Title
The Construction of Causal Schemes: Learning Mechanisms at the Knowledge Level

Journal
Cognitive Science, 38(5)

ISSN
0364-0213

Author
diSessa, Andrea A

Publication Date
2014-06-01

DOI
10.1111/cogs.12131

The Construction of Causal Schemes:

Learning Mechanisms at the Knowledge Level

Preprint (Oct. 2013): To appear in *Cognitive Science*.

Andrea A. diSessa

Graduate School of Education

University of California at Berkeley

Running Head: Construction of Causal Schemes

Key Words: Conceptual change, learning mechanisms, causal schemes, microgenetic analysis, science learning.

Abstract

This work uses microgenetic study of classroom learning to illuminate (1) the role of pre-instructional student knowledge in the construction of normative scientific knowledge, and (2) the learning mechanisms that drive change. Three enactments of an instructional sequence designed to lead to a scientific understanding of thermal equilibration are used as data sources. Only data from a scaffolded student inquiry preceding introduction of a normative model were used. Hence, the study involves nearly autonomous student learning. In two classes, students developed stable and socially shared explanations ("causal schemes") for understanding thermal equilibration. One case resulted in a near-normative understanding, while the other resulted in a non-normative "alternative conception." The near-normative case seems to be a particularly clear example wherein the constructed causal scheme is a composition of previously documented naïve conceptions. Detailed prior description of these naive elements allows a much better than usual view of the corresponding details of change during construction of the new scheme. A list of candidate mechanisms that can account for observed change is presented. The non-normative construction seems also to be a composition, albeit of a different structural form, using a different (although similar) set of naïve elements. This article provides one of very few high-resolution process analyses showing the productive use of naïve knowledge in learning.

1. Introduction

This research proposes to make a contribution in a notable lacuna in the conceptual change literature. Even though much conceptual change research is devoted to understanding changes that can be induced by instruction, reference models of concepts are generally diachronic, typically marking the naïve (entering) state on the one hand and the target, normative state on the other. In a classroom and in the design of curricula, however, learner attention must be managed and tracked on a much finer time scale. Teachers react on the fly to conceptions displayed by students in a discussion (Minstrell, 1982; 1989), and even written curricula must manage students' attention at a short time-scale, comparable to students' processing a sentence of text.

The more encompassing studies of conceptual change interpolate, for example, percentages of normative answers or even track empirical or theoretical trends in the data (Wiser & Smith, 2008; see also the broader literature on "learning progressions"). And yet, this literature is nearly devoid of analysis of real-time data, which could track student thinking in the regime where teachers and curriculum developers attempt to manage student reasoning in order to effect change. Surely establishing landmarks of change (naïve theories; normative thinking) is important. And yet, just as surely, the field needs, eventually, to establish the connection between thinking *during* learning events and long-term change.

The rest of this introduction positions the work here more carefully with respect to traditional work in conceptual change and, in addition, with respect to the complementary field of cognitive modeling. It then elaborates the critical theoretical terms "mechanism" and "causal scheme," and introduces the empirical methodology.

1.1. Modelers and non-modelers: Motivating "modeling at the knowledge level"

Computational modeling of cognitive structures and processes might well be regarded as the critical innovation that sparked the origin of cognitive science as a field. Landmarks include the symbolic modeling of Newell and Simon (e.g., work on human problem solving, Soar, unified theories of cognition), parallel and follow-on work by John Anderson (e.g., ACT* and ACT-R), and connectionist modeling (e.g., Churchland, Smolensky, and McClelland). Clark and Toribio (1998) provide a review.

One might think that cognitive modeling could provide the needed complement to diachronic analyses of conceptual change. However, in contrast to research areas such as problem solving and skill acquisition, the arena of conceptual change is nearly devoid of work in the paradigm of cognitive modeling. In the following quote, two important actors in conceptual change research, Inagaki and Hatano (2002), make this point, highlighting focus on beginning and end states, and comparative lack of relevant data and analysis of the intervening processes.

> Although the issue of *how conceptual change occurs* and the *specification of its mechanisms* is "one of the fundamental problems of cognitive psychology today" (Vosniadou, 1994, p. 3), the *available data are limited*. This is partly because recent developmentalists have focused on the specification of the initial form of a naive theory and, to a much lesser extent, on the description of the state after conceptual change, *without*

analyzing the process of change itself (e.g., Carey, 1985; Wellman, 1990). (p. 173; emphasis added)

More generally, one can look at the literature on conceptual change with two indices in mind. The first index is whether any process data are taken and analyzed near the time-scale needed for producing a cognitive model, say, in the range downward from a few minutes, to capture substructure of an event of learning. The second index is whether researchers seek to produce models or technical terms for mental objects and processes approaching the precision needed for computational modeling. Concerning the first index, my impression is that the conceptual change literature is very sparsely populated with process data and analysis. In order to check this impression more systematically, I engaged two strategies. First, I reviewed a significant number of papers on conceptual change by some well-known researchers in the area, including Susan Carey (and her students and colleagues, primarily Carol Smith and Marianne Wiser), Stella Vosniadou, and Michelene Chi. I found no process analyses. Second, I examined a recent handbook on conceptual change research (Vosniadou, 2013) looking for a subgenre of process analyses, microgenetic studies, which are easy to scan for. Only one of 31 chapters mentioned microgenetic methods. With respect to the second index, developing precise language and models, diSessa and Sherin (1998) survey the literature on conceptual change and make the case that even the concept of "concept" is, in general, severely ill-defined.

Developmental and diachronic approaches to conceptual change are important, of course. But interpolating to the actual processes of change is both scientifically and practically extremely important.

Why has cognitive modeling made few inroads into conceptual change? Historical critiques of modeling include that it involves overly homogenized languages (Norman, 1991) and too-idealized assumptions (Dennett, 1993). Along these lines, I would add that approaches to modeling typically emphasize a priori and general approaches to their choice of modeling language, as opposed to empirical, bottom-up study of the nature of knowledge elements and processes.

Given the ultimate value of cognitive modeling, yet respecting its putative limitations, I put forward a more bottom-up, but perhaps more realistic, program of research, which I call "modeling at the knowledge level." The program has the following properties.

1. Pressing toward modeling – The goal of explicit modeling is not rejected, but it is deferred and approached more slowly. Still, (a) precision in terms of describing entities (e.g., types of knowledge) and processes (e.g., learning mechanisms), and (b) ultimate runnability are persistent and visible concerns.
2. Process data – Unlike the mainstream of conceptual change data, the approach attempts to use data that deals with the structure of real-time processing.
3. A more empirical, bottom-up approach to knowledge ontology and mechanisms – Rather than immediately settling on a full modeling language, and completing a fully functioning cognitive model in it, modeling at the knowledge level seeks to uncover human-characteristic knowledge and mechanisms of change in particular regimes of intellectual performance via careful empirical work. By way of example, I have investigated and attempted to provide careful, data-based descriptions of

knowledge types such as physical intuitions (diSessa, 1993), technical concepts (diSessa & Sherin, 1998), and mental models (diSessa, 1991).

To review, modeling at the knowledge level takes cognitive modeling and its considerations seriously—particularly its concern for process data and the empirical visibility of mechanisms of cognition. Similarly, this mode of study aims for precision and completeness of descriptions of objects of mind and processes involving them. On the other hand modeling at the knowledge level defers actual implementation of models, approaching it more slowly, with extended empirical build-up indexed to particular regimes of human performance.

1.2. Mechanisms of conceptual change

"Mechanism" is a common term in science. However, given its importance to this work, I position the term more carefully in the context of modeling at the knowledge level.

Generally, a mechanism provides basic principles by which change happens. Specific descriptions of mechanisms depend on what is chosen as the basic level of change, and, more generally, how one describes cognitive state. Thus, particular mechanisms are deeply dependent on the relevant theoretical framework for describing knowledge. In cognitive modeling, basic learning mechanisms are generally built into the architecture, such as chunking, knowledge compilation, or adjustments of weights in connectionist models.

Mechanisms for modeling at the knowledge level are specified with respect to how the particular entities ("knowledge ontology") behave. With respect to learning (local conceptual change), one needs to specify how and when knowledge entities change, or how new ones come about.

In an article entitled "Thinking about Mechanism," Machamer, Darden, and Craver (2000) describe a theory of mechanism. The theory is explicitly dualistic, focusing on (a) *entities* and (b) *activities* in which they participate. Mechanisms are precisely a *productive* decomposition of a phenomenon into entities and so-called *reliable* activities (the latter are similar to, but distinct from, laws or principles). Details are not of concern, but one aspect of the model is highly relevant. Machamer et al. introduce the idea of *mechanism sketches*. A mechanism sketch is an early try at specifying a mechanism underlying some particular phenomena, including guesses concerning the relevant entities, the reliable activities in which they participate, and the functions that are necessary for those entities and activities to explain the phenomenon. Machamer et al. make the case that mechanism sketches are a natural part of the public process of scientific inquiry. They provide strong scaffolding for future work by exposing "black boxes" that need filling in, limitations concerning the current list of entities and reliable activities, and constraints on new or improved descriptions. They also provide empirical loci in which to work out improvements.

The "mechanisms" put forward here are precisely mechanism sketches in Machamer et al.'s terms. They are recognizably incomplete by the standards of cognitive modeling. However, they aim to provide solid empirical scaffolding for future work. For example, one could ask how a particular modeling language might implement some mechanism, which might result in a straightforward improvement in the precision and runability of the mechanism. On the other hand, it might also turn out that the modeling language is not perspicuous with respect to a well-motivated and empirically supported mechanism sketch, and hence the modeling language, itself, might become a focus for improvement.

1.3. What is a "causal scheme"; How might it be "constructed"?

Phenomenologically, what I mean by "causal scheme" is easily described. The idea is learner-centered: what a learner takes as general, explanatory, and predictive about a class of phenomena. In this case, the phenomena are physical, thermodynamic in particular. I also require that students have substantial awareness of their causal schemes. That is, one wants students to know that they have a particular way of explaining a class of phenomena; they should be able to state and defend the meaning and plausibility of the scheme. The assumption of awareness is a common one concerning professional scientific cognition, and it is also motivated by application to instruction, where students should have conscious access to, and be able to describe, important aspects of what they have learned. I also add the condition of "socially shared," which is again frequently taken as typical, if not definitive of professional science. Being socially shared has direct implications for instruction. For example, it helps rule out idiosyncratic explanations that are characteristic of one or only a few students.

It turns out that the particular causal schemes investigated here do not seem exotic. One is a near-normative description of thermal equilibrium; it is very close to a law announced by Isaac Newton. Another scheme that students developed is an intuitive competitor that offers similar predictions for some aspects of the process of thermal equilibrium, but not for others.

Going beyond this level of detail, of course, brings us to choosing particular theories or models for analysis. The model I use here extends one that has had substantial theoretical and empirical development (e.g., diSessa, 1983; 1988; 1993; Kapon & diSessa, 2012). I elaborate later, but offer a short introduction here. In this model, physical intuition consists of a large vocabulary of "p-prims" (phenomenological primitives), which are schemata that provide sensibility and naturalness to everyday phenomena. For example, one p-prim is that "increased effort begets increased result." P-prims are primitives in the dual senses of not needing further explanation or justification, and being evoked as a whole. P-prims have properties that allow them to be composed into compound structures, a point that is relevant to later parts of this article. Very roughly, if A causes B, and B causes C, then A causing C is judged to be natural precisely by recognizing the A→B and B→C causal links. This kind of composition is completely commonplace in the literature on causality. If the link from A to C turns out to be generally useful, becomes articulate, shared, and stable, then it constitutes a new causal scheme in the sense used here. One can say that this new scheme is constructed out of p-prims, and the relevant *mechanism* might be precisely this simple form of composition. At this stage, I note two important and congenial properties of p-prims. (1) P-prims operate within instructional timescales, over a few seconds to a few minutes. (2) They are not "scarce resources": There are hundreds or thousands of them (diSessa, 1993). Hence, p-prims make good candidates for use in small time-scale studies of innovation in conceptual competence.

1.4. An appropriate methodology

In order to study mechanisms of learning, one would like to have:

- process data that allows analysis of change in real time
- rich data, that permits triangulating on forms of knowledge, not just its presence or absence.

Traditional experimental studies are inappropriate. Instead, consider Siegler's (2006) characterization of microgenetic studies, which engage the difficult task of uncovering mechanisms of change: (1) focal episodes should include substantial change; (2) there should be multiple observations across each episode; and (3) one should make opportunistic use of all available data against which to generate and test hypotheses. These criteria align well with the desiderata for modeling at the knowledge level. The emphasis on theoretical innovation here implicates a subgenre of microgenetic study called microgenetic learning analysis (Parnafes & diSessa, 2013). However, I do not elaborate differences here.

The work of this article, then, is microgenetic in character. It studies notable change that occurs over periods of a few days, sampled in stretches of about fifteen minutes to an hour at a time. The present data are mainly verbal in form, and thus the lower-limit of time-resolution is a few seconds or a little less, during which students can display or explain a shift in thinking.

Microgenetic case studies are very rare in conceptual change research. As mentioned above, only one chapter in a recent handbook on conceptual change (Vosniadou, 2013) even mentions microgenetic study. (See also Rhodes & Wellman, 2013, for recent microgenetic work—although not a case study. These authors also advocate microgenetic methods, generally, in conceptual change research.) However, microgenetic studies of learning, broadly speaking, are not rare outside conceptual change work. In fact, they are a major mode of research for "design-based studies" (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003; Confrey, 2006; Design-based Research Collective, 2004), where analysis between iterations of instruction involves a serious theoretical and empirical review of how learning happened or did not happen (diSessa & Cobb, 2004). Many microgenetic studies of learning—and these here, as well—involve full-class discussions, with attendant strengths (e.g., ecological validity and direct relevance to educational concerns) and weaknesses, both those typical of case studies (representativeness, generality; see Yin, 2009 for a systematic treatment), and those particularly salient here (complicating factors such as social interaction, "dilution" of data on what any one participant is thinking). For examples of microgenetic case studies, see: Cobb (1999); Cobb, McClain, and Gravemeijer (2003); Lehrer, Schauble, Carpenter, and Penner (2000); and Confrey and Lachance (2000). Relevant to our focal concerns, classroom discussions can yield highly nuanced sequences of students' thinking. They can reveal stability or instability of ideas, include reactions (resonance or antipathy) to competing ideas provided by other students, and they often reveal levels of confidence that can be hidden in more formal data collection, such as correct and incorrect answers on assessments.

Unlike most microgenetic studies, one of the studies here benefits from a particular stroke of luck. Many of the knowledge elements (p-prims) that seem to be involved have been previously studied extensively via clinical interviewing (diSessa, 1993). That work triangulated over circumstances and subjects, which case studies in classrooms do not afford. Recognizing thoroughly documented elements in fragmentary classroom data implicates much more about what students are thinking than can typically be gleaned from on-the-fly analysis. See Kapon and diSessa (2012) for similar use of previously documented knowledge elements in analysis of real-time data.

The specific form of analysis here might be characterized roughly as "schema tracking." That is, one identifies knowledge schemata that students are using, seeking to understand their comings and goings, and what kind of change happens, including change in schemata, composition of schemata, and so on. The technique is similar to "model tracing" (e.g., Anderson, et al., 1990). VanLehn (1991) is a closer match, tracking rule change in detail in order to determine rule-learning processes. However, both of these comparison cases are more typical of computational modeling, rather than modeling at the knowledge level. Anderson's units (e.g., productions) and learning mechanisms (e.g., knowledge compilation) are typical of cognitive modeling and qualitatively different from the ones proposed here. VanLehn (1991) uses data to filter and augment standard machine learning mechanisms.

More detail on method is provided in section 2.5.

2. Setting up the study

This section (1) introduces the relevant theoretical framework, (2) elaborates the specific goals for analysis, (3) explains the instructed topic and (4) the instructional context, and (5) provides additional details on empirical methods.

2.1. Theoretical framework

The general framework behind this study is often described as "Knowledge in Pieces" (KiP: diSessa, 1988, 1993; diSessa & Sherin, 1998). It entails the following orienting assumptions:

KiP1: Complexity of the naïve state: Naïve knowledge in domains like physics is rich, complex, and diverse. Characterizing it as, for example, a "coherent theory," as is often done, is misleading (see also KiP3). Instead, a large number of loosely coupled knowledge elements are available and interact in various ways in various contexts to provide the surface phenomenology of "intuitive conceptions" or "intuitive theories."

KiP2: Productivity of intuitive knowledge: While the most prominent researched phenomena of conceptual change concern the appearance of conceptions that are at odds with normative conceptions, the naïve conceptual ecology also provides rich productive resources, building blocks for scientific understanding (Smith, diSessa, & Roschelle, 1993; Linn, 2006; Minstrell, 1989).

KiP3: Grain size and structure: Adherents to this general perspective usually focus on a smaller grain size than competitor perspectives, targeting elements and knowledge-structural relations below the grain size of full concepts or theories (diSessa, 2006; Clark, 2006). Opposing views tend to characterize naïve and scientific conceptions on the basis of "landmarks"—relatively global and assumed-to-be coherent conceptual positions, "naïve theories" being a prototype. For example, Vosniadou and Brewer (1992) characterize each of their subject's naïve ideas concerning the shape of the earth in terms of a single specified model, which all derive from a unitary "framework theory." In recent developmental work, Wiser and Smith (2008) do not position themselves as opposed to Knowledge in Pieces, and they also describe many complexities in students' performances. Still, the core of their work is a single, compactly describable landmark naïve theory of matter and material kind, as the starting point for conceptual change.

Following along with a smaller grain size, structural transformation of knowledge becomes the topic of more explicit concern, and also a much more complex affair in the Knowledge in Pieces view. Extensive structural transformations of large systems seldom admit of simple descriptions (Thagard, 2000).

KiP4: Learning on a small time-scale: Befitting a smaller grain size and a richer space of potential structural transformations, Knowledge in Pieces researchers skew toward micro-studies, often focusing on describing learning at multiple stages, in particular contexts, and learning that takes place in episodes on the order of just a few minutes in duration (Parnafes, 2007; Izsák, 2005; Wagner, 2006). Opposing views, partly because they usually have a more restricted space of theoretical possibilities marked out (e.g., "theories"), tend toward before and after studies of instructional treatments.

KiP4 (small time-scale study) obviously aligns with a principle theme in my motivation for modeling at the knowledge level—process data and analysis. KiP3 (small grain-size) alludes to the theoretical infrastructure needed to carry out such analyses. KiP1 (complexity in the naïve state) is a general finding within this theoretical framing. And KiP2 (productivity) is a presumption that is particularly distinctive of this perspective on conceptual change: Productivity of naïve knowledge is more plausible in this view, given that one has empirical and theoretical resources to investigate and describe many possible transformations from naïve thinking.

One of the knowledge types that has been developed within this perspective, p-prims, is, as mentioned, central to this article. The theory of p-prims includes descriptions of the nature of elements, their origins and development, their function, and system-level characteristics, such as the level of coherence among them. It also includes a set of observational principles that provide criteria for identifying and validating p-prim descriptions from empirical data. For present purposes, most of this is not immediately relevant. What is needed are descriptions of some particular p-prims that have been empirically established. These will appear later. The general characteristics of p-prims that are relevant are: (1) They are "atoms" of causality in that they provide a relatively impenetrable sense of obviousness or naturalness to phenomena. (2) They are not, in general, consciously accessible, hence not, in themselves, causal schemes. However, p-prims show in subjects' attention, in their descriptions of why things happen and in their judgments of plausibility or implausibility. Lack of conscious accessibility is connected to the fact that their connections to language are tenuous (diSessa, 1993; 2004b).[1] For example, there is no familiar lexicon for them. (3) Two major modes of development toward scientific competence have been hypothesized. (3a) The contexts in which p-prims are activated change (Smith, diSessa, & Roschelle, 1993; diSessa, 2004a); (3b) they become a part of a better-organized system (diSessa, 1993), for example, through composition. Unfortunately, careful direct empirical observations of (3a) and (3b) are rare, but see diSessa (1996) for an example. Both of these processes will be salient here. (4) Describing p-prims as "correct" or "incorrect" is a category error. They are plainly not scientific in many ways, but they often make correct predictions in everyday use; that is their ecological function. P-prims often appear as incorrect predictions and explanations ("misconceptions") in situations of importance in science, but this ignores both their frequent effective use in everyday reasoning and the fact that they may become effective parts of normative scientific understanding.

2.2. Specific goals for the analysis

In this subsection, I elaborate the particular research agenda for this article. To begin, I acknowledge arenas in which these studies do not break new ground. I focus on two particular conceptual constructions by students. There can be no pretention of covering the full range of naïve knowledge relevant to the conceptual topic generally (thermodynamic phenomena), or even to the particular task used here, if approached in other pedagogical framings. The grain size (KiP3) of the present analysis is similarly commonplace in certain circles of learning research. However, the following features are distinctive.

Repertoire (the set of engaged knowledge elements): The present work aims at a near-exhaustive analysis of the intuitive knowledge schemata that play roles in significant learning events concerning a difficult area of school science. Historical microgenetic analyses (Schoenfeld, Smith, & Arcavi, 1993, provides an excellent reference) usually have relied on on-the-spot characterizations of how students are thinking, rather than employing cumulative, careful descriptions of particular knowledge elements. Alternatively, microgenetic study may use a priori methods of knowledge analysis, such as partitioning the curriculum into small units consistent with expert decomposition of the field. See, for example, the apology for using "children's ideas, as we find them," in Steffe (2011), which is aligned with the perspective here. Finally, some microgenetic analyses (e.g., Nemirovsky, 1994) eschew units of knowledge (schemata, p-prims) in their analysis of naïve thinking, so "repertoire" is not an applicable term.

Productivity (positive contributions of prior knowledge): The vast majority of conceptual change studies characterize naïve conceptions in negative terms. Productivity is explicitly denied in framings such as ontological views (e.g., Chi, 1992), theory-theory views (Carey, 1991, 1999, 2009; Wiser & Smith, 2008) or in "misconceptions" studies (consult Smith, et al., 1993). As unproductive elements—the argument goes—naïve knowledge must be eliminated in conceptual change. Given the preponderance of negative characterizations, demonstrating productivity at all is a valuable research target.

While many have claimed that naïve ideas are often positively engaged in learning normative science and mathematics, there have been few studies that demonstrate such productivity in the details of particular learning events. Among those studies Wagner (Wagner, 2006; diSessa, 2004a) provided an extended case study of the development of the statistical principle known as "the law of large numbers," finding that the subject needed much learning in many different contexts (which engaged a number of naïve ideas) in order to encode and reliably apply the principle. On the other hand, the productive contribution of clearly defined prior elements was quite distributed in Wagner's data, the specific elements were not separately analyzed, and there was no attempt to track real-time development. Clark (2006) similarly charts long time-scale learning by the gradual heightening of use of "normative facets" of knowledge, gradual diminishing of the use of non-normative facets, and greater co-use (integration) of families of productive facets. But he does not track learning events.

In net, productivity is doubly important to the present analysis. First, it is a distinctive characteristic of the perspective on conceptual change that frames this work. Second, the use of careful process analysis to document productivity is rare even within the KiP perspective.

Mechanism: As suggested from the beginning of this article, one of its central aims is to develop descriptions of mechanisms of learning consistent with a focus on conceptual change. The pursuit here is greatly enhanced by the "stroke of luck" that many of the incoming student knowledge schemata appear to be previously well-documented ideas. With high-resolution and relatively strongly empirically supported descriptions of student incoming knowledge, one may much more easily see changes that are taking place during learning. Metaphorically, knowing puzzle pieces well at the start (incoming knowledge) can allow us to see even minor shaving (adjustment of incoming knowledge) that might be needed to fit them properly together. More literally, one wants to see as best one can what is new and what is old in a learning event. Without a good map of prior knowledge, it is very easy to assert novel compositions as whole "available resources," or to miss subtle changes in the use of a naïve idea, such as its activation in a new context, or a binding to novel attributes of the world. Others who have sought to uncover mechanisms of learning from similar perspectives to that here include Parnafes (2007) and Izsák (2005).

It is worth remarking that surprise at having a good existing map of incoming student knowledge is at least somewhat characteristic of Knowledge in Pieces, where the general claim is that many elements are involved in intuitive understanding, and there is no expectation that all of those elements have been documented. Other points of view might well have it that the entire set of student ideas may already be mapped, for example, as a specified "intuitive theory."

2.3. Instructional topic

The subject matter at issue would be called "dynamical systems theory" (DST) by scientists. DST is a highly mathematized discipline that emerged mainly in the twentieth century as a generalization of specific dynamical theories, notably Newtonian mechanics. The theory deals with a very wide range of systems that can be characterized mathematically in terms of parameters and laws of change. It includes the study of simple systems, such as a pendulum, but also complex systems of many interacting parts. Typical foci of attention are generalized parameter spaces that characterize such systems (phase space), the properties of trajectories (orbits) in those spaces (stability, quasi-repetitive motion, etc.), and things that stay the same (conservation) or change regularly or unpredictably (chaotic motion). Recent attention in education to "complex systems" and the phenomenon of emergence (see Jacobson & Wilensky, 2006, for a review) is a subfocus of dynamical systems theory. Our[2] own attention, in contrast, has been on the simplest systems. Hence our work does not much intersect existing cognitive studies of DST. Our instructional interest is in especially accessible aspects of DST that would make a good introductory curriculum in the middle school to early high school range. This leaves out a lot of the mathematical core of the discipline, at least in early stages of instruction. In addition, we concentrate on particularly important examples of patterns, such as threshold, oscillation, equilibration, pumping, and resonance, rather than general laws or methods.

All of the data here concern our instruction of one pattern, equilibration. A prototypical phenomenon is temperature equilibration, where two bodies of different temperature in contact settle gradually on a joint equilibrium temperature. Alternatively, a single body equilibrates to the temperature of its ambience. However, equilibration, as a pattern and general mechanism, characterizes a very wide range of systems, particularly dissipative systems, where energy is gradually lost or distributed.

Newton first provided a description of temperature equilibration, often called Newton's law of heating (or cooling), which is an excellent approximation to our instructional goal for this pattern. Newton noted that the rate of change of temperature of a body is proportional to the difference of temperatures between it and its "partner" object of differing temperature, or between the object and its ambience. In modern notation, this would be expressed by the following differential equation.

$$\frac{dT}{dt} = k\ (T_{ambient} - T)$$

In words, the rate of change of temperature with respect to time is proportional to the relevant difference of temperatures. The difference term is sometimes called the "thermodynamic driving force," although it is not a force in the ordinary Newtonian sense. In my analysis, I will show that students sometimes spontaneously develop conceptions that approximate "driving force" for the role of the difference in temperatures. The equation expresses a particular causality[3]: that a change comes about from a difference between the relevant changing quantity (*T*) and some influencing quantity ($T_{ambient}$).

Our instructional goals can be summarized thus: We want students to come to feel that "temperature difference drives rate of temperature change" expresses the essence of what happens in a broad range of situations, and also we want them to understand its consequences—for example, that temperature generally settles in gradually (technically, following an exponential decay) to its target temperature. Fig. 1 shows the characteristic pattern of equilibration in a graph of data taken by students in one of our experimental courses.

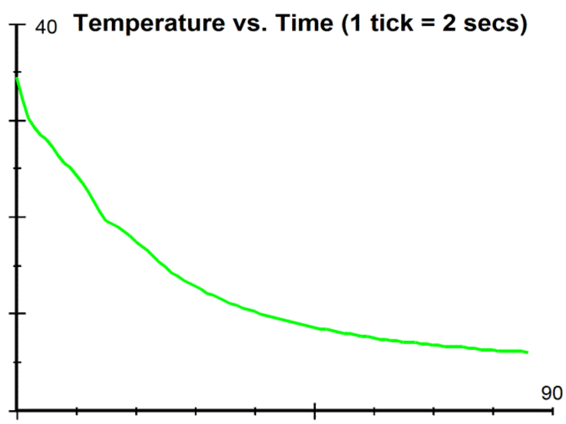


Fig. 1. Characteristic "settling in" of an equilibration (exponential decay).

Analysis here will emphasize that Newton's law of thermal equilibration, while not easy, is accessible (and it suggests that a popular alternative, a molecular model, is less accessible). Analysis will also show that Newton's law can come to be taken by students as causal, and, furthermore, the relevant learning shows at least one core earmark of conceptual change: It is not easily won. Thus, like other studies in conceptual change, equilibration understood in this way constitutes a good and practical target of study in cases of both success and failure.

2.4. Instructional context

The lesson from which our case study data come has been executed in three enactments in a relatively consistent way. In two of the enactments the lesson was part of a larger course. The course's topic was "patterns of change and control," which is the informal name we use for dynamical systems theory. Each edition was run over five 3-hour weekly sessions in the Academic Talent Development Program (ATDP), a summer enrichment program run by the University of California at Berkeley Graduate School of Education (I use the labels ATDP 2004 and ATDP 2006). Participants were volunteers—whoever signed up for our advertised course—and spread across early to mid high school ages (grades 8-10). The classes were small (about a half-dozen participants) to medium sized (about a dozen participants), and ethnically diverse. As an enrichment program, most students were academically oriented and good achievers in school. However, we did no screening of any sort during the process of advertising and enrolling students.

The curriculum in the course emphasized in-class inquiry into patterns. We introduced situations involving several important patterns (threshold, equilibration, oscillation) and had students consider commonalities and differences among exemplars of each pattern, and how the relevant pattern might be described, explained, and applied to widely varying examples. A separate, broader, and more student-oriented inquiry—into whatever the students themselves identified as patterns—was a leitmotiv in the lessons.

In the third enactment, the equilibration unit (one three-hour session in the summer classes) was run stand-alone, without the other lessons in the full patterns class. This break-out study (labeled Urban 2005) entailed 3 roughly hour-long classes, rather than the one 3-hour session used in the summer enrichment courses. The context was a private high school, recruited on the basis of contact with teachers and interest on the part of one teacher in helping our research. Students were roughly in the same age ranges, and I would judge the students, on average, were relatively highly motivated and accomplished, perhaps more so than our summer enrichment class. This class was more homogeneous, consisting mostly of White students, most of whose parents could afford a private school.

Case Study I comes from the break-out class on equilibration (Urban 2005). Case Study II comes from the first summer enrichment program (ATDP 2004).

The teaching sequence for the equilibration class occurred in the following phases:

1. We began with an open discussion concerning what happens to a glass of milk when it is pulled out of a refrigerator and put on the kitchen table. All students felt the milk would converge to room temperature, although in eliciting explanations we always encountered a wide range of competing ideas, essentially none of which was close to normative explanation.
2. We scaffolded continuing discussion of temperature equilibration by asking students to consider what a graph of temperature over time would look like. Unlike the more open initial discussion, classes sometimes converged on a favorite predicted graph and explanation of it, but different classes converged on different graphs. Only one class (Case Study I) converged on a graph resembling the normative graph, Fig. 1.

3. We asked students to do a family of experiments with hot or cold water in a test tube equilibrating in a bath of room-temperature water. They collected and graphed data on a computer using a temperature probe.
4. On the basis of the data they obtained, we asked students to notice features of the resultant graphs and to try to explain them. One focal feature was the gradual slowing of temperature change over time; another was heating compared to cooling, and heating/cooling from different temperatures to the same ambience.
5. Each class was led in a scaffolded construction of a computer model of temperature equilibration. This section was conducted so as to include a version of the normative model. See diSessa, 2008b, for an account.
6. The final phase was a discussion of the generality of equilibration and a consideration of widely differing examples.

Case Study I came from phase 4, discussion after the lab experiment. Case Study II came from phase 2, open discussion of heating/cooling via graphs. Both cases occurred before we introduced the normative model, at phase 5, so they were necessarily substantially student constructions, granted context setting by the instructors.

2.5. Methodological narrative

I identified the method of analysis here as microgenetic analysis, focusing on identifying and tracking knowledge schemata involved in students' developing a causal scheme. Here, I add further detail.

All of our project's classes are videotaped. During analysis, important segments are transcribed. However, most of the analysis is done directly from video in order to make available aspects of interaction, prosody, pacing, student gaze, drawing, or gestures that are difficult to transcribe.

The first main task here was to identify episodes during which causal schemes were developed. The criteria used were: (1) the students considered the scheme explanatory; (2) the scheme was stable and used consistently in several contexts; (3) students were articulate about the scheme, consciously aware of it, and capable of explaining and defending it; and (4) the scheme was socially shared—all or most of the students in the class gave evidence that they understood and applied the scheme in the same way. These conditions are stringent, imply a lot of visible data in student talk and action, and, in our case, made detection of causal schemes easy.

I found only two exemplars of student-generated causal schemes in the data corpus. At this stage, I had no particular presumptions about how to parse the episodes into knowledge schemata and changes (mechanisms), nor even a firm expectation that that could be done. What followed was an opportunistic bootstrapping, tentatively identifying schemata, articulating their properties, and checking against whatever data I could find that bore on the validity of particular descriptions. It gradually became clear that some p-prims might account for some of the data. Since those p-prims had been previously documented, I retrieved published descriptions, systematically matching them against data. This matching involved: (1) *Conditions for activation*. The published descriptions all contain descriptions of contexts under which the relevant p-prim is typically used, and conditions under which it is not activated. Conditions include the existence of objects (e.g., "agentive entities," "a resulting action"), and also situation characterizations such as "spatial symmetry." Often relations among several entities or characteristics are entailed, which were also checked. For example, an "agent" is connected to a "result," through an identified "moderator" (resistance). Is there evidence in the data of three such loci of attention, and sensitivity to relations among them? (2) *Causal implications*. Each p-prim typically stipulates that under given conditions, certain things happen "naturally" and without need of further explanation. These stipulations were also checked against the data. Any mismatches, including student attention to attributes or aspects of the situation not implicated in the published specification of the p-prim, counted as disconfirmatory.

The analysis sections here include sketches of the case for identifying the use of each particular p-prim. Evidence for the use of particular p-prims seemed, in most cases, very good. Conditions of application mostly matched those in published descriptions (exceptions are noted), and so did loci of students' attention and the causal implications they drew. I suggest that matches along several well-articulated and extensively empirically supported dimensions for each of many knowledge elements identified in a substantial data corpus is an uncommon but helpful practice in conceptual change research.

A number of schema candidates resisted characterization as documented p-prims. In these cases, I provide the best ad hoc analysis that I could develop. For some of these, relevant literature was located, enhancing those analyses.

If students learn—and I will argue extensively that our subjects did learn—then, at some point, their incoming knowledge schemata (e.g., p-prims) and/or their use of them (say, in new and stable compositions) are very likely to change. Careful descriptions of these changes implicate potential mechanisms. In some cases, changes are implied in the first use of a scheme, using prior, published descriptions as reference points. For example, a p-prim might be applied, and continue to be applied, in an unusual context. In such a case, the fact of use suggests change, for example, in the conditions for activation of the schema.

Some readers may be concerned that most of the relevant prior work in identifying knowledge schemata was done within the Knowledge in Pieces paradigm. Are other knowledge elements, identified from other theoretical perspectives, being ignored? Of course, the success of prior work in finding useful application here tends to affirm the value of the general framework. And specific critique of the precision and completeness our analyses is always a scientific option. Still, many of the vast array of intuitive ideas identified in the literature on thermal physics *should not be expected* to apply here. Our instruction did not, for example, introduce or attempt to teach the distinction between heat and temperature, which forms the basis of a large literature on students' ideas about thermal equilibration. Nor did we engage phenomena related to the idea of specific heat, the perception of temperature, or similar phenomenological foci that have been studied in other work.

The second case study involves fewer well-documented elements and therefore a lesser degree of certainty concerning descriptions. An informal review—provided later— of intuitive schemata implicated in the many student attempts at explanation that *did not* result in a causal scheme reveals a broader range of intuitive schemata. However, those are not systematically analyzed there.

After analysis of the episode in which the relevant causal scheme emerged, I examined related episodes that occurred before and after. These help illuminate the lead-up to the event of construction. Following episodes in one case study show further development toward normative understanding. Some such episodes triangulate on claims for the involvement of one or another intuitive schema. For example, one episode shows a striking parallelism between developments in independent sub-groups of the class. The parallelism helps confirm one of the mechanisms put forward.

3. Preparation for case study I: Repertoire

I begin the first case study by outlining what is known on the basis of prior study about the intuitive ideas that appeared in the study.

3.1. Equilibration

Intuitive ideas concerning equilibrium are underemphasized in the broader literature on conceptual change in physics. Force and motion (Newton's laws of mechanics) have had much more study. However, earlier work in the tradition pursued here uncovered a fairly extensive and apparently important reservoir of intuitive ideas in the vicinity of equilibration. What I describe is a central part of the account in diSessa (1993, pp. 135-142). All of these schemata are identified as p-prims, and, in general, I use the terminology of that paper. Italics identify p-prim names. Quoted phrases, below, are glosses of intuitive schematizations and student language, not proper physics.

Abstract balance is a p-prim that asserts the equality of quantities that "should" balance each other. Spatial considerations seem to be critical in cuing and sustaining this primitive. For example, one question used in the work cited above concerned a symmetrical situation, where a monkey and an equal dead weight, connected by a rope through a pulley, start "balanced" at the same height on either end of the rope. What happens if the monkey begins to climb the rope? A common intuitive response is that "the balance is maintained," and weight and monkey rise in parallel. This is, in fact, the correct prediction, but its proper justification has nothing to do with a necessary balance. The correct mechanism maintains an imbalance if the weight and monkey start at different heights. In the 1993 paper, the primary examples of abstract balance all involved spatial symmetry.

Abstract balance may be temporarily broken, resulting in *abstract imbalance*. For example, a pan balance may be put "out of equilibrium" by pushing with a finger. In such a case, removing the perturbation results in the *abstract balance* reasserting itself: The pan balance returns to spatial "equilibrium." The return motion is called *equilibration*. The typically assumed pattern of motion during equilibration is a gradual decrease in speed, leading to stop. This is *slowing equilibration*. A second and much less frequently evoked pattern of motion involves overshoot, one or more times, with eventual settling in on the equilibration point. This version is called *overshooting equilibration*.

One of the stark disjunctions between professional and intuitive physics is that intuitive *equilibration* is typically interpreted naively without agency. (See the discussion of agency, directly below.) That is, there is no agent that is responsible for the equilibrating motion. *Abstract balance* simply exists as a tendency to return to “balanced” configurations. As a consequence, subjects queried about how or why a situation returns to balance often deny that any forces (instruments of agency) exist to re-establish the balance. In a professional view, re-equilibration (say, of a pan balance) needs an agentive cause; that is, it needs a force or torque to drive it. The distinction here is analogous to the distinction that Aristotle made between violent (forced) motions, and natural ones. In the latter category, gravity involves objects’ returning naturally to their “home” place, the center of the earth. No external agent (force) is required for this motion.

3.2. Agency and Ohm’s p-prim

This section parallels the equilibration section, but it sketches intuitive schemata having to do centrally with agency (force and motion). Although agentive phenomenology seems even richer than equilibration, for purposes here one needs only some of the most prominent intuitive elements. See diSessa (1993, pp. 125-133; 151-156) for details.

Agency has been imputed in the development of physical ideas, force in particular, at least since Piaget. Piaget himself credited philosopher Maine de Biran, a hundred years earlier, for hypothesizing that the child’s personal sense of agency and effectiveness rooted all causality (Piaget, 1930, p. 126). Indeed, a good fraction of a century before de Biran, David Hume (1995; p. 644) also claimed that personal agency was a root model for all causality.

In general, agency becomes involved in complicated ways with the Newtonian concept of force. On the one hand, the root push or pull metaphor used by essentially every physics textbook is a good anchor for Newtonian force. On the other hand, physical forces often have no direct connection to perceived agency. The results are a bevy of persistent misconceptions. For example, Brown and Clement (1989) studied the naïve belief that a book may push down on a table, but there is no way (novices say) that a table can push up on the book; it merely supports it, or blocks its fall. The table lacks any signs of agency that are required, naively, of the source of a force. Minstrell (1982) and others (e.g., Brown & Clement, 1987; 1989) document that students’ perceptions of the relative strength of forces in action-and-reaction pairs—which are all equal in virtue of Newton’s Third Law—may be extensively manipulated by changing students’ perceptions of agency. For example, the smaller of a pair of objects in a collision is usually thought to exert a smaller force on the bigger (smaller agents are generally viewed as “weaker”). However, that perception can change if the smaller object is literally an agent, say a person, who acts “forcefully” (say, “trying hard” in kicking the larger object).

Among the most important of p-prims, *Ohm’s p-prim,* relates the “effort” that an agent exerts in a situation and the “resistance” that it encounters to the “result” obtained. More effort begets greater result, other things being equal. But a greater resistance reduces result, or a lesser resistance increases result. Interviewees will report that when the action of a vacuum cleaner is interfered with—a hand is placed over the intake nozzle—the higher pitch of the vacuum is accounted for by the motor’s “working harder” to compensate for “increased resistance.” This is incorrect, although plausible on the basis of *Ohm’s p-prim* (see diSessa, 1993, p. 126 ff.). In less exotic contexts, *Ohm's p-prim* interprets daily events such as throwing harder to make an object go farther, and the fact that a bigger object (which offers "greater resistance") cannot be thrown as far. Speed is often interpreted as the result of effort, although distance (the net result of a throw) and other situational attributes may also be interpreted as results. When speed is interpreted to be the relevant result, one of the oldest reported misconceptions pertains, that forces cause velocity (Viennot, 1979) (rather than force causing acceleration, per Newton's $F = ma$). *Ohm's p-prim* seems to have extraordinary scope, and it likely also interprets, for example, personal intellectual effort ("trying harder") and its results (say, better or more accomplishment in school).

4. Case study I: "Freaking out"

I start by going directly to the heart of the matter with an analysis of a key contribution by one student. Afterward, I discuss other matters, including, prominently, the prior and continuing development and use of the scheme.

4.1. The focal explanation

After 5 minutes of discussion in a session that began phase 4 listed above—review of experimental data with an eye toward explaining patterns in the data—one of the students in this class produced the following explanation of why the graph of temperature versus time started steeply but became less steep over time. The students in this class, similarly to our other classes, seemed comfortable and confident reading graph slant as the rate of heating/cooling, even though none of them had had calculus. It is important to keep in mind that up to this point, we made no attempt to explain or draw out a normative explanation for any temperature phenomena. Until we introduced the normative model, none of the other original contributions by any students in any of our classes came very close to normative, scientific explanation of heating/cooling phenomena.

I divide the contribution by the student, called "W," into chunks, and enumerate them for reference. Transcription conventions follow:

// — A break in speech, typically including a pause, then restart or new direction.
/ … / — Interruption or parallel speech.
< ... > (Italics) — Interpretive commentary.
[…] — Uncertain or inaudible speech. Text indicates our best guess.
For simplicity, I usually delete pause markers (e.g., "umm,"), colloquial hedges (e.g., "like"); and repeated words in pause-and-restart sequences). In addition, I "timestamp" each video segment by [<*class*>, <*video number*>, <*minute : seconds*>]

[Urban-05, 2, 7:00]
W:
1. I think that the liquids like to be in an equilibrium.
2. So, when one is way off, they sort of freak out
3. and work harder to reach equilibrium.
4. And when it's closer to equilibrium, they're more calm,

5. so they sort of drift slowly towards equilibrium.
6. So maybe that's why it moves fast at first, because it's like freaking out.
7. But then it just calms down as it approaches the right temperature.

This account seems intuitively easy to understand, which was corroborated in the reactions of other students (below). In the following analysis I explain how this is so, and on what prior knowledge it relies.

4.1.1. Segment 1: Equilibrating

In segment 1, W appears to be invoking the *abstract balance* p-prim in asserting that liquids like to be "in an equilibrium" with each other. Consistent with intuitive *equilibration*, the implication seems clearly that they will return to equilibrium if left to their own devices. Data exposed below support the contention that this student came to his explanation with *equilibration*, in particular, *slowing equilibration*, in place as his view of this situation. It is important to note that *abstract balance* and *equilibration*, by themselves, are not yet close to the target conceptualization. They leave out the central element of causality, that the difference of temperatures drives temperature rate of change proportionally: greater gap in temperature implies a greater speed of return to equilibrium. Indeed, the rest of the contribution, described below, supplies exactly this missing piece.

W's language is notable. It appears anthropomorphic: "liquids like to …." I believe that "liquids like to …" is an example of innocuous anthropomorphism, forced by the problematic relationship of intuitive schemata and language noted above. That is, there is no standard lexicon for the intuitive perception of agency in inanimate objects, so anthropomorphism is recruited. Phrasings such as "objects want to do something" were common in our classes (dozens or scores of occurrences), concerning not only temperature, but also many other physical phenomena. For example, students used such language to explain how springs or springy objects work: "They want to retain their shape." Neither students nor instructors reacted as if this language was in the least unusual. It is worth noting that innocuous anthropomorphism is commonplace in physics classes on the part of instructors, as well as students. The effect of momentum is frequently explained as "objects want to continue moving."[4]

In net, I propose that W's anthropomorphism here, "liquids like to…," is innocuous and not at all unusual. This stands in contrast to other highly agentive language, analyzed below (although, it may be a small step in that direction, away from previous non-agentive slowing equilibration talk). Later analysis will show the critical productive function of W's striking anthropomorphism.

There are, however, unusual aspects of this initial appeal to equilibration. In the first instance earlier work on *abstract balancing* (cited above) imputed spatial symmetry as an important cuing and sustaining attribute. That is, balance usually appears in the form of spatial symmetry, and imbalance as spatial asymmetry. Here, there is no such spatial component, and what is being balanced (equilibrated) is highly abstract: temperatures of different objects. The suggested conclusion is that this student is making an unusual move; it may well be a move toward expanding the use of the naïve schemata (*abstract balance* and *equilibration*) to a broader range of circumstances. As mentioned, shifting contextuality is a commonly cited mechanism of change within the Knowledge in Pieces perspective.

That it is an unusual and creative move to use *abstract balance* and *equilibration* in the case of non-spatial quantities is reinforced by our experiences in other classes. Although it may be surprising since it is part of the scientific lexicon, "temperature equilibration" was not mentioned by any student in any other class during the pre-normative instruction phases. As far as we have analyzed the corpus, the only other persistent and socially shared use of an equilibrating principle concerning any of our examples of equilibration was by a single pair of other students. However, these students were reacting to a spatial situation, where beans shaken in a two-chamber container with a gap between (the gap allowed beans to move randomly from one chamber into the other) were seen to equilibrate (our language). These students systematically proposed "things balance out" (their language) to explain a host of questions about what they observed in shaking the bean container, and why it happened. Similar to my descriptions of intuitive *equilibration* and *abstract balance*, this pair of students seemed to have no sense at all about how and why equilibrium was restored, which is the contribution that the remainder of W's explanation makes.

4.1.2. Segment 2: Freaking out

The language that this student introduces in segment 2, "freaking out," requires discussion. Unlike most of our analyses, this one cannot rely on prior study of intuitive schemata. "Freaking out" is plainly a culturally shared linguistic expression; it is not a p-prim. However, in order to engage the level of detail provided by other analyses here, it would serve well to schematize the meaning of this expression. What follows are elements of an ad hoc but plausible description of the meaning of "freaking out." I try to be conservative but specific enough to engage the rest of our discussion of what is happening in this student's explanation.

1. Highly agentive state: The state of "freaking out" is strongly agentive and active. It implies that the agent is highly aroused and probably acting overtly (shouting, gesturing, etc.), possibly uncontrollably. A person who is freaking out doesn't characteristically listen to reason or calm easily. The freaking out state is in marked contrast to the "normal" state of modulated, reasoned control over actions.
2. Dichotomous: The default stance evoked by "freaking out" is dichotomous in the sense that it highlights a categorical shift from a "normal" state to the highly agentive state, described above.
3. I believe that freaking out is often or usually interpreted to be a threshold phenomenon. That is, some parameter (e.g., level of irritation) may gradually change or accumulate, and at some point, a categorical shift in mode occurs to a persistent (it cannot be easily undone), less controlled state—that of freaking out. Freaking out can, thus, focus ambiguously on the mode shift ("In response to his son's continuing irresponsible actions, he finally just freaked out."), or on the final state ("He was completely freaking out.").

Returning to the analysis, in segment 2, "so, when one is way off, they sort of freak out," one sees the assertion of freaking out as a mode change. In this case, the first clause, "so when one is way off," implicates the parameter (temperatures are "way off"; temperature is first mentioned by name in segment 7) that causes freaking out. This seems consistent with the default interpretation of freaking out as dichotomous ("when" suggests a particular time or parameter value), although that will be explicitly amended shortly in W's description of liquids' equilibration.

Anticipating the analysis just below, I claim that the major functions of the freaking out language are two-fold: (1) to invoke agency in an ostentatious way, attributing it to the liquids, in a situation (*abstract balance* and *abstract imbalance*) where agency is typically not invoked. This provides the central "hook" for the causal scheme that is being developed in W's contribution. The second function of W's freaking out statement is: (2) to connect the causal agency to what happens to be the normative driving parameter, the difference in temperature between the liquids. That is, the level of activation is a result precisely of the difference in temperatures. "When one is way off" is, on its own, very likely a statement concerning difference in temperatures (confirmed in segment 7), and it is imputed as the condition for freaking out. The interpretation of temperature difference as a causal factor later becomes explicit and socially shared, even if it might be somewhat ambiguous here.

In contrast to the innocuous anthropomorphism of "objects *like* or *want* …," discussed relative to segment 1, the anthropomorphism of freaking out is jarring. Other students laughed at W's contribution not only when he first said it but nearly every time he or others used "freaking out" on the first half-dozen or so occasions. The teacher's first reaction was also to mark the unusual nature of the attribution. She said, "That's very descriptive, but I'm not sure what it means!" Neither W nor any other student provided an articulate justification or explanation for the use of freaking out.

Here are two complementary explanations for the fact that the students perceived this anthropomorphism differently than the more innocuous sort. First, freaking out is highly specific, psychological, and human-characteristic. In contrast, liking or wanting seems to have become an almost conventional way of talking about certain tendencies in inanimate objects and systems. Furthermore, W is dramatically violating the standard intuitive conception of *equilibration*. Instead of a natural and spontaneous occurrence, W has introduced an agent where none is typically invoked. That invocation is, however, the central conceptual innovation that he is providing in his causal model of equilibration. He is moving from *equilibration* as a "black box" pattern of change, to one whose rate is driven in proportion to a difference in temperatures, as will be seen in continuing analysis.

Below, I make reference to further data that reinforces the interpretation that W is strongly evoking agency where it is not usually seen, but that, once evoked, it is judged sensibly apt. While students reacted to W's language with laughter, they expressed approval of the explanation and used it to explain other situations and phenomena of equilibration. No one ever explicitly objected to freaking out as either wrong or inadequate in providing an explanation of temperature equilibration.

4.1.3. Segment 3: "Working harder" and Ohm's p-prim

Segment 3 (the liquids "work harder to reach equilibrium") invokes the active agency—imputed in segment 2 and there connected to the normative control parameter, difference in temperatures—and channels it into an effort aimed specifically toward achieving a certain end (to reach equilibrium). The effort of an agent is the prototypical control parameter in *Ohm's p-prim* and, in fact as discussed above, "trying harder" or W's "working harder" is a typical expression associated with increases in this parameter.

Segment 3 directs the agent's harder work toward reaching equilibrium ("work harder *to reach equilibrium*"). Now, however, within the regime of *Ohm's p-prim*, continuous parameters in terms of effort and result are highlighted. This is in contrast to the default dichotomous perspective of freaking out and W's first phrasing implicating a point of transition ("when one is way off"), but more aligned both with the obviously continuous control parameter [being closer or farther from equilibrium (segment 2)], and the evidently continuously varying effect seen in the different rates of return to equilibrium (Fig. 1), which are the focus of attention in this discussion. The appendix provides a bit more analysis of the language phenomena identified here.

Up to this point, the result connected to the effort in *Ohm's p-prim* has remained implicit. W does not explicitly mention rate of temperature change. However, given the pragmatics (particularly the phenomenon that is the focus of explanation), it seems clear that he intends to interpret the effect of the agent's effort as the rate of temperature change toward equilibrium. In the next segments, W explicitly makes this connection. In my discussion of *Ohm's p-prim*, I noted that the interpretation of effect varies; sometimes it is spatial net result (effort controlling the distance of a throw), sometimes it concerns rate (a harder throw results in greater speed). Here, W seems unambiguously connecting effort to rate of change of temperature. Note that this connection is highly abstract. In the physics conceptual change literature, the well-known misconception that force begets velocity (rather than acceleration) focuses on an object's spatial motion, which might be regarded as "concrete" in the sense of easily visible. Here, W uses rate of change of temperature as a result. Unlike the invocation of agency in the causality of equilibration, W's extension to rate of change of an abstract parameter, temperature, seems less novel since, as I pointed out, *Ohm's p-prim* gives every sign of being quite general in its application to many sorts of "efforts" and "results."

4.1.4. Segments 4 and 5: Emphasizing continuous control

Segments 4 and 5 follow the causal chain just announced in the differential case of a smaller "driving force," a smaller difference between temperatures of the liquids. Here, the continuous interpretation of *Ohm's p-prim* is obvious in comparative language concerning the difference in temperatures, the level of activation of the agency, and the resulting "effect," which is the rate of change of temperature, connected via *Ohm's p-prim* to the agent's effort. "When it's closer to equilibrium <*i.e., when the two temperatures are closer*>, they're <*the liquids are*> more calm <*lower level of activation or agency*>, so they sort of drift slowly <*lower activation is linked to lower effect and lower rate (drift) of temperature change*> towards equilibrium." For completeness, I note that W has elided explicit mention of the "channeled" version of level of activation, working harder *in order to achieve equilibration as an end*. Instead, the level of activation ("more calm") is directly connected to the result (slow drift towards equilibrium).

4.1.5. Segments 6 and 7: Summarizing the focus of the analysis and its logic

In segments 6 and 7, W masterfully reviews the focus of his causal explanation (why temperature changes rapidly at first in equilibration and slower later on) and his causal explanation of it by contrasting the high agency (big temperature difference) situation and its causal result (faster "movement" of temperature), with the low temperature difference (closer to equilibrium), slow results ("then it slows down"). "So, maybe that's why it moves fast at first, *because* it's like freaking out. But then it slows down *because* it's approaching the right temperature." Italics have been added to emphasize the causal links that W is imputing to this situation. Again, agency or activation is highlighted in his explanation, and effort, per se, is elided.

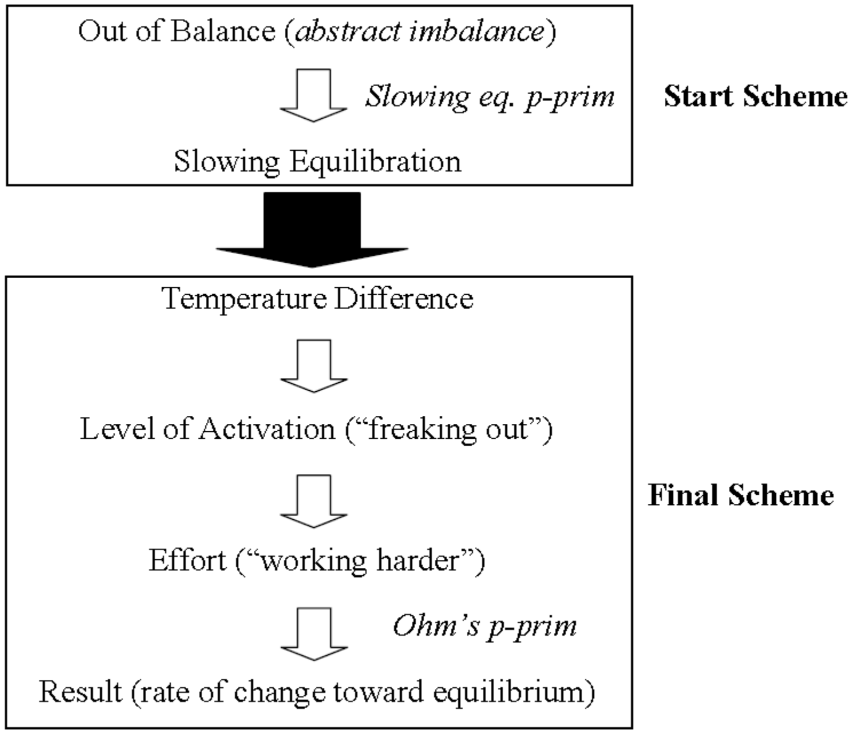


Fig. 2. The overall transformation in W's "freaking out" explanation. Slowing equilibration (top panel) is replaced by an extended causal chain (bottom panel) featuring Ohm's p-prim as a critical link.

Fig. 2 reviews the overall transformation in W's explanation, from a simple *slowing equilibration* scheme (shown in the top panel; again, see continuing data and analysis of W's conception of equilibration coming into this new construction), which gets transformed (bold arrow) into an extended causal chain (bottom panel). The behavior over time (slowing change) emerges from "running" this new causal scheme—or at least comparing two stages of its application—which I do not separately analyze.

I briefly discuss the nature of the first two causal links in the bottom panel of Fig. 2. Both of these links seem to be transient anthropomorphic metaphors. In the first case, it seems natural that a person might "freak out" because s/he is not getting what s/he wants. In this case, the liquids are "way off" from the equilibrium they want, so they freak out. In the second case, if one has reason to freak out at a condition, such as being far from what one wants, then, naturally, one may "try hard" to close the gap. I propose to view these, complete with their bindings, as on-the-spot constructions rather than, for example, pre-existing p-prims. I believe they are forced, locally, in order to complete the overall structure of the explanation (see discussion of intuitive completeness, below). However, continuing data analysis will show them disappearing from student explanations.[5]

4.2. Ancillary claims and analyses

This section fills out our analysis of W's explanation by: (1) elaborating the claim that W's explanation is nearly normative, but explaining the ways in which it remains non-normative; (2) supporting the claim that his explanation was stable and became socially shared (primary criteria for selecting student explanations for analysis in this article); (3) showing a bit more of the development of the scheme during its use.

4.2.1. W's explanation is nearly normative

W's explanation starts with precisely the normative cause, amount of temperature difference, and traces it through to precisely the normative effect, rate of change of temperature. See the first and last elements in the causal chain displayed in the second panel ("Final Scheme") of Fig. 2.

It is worth noting that very few other students in our studies were even at the right level of abstraction to produce this scheme. Students sometimes proposed much more specific and, one might say, mechanistic explanations for equilibrium (invoking molecular motion and energy), but these were vague, seemingly unstable, and never became socially wide-spread. Note, also, that "work harder" is abstract in the sense that it does not include any detail at all as to how liquids could work hard to accomplish the desired end. In contrast, sometimes students proposed apparently mathematical and causality-free explanations, which seem aptly described as *more* abstract than the normative model. One student proposed that the temperature changed according to the rule that, in each equal interval of time, the temperature was multiplied by a constant. The second socially wide-spread model, which is analyzed below, contained one phase that seemed to be defined by a similarly simple and highly abstract mathematical model (linearity), possibly with no causal explanation as to how it came about.

4.2.2. Non-normative aspects of W's model

Obviously, the language of W's model is non-normative. Even if public language is created by asserting a convention, probably no self-respecting physicist would choose such an overtly anthropomorphic label ("level of freaking out") for a technical term. However, similar incorporation of metaphorical (if not anthropomorphic) terms into a technical vocabulary is quite common. For example, concerning the normative Ohm's law—current is proportional to voltage, and inversely proportional to resistance—"resistance" is arguably based on the kind of resistance utilized in the intuitive version of the law (*Ohm's p-prim*), and the British term for voltage, "tension," would seem to have similar metaphorical origins. (See diSessa, 1993, p. 221.)

Whatever its roots, technical language of physics becomes functionally stripped of intuitive meanings for practical purposes ("practical" meaning "in scientific practice"), in favor of being embedded in a more explicit and operationally defined matrix of terms and concepts. Some observations, below, show students moving in this direction.

Finally, I believe that it is appropriate to characterize W's link between temperature difference and rate of temperature change as qualitative proportionality (Sherin, 2001; Forbus, 1984), distinct from strict mathematical proportionality. Qualitative proportionality implicates only that as one quantity gets bigger, another does also; linearity is not necessarily entailed. This leaves room for further mathematical development, and, in later analysis, I will point out that, indeed, students made such progress.

4.2.3. Development, use, and social spread of the "freaking out" model

Because stability and social spread are defining characteristics of the selection of student-generated causal schemes, I document these here. I also remark on important precursors to the above explanation and on signs that the scheme continued to develop in the group over a period of time. I discuss these in chronological order. The most striking aspect of this development is the quasi-independent emergence of essentially the same scheme by another student. That helps reinforce the claim that this scheme is supported by commonly available intuitive resources, rather than idiosyncratic to W or to the social context. It also reinforces claims made above concerning the function of W's overt anthropomorphism.

Precursor 1 (Instructional Phase 1): W's first explanation of the milk's temperature behavior (during the initial open discussion about what happens when cold milk is left on the kitchen table) introduced the word "equilibrium." W described the room and the glass of milk [Urban-05, 1, 12:56] in a kind of "battle to reach dynamic equilibrium," in which the glass of milk gets "beaten by the room," and temperatures of both come to match up and "exist in harmony." He explained that in the battle, "the stronger one affects the weaker one more."

There are two striking points of commonality with his later model. First, he introduces anthropomorphic agency dramatically: "battle." So this is, plausibly, the point at which agency, per se, was introduced into his and others' thinking, even if it is not yet properly bound to features of the world (temperature difference, and rate of temperature change). Further, this is the point at which W introduced the word "equilibrium" into the discussion. Whether or not W's introduction of the term was part of the cause, the word equilibrium played a much larger role in this class, compared to the others.

There are also points of discontinuity between this explanation and his "freaking out" explanation. Centrally, notice that he has dual agencies (room and milk) as competing influences, whereas the final model has temperature difference as a single "driving force" (possibly acting independently in each liquid: "They [the liquids] sort of freak out."). Competition among dual agencies turns out to be a central player in intuitive physics (see discussion of the intuitive attribute of "strength" and of *dynamic balance* and related p-prims in diSessa, 1993, p. 132, p. 138 and following). But, W's final model does not have the cold liquid competing against the warm liquid. Asymmetrical influences that depend on size or strength ("the stronger one affects the weaker more") also play no role in the final model.

There appeared to be no social uptake of these battling ideas in the class.

The next three precursors took place in the second phase of instruction, where we encouraged students to draw graphs of the expected behavior of temperature over time. Our general experience is that the range of explanations given in this phase is reduced, compared to the open discussion of "what happens" in phase one. Stripped of "physicality" in view of the graphical representation, changing temperature seemed mostly explained, if at all, by the very abstract equilibration p-prims. However, another student, R (in precursor 4) offered an important agentive enrichment, pointing toward what emerged strongly later on.

Precursor 2 (Instructional Phase 2): The first person to the board to explain her proposed graph, C, strikingly offered *overshooting equilibration* (*slowing equilibration's* less "powerful" partner) as a primitive explanation. C was quite tentative in her explanation ("I recalled in some distant memory…"), consistent with the low priority previously documented for *overshooting equilibration*. She uses the phrase "finding balance" to convey, in my interpretation, the explanatory force of *equilibration*. Obviously, she cannot be remembering that this happens (since it does not happen), which supports the p-prim interpretation.

[Urban-05, 1, 24:33]
C: I recalled in some distant memory <*laughing*> maybe it would go like this <*draws overshooting graph*>, and it would heat up more than it should, and then go back down to room temperature where it would find its balance.

Precursor 3 (Instructional Phase 2): W was the next student contributor after C. W backed out of his agentive "battling" try at equilibration, given during the prior instructional phase, and appeared to assume a *slowing equilibration* perspective. *Slowing equilibration* is in the same family, but slightly different than C's *overshooting equilibration*. He drew a normative-looking graph, but said nothing whatsoever about why that would happen, which is entirely consistent with *slowing equilibration*'s serving, at this stage, as a primitive, holistic explanation. Here is W's description, which I would describe as behavioral, not causal, consistent with a focus on intuitive *equilibration*.

[Urban-05, 1, 26:43]
W: The glass goes toward the room temperature quickly at first, but then slows down and slows down. But eventually, at a very slow pace, it reaches the room temperature. This straight part <*the final part of the graph*> represents the equilibrium.

Precursor 4 (Instructional Phase 2): The next presenter at the board, R offered an apparently normative graph of the milk's temperature, explicitly agreeing with W. When questioned about why the slowing that she depicted might happen, R offered:

[Urban-05, 1, 31:05]
R: Just because it's such a cold object and such a hot environment that it would just heat up really quickly, because the difference [that] it's trying to reach equilibrium.

R is now focusing on the difference in temperatures as the causal source for speed of temperature change. She both marks the contrast of hot and cold, and speaks of "difference." R also speaks in anthropomorphic terms, introducing agency: "It's trying to reach equilibrium." This appears to be anthropomorphism of the innocuous kind—a way of marking a tendency, as discussed above; there were no public marks of surprise, such as laughter.

R's remarks met with no commentary and no visible social uptake. R's explanation here is a little less elaborated, but not particularly distinguishable from W's eventual "freaking out" explanation, although W's more dramatic language could have been influential. In view of the rapid uptake of W's "freaking out," it still seems anomalous that there was no uptake at all of R's explanation.

Precursor 5 (Instructional Phase 3): During the following instructional phase, where students had broken into two groups to collect data on equilibrating temperature, both groups were asked independently to explain the shape of the resulting curve. R can be heard in the background, while the camera focused on the other group.

[Urban-05, 1, 47:45]
R: The hot water is like shocked […], but, umm, the colder water that you put it into […] causes it to cool down really quickly. But once it's at a lesser temperature, it starts to slow down as it reaches [equilibrium].

R seems to be following up on her explanation, just above, except that she has moved to mark the agency with more dramatic language, "shock." This is strongly consistent with the interpretation that W's blatant anthropomorphism serves to underscore the existence of agency, which is appearing in an unusual situation, *equilibration*. R has, independently, moved from less to more dramatic anthropomorphism in her explanation. It may also be worth noting that R has the cold water (bath) acting on the hot water ("the colder water… causes it to cool down…"), rather than the cleaner version that W offers later, where the difference in temperatures is, essentially, the causal agent.[6] The general and important point, to which I will return, is that finding an appropriate binding of intuitive slots, such as "the causal agent," is a critical part of the construction of the eventual causal scheme.

Unfortunately, the discussion in R's two-person group is inaudible before and after this comment. It is important to note that W was not in R's lab group, and as far as one can tell from the video, neither he nor other members of his group attended at all to what was going on in R's group. W and his colleagues can be seen and heard on the video actively engaged in their own discussion of the experiment while R and her partner discuss the "shocked" water. In general, R's quasi-independent development of a model very much like W's "freaking out" provides suggestive data that the path of development is specific (e.g., involving explicitly enhancing agency), not idiosyncratic to particular students, and comparatively easy to achieve.

Now I turn to the social uptake of W's explanation and its further development.

Uptake and development 1 (Instructional Phase 4): With one intervening turn after W's announcement of his model (a teacher's turn), R displayed both understanding of W's model and affiliation with it. The interpretation presented here is that she had created nearly the same model as W, respecting a difference in language.

[Urban-05, 2, 7:51]
R: Well, I agree with W, and the way I was trying to describe it was: Once the really cold, for example, water gets put into the much warmer water, it experiences like a shock because they're so drastically different. So it gets closer to that temperature faster in the beginning and stops freaking out and calms down as it approaches equilibrium.

R appears to believe W's model is the same as hers, with the exception of language (shock vs. freaking out). She starts by using shock in her contribution here, but then uses freaking out near the end of her explanation, implicitly aligning her language with W's, simultaneously appearing to suggest that the difference in language does not make much difference.

In terms of appropriation, some aspects of mutual influence of R and W are hard to sort out. R could have been following up on W's first introduction of the term equilibrium when she first explained the equilibration pattern (precursor 4). Even if she did, however, she did not use W's dual agency "battle" framing on her first announcement of her model (precursor 4, again, which did not use "shock" terminology). It seems plausible that W followed up on R's sketch of agentive causality in his construction of the freaking out model, although, again, it does not seem at all likely that he could have heard her development of the idea with the word "shock." Without better data, deciding such possibilities is speculative. What seems indisputably clear, however, is that W and R now share a relatively common understanding and language.

Uptake and development 2 (Instructional Phase 4): After both W's and R's contributions, the teacher reviewed the logic of the argument while pointing to different parts of equilibration graphs that appear on the board. She uses both "shock" and "freaking out," following R's lead by suggesting implicitly that the schemes are the same. Uptake, of course, often can be aided by teachers' work in repeating students' ideas, thus providing positive sanction. However, this move follows a common practice in our classes of repeating student ideas, right or wrong, and asking other students for agreement, disagreement, or commentary.

[Urban-05, 2, 8:23]
T: Okay, so it sounds like you guys are saying there's a lot more freaking out or shock value or something up here <*pointing to initial portions of graph*> or here, but then as it reaches closer to the equalizing part, the freaking out is less?

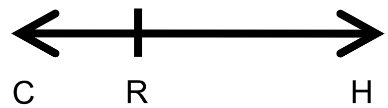
Later on, more direct sanction was introduced, asking students to explain various phenomena in terms of W's model.

Uptake and development 3 (Instructional Phase 4): After the teacher's restatement of W's and R's model, both Z and R can be seen nodding their heads in affirmation. Z was R's lab partner, hence he probably benefitted from R's presenting her explanations, using the term "shock," during the lab. The teacher had framed her own turn as if to say "what I hear you saying is …; is that correct?" It appeared she interpreted positively nodding heads as a sufficient and affirmative response.

Uptake and development 4 (Instructional Phase 4): A short time later, the teacher returned to a comment that another student had made earlier, that a cooling graph of hot water appears steeper than a heating graph of cool water. It happened that the hot water was farther above room temperature than the cool water was below it. The teacher asked for an explanation, and W volunteered:

[Urban-05, 2, 10:30]
W: Maybe that was more degrees hotter than room temperature than the cold one was colder. So, / T: What do you mean by that? / so, it was farther away temperature-wise, in degrees, from room temperature, which meant it like had a bigger freak out. <*Laughter*>
…
T: So, what do you mean by farther away?
W: I mean in terms // like, if you had a number line with negative and positives.
T: Can you come up and just draw that for us?

<W drew and narrated, as follows.>



W: The hot temperature is farther away from room temperature than the cold temperature is.

The importance of this episode is as (1) a re-use of the model in a different context (comparing different graphs rather than comparing slopes at different places in the same graph), and (2) a move toward mathematizing the model with a clear invocation of numerical differences as being relevant. If there ever were ambiguity that “farther from equilibrium” implicated temperature difference, there is none here.

Uptake and development 5 (Instructional Phase 5): Thirty-five minutes later, the students had produced a computer model that matched their heating and cooling data. The model iteratively added or subtracted a “temperature-change” variable from the current temperature, while changing “temperature-change” by multiplying it on each iteration by a constant a bit less than one. The teacher noted that the two graphs (cooling and heating) started with different temperature-change values, and asked why. No response was forthcoming. The teacher asserted that this was the same question about steepness that they had been asked a while ago. After a long pause, C offered, [Urban-05, 2, 44:55] “Would it have to do with W’s number line thing?” The teacher then scaffolded production of an appropriate use of W’s “number line thing.” This seems obviously to be a more complex context, where students need to see their scheme at work in the code of a computer program. But, with help, they succeeded in making the connection. Freaking out was not mentioned. It is also telling in terms of social spread that C, and not R nor W, initiated this application of the model.

Uptake and development 6 (Instructional Phase 5): Near the beginning of the next class, two days later, a similar comparison of different graphs came up. When the teacher asked why one graph was steeper than the other, C quickly offered [Urban-05, 3, 5:50], “Because the hot one started farther away than the cold one.” Again, freaking out was not mentioned. The progression toward these last two mentions of the model, where freaking out seems no longer salient, suggests an appropriate move toward mathematization of the model, stripping away anthropomorphic language and, indeed, stripping away the agentive middle nodes in the causal chain.

All in all, one sees a rather extended process. The model begins apparently at least somewhat independently with R and W’s explanations, although some mutual influence is plausible. Both R and W appeared to start with a *slowing equilibration* model, then introduced agency to elaborate. R used the agentive model independently in her lab group, stepping up to more radical language: “shock.” W’s formulation, also emphasizing agency with dramatic language, became the class’s explicit reference point. W also contributed an extension toward mathematization; he introduced a number line. All of the implications that I could glean from the data are consistent with all the students becoming reasonably conversant with the scheme, and they used it more and more fluently in different contexts (comparing different graphs, as opposed to differing slopes in the same graph, in the context of a computer model). Overt anthropomorphism was gradually eliminated from their ways of talking about the model.

4.3. Thematic review

I review and synthesize the lessons of the "freaking out" case study. I focus primarily on W's construction, and not much on its development and social spread. I do this using the list of specific goals for the analysis, described earlier.

4.3.1. Repertoire

Perhaps the most striking result of our analysis is that W's model (and supporting prior segments) invokes a number of previously identified intuitive schemata; the data rather unambiguously accord with prior descriptions of them. Elements include: (a) *abstract balance* and *abstract imbalance*; (b) *equilibration* (primitive, non-agentive return to *balance* from *imbalance*—in two forms: *overshooting equilibration* and *slowing equilibration*); (c) *Ohm's p-prim*, where "effort" is consequent to the degree of "freaking out," and "result" is interpreted as a rate of change of temperature, a generalized kind of speed.[7]

A striking feature of the model W developed was its language. "Freaking out" (or, in R's case, "shock") seemed to play an important role, probably both individually, and in the social exchange. Language and the ready models it supports, of course, can be considered a resource that fed into this construction, although it is of a different sort than the availability of particular p-prims. Consult the appendix.

4.3.2. Productivity

I argued that the freaking out scheme was nearly normative, with the notable exception of language. It causally connected the normative parameter, temperature difference, to the normative "result," rate of temperature change. Students used the scheme to explain several phenomena of heating and cooling consistently with the normative model. The dynamic of discussions seems convincing that students felt this scheme to be explanatory. They were not parroting instructed explanations, and, indeed, the students seemed almost completely in control of the development of the scheme. For example, the language was provided by the students, and the teacher initially positioned W's construction honestly as something she did not understand.

Putting these students in the context of a full, normative understanding of Newton's law of heading/cooling, my analysis suggested that significant progress was made beyond the initial model. The students stripped away anthropomorphism, and began mathematizing the scheme. Mathematical limitations remained (diSessa, 2008b), but these are not described here.

4.3.3. Mechanisms

This subsection schematizes several aspects of the "goings on" in the freaking out learning sequence so as to make them plausible candidates for general learning mechanisms.

1. **Composition**: The coordination of a cluster of elements to create a composite is, as pointed out, an obvious and generic principle of developing new conceptualizations out of old ones. Composition is the basic form of analysis provided of the freaking out

explanation. In particular, the composition is in the form of a **causal chain**: X causes Y causes …. (Difference in temperature causes increased freaking out, which causes increased effort, which causes increased rate of temperature change.) The second case study will show a structurally different kind of composition.

2. **Shift of context**: Several shifts of contexts of invocation were observed. The central one was that agency was invoked in a context where, typically, no agency is seen or needed: *equilibration*. Whether the actual shift makes the relevant knowledge element (e.g., *Ohm's p-prim*) more likely to be activated in different circumstances in general, or whether its activation is tied only to the contexts in which the new composite applies, is outside the reach of available data.

   Brown and Clement's work (1987, 1989) on teaching students to see that a table supplies a force to support a book laid on it provides a historical precedent for "invocation of agency" as a key shift in learning Newtonian mechanics. In particular, part of their successful instruction is to replace the table with an obvious agent—a person, the analog of the agentive implications of "freaking out."

   Another example of shifting context is the fact that *abstract imbalance* and its associated pattern of change, *equilibration*, came to be invoked in a situation where there is no spatial representation of balance. Spatial symmetry seemed typical and important for the invocation of abstract balance in prior study. Data do not allow us to determine whether *abstract balance/imbalance*, per se, now is more likely to be invoked in situations without spatial representation of balance, or whether a much more specialized increase in range of application is entailed, limited to this particular context and conceptual construction.

3. **Binding**: The real world features to which an intuitive attribute (schema slot) can attach may be ambiguous, or they may change in learning. I pointed out that *Ohm's p-prim* sometimes implicates rate of change, and sometimes amount of change. In terms of learning, liquids at different temperatures (or a table in Brown and Clement's work) may come to be seen as agentive, even if they are not initially perceived to be so, hence they can fit into the "agent" slot in an intuitive schema. Two bindings here were most critical in producing a near-normative causal scheme, the "input" and "output" of the scheme. First, differences in temperature needed to be identified as a controlling influence. This binding was instigated by an anthropomorphic response (freaking out) to temperature differences. Second, the terminal "result" slot (as it turns out, of *Ohm's p-prim*) needed to be bound to rate of temperature change. At the end of the learning sequence, difference in temperature was, as far as one could tell from verbal data, directly connected to the terminal "result" slot.

   Izsák (2005) also described the search process of finding an appropriate binding as a mechanism. In his case, the bindings were the meanings of variables in algebraic expressions.

   Bindings at intermediate steps in the causal chain might have been important for stability and conviction. Recall that R's meandering and uncertain binding of agency (the hot water was shocked, but the cold water was the agent that cooled) seemed upgraded in confidence and stability as soon as W produced his particular bindings.

4. **Causal interpolation**: The top-level generic pattern demonstrated here is that a primitive expectation (*equilibration*, as in W's early assertion of *slowing equilibration* as how heating works) is replaced by a more elaborate and articulated system that

reproduces the same behavioral "net result." The more elaborate system is, in this case, the freaking out scheme, which reproduces the effect of *slowing equilibration*. This learning mechanism was anticipated as fairly routine in diSessa (1993). (See the discussion concerning the "principle of impenetrability" on p. 121.) In addition, Kapon and diSessa (2012) identify several instances of this happening in empirical data. Other classes and other individuals in this class (besides W, R, and arguably C, who introduced *overshooting equilibration*) never included *equilibration*, the starting conception in the transformation to the freaking out scheme, in their considerations. Thus, although bare *equilibration* was far from the endpoint, it likely provided an important focus that seeded subsequent development.

5. **Causal elision**: I pointed out that W's initial model employed a fairly long chain. Large differences in temperature provoke higher agency (freaking out); the higher agency is "channeled" as effort toward reducing temperature difference; and the effort results in a proportional speed of temperature change toward equilibrium. W quickly elided explicit mention of effort in his description. Later, students shortcut the description even more: Rate of temperature change is spoken of as the direct result of differences in temperature. Thus, agency and effort seem critical to the conceptual development here, but also transient. This is, perhaps, one reason why the productive use of intuitive knowledge is seldom documented in detail, especially in diachronic studies; it does not "hang around" long, at least in overt explanation. Whether and to what extent those elements remain surreptitiously present is beyond the data available here.

4.3.4. "Emergence" as a meta-mechanism

This section deals with a more speculative mechanism, and one of a different order than those above. The analysis of freaking out noted a fairly large number of coordinated changes entailed in W's construction, for example, bringing together a particular set of naïve conceptual elements, making some changes in them, and making a particular choice of bindings. Is there anything to say about what instigated and guided those changes? Coarsely speaking, it seems sensible to implicate the large-scale teleology of the activity in which this class engaged. This discussion will, in fact, sketch a "meta-mechanism," that is, a broad mechanism with many probably different instantiations. It is "meta" specifically in the sense that it may manage and integrate each of the above-listed mechanisms. For example, the meta-mechanism helps identify why a particular set of elements is composed, and so on. It is abundantly clear that this teleological meta-mechanism provides only a very partial explanation; other factors, including social dynamics, spontaneous activation of elements relevant to a particular construction, and so on, explain things that this meta-mechanism will not explain. Yet, it proposes to answer, if only in a very rough sketch, things that the separate mechanisms above in principle cannot answer.

The students are focused on a very particular class of phenomena, thermal equilibration, and are trying[8] to achieve an articulate, stable, and socially shared causal scheme. They are quite evidently trying out many possibilities, rejecting (mainly by dropping or neglecting) most of them, settling in and elaborating one of them. This involves (as is evident in the data) generating hypotheses involving particular intuitive schemata, ways of describing them in words, different relevant bindings to real-world properties, and, ultimately, a judgment that the causal scheme is at least "pretty good." One can think of this teleology as a kind of driving force that pushes toward invoking ideas that are not obvious in the situation, putting them in unusual combinations, and binding them to non-obvious and, at first, ignored features of the situations to be explained. That is, I nominate "emergence" in this sense as a general motor that drives and guides all of the above more particular mechanisms.

Four aspects of emergence seem salient and somewhat supported by this data. The first (E1) is merely to name the teleology. W's model depended on the particular intent of student participants; it had a particular **explanatory goal**. It emerged in the context of trying to explain a very particular class of phenomena that are, not incidentally (this was our instructional intent!), focused on key phenomena that Newton's laws of heating/cooling explain. Thus, it is completely sensible that students measured the model in significant degree by how well it worked, altogether and in their view, to explain these focal phenomena. Emergence may, thus, have "caused" changes in those schemata in this context of use precisely so as to combine them effectively into a workable explanation. The rest of the features name effects and conditions of the teleology.

Selecting pieces, changing and configuring them to fit a working whole is the second aspect of emergence (E2). In discussing mechanisms, a rather large number of innovations from the naïve state were listed, presumably motivated and stabilized by the particular explanatory goal. I put binding of intuitive attributes to particular and unusual world features under the rubric of "changing the pieces."

The third aspect of emergence (E3), which one might call **intuitive completeness**, is the necessity of *all* the assembled pieces in order to produce an acceptable (to the students) explanation. On the one hand, the existence of intuitive completeness would seem to be a relatively unproblematic claim. W's model contains: difference in temperature drives "freak out" level; "freak out" level drives effort and, via *Ohm's p-prim*, speed of temperature change. One cannot get from one end of the chain (temperature difference) to the other (rate of temperature change) without the all the links in the chain. After establishing that completeness, however, small and, eventually, quite severe elisions occurred, moving toward more normative understanding. Thus, intuitive completeness is a time-localized constraint for students' initial judgments of quality of the explanation, even if it is critical during those early phases.

Finally, (E4) it is worth explicitly calling out the capacity and use of some fairly global **judgment** that one explanation is adequate, or more adequate than others. Students abandoned many attempts at explaining equilibration, and persisted and, indeed, improved others. Some aspects of this explanatory judgment are either evident or highly plausible. Re-use of felt-to-be reliable p-prims is one aspect. Students are almost certainly sensitive to at least some ambiguities, such as what, exactly, does freaking out mean (although this did not come up in these students' discussion!).

5. Case study II: The three-phase model

This second case study of the development of a relatively stable and socially shared causal scheme makes several contributions. First, it provides a second example where composition of naïve elements is arguably the primary mechanism of learning. Hence, it contributes to a sparse research literature delineating specific, element-by-element conceptual constructions. But it also has some contrasts, helping to provide a better sense of the range of ways learning may happen.

1. The naïve elements involved are different in content, and, arguably, different in form (different knowledge types) as well. For example, we already considered that students may use strictly mathematical (as opposed to causal) regularities. They might also use empirical generalizations specifically concerning thermal phenomena, rather than very general and abstract schemata like *Ohm's p-prim*.
2. The form of composition that was used is not causal chaining, but time sequencing.
3. Language, anthropomorphic metaphor, and agency do not play nearly so central roles.

The second case study also makes implicit methodological contributions by showing several failures of p-prims to explain features of this scheme. The data and analysis are sufficiently strong to rule out a number of at first plausible interpretations.

This second case was developed out of the data in our first Patterns class, in the summer of 2004. It emerged during phase 2 of the instructional plan, when students were prompted to draw graphs of how temperature changes in the cold glass of milk as it moves toward what all agreed would be its final state, room temperature. This was the third of five 3-hour classes in the course. Although this *three-phase model* seemed to be a composition, similar to the freaking out model, understanding the elements that were composed is less well prepared by prior work.

Unlike the freaking out scheme, this one developed overtly and gradually in the group, with multiple students contributing, over a continuous 22-minute long discussion. Although I made plain that the freaking out model had precursors and possibly inputs from multiple students, the three-phase model had a more apparent, continuous, and more public construction. I first present the scheme and our best explanation of its elements. Then I return to the dynamic of its creation.

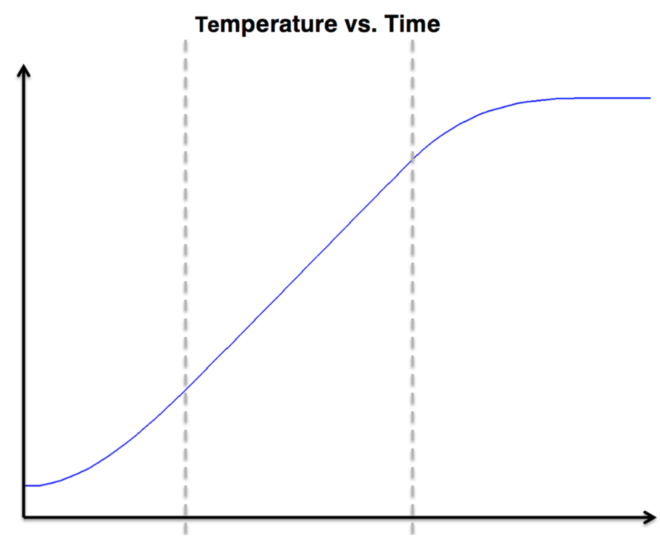


Fig. 3. Equilibration in the three-phase model. First phase is "change takes time"*; second phase is linear change; third phase is "settling in."*

Fig. 3 shows the graph associated with the students' three-phase model of temperature equilibration. During the first phase, temperature change gradually speeds up. The second phase is linear, a constant change in temperature across equal intervals of time. The students were definitive and articulate about linearity. (In contrast, our "freaking out" students explicitly rejected the teacher's suggestion of linearity.) These students offered quantitative estimates of rates and calculations of duration consistent with a linear process during the second phase. The final phase is a gradual "settling in" of the temperature to its final state.

That students offered three clearly distinct phases, which were each explained differently, may be surprising. However, it is consistent with general principles of the Knowledge in Pieces point of view, particularly KiP1 (complexity) and KiP2 (grain size and structure). Indeed, previous research (diSessa, 1996; Sherin, 2001) noted, for example, that students often conceptualize the upward part of a toss quite differently from the downward part, despite the fact that they are treated uniformly, with a single explanatory structure across the two phases, in normative Newtonian mechanics. In general, then, finding a phased explanation did not surprise us.

5.1. Phase 1: Change takes time

The first phase seems to be guided by the *change takes time* p-prim[9] (diSessa, 1993, p. 133), which prescribes that effects often or always take some time to blossom fully. A good example of this phenomenon is that, when you put a car's accelerator down farther, it takes time for the car to reach its ultimate ("proper") speed. To a physicist, this is the expected pattern when a force is changed in a context of a viscous resistance—for example, when resistive force is proportional to velocity. In the case of an accelerating car, air resistance is the prominent viscous resistance. As is typical of p-prims, and just like *slowing equilibration*, the *change takes time* pattern is typically primitive. That is, it has no articulated explanation, unlike what a physicist can supply.[10]

P provided a clear explanation of the conditions under which the gradual increase in rate of temperature change occurs. At one point, where the class was discussing whether Fig. 3 or a simple linear graph would be more appropriate, P asked whether the glass of liquid was "already increasing in warmth" or not. If it was, he explained that the curve would not have the first phase, but proceed linearly and then "settle in" as prescribed by the second and third phases in Fig. 3. The teacher asked, "What if it were in the cooler all night?" P confirmed that in that case, starting from "rest," as it were, the full Fig. 3 pattern would be enacted. Incidentally, these students were articulately aware that their explanations were in terms of different phases, each with a different explanation. They used the term "stages." Both P and Ch talked about the possibility of "skipping stages" under certain circumstances.

There is nothing particularly unexpected in the students' apparent application of the *change takes time* p-prim to this context, with the possible exception that the context is not mechanical (change of position), but an abstract change, that of temperature. However, as noted earlier, many p-prims, including *Ohm's p-prim,* seem to have wide scope. More generally, as analysis will continue to show, this causal scheme seems to be a more straightforward and less innovative construction than freaking out.

5.2. Phase 2: Linear change of temperature with time

Although p-prims may appear to embody linearity, students' use of this idea has several aspects that rule out a pure p-prim interpretation. In short, students' use of linearity is too complex (e.g., applied in different stages, across time) and reflectively available to count as a p-prim. Most obviously, students have elaborate, extended, and explicit mathematical strategies to implement linearity (see later data). If they do not know a name like "linearity" for the phenomenon, still they recognize it and can speak fluently about it.

Instead of a pure p-prim interpretation, I can offer three non-exclusive explanations for its origins and status. First, students do sometimes and self-consciously take simple mathematical patterns as explanatory. This likely has a reflectively accessible history in school and involves distinctive and extended reasoning patterns; students know quite a number of different things to do, such as checking a hypothesized pattern, unlike p-prims per se. While reasoning about some mathematical patterns may build on intuitive proportional reasoning that children use in many situations—such as the proportionality of quantities like distance and time in uniform motion—the resulting knowledge system is a more complicated system than a p-prim. Van Dooren et al. (2005) document the extensive and extended use of linearity in mathematics learning in both appropriate and inappropriate circumstances. Levin (2012) provides an account of "strategy systems," which provide a better model for linearity, as described here, than p-prims.

The second explanation of students' invocation of linearity is that students implicate some unspecified process that creates temperature change and, by default, assume the process is uniform. This is a little more complicated than typical p-prims, but is not clearly of a wholly different kind. There is some data supporting this explanation of the appearance of linearity. Ch provided a suggestive description of the first two phases: "If you just take the glass of milk out of the refrigerator and set it on the table, it's going to start a little slow [the first phase], but then, *once the // whatever process gets going*, *it's going to have a consistency*, or whatever." [ATDP-04, 3A, 31] Italics are added to mark what I take to be a critical part of the explanation. The particular "consistent" process at issue, however, was never substantially elaborated by any student in this class. (But see Ch's anticipation of this process explanation, discussed in the next section.) Still, referring to a process that is subject to further inquiry, and also attributing a separate property of it, "going to have a consistency," seems beyond the encoding presumptions of p-prims.

The third explanation of the linear phase has more tenuous support from the present data corpus. The data that support this interpretation come mostly from occasions other than the period of time that the three-phase model was being developed. Hence it lacks proximity to my imputed use of those ideas. *Ohm's p-prim* prescribes a constant effect, given a constant effort. Again, one prototypical application, which results in a well-documented misconception, is that constant force results in constant speed. In several other exchanges, students implicated a kind of force or influence in the change of temperature. Here, D provides a clear invocation and explanation of the force.

[ATDP-04, 3A, 40]
D: The change in temperature is dictated by the forces that are causing it. Under a heat lamp, that would cause it to rise rapidly.

Teacher: So, it's your source of heat that determines how quickly it'll change?
D: Yeah.

D's conception is consistent with Wiser's (1995) source-recipient model, claimed to be the naïve theory of heat and temperature; this is the closest approach to that model that we could see in our data. Note that, whether or not one takes *Ohm's p-prim* as the source of the linearity relationship, per se, the whole model, specific to heating and cooling, is implicated, not just one p-prim. In the same way, *Ohm's p-prim* is part of the freaking out scheme, not all of it.

Note in passing that the source-recipient model is inconsistent with the freaking out scheme. With source-recipient, only one of the partners in an equilibration process (involving two objects in contact, or an object and its ambience) is taken to be responsible for the change; no relational parameter, such as a difference in temperatures, is involved. D's schematization can explain the first two phases of the present three-phase scheme (given the use of *change takes time* in the first phase), but it would seem to require some unknown ancillary principle to be extended into the third stage: Does the instigating "force" or its effect somehow just fade away toward the end of the equilibration process?

Although, of course, additional data might help decide which of these explanations for the linear model obtains, it seems likely that they all may provide part of the perceived sensibility of the model, or different students may be sensitive to different support for the same conclusion: there is a linear phase.

5.3. Phase 3: "Settling in"

The heritage of "settling in," the final phase of the three-phase model, is uncertain. One possibility, of course, is that it represents the pure, non-agentive *slowing equilibration*, which W and some of his colleagues invoked before the appearance of the causally interpolated version, freaking out. Why *equilibration* would follow a phase of linear effect, or, possibly, constant driving force (the third explanation of the second phase, above), is unclear. Indeed, no students in this class suggested *slowing equilibration* pattern for the whole course of temperature change, and at least one student explicitly rejected the normative graph as a possibility (data will be presented below). In contrast, the first two phases of the three-phase equilibration model constitute a natural sequence—"warming up" to a constant rate of change.

A second, remoter possibility for explaining the third phase is that it represents a *dying away* p-prim (diSessa, 1993, p. 133). *Dying away* is an abstraction of cases where an impetus creates an effect, which, however, diminishes gradually after the impetus is removed. Examples include the slowing to stop after launching a ball across the floor, or the fading sound of a struck bell. However, these examples suggest a mismatch between dying away and the present circumstances. *Dying away* applies when no maintaining impetus is present, and so would seem to implicate ceasing of the unnamed uniform process (or the driving force of a source of heat) that created the linear phase. Furthermore, the characteristic slowing of rate shown in the third phase does not appear, typically, in *dying away* (at least, it is not cited or documented in diSessa, 1993). If *dying away* explains the third phase here, it would seem to be applied in an anomalous context. And yet, students do no explicit work to prepare the context for *dying away*. For example, they might have explained that the "consistent" (uniform) process of the second phase "wore out" or stopped, for some reason. This lack of explicit preparation contrasts with the freaking out language that, I interpreted, served to invoke agency in an unusual context in Case Study I. A suggested generality, which I will not pursue, is that students simply do not concern themselves much with aligning parameters and causes across changes in phase.

A third possibility for the final phase is that some as-yet undocumented p-prim lies behind this phenomenon. Indeed, the only student comment on this phase, other than to note its existence, suggests either the existence of a p-prim, or possibly some empirical generalization related specifically to temperature change. Ch, continuing her explanation of the linear phase [ATDP-04, 3A, 31] (described above), says of the final "settling in" phase: "And then probably like, when you boil water, it always takes a little longer to get to the right temperature you want <*toward the end of the warm-up to boiling*>." Since it is very unlikely that Ch has actually observed a slowdown in speed of heating close to boiling (beyond the fact that the physics prescribes no slow-down, to a close approximation, it is also difficult to imagine how temperature rate of change can be "observed" with any reasonable precision in the first place), one may well suspect that a p-prim is being applied in a somewhat unusual context (where most other students would not apply it) and one where it leads to a normatively incorrect conclusion. Recall that, in the other case study, C claimed to remember, if vaguely, that temperature overshoots equilibrium before settling in; C imputed "memory" as the source of her idea, whereas it seems much more likely that it is simply a p-prim applied in somewhat unusual circumstances. Ch may be doing the same thing with regard to the third phase of equilibration.

5.4. Development and social spread of the three-phase model

It is possible to imagine that W substantially created the freaking out model more or less completely and on his own, granted some anticipation of elements in the explanations of other students, notably R. However, as mentioned, the three-phase model developed gradually in the course of a twenty-odd minute discussion among several of the students. Four students were substantially involved; an additional student entered the class late, and his participation was minimal. The discussion occurred in a round of sharing the graphs that students had produced in the second phase of the instructional plan. In the following, I recount, briefly, the chronology of development of the three-phase model. In doing so, I display more about the nature of pre-instruction elements involved, the nature of phased models, and also some of the social work that went into developing a shared model.

There were two minor anticipations of the three-phase scheme in the prior (first) instructional phase, open discussion of the warming glass of milk. First, Ch anticipated her process explanation. She said, "It's some kind of process, or whatever." When queried about what kind of process, she responded only "when the heat exchanges, or something like that." [ATDP-04, 3A, 15] D also seemed to initiate his conceptualization of heating as the result of an external force in this early discussion. He had mentioned the possible role of your hand in the cold-milk-on-table situation. When the instructor asked for a clarification, he talked about how your hand would have more effect than air, because it is solid and because "it's [an] applied force." [ATDP-04, 3A, 17]

At this point, the instructor proposed, for definiteness, to imagine warming beginning at 2 degrees Celsius, at 9:00, and the room temperature would be 22 degrees Celsius. After students were given a few minutes to draw their graph and prepare to explain it, each of the students, in turn, explained their production. P began with a strikingly mathematical model.

[ATDP-04, 3A, 24:25]
P: I just really thought that temperature would increase by // by like, by // would multiply itself by 2 each minute. <*shows his graph*> … It goes from 2 Celsius to 4 Celsius, to 8 Celsius, and in between 9:03 and 9:04, it goes to 22 Celsius.

This explanation reinforces the notion that students find mathematical regularities at least somewhat explanatory of naturally occurring patterns. Notice that this pattern is curved oppositely from the normative graph; it shows increasing rate of change as the temperature approaches its final temperature. On another occasion P would explicitly criticize the normative graph, slowing toward the end of heating/warming. P also seemed completely insensitive to the sharp change of rates expressed by his own pattern. Just before it reached its final, constant temperature of 22 degrees, it is increasing rapidly. The instructor asked why it would not go on to 32 degrees, and P repeated his claim it would just stop increasing at 22 degrees. The instructor responded, "Oh, so it would keep doubling until it reaches room temperature, and then it would just stop?" P responded, "Yes." P then extended his graph (originally a bar graph) with a flat, horizontal segment at the point where his doubling graph reached 22 degrees. P appeared to be comfortable with an instantaneous, discontinuous shift from a doubling phase to a constant phase. This is consistent with my suggestion that students may not be particularly concerned with continuity across phases.

The second student to present, D, showed a straight-line graph with points marked off at one-minute intervals. [ATDP-04, 3A, 25:43] He explained that he thought the temperature would go up about 5 degrees in 5 minutes, thus 1 degree in 1 minute. He calculated (after correcting an arithmetic mistake) that the heating (from 2 to 22 degrees) would take 20 minutes. Recall that D appeared to be the most articulate defender of the idea of thermal "force," and that rate of temperature change would be proportional (a la *Ohm's p-prim*) to the temperature of the source. So, these data are consistent with the explanation of the linear phase as resulting from a constant-force and *Ohm's p-prim* interpretation of temperature change.

In continuing, however, D explained that, "actually," temperature change would start off slower, and then increase. His explanation was almost incomprehensible, but he ended by saying that the first 5 minutes would be much slower than the rest. The instructor, in an attempt to clarify D's explanation, showed him a graph that included an initial phase of "warming up" before a linear phase, and asked him if, after the gradual start, the graph would go "straight up to room temperature and then stop." D said, "Yeah." Speculatively, one might imagine that linearity was D's first understanding of the situation. However, in reconsidering the system, a slowed start became a salient possibility. Such a pattern is typical of a Knowledge in Pieces point of view: Students have many ideas to apply, and in slightly different contexts or merely under repeat consideration, different elements or different combinations may be activated.

The third student, G, emphasized that she thought it would take a long time to reach room temperature, but she explained that it would go up about 6 degrees every 10 minutes. [ATDP-04, 3A, 30] The instructor queried "it kind o' goes up in a straight line?" and G responded, "Yeah." G's graph appeared to be linear.

The final presenter, Ch, had produced a linear graph. Her production of it is visible in the videotape, and she is seen marking a few points in a straight line (presumably iterating a constant temperature change), and then sketching line segments, in sequence, between the marked points. However, strikingly, her explanation included both D's *change takes time* phase, and a final "settling in" phase. The first phase had no precedent before D's just-prior invocation of it. Settling in was not anticipated in anything any student had said to that point. I already briefly presented Ch's three-phase rendering, but repeat it here, in full, for the reader's convenience.

[ATDP-04, 3A, 31:25]
Ch: If you just take the glass of milk out of the refrigerator and set it on the table, it's going to start a little slow <*the first phase*>, but then, once the // whatever process gets going, it's going to have a consistency, or whatever <*the linear phase*>. … And then probably like, when you boil water, it always takes a little longer to get to the right temperature you want <*"settling in," toward the end of the warm-up to boiling*>. So it's going to be slow here <*gestures toward first part of graph*> and slow here <*gestures toward end*>, but then it's going to be constant in here <*gestures toward the linear middle of her graph*>.

In response to the instructor's question about how "slow" is represented, Ch draws a segment of gradually decreasing slope, apparently showing the "settling in" segment (although this segment is freestanding, and not connected to the linear segment).

After these presentations, the instructor prepared a set of graphs to consider and vote on, mostly using the students' productions. He had a linear graph, the normative graph, the three-phase graph, and also he added a graph that overshoots the final temperature. In asking about the plausibility of the overshoot, D produced his "it would need an outside force" explanation to reject the possibility. No one contested D's denial of plausibility. Again, *overshooting equilibration* seems generally to be a weak intuition.

The teacher called for a vote on how heating occurs. All students voted for the three-phase graph, although D and P said that under certain circumstances, the first stage might be skipped. Ch admits some plausibility to a skipped stage, but she said that the temperature would have to be very high for that to occur. She added that, while the stage might be smaller, it would always be there.

Finally, the instructor summarized:

[ATDP-04, 3A, 47]
Teacher: So does everyone pretty much agree with the purple one <*three-phase graph*>? Does anyone not like the purple one? <*no evident response*> OK.

Apparently, all the students agreed on the three-phase model, and three of the five students present during this part of the discussion have been vocal in proposing and explaining the main aspects of it.

5.5. Review of the three-phase model

The first contribution of the analysis of the three-phase model is to make it clear that students may construct different models out of their intuitive resources. There is no reason to believe that the freaking out scheme and the three-phase scheme exhaust the possibilities. The three-phase scheme is, like the freaking out scheme, primarily a composition. However, in contrast to freaking out, the principle of composition is temporal chaining, rather than causal chaining. Temporal chaining, thus, should be added to our list of learning mechanisms.

Secondary contributions of this case include an enlarged ontology for the elements that are composed in the construction. In the freaking out case, elements seem mainly to be p-prims. In the three-phase case, mathematical regularities (linearity, or iterative multiplication by a constant), more articulate ideas (such as the "consistent process" mentioned by C), and thermal-specific empirical generalizations (temperature change slows down near a "final" state, such as approaching boiling) should be added to the list of possible elements used in composition.

6. Other explanations of temperature equilibration: Multiplicity and contextuality

Because the case studies were selected on the basis of successful construction of a causal scheme, they ignore the majority of ways our students construed thermal equilibration precisely because students could not complete a successful scheme with those as starting points. According to the Knowledge in Pieces perspective, one of the central difficulties of conceptual change is the hyper-richness (diSessa, 2007) of alternative ways of construing scientific phenomena. To make this point, I very briefly mention some of the more prominent issues concerning hyper-richness.

Molecular models: A good proportion of our students knew that heat concerned motion of molecules. Given that fact, it is notable that no causal schemes emerged from those ideas. Indeed, it often seemed that molecular models were more prone to non-normative interpretations, than progressing toward normative ones. With regard to difficulties with molecular models, see Linn, et al. (2010) and references therein. In our data corpus, for example, one student noted the fact that objects (molecules) are easier to slow down than to speed up, so that cooling was more natural and would proceed more quickly than heating up. Normatively, there is no such asymmetry.

Another student seemed to analogize heat with the behavior of a gas. He claimed that it was easy for heat to expand outward into the environment (cooling), but it was more difficult to get the heat "concentrated."

Contextuality: Students often considered that different things would happen in different situations. The three-phase model might skip phases if temperature differences were very high. Or, as mentioned, heating would not follow the same curve if an object started from a steady temperature over time, or, in contrast, if it were already heating up. Normatively, there is no such "momentum" effect.

One student contested the comparability of the lab experiment, a test tube of water immersed in an ambient water bath, with the initial case that was considered, a glass of cold milk sitting on table. Indeed, there is normative justification for this distinction. Convection—in which the relation between temperature difference and rate of temperature change is not linear—is a significant factor in the case of the glass of milk on the table, but less so for the test tube.

Overall, then, there is ample evidence of wide-ranging ways of construing thermal equilibration phenomena. In fact, there was very seldom agreement among *any* students on these scheme fragments, which is very likely partly why causal schemes did not emerge more frequently. These observations highlight the "lucky stroke" character of the freaking out case (that it involved mostly previously documented elements), and also make the observation that no student ever questioned or provided an alternative to ultimate equalizing of temperature rather more surprising.

7. Summary and final remarks

7.1. Highlights: Productivity of intuitive knowledge

This work entailed a microgenetic analysis of two student constructions of causal, explanatory schemes for temperature equilibration. Both were largely autonomous in that they occurred spontaneously while inquiring about simple thermal phenomena, the settling in to ambient temperature of an object that is initially hotter or colder than surroundings. They both occurred before any explicit instruction in the normative model, Newton's law of heating/cooling.

Both schemes seemed principally fairly straightforward compositions of a number of intuitive schemata. However, the compositional principles were different in the two cases. In the first, the composition was causal chaining: x causes y, y causes z, …. In the second case, the composition was temporal chaining: one kind of phase was followed by another, and then another. In the first case, central elements were previously described and empirically documented p-prims, such as *Ohm's p-prim* and one or another *equilibration* p-prim. The second case seemed less about p-prims. While the first phase in the three-phase model appeared clearly to match a previously documented p-prim (*change takes time*), the linear phase seems more complex in its origins, perhaps building on the idea that simple mathematical patterns are frequently explanatory (in school, at least), and perhaps (at least for some students) itself a composition of a constant thermal driving force (that of a hot object, for example) and thus (via *Ohm's p-prim*) a constant rate of change. The third phase might have been due to an undocumented p-prim, or to an empirical generalization about heating objects.

Within the literature on conceptual change, the first case study is especially notable as a meticulous documentation of the productive use of naïve schemata, achieving a good approximation of the normative Newton's law of thermal equilibration. The second analysis shows the construction of a "misconception" out of different, if somewhat similar, knowledge elements. Together, these studies reinforce a general orientation of Knowledge in Pieces, that both the construction of some non-normative and normative understandings can be understood as involving compositions of naïve knowledge elements (diSessa, 2008a).

In the development of the freaking out scheme, language had a prominent role. The main function for the freaking out language seems quite clear and specific. The core, relevant causal link, *Ohm's p-prim*, needs an agentive cause to be activated and used. As I emphasized, *equilibration*, which was the starting place for the scheme, is an intuitive causal schema markedly without agency. So, the terminology "freaking out" serves the function of expressing and communicating an unusual and critically innovative agentive stance toward a situation of equilibration. R's independently developed language, "shocked," which evolved from a more benign use of anthropomorphic language as the student struggled to understand equilibration, reinforces my interpretation of the functionality of the language.

7.2. Mechanisms

While much of this work extends and exemplifies claims made in other places, notably those concerning p-prims in diSessa (1993), arguably the most novel and potentially general of our results is developing, directly from data, a list of knowledge-level mechanisms of conceptual development. I list these here and briefly comment on their status.

1. **Composition**: The basic form of conceptual innovation in both of our case studies is the composition of other, mostly prior knowledge elements. There were two different structural forms. Both are linear (chains), although, in principle, nets or other structures are possible. In the first study, the composition took the form of a **causal chaining**, x causes y, causes z, .… In addition to p-prims, I noted transient anthropomorphic metaphors as constituting causal links. In the second case study, the principle of composition was **temporal chaining**, where more-or-less independent phases followed one another to explain an overall temporal pattern. This kind of composition seemed loose, and students did not carefully articulate reasons for, or processes during, transition from one phase to the next.
2. **Shift of context**: The "big story" of the freaking out model is that agency gets introduced in a felicitous way into a construal of a situation (equilibration) where, typically, agency is not implicated. It is worth emphasizing again that the central causal element, *Ohm's p-prim*, is the core of what I argued was essentially a normative construction. In contrast, in cases where *Ohm's p-prim* is used to relate force to another kind of result, speed of motion, its use results in "a misconception." Context of use can make a huge difference in the productivity of the same intuitive knowledge element.
3. **Binding**: Scientific explanation must connect to relevant features of the world. Intuitive schemes are often ambiguous as to what aspects they connect to. Students' choice of world features to bind to intuitive schemes can make a huge difference in the correctness of a causal scheme; students may ignore key, relevant aspects, or focus on completely irrelevant ones. In the case of the freaking out model, the controlling attribute for agency (or effort) stems from a difference in temperatures, which is precisely the normative cause.
4. **Causal interpolation**: In the class that developed the freaking out model, some students (both R and W) appeared to start building the freaking out model with a sense that a state of *abstract imbalance* leads, without any mediating structure or processes, to a return to balance (*equilibration*). The freaking out model interpolated a series of causes that, all together, explained *equilibration*.

5. **Causal elision**: After enriching non-agentive equilibration with a causal chain in producing the freaking out model, students began immediately to drop or at least background some aspects of it. In particular, the effort ("working harder") explicitly described on first mention of "freaking out" was immediately dropped, never to reappear. The last use of the freaking out scheme in this corpus connected difference of temperature directly to rate of change of temperature, with no anthropomorphic language, no evident agency, and no interpolating causal chain. The overall process seems an efficient use of prior knowledge. First, a starting interpretation is enriched with causal interpolation to provide "hooks" to connect to normative parameters, like the difference in temperature. Then, gradually, the ancillary intuitive attributes that are unnecessary from a scientific point of view, agency and effort, get dropped.
6. **Emergence**: "Emergence" is a general principle of teleology and local coherence that may drive the construction and stabilization of particular compositions and other innovations in the construction of causal schemes. In particular, the net effect of a composition in providing a satisfactory causal explanation of a particular phenomenon may help select relevant elements and "motivate" changes in the individual element and their particular composition. The mere fact that students converged on these schemes and rejected or left behind other attempts at explaining thermal equilibration indicates that they could and did exercise a global judgment on whether the explanations they created sounded right to them.

7.3. Broader considerations

In Section 2, I described the general view of conceptual change that is used here for analysis with a set of principles. I comment briefly on the principles' appearance in this data and analysis.

KiP1: Complexity of the naïve state: This principle was evident in many aspects of the data and analysis. For example, two different causal schemes were developed by two different classes, suggesting a degree of flexibility and generativity in underlying intuitive resources. Each scheme drew on particular, different elements of the underlying ecology. Many partial attempts at explanation, beyond the two causal scheme case studies, show that still more resources are available, whether they can be formulated into stable and socially-sharable schemes, or not.

KiP2: Productivity of intuitive knowledge: Because of its importance for instruction, the fact that students in Case Study I developed a near-normative model, approaching Newton's law of thermal equilibration, is especially notable. I made the point earlier that the literature is very sparse with studies that document in detail the general proposition that prior conceptions are useful in constructing scientific understanding. In this analysis, I documented not only that prior knowledge elements were useful, but *how* they came to be so (e.g., as components in a composed structure). This included some changes in the elements (e.g., their activation and binding) that were required for this use of them.

KiP3: Grain size and structure: The use of many particular sub-conceptual elements (p-prims and similar structures) in very particular combinations suggest both a general adequacy of the grain-size chosen for analysis of this data, and the rich space of possibilities for change during learning.

KiP4: Learning on a small time-scale: Case Study I took place almost entirely within a 1-hour class, and the critical events constituted only a small fraction of that time. Case

Study II described, primarily, what happened in a continuous 22-minute stretch of discussion. The data, however, are quite rich and I believe they strongly constrain interpretation, especially if one includes events that preceded and succeeded the focal constructions.

7.4. Qualifications and limitations

Beyond from previously mentioned limitations, those of case studies, which this study inherits, and the fact that we focus only on learning one particular topic, in one particular instructional frame, I want to underscore several more distinctive qualifications and limits.

7.4.1. Is this a matter of conceptual change?

Calibrating how much of what kind of learning has happened in a relatively short period of time is a difficult task. Because of very different theoretical framings, different researchers put different earmarks on true conceptual change. Carey (1991) uses extended argumentation to justify the label "conceptual change" for the development from childhood "biology" to adult common-sense biology. See, also, Chi (2005) for judgments on what true conceptual change means. Because a relatively standard base of problems (end-of-chapter problems) constitutes a good test of understanding schooled ideas such as the concept of force—and in addition, the research community has accumulated a decent set of diagnostic problems to show up problematic, intuitive conceptions in certain cases (see, for example, Hestenes, Wells, & Swackhamer, 1992)—one can set rough but useful bounds on how much a student has surpassed intuitive ideas and entered the normative scientific world, *in certain cases*. With these standards in mind, I would not claim that either of these case studies manifestly represents deep conceptual change.

Still, in the literature that shows what kind of learning can take place on the basis of limited-time interventions, these cases are clearly in a middle, if not high end of "amount of change." The Brown and Clement (1987; 1989) intervention, using bridging analogies, which I have referenced several times in this article as a comparison and rough standard, seems to show a comparable amount of change in a comparable period of time. In addition, as mentioned, I view the learning that Brown and Clement's interventions precipitated as structurally comparable to what happened in Case Study I: The primary innovation in Case Study I and in Brown and Clement's work is that students come to perceive agency in situations (a table pushing up; thermal equilibration) where they do not at the outset. The extent of conceptual composition and re-organization is also comparable to that hypothesized in the analysis of one student's construction of the "impetus misconception" (diSessa, 1996). A recent paper (Kapon & diSessa, 2012) provides extended analyses in which to compare Brown and Clement's theoretical perspective with the one here in terms of accountability to microgenetic data.

7.4.2. How much learning looks like this?

These two cases appear to be learning of a particular type: the construction of a relatively complete and stable scheme for a key scientific phenomenon. How much of school learning of science looks like that? Given the paucity of microgenetic analyses of learning, the question is difficult to answer. However, in view of the need for genuine conceptual change concerning heat and temperature that one might expect in a school course on the subject, it seems safe to assume that there are other, more extended modes of learning involved in unambiguous conceptual change. Extensive literature shows that an extended learning trajectory, involving many different contexts, is necessary to achieve conceptual change. See, for example, Levrini and diSessa (2008) and Wagner (2006). Both of these studies, however, involve a conceptual grain size consistent with that here, and the learning events that are identified are, grosso modo, similar to that which happened here.

7.4.3. How would this learning appear in competitor theories of learning and conceptual change?

It would not be reasonable, as part of this article, to develop interpretations of the learning exhibited in these cases within other perspectives on conceptual change, and then to pursue competitive argumentation. Developing such an interpretation accountable to the level of detail in data presented here is not an incidental task. However, one can make the following points, mainly in contrast to theory-theory or landmark models of conceptual change:

Grain-size. The grain size of conceptual elements used here (many relatively simple elements, limited, if malleable, range of invocation) seems highly adapted to the empirical data. Composition, and, indeed, gradual composition as demonstrated here (particularly in Case Study II, where the development of the three-phase model took place, in increments, over about 20 minutes of work) would seem to require at least as fine a grain size as that here. In contrast, the data may provide challenges for the assumption that intuitive knowledge comes in large and integrated chunks, "naïve theories." How can one handle the important intermediate states in both case studies, which are neither the normative theory nor the naïve theory? If, indeed, composition is a central process, how are the elements that are composed extracted and separated from the naïve theory, or where else might they come from? How can one understand the contextuality and diversity of student conceptions outside the focal case studies—the majority of ideas that students put forward, which did not result in a stable causal scheme?

Reasons for difficulty. In contrast to a common assumption in the conceptual change literature that a small set of ideas (e.g., a "naïve theory") explains difficulty in learning normative science, this study adds to literature on hyper-richness (diSessa, 2007) suggesting that the very abundance of naïve knowledge can be a problem. Of all the different ways that students have to begin considering thermal equilibration, only a very particular set of naïve elements, related in a particular way, with suitable bindings, can constitute a causal scheme that is also normative.

Specificity. The analyses here are not of the "generally speaking" sort. They involve highly specific elements and relations in composition. The properties of many of the particular elements were triangulated from previous study. I believe that this constitutes a high standard of specificity of analysis, approaching (but not reaching) the level of a computational model.

Domain-general elements. Historical study of conceptual change has emphasized the independence of domains (e.g., Hirshfeld & Gelman, 1994; Inagaki & Hatano, 2002; Wellman & Gelman, 1998). "Naïve theories" are almost universally considered domain specific (e.g., mechanics, thermodynamics, psychology, biology). Two aspects of this study run counter to domain specificity. First, the many previously documented elements were all extracted from student conceptions of mechanics. Thus, their application to thermal phenomena, as found here, is anomalous, in view assumptions of domain specificity. Similarly, one of the most visible elements of the analysis here is in invocation of psychological ways of thinking ("freaking out," "wanting to…"), which seemed directly on the path (facilitating the invocation of ideas more typically associated with agentive situations) to the normative view that Case Study I students reached.

Richness of data. Typical of microgenetic analysis, I tried to use all data that was accessible during the case studies at issue. Of course, there are subtleties and things cannot be explained by these accounts. In particular, the mechanisms are, indeed, mechanism sketches. However, process data like that presented here provide a stern test for any model of leaning that claims to be close to tracking moment-by-moment change.

Primitive and sophisticated forms of explanation. Finally, a theme in many conceptual change researchers' work is that directed, agentive causality is a primitive form of explanation, which universally impedes more expert understanding (Chi, 2005; Perkins & Grotzer, 2005). In view of those claims, there is a highly focused irony in the fact that, here, importing agency into a situation where none is initially seen, seems to provide the key insight that helped the class in Case Study I develop a normative causal scheme for thermal equilibration. Indeed, the class seemed to have little difficulty leaving agentive features behind (causal elision)—once they bootstrapped particular features (e.g., temperature difference) and relations into view—to produce more normative verbal accounts of thermal equilibration (temperature differences cause a proportional rate of temperature change). Even if direct, agentive causality is "bad," it seems easy to escape. Suppressing it a priori would have undermined a productive path to normative understanding.

7.4.4. Status of learning mechanisms

I take it to be obvious that there are contingencies in what happened here beyond the level of detail provided in our list of learning mechanisms. For example, how do we know when these mechanisms will be invoked, and when they will not? With respect to the "shift of context" mechanism, one could ask when an element or attribute may be, on the one hand, activated in a new context where it is seldom or never used before, and when, on the other hand, no one will think of using the element and might even reject its relevance if proposed. These details are beyond the current specification. Similarly, the details of when and how causal interpolation or elision may happen are unspecified, as are how bindings are made between world attributes (such as temperature difference) and intuitive attributes (such as agency).

One would think that at least some of these questions would require a more detailed, possibly neural model where, for example, plasticity of activation, or the detailed mechanisms underlying composition, might be represented explicitly.

As discussed previously, there are no grounds to argue for the completeness of the list of mechanisms. Indeed, one should expect more such to be uncovered in future work.

Given these limitations, I still propose that the present analysis makes good case for pursuing knowledge-level mechanisms. First, these mechanisms seem insightful of real-world learning and might be used at least heuristically to motivate and explain effective instructional possibilities. With respect to symbol-level or neural-level cognitive modeling, these mechanisms constitute goals for reductive modeling. Given a proposed low-level model or modeling language, can it perspicuously implement these mechanisms and, indeed, explain unexplained contingencies, such as those mentioned above? Finally, these mechanisms also challenge coarser-grained mechanisms of learning, such as equilibration or reflective abstraction (Piagetian or neo-Piagetian models), rational models of conceptual change (Posner et al., 1982), or some mechanisms proposed by developmental psychologists (compare Carey, 1999, 2009). Can those models effectively accommodate and possibly unify the specifics of the different mechanisms proposed here, and the contexts and particulars of their documented use?

Acknowledgements

This work was supported by grants (numbers 200500036 and 201100101) to the author from the Spencer Foundation. Their continuing support is gratefully acknowledged. I thank the Berkeley Academic Talent Program and the Urban School of San Francisco for hosting our experimental classes. Some of the analysis here was done in consultation with the Patterns Research Group, including Lauren Barth-Cohen, Jeanne Bamberger, Rozy Brar, Amy Bullis, Janet Casperson, Emily Chan, Steve Hassani, Patrick Lee, Colleen Lewis, and Sandhya Rao. The instruction was designed and data was collected by present and past members of the Group, including Janet Casperson, Lauren Barth-Cohen, Zach Powers, Karen Chang, Bradford Hill, and the author. Researchers who served as instructors at various times in the corpus of data analyzed here included Janet Casperson, Karen Chang, Bradford Hill, and Zach Powers. Helpful commentary on the content and form of this article was provided by the Patterns Research Group, Melinda diSessa, Douglas Clark, and Mariana Levin. Allan Collins provided excellent, focused editorial advice that resulted in an improved paper. Two anonymous reviewers, along with Allan Collins, provided very positive support in fleshing out the current project, and also excellent pointers for future work.

Notes

1. However, schemata similar to p-prims do seem to appear *implicitly* in language (Talmy, 1988), but apparently not as lexicon, per se, nor self-consciously. See discussion of Talmy's work in diSessa (1993).

2. In general, I use "we" or "our" to refer to my research group.

3. For science education researchers, what constitutes "causal" is an important and debated issue. One view of causality would insist that Newton's thermal law is at the macro level, and not causal; causality is seen only at the micro level of collisions. Counter arguments include: (1) Other examples of temperature equilibration do not need analysis at the micro level, such as equilibration by radiation. (2) Physicists regularly use this level and speak of it as causal. For example, a classic graduate textbook on mathematical physics, Morse & Feshbach (1953), describes the causality of diffusion (a more general class of phenomena, but one including temperature equilibration), with no necessary reference to micro models (p. 173). (3) When Newton formulated his law, the micro-level analysis of thermal equilibration was not possible. Similarly, if one excludes the marco level from the regime of causality, then Newtonian gravity also is not causal.

4. I do not know of published documentation of this fact. However, it is consistent with articulate lore among physicists, that putting oneself in a situation often provides insight (e.g., Einstein's famous thought experiment of "riding a light ray"). It is also consistent with my own experience listening to (and giving) freshman physics lectures. Several physicists I checked with informally confirmed this impression from their own experience. See also what appears to be a related kind of anthropomorphic projection in professional talk in Ochs, Gonzales, and Jacoby (1996).

5. Hatano and Inagaki's "two-phase" description of anthropomorphic reasoning (Inagaki & Hatano, 1991) is helpful, here. These authors view anthropomorphism as heuristic in instigating ways of thinking that are then corrected or modified by case-particular knowledge. In the present situation, it appears anthropomorphism initiates a productive line of thinking (toward the "freaking out" model), but then students spontaneously strip the overt anthropomorphism from later consideration.

6. There is a suggestive progression in R's attribution of agency. Her first use of agency (precursor 4) was ambiguous. Even though "it's <*the milk is*> trying to reach equilibrium," the causal consequent is more neutrally described, "it would just heat up really quickly." Here, the hot water is "shocked," indicating potential agentive status, but it turns out that the cold water is acting on (cooling) the hot water. Eventually, W's "freaking out" stabilized R's attribution of agency; the "shocked" agent causes its own equilibration. Thus, a sense of the relevance of agency gradually finds stable bindings in the chain of cause and effect.

7. Because the effect of resistance, the third part of *Ohm's p-prim*, is not highlighted in, or necessary to explain any of the phenomena discussed by the students, it is not surprising that it does not appear explicitly in their explanations. This lends support to the role of explanatory purpose, articulated below in section 4.3.4. At the end of this class, however, students were asked what things the rate of heating/cooling depended on, and they appropriately mentioned mass and density as mediating variables. These would, in fact, serve the role of "resistance" to the "thermal driving force."

8. "Trying" obviously begs further analysis. Coming to a consensus was never mentioned by the teacher and, in fact, was not expected at this stage of instruction. Instead, it probably, at best, emerged as a public goal as several students, for several reasons (including intuitive resonance with others' ideas along with affiliative instincts) began to align their own ways of thinking.

9. Ironically, this p-prim is referred to as *warming up* in diSessa (1993). That terminology would be confusing in application to the first phase only of equilibration, so I use here the alternate formulation from the original paper, *change takes time*.

10. Immanuel Kant (1960), in his *Critique of Pure Reason*, used this idea articulately and very generally in describing why people think effects come after causes, whereas he contended that effects are always simultaneous with causes. He said that, very often, the *full* effect of a cause takes time to develop.

References

Anderson, J. R., Boyle, C. F, Corbett, A. T, & Lewis, M. W. (1990). Cognitive modeling and intelligent tutoring. *Artificial Intelligence*, *42*, 7-49.

Brown, D. E., & Clement, J. (1987). Misconceptions concerning Newton's law of action and reaction: The underestimated importance of the Third Law. In J. D. Novak (Ed.), *Proceedings of the second international seminar: Misconceptions and educational strategies in science and mathematics* (Vol. 3, pp. 39-53). Ithaca, NY: Cornell University.

Brown, D. E., and Clement, J. (1989). Overcoming misconceptions via analogical reasoning: Abstract transfer versus explanatory model construction. *Instructional Science*, *18*(4), 237-261.

Carey, S. (1991). Knowledge acquisition: Enrichment or conceptual change? In S. Carey & R. Gelman (Eds.), *The epigenesis of mind* (pp. 257-291). Mahwah, NJ: Lawrence Erlbaum Associates.

Carey, S. (1999). Sources of conceptual change. In E. K. Scholnick, K. Nelson, S. A. Gelman, & P. H. Miller (Eds.), *Conceptual development: Piaget's legacy*. Mahwah, NJ: Lawrence Erlbaum Associates.

Carey, S. (2009). The origins of concepts. Oxford, UK: Oxford University Press.

Chi, M. T. H. (1992). Conceptual change across ontological categories: Examples from learning and discovery in science. In F. Giere (Ed.), *Cognitive models of science: Minnesota studies in the philosophy of science* (pp. 129-160). Minneapolis, MN: University of Minnesota Press.

Chi, M. T. H. (2005). Commonsense conceptions of emergent processes: Why some misconceptions are robust. *Journal of the Learning Sciences*, *14*(2), 161–199

Clark, A., & Toribio, J. (1998). *Cognitive architectures in artificial intelligence: The evolution of research programs*. New York, NY & London, UK: Garland Publications.

Clark, D. B. (2006). Longitudinal conceptual change in students' understanding of thermal equilibrium: An examination of the process of conceptual restructuring. *Cognition and Instruction*, *24*(4), 467–563.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in education research. *The Educational Researcher*, *32*(1), 9-13.

Cobb, P. (1999). Individual and collective mathematical learning: The case of statistical data analysis. *Mathematical Thinking and Learning, 1*, 5-44.

Cobb, P., McClain, K., & Gravemeijer, K. (2003). Learning about statistical covariation. *Cognition and Instruction, 21*(1), 1-78.

Confrey, J. (2006). The evolution of design studies as methodology. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 135-151). Cambridge, UK: Cambridge University Pres.

Confrey, J., & Lachance, A. (2000). Transformative reading experiments through conjecture-driven research design. In A. E. Kelly & R. A. Lesh (Eds.), Hand*book of research design in mathematics and science education* (pp. 231–266). Mahwah, NJ: Lawrence Erlbaum Associates.

Dennett, D. (1993). Book review: Allen Newell, Unified Theories of Cognition, *Artificial Intelligence*, *59*, 285-294.

Design-based Research Collective (2003). Design-based research: An emerging paradigm for educational inquiry. *The Educational Researcher*, *32*(1), 5-8.

diSessa, A. A. (1983). Phenomenology and the evolution of intuition. In D. Gentner and A. Stevens (Eds.), *Mental Models* (pp. 15-33). Mahwah, NJ: Erlbaum Associates.

diSessa, A. A. (1988). Knowledge in pieces. In G. Forman and P. Pufall (Eds.), *Constructivism in the Computer Age.* Hillsdale, NJ: Lawrence Erlbaum, 49-70.

diSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and Instruction*, *10*(2-3), 105-225.

diSessa, A. A. (1991). Local sciences: Viewing the design of human-computer systems as cognitive science. In J. M. Carroll (Ed.), *Designing Interaction: Psychology at the Human-Computer Interface* (pp. 162-202). New York, NY: Cambridge University Press.

diSessa, A. A. (1996). What do “just plain folk” know about physics? In D. R. Olson & N. Torrance (Eds.), *Handbook of education and human development: New models of learning, teaching, and schooling* (pp. 709–730). Oxford, UK: Blackwell.

diSessa, A. A. (2004a). Contextuality and coordination in conceptual change. In E. Redish and M. Vicentini (Eds.), *Proceedings of the International School of Physics “Enrico Fermi:” Research on physics education* (pp. 137-156). Amsterdam, NL: ISO Press/Italian Physics Society.

diSessa, A. A. (2004b). How should we go about attributing knowledge to students? In E. Redish and M. Vicentini (Eds.), *Proceedings of the International School of Physics “Enrico Fermi”: Research on physics education* (pp. 117-135). Amsterdam, NL: ISO Press/Italian Physics Society.

diSessa, A. A. (2006). A history of conceptual change research: Threads and fault lines. In K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 265-281). Cambridge, UK: Cambridge University Press.

diSessa, A. A. (2007). Systemics of learning for a revised pedagogical agenda. In R. Lesh, E Hamilton, & J. Kaput (eds.), *Foundations for the future in mathematics education* (pp. 245-261). Mahwah, NJ: Lawrence Erlbaum Associates.

diSessa, A. A. (2008a). A bird's eye view of "pieces" vs. "coherence" controversy. In S. Vosniadou (Ed.), *Handbook of conceptual change research* (pp. 35-60). Mahwah, NJ: Lawrence Erlbaum Associates.

diSessa, A. A. (2008b). Can students re-invent fundamental scientific principles?: Evaluating the promise of new-media literacies. In T. Willoughby, & E. Wood (Eds.), *Children's learning in a digital world* (pp. 218-248). Oxford, UK: Blackwell Publishing.

diSessa, A. A., & Cobb, P. (2004). Ontological innovation and the role of theory in design experiments. *Journal of the Learning Sc*iences, *13*(1), 77-103.

diSessa, A. A. , Hammer, D., Sherin, B. & Kolpakowski, T. (1991). Inventing graphing: Meta-representational expertise in children. *Journal of Mathematical Behavior*, *10*(2), 117-160.

diSessa, A. A. & Sherin, B. (1998). What changes in conceptual change? *International Journal of Science Education*, *20*(10), 1155-1191.

Forbus, K. D. (1984). Qualitative process theory. *Artificial Intelligence, 24,* 85–168.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, *30*, 141-158

Hirschfeld, L. A., & Gelman, S. A. (Eds.) (1994). *Mapping the mind: Domain specificity in cognition and culture*. Cambridge, UK: Cambridge University Press.

Hume, David (1995). An enquiry concerning human understanding. In Steven M. Cahn (Ed.), *Classics of Western philosophy* (pp. 613-699). Indianapolis, IN: Hackett Publishing Co.

Inagaki, K, & Hatano, G. (1991). Constrained person analogy in young children's biological inference. *Cognitive Development*, *6*, 219-231.

Inagaki, K, & Hatano, G. (2002). *Young children's naïve thinking about the biological world*. New York, NY: Psychology Press.

Izsák, A. (2005). "You have to count the squares": Applying knowledge in pieces to learning rectangular area. *Journal of the Learning Sciences, 14*(3), 361–403.

Jacobson, M. & Wilensky, U. (2006). Complex systems in education: Scientific and educational importance and implications for the learning sciences. *Journal of the Learning Sciences*, *15*(1), pp. 11-34.

Kant, I. (1960). Critique of pure reason (abridged). In Monroe C. Beardsley (Ed.), *The European philosophers from Descartes to Nietzsche* (pp. 370-469). New York, NY: Modern Library.

Kapon, S., & diSessa, A. A. (2012). Reasoning through instructional analogies. *Cognition and Instruction*, *30*(3), 261-310.

Lehrer, R., Schauble, L., Carpenter, S., & Penner, D. E. (2000). The inter-related development of inscriptions and conceptual understanding. In P. Cobb, E. Yackel, & K. McClain (Eds.), *Symbolizing and communicating in mathematics classrooms: Perspectives on discourse, tools, and instructional design* (pp. 325-360). Mahwah, NJ: Lawrence Erlbaum Associates.

Levin (Campbell), M. E. (2012). *Modeling the co-development of strategic and conceptual knowledge during mathematical problem solving*. (Unpublished doctoral dissertation). University of California, Berkeley.

Levrini, O., & diSessa, A. A. (2008). How students learn from multiple contexts and definitions: Proper time as a coordination class. *Physical Review Special Topics: Physics Education Research 4*, 010107.

Linn, M. C. (2006). The knowledge integration perspective on learning and instruction. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 242–264). Cambridge, UK: Cambridge University Press.

Linn, M. C., Chang, H.-Y., Chiu, J., Zhang, H., & McElhaney, K. (2010). Can desirable difficulties overcome deceptive clarity in scientific visualizations? In A. Benjamin (Ed.), *Successful remembering and successful forgetting: a Festschrift in honor of Robert A. Bjork* (pp. 239-262). New York, NY: Routledge.

Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science*, *67*, 1-25.

Minstrell, J. (1982). Explaining the “at rest" condition of an object. *The Physics Teacher*, *20*, 10-14.

Minstrell, J. (1989). Teaching science for understanding. In L.B. Resnick & L. Klopfer (Eds.), *Toward the thinking curriculum* (pp. 133-149). Alexandria, VA: Association of Supervision and Curriculum Development.

Morse, P. & Feshbach, H. (1953). *Methods of Theoretical Physics: Part I*. New York, NY: McGraw-Hill.

Nemirovsky, R. (1994). On ways of symbolizing: The case of Laura and velocity sign. *Journal of Mathematical Behavior*, *13*, 389–422.

Norman, D. (1991). Approaches to the study of intelligence, *Artificial Intelligence*, *47*, 327-346.

Ochs, E., Gonzales, P., & Jacoby, S. (1996). When I come down I’m in a domain state: Talk, gesture, and graphic representation in the interpretive activity of physicists. In E. Ochs, E. Schegloff, & S. Thompson (Eds.), *Interaction and grammar* (pp. 328-369). Cambridge, UK: Cambridge University Press.

Parnafes, O. (2007). What does "fast" mean? Understanding the physical world through computational representations. *Journal of the Learning Sciences*, *16*, 415-450.

Parnafes, O., & diSessa, A. A. (2013). Microgenetic learning analysis: A methodology for studying knowledge in transition. *Human Development*, *56*, 5-37.

Perkins, D.N., & Grotzer, T.A. (2005). Dimensions of causal understanding: The role of complex causal models in students' understanding of science. *Studies in Science Education, 41,* 117-166.

Piaget, J. (1930). The child's conception of causality (tr. M. Gabain). London, UK: Kegan Paul.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education, 66*(2), 211-227.

Rhodes, M., & Wellman, H. (2013). Constructing a new theory from old ideas and new evidence. *Cognitive Science 37*(3), 592–604.

Schoenfeld, A. H., Smith, J. P., & Arcavi, A. (1993). Learning: The microgenetic analysis of one student's evolving understanding of a complex subject matter domain. In R. Glaser (Ed.), *Advances in Instructional Psychology* (pp. 55-175). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sherin, B. (2001). How students understand physics equations. *Cognition and Instruction*, *19*(4), 479-541.

Siegler, R. (2006). Microgenetic studies of learning. In D. Kuhn & R. Siegler (Eds.), (W. Damon & R. Lerner, Series Eds.), *Handbook of Child Psychology: Vol. 2. Cognition, Perception, and Language* (6th edition). Hoboken, NJ: Wiley.

Smith, J. P., diSessa, A. A., & Roschelle, J. (1993). Misconceptions reconceived: A Constructivist analysis of knowledge in transition. *Journal of the Learning Sciences*, ***3***(2), 115-163.

Steffe, L. (2011). A theory bite on learning through mathematical activity. *Cognition and Instruction*, *29*(2), 256-263.

Talmy, L. (1988). Force dynamics in language. *Cognitive Science*, *12*, 49-100.

Thagard, P. (2000). *Coherence in thought and action*. Cambridge, MA: MIT Press.

Van Dooren, W., De Bock, D., Hessels, A., Janssens, D., & Verschaffel, L. (2005). Not everything is proportional: Effects of age and problem type on propensities for overgeneralization. *Cognition and Instruction, 23*(1), 57-86.

VanLehn, K. (1991). Rule acquisition events in the discovery of problem-solving strategies. *Cognitive Science*, *15*, 1-47.

Viennot, L. (1979). *La raisonnement spontané en dynamique elementaire*. Paris: Hermann.

Vosniadou, S. (2013). *Handbook of conceptual change research* (2nd ed.). New York, New York: Routledge.

Vosniadou, S., & Brewer, W. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology, 24*, 535–585.

Wagner, J. F. (2006). Transfer in pieces. *Cognition and Instruction*, *24*(1), 1–71.

Wellman, H. M., & Gelman, S. A. (1998). Knowledge acquisition in foundational domains. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology* (pp. 523 – 573). New York, NY: John Wiley & Sons.

Wiser, M. (1995). Use of history of science to understand and remedy students' misconceptions about heat and temperature. In D. Perkins, J. Schwartz, M. West, and M. Wiske (Eds.), *Software goes to school: Teaching for understanding with new technologies* (pp. 23-38). New York, NY; Oxford, UK: Oxford University Press.

Wiser, M., & Smith, C. (2008). Learning and teaching about matter in grades K-8: When should the atomic-molecular theory be introduced? In S. Vosniadou (Ed.), *Handbook of conceptual change research* (pp. 205-239). Mahwah, NJ: Lawrence Erlbaum Associates.

Yin, R. K. (2009). *Case study research: Design and methods* (4th edition). Los Angeles, CA: Sage.

Appendix: The role of language in case study I

The role of language is very much evident in Case Study I. One might, in fact, try to schematize the kinds of work language does as mechanisms. On the other hand, the analyses of language presented here were ad hoc and certainly less well-prepared theoretically than other aspects of the analysis. Nonetheless, this appendix briefly points to two contributions that language seemed to make to the development of the freaking out causal scheme.

Meaning shift from categorical to continuous: I argued that freaking out is typically interpreted as a dichotomous state change, from "normal" to a highly agentive state. However, the central causal relation, *Ohm's p-prim* requires a continuous interpretation of level of "effort." It appears that this transition is easy to make, and students and teacher quickly came to speak in continuous terms, for example, speaking of "degree of freaking out," and even introducing a number line to express the relevance of interpreting difference in temperature continuously. They tellingly did not establish a threshold of temperature difference at which point freaking out occurs. In W's first description of freaking out, what he says after line 2 is consistent with a continuous driving force and result, and less so with a discrete change or threshold. In prior work on student representations of motion (diSessa et al., 1991) our group noted that students often represented changes of motion as discontinuous, but they seemed conceptually fluent in interpolating to continuous change, if necessary. It may be that this shift in meaning from categorical to continuous reflects an underlying conceptual competence, and nothing particularly linguistic, even if ready lexicon, “degree of …,” smoothes the transition.

Stabilizing and communicating function: I treated the use of “freaking out” as a language issue, distinct from, for example, invoking a particular p-prim or another intuitive schema. However, the function of the term appears to be to introduce agency forcefully into a situation that, typically, would not evoke such an attribution. It seems likely that the use of the term (or R’s “shocked”) might derive from the feeling that the attribute of agency might work here explanatorily (see the discussion of emergence, in the text). After the initial invocation, remembering and using this language may well stabilize a student’s own conceptualization, re-introducing agency where the situation might not re-evoke it on a return to consideration. Socially, the freaking out language seemed clearly to convey the presence of agency to other students.

While (1) providing smooth transitions from one way of thinking (categorical view) to related one (continuous view) (2) and invoking and stabilizing relevant ideas (agency) are not at all surprising or controversial functions for language, they are still worth noting for the important role they seem to play in Case Study I.