of imagistic simulation. Italicized type above in sections 7 and 8 identifies examples of several imagery-related observation categories, in the following order: **personal action** projections (spontaneously redescribing a system action in terms of a human action, consistent with the use of kinesthetic imagery), **depictive hand motions**, and **imagery reports**. The latter occurs when a subject spontaneously uses terms like "imagining","picturing", a situation, or "feeling what it's like to manipulate" a situation. In this case it is a **dynamic imagery report** (involving movement or forces). None of these observations are infallible indicators on their own, but I take them as evidence for imagery, and this is reinforced when more than one appear together. Such indicators also appear alongside new predictions at many earlier points indicated by italicized type in the protocol segments. (There is not space for a review here, but an increasing variety of studies of depictive gestures suggest that they are expressions of core meanings or reasoning strategies and not simply translations of speech. Others indicate that the same brain areas are active during real actions and corresponding imagined actions.)

One can draw on the precedent of motor schema theory [Schmidt, 1982] in hypothesizing that perceptual motor knowledge structures that can control real actions over time (e.g. a schema for “twisting” objects) are involved here. The observations in sections 7 and 8, for example, can be explained via what I have called schema–driven *imagistic simulations* wherein: (1) the subject has activated a somewhat general and permanent perceptual motor schema that can control the action of twisting real objects; (2) the schema assimilates an image of two rods of different lengths that is more specific and temporary; (3) the action schema “runs through its program” vicariously without touching real objects, generating a simulation of twisting the two rods, and the subject compares the anticipated effort required for each. Such a simulation may draw out implicit knowledge in the schema that the subject has not attended to before – e.g. in this case the simulation may draw out knowledge embedded in analog tuning parameters of a motor schema to anticipate differences in the effort required to twist a long and short rod. In other words, a hypothesis can be made, with initial grounding in data such as that in protocol sections 7 and 8, that the subject is going through a process wherein a general action schema assimilates the image of a particular object and produces expectations about its behavior in a subsequent dynamic image, or simulation [Clement, 1994; Clement, 2003]. The knowledge being used there is “embodied” in this sense.

Such domain specific schemas that generate simulations can work in concert with more general spatial reasoning skills – reasoning operations that embody spatio-temporal constraints on **any** system of objects, such as the constraint that solid objects may not occupy the same space, or that the face of an object turning on a vertical axis will disappear and reappear (cf. [Shepard and Cooper, 1982]). Given these elements the ability to run thought experiments is most plausibly explained using a framework that includes (a) flexible perceptual motor schemas that can run imagistic simulations via the extended application of the schema outside of its normal domain of application (”outside” means that the schema is being applied to either: an unfamiliar situation; or a familiar situation along with question that has never been asked before about the outcome); (b) converting implicit into explicit knowledge and/or (c) spatial reasoning such as that involved in section 5 in imagining whether the contributions to stretching in each side of the square coil add or cancel [Clement, 1994]. This last case also involves a (d) compound simulation wherein multiple simulations (of bending or twisting in each side) are performed on the same image and the results added together imagistically. One can point to the above sources (a) through (d) as potential origins of conviction in thought experiments, to help us begin to explain the fundamental paradox. They can also explain the effectiveness of the extreme case at the end of the transcript above as an example of what I call "imagery or simulation enhancement", a phenomenon difficult to explain in other ways [Clement, 1994]. The extreme case leads to the same prediction with a much higher degree of confidence. I infer that this comes from increasing the differences between the two images being compared and making that difference more detectable under inspection of the images – here the kinesthetic difference in the torque or twisting force applied to a "normal" rod and a very short rod in order to put a certain amount of twist in it. Thus, I hypothesize that this is a case of "simulation enhancement" and that the role of this extreme case is *to enhance the subject's ability to run and compare imagistic simulations with high confidence.* In this case the main source of conviction in the simulation is the tapping of implicit knowledge and its conversion into explicit knowledge. The ability of the theory to explain why the extreme case helps in this case lends some initial support to the theory.

Another subject, S6, attempted to imagine the direction in which the wire would twist in a spring coil when it was stretched, but found it difficult. By adding "little paint dots" to his image of thick spring wire, he was eventually able to do this. This appears to be another kind of imagery enhancement strategy.

In summary, there is evidence that the TE's documented here involved the use of imagery. The successful use of TE's above can be explained by hypothesizing that they involve an imagistic simulation process wherein a perceptual motor schema generates dynamic imagery that can be interrogated. Four possible sources of conviction in TE's were suggested that can begin to explain the fundamental paradox.

## Evaluative Gedanken experiments

*Band spring.* A use of the term "thought experiment" in a *narrower sense* is what I call an *evaluative Gedanken experiment.* 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 its behavior. An evaluative Gedanken experiment is usually more complex than a simple thought experiment. In the cases discussed here the subject considers a new system for which the present model is predictive, but for which another source of knowledge is also predictive, giving the potential for conflict or consonance between the two prediction methods. An example from subject S2 is the case of a spring made of a vertically oriented band of material shown at the bottom of Figure 4a. (the reader might imagine the metal strip unwound from a coffee can, with coils that are reshaped to make a spring, say, 3" wide.)

**Section 9**: "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. Or isn't particularly relevant at all maybe somehow... 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 14.4) ... 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 (*motions as if twisting an object*)."

Given the initial imagery reports here, I interpret this to mean that the subject imagined that such a spring would still be quite stretchable even though the band "cannot bend in the up–down direction", challenging the necessity of bending as not "particularly relevant at all" to stretching. In this type of evaluatory Gedanken experiment he designs a special case where the bending model yields a prediction, (predicts no stretch) but where he also has some other independent source of information that can evaluate that prediction (physical intuition predicts that it will stretch), as shown in Figure 4a. This is a Gedanken experiment because it is designed to help test a model. It may actually involve two thought experiments, so it is a more complex reasoning pattern than a simple thought experiment. Either of the competing predictions just mentioned and represented by vertical arrows in Figure 4a may occur via a simple thought experiment.

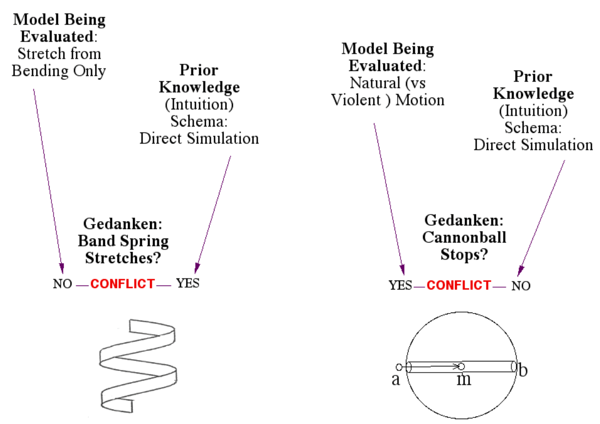


Figure 4. a and b.

*Galileo.* A similar pattern can be seen in the Dropping the Cannonball Through a Hole in the Earth experiment shown in Figure 4b, used by Galileo [Galilei, 2001] in his *Dialogues on the Two Chief World Systems.* (Actually this thought experiment can be traced back at least as far as his predecessor Tartaglia.) The larger issue at hand is whether Aristotle's model of "Natural" and "Violent" motion, as two different types of motion, is valid. Here "Natural" motion is motion toward the center of the Earth, where as throwing a ball upward is "Violent" (unnatural) motion. Drilling a hole through the Earth and dropping a cannonball into it appears to produce a smooth transition from natural to violent motion at the center of the Earth. And yet the cause of upward motion seems eminently "natural".

> Salviati: But [the cannonball] having arrived at the center is it your belief that it would pass on beyond..?
>
> Simplicio: I think it would keep on going a long way.
>
> Salviati: Now wouldn't this motion beyond the center be upward, and, according to what you have said, preternatural and constrained? Let me see you find an external thrower who shall overtake it.. to throw it upward.

The inference that the cannonball appears to pass through the center absolutely smoothly without any discontinuous change attacks Aristotle's central distinction between Natural and Violent motion and helps discredit the model in the dialogues. Again, as shown in Figure 4b, the Gedanken case evokes one prediction from the model and a conflicting prediction from another source, here one's intuitions about the strong momentum of a speeding cannonball. Thus the pattern of reasoning in the band spring experiment from the expert protocol appears to have the same form as an evaluative Gedanken experiment from Galileo's dialogues.

*Shear forces in spring.* A similar pattern was seen when another subject, S15 (interviewed after the original set of subjects discussed at the beginning of this paper) generated a microscopic model of shear forces and displacements in the spring wire that shows one element being displaced downward relative to the previous element. He drew these wire element displacements as a kind of slowly descending, spiral staircase. However, by running a Gedanken experiment he discovers a contradiction: this model predicts that stretch in a loosely coiled wire and a tightly coiled wire of the same length should be the same. But this is counter to his intuition since in running this comparison macroscopically without analyzing it, he imagines that the tightly coiled wire would stretch less. This case is not the same as the original problem because the wire length in the newly proposed experiment is the same for both springs. It is a Gedanken experiment that pits the shear model against strong intuitions about the behavior of loosely and tightly coiled material, and it yields a contradiction to his model. Multiple instances of depictive hand motions in this episode provide evidence for imagistic thought.

> **S15**: "(Draws spring like coil with elements descending in steps so that each element is slightly displaced vertically with respect to the next) The mechanism of force communication is this sort of shearing-like property (h*olds both hands side by side oriented vertically, palms open and facing each other* ) But now the next think I want to try to understand is, ok,... given that mechanism, explain why then that the very loosely wound coil (*makes a wide tracing of a large coil with his right hand*) and the tightly wound coil have different elasticities. The net displacement is proportional to how much relative motion (*holds palms flat, face down, so index fingers are touching and slides one hand upward and one down*) the communication of force... calls for...
>
> But now I could take this thing, [the same piece of wire] and I could wind it tightly, or I could wind it loosely, and I have the sense that if my model's correct, the rate that I wind it is gonna be (pause) ok. Now I'm extremely unhappy – The stiffness of the spring is just a function of the length of the material – (*holds palms flat, face down, so index fingers are touching and slides hands in opposite directions*) – the contradiction [is]... that I could wind this thing as tight as a spring... or as a very loose spring and it would still have the same displacement".

Here again in this case, his current model predicts one result (equal stretching of the two coils) while a powerful intuition schema predicts the opposite result, discrediting the model strongly.

## Conclusion

In this paper I have attempted to show that it is possible to document the use of thought experiments in think aloud protocols. We saw that for the three subjects discussed here, thought experiments can be either supportive of or conflicting with an explanatory model. By this I do not mean confirming or rejecting a hypothesized model with certainty. However, the present paper argues that at the very least, TE's can add or detract support for a model that may be competing with other models in the mind of the scientist as a plausible theory.

The Fundamental Paradox of TE's: "How can findings that carry conviction result from a new experiment conducted entirely within the head?" is a long standing problem for philosophy of science as well as cognitive science. It appears to apply to all Untested TE's as defined here in the broad sense as the act of predicting the behavior of an untested, concrete, system (the "experiment"). Since this definition for untested thought experiments is quite broad, this suggests that the paradox applies surprisingly broadly. And in fact, the examples of TE's presented in this paper occurred within a broad variety of types of reasoning: Elemental TE's, Analogy, Extreme Case Use, Running a Model, and Gedanken Experiments.

*Toward explaining the paradox.* Instead of one source of conviction that might explain the fundamental paradox, a number of sources were hypothesized to account for the transcript data. However, a central theme is that these all depend on imagistic representations. There was evidence that the TE's in the protocols used imagistic simulations. Multiple indicators were identified, including: predicting changes in a system, imagery reports, personal action projections, depictive hand motions, and dynamic imagery reports. The hypothesized sources of conviction in thought experiments utilizing such imagistic simulations included: (a) flexible perceptual motor schemas that can run imagistic simulations via the extended application of the schema outside of its normal domain of application; (the Bending Spring, and Vertical Band Spring cases) (b) converting implicit into explicit knowledge (twisting rod case); (c) spatial reasoning (increased stretch in wider square coil case; and non–cumulative slope in square coil); and (d) compound imagistic simulation (also used in the these two square coil cases).

Are thought experiments primarily based on prior experience or creative reasoning? In an early paper on thought experiments, Gooding [1992] took a position that was intermediate between the empiricist and rationalist ends of the spectrum, concluding that: "What is needed is a combination of empirical knowledge and the ability to reason with it." The present analysis fleshes out details in Gooding's general position and provides evidence that supports it. The first two sources above are developmentally experiential to the extent that they have historical roots in experience. In (a) using an old schema in a new situation outside its normal domain of application can mean that there is some uncertainty in the application, but there may also be a fair amount of conviction. In (b) one uses stored knowledge, possibly from an old empirical source, that is encoded in analog form and needs to be made explicit. Both of these processes use prior knowledge schemas but must extend them via reasoning since, by definition, a TE deals with an unfamiliar aspect of a case.

Sources c and d above primarily utilize additional reasoning capabilities to make further extensions of thought. In (c) spatial reasoning, about perceptual transformations or the way movements add together for example, is a type of plausible reasoning that is presumably ubiquitous but largely unrecognized in philosophy of science. And in (d) the subject runs two schemas on the same case in a way that involves sequencing or coordinating and predicting the effects of a new combination of actions, a related form of generative reasoning via imagistic simulation. Thus it can be hypothesized that TE's can utilize specific experiential knowledge; but this imagistic knowledge is extended via reasoning. They can use both prior knowledge schemas and reasoning processes which extend the range and types of inferences made from using the schemas in new or novel situations. Thus I believe that analyzing the contribution of the above four aspects of imagistic simulation can begin to give us a theoretical picture of how thought experiments work. This may not be the only mechanism used, but it appears to be a very important one.

*Gedanken experiments.* I also introduced the concept of Evaluatory Gedanken Experiments: (TE's in the narrow sense): these are designed to help evaluate a theory, model, or concept [Clement, 2003]. These are more similar to real experiments, and can include something like the control of variables. Their argument structures can be more complex than that of a simple untested thought experiment. The structure of evaluative Gedanken experiments observed in the expert protocols here appears to be the same as that of a thought experiment discussed by Galileo. This leads us to conjecture that the concepts derived here may be applicable to certain thought experiments in the history of science. It is hoped that the descriptors generated here will help us to develop a theory of TE's based on evidence from naturalistic observations of behaviors, and that this can interact fruitfully with the analysis of historical TE's.

*Educational experiments.* A relatively small number of specific Gedanken experiments (in the narrow sense) in the history of science have been recognized for their pedagogical value wherein the scientist uses the experiment to convince others of the validity of his or her theory (cf. [Reiner and Gilbert, 2000]). Because TE's in the broad sense appear in many types of reasoning, they are more ubiquitous than previously thought. This suggests that they may play a more widespread role in instruction. Therefore I believe understanding the mechanisms for TE's in scientists will lead to important payoffs for education as well.

In the framework developed here, expert TE's can make use of embodied analog perceptual–motor systems and schemas used in mental simulations. Both the broad and narrow concepts of thought experiment as defined here appear to be useful, and both can be documented in think aloud protocols.

The broad concept is appropriate for expressing the fundamental paradox. The narrower concept of an evaluative Gedanken experiment encompasses Galileo's use of the "hole through the earth" thought experiment, impressive in that it was designed to contribute to eliminating an established theory.

## BIBLIOGRAPHY

[Clement, 1988] J. Clement. Observed methods for generating analogies in scientific problem solving. *Cognitive Science*, 12:563–586, 1988.

[Clement, 1989] J. Clement. Learning via model construction and criticism: Protocol evidence on sources of creativity in science. In J. Glover, R. Ronning, and C. Reynolds, editors, *Handbook of Creativity: Assessment, Theory and Research*, New York, 1989. Plenum.

[Clement, 1994] J. Clement. Use of physical intuition and imagistic simulation in expert problem solving. In D. Tirosh, editor, *Implicit and Explicit Knowledge*, Norwood, NJ, 1994. Ablex.

[Clement, 1998] J. Clement. Expert novice similarities and instruction using analogies. *International Journal of Science Education*, 30(10):1271–1286, 1998.

[Clement, 2003] J. Clement. Imagistic simulation in scientific model construction. In *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society, 25*, Mahwah, NJ, 2003. Erlbaum.

[Clement, 2004] J. Clement. Imagistic processes in analogical reasoning: Conserving transformations and dual simulations. In K. Forbus, D. Gentner, and T. Rieger, editors, *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society*, pages 233–238, Mahwah, 2004. Erlbaum.

[Galilei, 2001] G. Galilei. *Dialogues Concerning the Two Chief World Systems*. Random House, New York, 2001.

[Gooding, 1992] D. Gooding. The procedural turn: or, why do thought experiments work? In R. Giere, editor, *Cognitive Models of Science*, pages 45–76, Minneapolis, 1992. University of Minnesota Press.

[Gooding, 1998] D. Gooding. Thought experiment. In E. Craig, editor, *The Encyclopaedia of Philosophy*, London, 1998. Routledge.

[Griffith *et al.*, 2000] T. W. Griffith, N. J. Nersessian, and A. Goel. Function-follows-form transformations in scientific problem solving. In *Proceedings of the Cognitive Science Society 22*, pages 196–201, Mahwah, N.J., 2000. Erlbaum.

[Ippolito and Tweney, 1995] M. F. Ippolito and R. D. Tweney. The inception of insight. In R. J. Sternberg and J. E. Davidson, editors, *The Nature of Insight*, Cambridge, MA, 1995. MIT Press.

[Kuhn, 1977] T. W. Kuhn. A function for thought experiments. In T. Kuhn, editor, *The Essential Tension*, Chicago, 1977. University of Chicago Press.

[Magnani, 1999] L. Magnani. Model-based creative abduction. In L. Magnani, N. J. Nersessian, and P. Thagard, editors, *Model-Based Reasoning in Scientific Discovery*, pages 219–238, New York, 1999. Academic/Plenum Publishers.

[Nersessian, 2002] N. J. Nersessian. Maxwell and "the method of physical analogy": model-based reasoning, generic abstraction, and conceptual change. In D. Malament, editor, *Essays in the History and Philosophy of Science and Mathematics*, pages 129–166, Lasalle, IL, 2002. Open Court.

[Reiner and Gilbert, 2000] M. Reiner and J. Gilbert. Epistemological resources for thought experimentation in science learning. *International Journal of Science Education*, 22(5):589–506, 2000.

[Schmidt, 1982] R. A. Schmidt. *Motor Control and Learning.* Human Kinetics Publishers, Chmpaign, IL, 1982.

[Shepard and Cooper, 1982] R. Shepard and L. Cooper. *Mental Images and their Transformations*. MIT Press, Cambridge, MA, 1982.

[Trickett and Trafton, 2002] S. Trickett and J. G. Trafton. The instantiation and use of conceptual simulations in evaluating hypotheses: Movies-in-the-mind in scientific reasoning. In W. Gray and C. Shunn, editors, *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society*, pages 878–883, Mahwah, NJ, 2002. Erlbaum.

John J. Clement
University of Massachusetts
Amherst, USA
Email: clement@srri.umass.edu