Version Revised by Slotta in 3/99

How physics novices can overcome robust misconceptions through ontology training

James D. Slotta
slotta@socrates.berkeley.edu
http://kie.berkeley.edu/people/slotta
Graduate School of Education
University of California at Berkeley

Michelene T. H. Chi
Chi@vms.cis.pitt.edu
http://www.pitt.edu/~chi
The Learning Research and Development Center
University of Pittsburgh

This research was supported by the Mellon Foundation. Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors, and do not necessarily reflect the views of the Mellon Foundation. We are grateful for the help from Stephanie Siler in coding some of the protocols. Reprints may be obtained from the WEB or requested from Micki Chi, Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA 15260.

Introduction

Towards a Cognitive Theory of Instruction

A student's naive science knowledge is often very different from the concepts she learns in science class. Concepts such as force, light, heat, and electricity are among the most difficult to instruct, because students' naive views of these concepts are well-established and quite distinct from the conventional scientific views offered by instructors. For decades, cognitive and science education research have examined the science knowledge of novices and experts in a widespread effort to identify and characterize preconceptions of various science concepts. Many of the earliest studies of naive science conceptions (e.g., King, 1961; Kuethe, 1963; Doran, 1972; Viennot, 1979; Minstrell, 1982; Shipstone, 1984) sought to document the existence of firmly held preconceptions or misconceptions. (The term “misconceptions” will refer to preconceptions that are robust.) Not surprisingly, a consensus emerged that "young children do have firmly held views about many science topics prior to being taught science at school" (Osborne and Wittrock, 1983, p. 489). This simple, but important, statement is reflected by nearly 2000 published studies of students’ misconceptions and instructional attempts at their removal (Pfundt & Duit, 1988).

With the realization that these misconceptions could be partly responsible for the difficulty experienced by students in many science domains, research began to focus on trying to find instructional approaches that took students' misconceptions into account. One basic approach involves building upon students’ misconceptions, assuming that they are the beginning point of some gradual progression towards accepted scientific conceptions (e.g., Clement, 1987; Joshua and Dupin, 1987). These efforts were based on a cognitive theory that treated misconceptions as discreet (and perhaps independent) elements (e.g., diSessa, 1988, 1994), and conceived of conceptual change either as a gradual transition from a naive conception to a scientific one, through successive revisions, or as wholesale replacement (i.e., in the sense of being "over-written"). Thus, early instructional approaches either built upon existing misconceptions or simply "targeted" misconceptions by trying to design instruction that exposed the flaws in naive models, perhaps by leading students through experiments where their models predicted contradictory results to those observed (e.g., Champagne, Klopfer, and Gunstone, 1982; Tasker and Osborne, 1985). As a result of the limited view of conceptual change embraced by such instructional approaches, the "progressive construction" approach and the "cognitive conflict" approach obtained minimal success, with negligible transfer results.

Another instructional approach adopted a more holistic view of conceptual knowledge, in which concepts are seen not as discreet associated elements, but as firmly embedded or even distributed within a broader coherent framework (McCloskey, 1983; Carey , 1985; Keil, 1989). In this view, a discreet concept (e.g., a "mis-conception") is difficult to identify or isolate, much less remove. These complex and mostly ill-defined conceptual frameworks have often been referred to as "explanatory frameworks" (Ausubel, 1968; Driver and Easley, 1978) or "theories" (e.g., McCloskey, 1983; Keil, 1989; Carey, 1985) because of their intrinsic coherence and domain specificity. These researchers generally distinguish between gradual conceptual change, which constitutes ordinary conceptual development, and so-called radical conceptual change, in which a theory is somehow fundamentally unable to handle new phenomena, and must be completely replaced or restructured (e.g., Vosniadou & Brewer, 1992, who propose the two processes of "tuning" vs. "restructuring"; Carey proposes "weak" vs. "strong conceptual change"). However, this theoretical approach has not yet specified clearly what constitutes “restructuring” or “strong conceptual change”, nor demonstrated empirically the processes of radical conceptual change. Instead, empirical demonstrations have focused largely on concepts that only require changes that occur within a framework-- that is, of gradual conceptual change.

The term "theory" is consistent with another popular view of knowledge development, in which the learner was compared with an empiricist or "scientist" who possesses a preferred "theory" of objects, experiences or perceptions, but is aware of possible contradictions to the theory, and slowly constructs, revises, or replaces the theory as new phenomena contradict beliefs or suggest more inclusive principles. Posner, Strike, Hewson, and Gertzog (1982) provided an early statement of this "student as scientist metaphor" (a term offered by Driver and Erickson, 1983) in which they talked about novel experiences as being either "assimilated" or "accommodated" (borrowing the terms apologetically from Piaget). Assimilation in this case occurs when a learner's existing theory allows her to understand or explain a novel phenomenon; accommodation occurs when the theory is unable to account for the new phenomenon, and the student must revise, reorganize, or replace the theory. (See Chi, in press-a, for an example of “accommodation”.) However, such a “theory” view sidesteps the critical issue of why anomalous data are often ignored and/or rejected by students or scientists even though they conflict with the students’ strongly held views (Chinn & Brewer, 1993)**.** The theory view is also mostly silent on how theory-like knowledge is organized or structured, and has failed to provide a tangible and empirical assessment of learning that involves radical conceptual change.

Thus, the absence of any explicit theory of conceptual change contributes to the lack of progress towards a cognitive theory of instruction. Without such a theory, educational research is forced into a strategy of simply trying different instructional approaches, in contrast to predicting a specific approach based on some theory of conceptual change. Only through such a theoretically motivated approach can cognitive science inform the instruction of these notoriously difficult science topics, as forecast by Driver and Erickson (1983) in the following passage:

> With a growing interest in this field of inquiry we see a danger in a proliferation of 'natural history' studies of student ideas (in individual or group situations) being documented in the absence of any systematic rationale. Although case studies of this kind have a value in developing an awareness of the complex issues involved in classroom learning, we suggest that the most useful information will be obtained from studies of instructional programmes which are undertaken from an explicit theoretical perspective. In this way theories and observations can inform one another in a developing programme (Driver and Erickson, 1983, p. 41).

Clearly, a first step towards an instructional theory is to formulate a theoretical framework of conceptual change that affords (1) operational definitions, (2) instructional guidelines, and (3) assessment methodologies. Finding a method to assess conceptual change without a theory has been a difficult problem because in a domain such as physics, it is entirely possible that a student may learn the "recipe" for correctly solving certain types of problems (e.g., Newton's second law problems), but retain misconceptions of the basic underlying concepts (e.g., as described by Chi & VanLehn, 1991, for the concept of weight). Thus, it is insufficient and perhaps inaccurate to claim that improved problem solving scores can be taken as a measure of conceptual change (i.e., the extent to which student misconceptions are removed). Successful research requires a clear definition of conceptual change, and a corresponding methodology of assessing whether and to what extent conceptual change has actually occurred.

The present research builds on recent progress towards an understanding of naive science conceptions, and a resulting theory of conceptual change (Chi, 1992) to offer a new instructional approach that targets and assesses conceptual change. After briefly discussing the nature of students’ misconceptions and outlining the theory of conceptual change, we present research involving a novel approach to science instruction, and the accompanying assessment of conceptual change. Our results provide further support for the theory.

The nature of students' misconceptions

Chi (1992) has observed that some misconceptions are easily removed in the course of instruction, while others are characteristically robust, meaning that they survive even when directly confronted by instruction. She hypothesized that these misconceptions are robust because they require a process of conceptual change that is unnecessary in learning other concepts. That is, students undoubtedly enter the science classroom with preconceptions of many concepts, but only encounter serious difficulties in learning those concepts that require profound qualitative change. Thus, a student may have flawed notions of the concept, "molecule", but will not experience great difficulty in learning the accepted view of this concept if her initial ideas are qualitatively related to the target ideas (e.g., as long as she thinks of a molecule as "some tiny piece of matter"). In contrast, "robust misconceptions" are typical of many science concepts (e.g., light, heat, force, and electricity) where students' initial conceptions are often very well established (from everyday use) and yet very different from those of the science instructor. This class of misconceptions may be largely responsible for the difficulty experienced by both students and teachers in the physics classroom, as students are unwilling or unable to undergo the required process of conceptual change. For example, more than half of the university engineering students in one study, all of whom had completed a semester of university physics, were still plagued by misconceptions of the basic concept of force (McCloskey, 1983).

For good reason, research on student misconceptions has focused on those concepts that give students the most trouble. In a broad review of such studies, Reiner, Slotta, Chi, and Resnick (in press) found that students often attribute difficult concepts with materialistic properties. They scrutinized the literature for details of naive conceptions of force, light, heat, and electricity (all of which are notoriously difficult concepts), and found clear similarities across concepts. Reiner et al., (in press) observed that robust misconceptions tend to be of a form similar to material substances. That is, physics novices tend to think of these concepts as if they are material substances, or have certain properties of material substances. This conclusion was based on the arguments of the various authors reviewed, as well as on inferences from particular attributions in the misconceptions they reported. Thus, if a novice reasoned that a moving object slows down because it has "used up all its force," this was taken as evidence of a *substance-based* conception of force (in contrast to the more conventional view of force as a *process* of interaction between two or more objects). Similarly, conclusions about naive conceptions of heat were drawn from reasoning that involved heat (or cold) being "blocked" or "trapped," which suggests a substance-like property of heat.

Many authors of the studies cited by Reiner et al. (in press) either offered anecdotal evidence for the materialistic view underlying novices' conception or else they made such claims in their arguments. The contribution offered by the Reiner et al. review was twofold: First, it established a pattern across concepts, which suggests a common mechanism or disposition behind the formation of naive conceptions; second, it demonstrated a method of analyzing the nature of naive conceptions that involves drawing inferences from specific attributions (e.g., if something can be blocked by a wall, then it must have at least some properties of a material substance). The first contribution is important theoretically, because it offers a common thread amongst all the physics concepts that are notoriously prone to robust misconceptions. Any theory of conceptual change should be able to account for this important feature of naive conceptions. The second contribution of the Reiner et al. review is important methodologically, because it suggests a means of assessing true conceptual change: If, after instruction, naive reasoning about these concepts is examined in terms of such attributions, and no evidence of substance-like properties can be found, then the substance-like view of the concept must have been changed or replaced in the course of instruction; if not, then conceptual change has yet to occur. Slotta, Chi, and Joram (1995) developed such a methodology for assessing conceptual change, which will be discussed below.

A theory of conceptual change

By applying the idea of ontological boundaries (Keil, 1981) to account for observed patterns in science misconceptions, Chi (1992; 1993; 1997; in press-b) and her colleagues (Chi & Slotta, 1993; Chi, Slotta & deLeeuw, 1994; Ferrari & Chi, in press; Slotta & Chi, 1996; Slotta et al., 1995) have advanced a theoretical framework that accounts for why certain concepts are more prone to "robust" misconceptions than others, and why these persistent misconceptions are often substance-like in nature. The theory begins with the assumption that people associate all concepts with distinct ontologies, such as *processes, ideas,* and *material substances*[1], to name a few. When a new concept is learned, it is associated with some ontology, which helps the learner understand what kind of concept it is, and what ontological attributes it may possess. Thus, in learning about a new concept such as "osmosis", a person may attribute the concept with a *process* ontology,[2] which implies such attributes as "occurs over time", etc. Misconceptions arise, however, when a person associates the wrong ontology with a newly learned concept. In learning about the concept of "heat", for example, many children wrongly assume a *material substance* ontology, perhaps because of language conventions such as "close the door, you're letting all the heat out." In fact, the concept of heat is more appropriately associated with a *process* ontology, as it is best thought of in terms of the transfer of molecular kinetic energy. Once an ontological commitment is made with respect to a concept, it is difficult for this to be undone. We rarely, through any stages of mental transformation, come to change our conception of something from a *substance* to a *process*.

Specifically, Chi has proposed that a particular ontological class of science concepts, which will be referred to here as an *equilibration process* (i.e., those processes involving *complex dynamic systems* [3]), is a sub-class of the broader *process* ontology, and pertains to concepts that typically involve constraints such as the equilibration of certain system properties (e.g., inside and outside temperature; voltages; air pressures; etc.), which are often difficult for a physics novice to perceive. This class of concepts are characteristically mistaken by novices as possessing the ontology of *material substances*. When Slotta, et al. (1995) asked physics experts to explain the same conceptual problems that were given to the novices (problems concerned with topics of light, heat and electricity), their explanations of these problems were consistent with an *equilibration process* ontology (specifically, that of *complex dynamic systems*), and not a *material substance* one. These results suggest that there are indeed ontological differences between the conceptions of novices and experts with regards to their knowledge of certain physics concepts.

Chi’s account has implications for the assessment of conceptual change. First, any determination of a physics novice's misconceptions must include the ontological nature of those misconceptions. Without knowing this initial state of the knowledge, we would not be in a position to determine whether conceptual change has actually occurred in the classroom or laboratory. It will be important, therefore, to develop reliable assessment tools which are sensitive to the ontological associations of naive conceptions. Similarly, it will be important to measure the ontological status of the student's conceptualization when she has completed any relevant physics

---

[1] Throughout the paper, ontological categories and attributes will be italicized.

2 We are not addressing the issue of how such ontological associations are actually formed.

[3] Note that in developing this conceptual change theory, Chi had used various terms to refer to this class of ontological concepts. Previous names, such as events (Chi, 1992), acausal interactions (Chi & Slotta, 1993), constraint-based interactions (Chi, 1997), were misleading and do not convey the full meaning of this class.

instruction. Under Chi's proposed framework, only by contrasting the initial and final ontologies of a student's conceptions can we make valid claims about conceptual change. By performing a protocol analysis of the language used by novices and experts when they explained physics problems, Slotta, et al. (1995) were able to draw inferences about their differing ontological commitments. As these expert-novice differences should correspond to the shift in ontology that marks conceptual change, such an assessment tool could be used to confirm that a student's ontological commitments have indeed changed in the course of instruction.

By identifying a specific ontological boundary between two existing explanatory frameworks, and relating it to real phenomena of misconceptions and conceptual change, Chi’s theory has contributed a new level of explicitness to the discussion of explanatory frameworks. This level of explicitness affords an account of empirical patterns within the misconceptions literature (Chi, 1992; Chi et al., 1994), and also provides testable predictions. The present research thus applies Chi's theoretical framework within an instructional manipulation (in keeping with the spirit of the Driver and Erickson quote above), but also tests a strong prediction of the framework, providing it some important empirical support. The main theoretical idea supported by this research is that conceptual change need not be a process of continuous transition; indeed it cannot be in certain cases, according to Chi's account, and the consensual definition of radical conceptual change (e.g., Carey, 1985; Keil, 1987; Thagard, 1989; Posner et al, 1982). Thus, certain science misconceptions need not be addressed, targeted or "removed" as a prerequisite to conceptual change, and instruction need not build bridges to or from them.

The theory affords some predictions about how instruction can best proceed in addressing persistent misconceptions, and it is upon one such prediction that the present research is founded.

Assessing conceptual change

Slotta et al. (1995) performed a direct test of the hypothesis that physics novices prefer *substance*-like conceptions, while experts prefer *process* ones. When they asked physics novices to solve conceptual problems involving light, heat, and electric current, they observed a clear bias towards substance-like mental models (e.g., reasoning about electric current in a wire as if it were a fluid flowing inside a hose). This result was determined by presenting subjects with isomorphic pairs of problems, one of which was a "physics concept problem" (involving light, heat, or electric current), and the other a direct “material substance” isomorph of that problem, assuming the relevant physics concept was viewed as a material substance. For example, a problem involving an electric circuit with several bulbs in series would resemble a problem involving water flowing through a hose with several sprinklers in series -- assuming the physics novice thought of electric current as something like a flowing fluid. Such isomorphic pairs of problems were constructed so that similar answers to the problems would reflect similar conceptual reasoning (e.g., "the bulbs closer to the battery come on before the bulbs farther away" is similar to "the sprinklers closer to the faucet will come on before the sprinklers farther away")[4]. The results of this study showed that novices preferred substance-like models, in the sense that they chose the wrong answer to the physics concept problem, consistent with the right answer to the corresponding substance isomorph problem. Still, this pattern of errors did not unambiguously imply that the subjects had similar conceptions of electric current and water. That is, similarities of multiple choice responses was necessary, but not a sufficient measure to address such a question with authority.

Slotta et al. (1995) developed a second, more sensitive measure of the nature of their subjects' conceptualizations, based on the verbal explanations subjects provided for each problem. A method of analysis was developed which examined patterns of verbal predication in the language used by physics novices and experts as they reasoned about these problems. Because their hypotheses were specifically concerned with subjects' ontological associations, Slotta et al. (1995) drew inferences from the presence of particular verbal predicates about corresponding ontological attributes. For example, if a subject said, "The current comes down the wire and gets used up by the first bulb, so very little of it makes its way to the second bulb", then these four (underlined) predicates were taken as evidence that subjects conceptualized current as a substance-like entity which (1) Moves, (2) can be Consumed, (3) can be Quantified, and (4) Moves, respectively. The ontological attributes to be coded were generated a priori, based on prior observations of the properties of *substances* (Reiner et al., in press) and *process* (Chi, 1992; Chi, et al., 1994). Slotta et al. (1995) measured the degree to which subjects used these attributes in explaining their answers to the Physics Concept problems. By analyzing specific attributes of a concept (based on corresponding verbal predications) and summating over multiple problems, it became possible to quantitatively address the question of ontological association.

Whereas novices relied on substance attributes almost exclusively, it was found that experts used predominantly *process* attributes for their descriptions of electric current (Slotta et al.,1995). Figure 1 displays the average level of process and substance predication used by experts

---

[4] The physics concept problems used by Slotta et al (1995) were always presented as a block before the material substance isomorphs, so as to avoid making any suggestion of the faulty model.

and novices in the Slotta, et al. (1995) study. Furthermore, the level of substance predication, as well as the distribution between attributes (e.g., across all the *substance* attributes assessed: Moves, Consumed, Quantified, Blocked, etc.) was identical for novices between the physics concept problems and their material substance isomorphs. Thus, not only was a novice's multiple choice response similar between the two members of an isomorphic pair of problems, so was the pattern of verbal predication within her explanations (see the parallel solid lines in Figure 1). This analysis provided some fairly direct evidence that certain naive physics conceptions are attributed with the properties or behaviors of material substances, while expert conceptions of the same topics show no sign of a substance-like nature, but rather appear to be consistent with Chi's (1992) suggested (*process*) ontology of *complex dynamic systems.*

Insert Figure 1 about here

Instructional implications of Chi's framework

A theory of conceptual change should be able to account for why some physics misconceptions are robust and others are not, and why there is an apparent preference for materialistic misconceptions. It should also account for the qualitative differences between novice and expert conceptions, and suggest methods of analysis that can be used to determine whether or not conceptual change has occurred. Chi's (1992) theoretical framework will be useful to the extent that its claims about ontological associations in novice and expert conceptions can be drawn upon as a source of insight about how to instruct these difficult concepts. In addition, any direct predictions about effective instruction can be interpreted as important empirical tests of the framework itself.

Because the present research will be concerned with the concept of electric current, it is useful to consider an example of how Chi's framework relates to the issues of conceptual change discussed so far. Thus, if a student has wrongly associated the concept of electric current with the *material substance* ontology, she might consistently talk about it as "shooting out of the battery", or "leaking out of the wires", even when told explicitly that such descriptions are incorrect. Because of her strong ontological commitment, the student will experience great difficulty in learning about electric current in its scientific sense: as a type of an *equilibration process* where free electrons in the circuit move simultaneously and independently with no simple explanation, in a dynamic state that has no clear beginning or ending. Chi (1993; Chi & Slotta, 1993; Chi, et al., 1994) has called this barrier to conceptual change "an incompatibility" between the student's ontological association and the appropriate one.

Based on the theoretical discussion above, students would be less likely to make faulty ontological associations if (a) they were prepared with some knowledge of the appropriate ontology before they first encountered these concepts, and (b) their initial experience with (or instruction about) the concepts did not provide any suggestions of the wrong ontology, so as not to promote a misclassification. But children of almost any age have had some exposure to concepts such as heat, light, and electricity, and those initial exposures were more likely suggestive of *substance* ontology than that of *equilibration processes*. That is, young children probably have had scarce exposure to the *equilibration process* ontology, and have almost certainly been exposed to substance-based language and conceptualizations regarding topics like electricity. Thus, it is likely that most students are already in possession of the robust misconceptions with materialistic notions that we wish to avoid, as shown in the Reiner et al. (in press) review.

We must therefore ask whether points (a), providing some knowledge of the appropriate ontology before instruction of a specific concepts, and (b), barring any association with the wrong ontology, would also apply to the removal of misconceptions and not just their prevention. Indeed they must, if there is any hope of removing or replacing these robust materialistic conceptions (and we should take the occasional success of physics instruction as hopeful evidence that this "removal" can occur)[5]. So given a student who possesses a materialistic conception of a physics concept, we should succeed in helping this student to learn the correct ontological association for the concept to the extent that we (step 1) provide the student with some relevant knowledge of the target ontology before they receive any further instruction about the concept, and (step 2) avoid any use of terms or analogies which suggest the material substance ontology. Additionally, it seems that direct repudiation of the material substance ontology could only be helpful, as well as direct endorsement of the *process* ontology -- once step 1 is complete, and students have some knowledge of the new ontological class.

---

[5] Another open question concerns whether or not any naive conception is actually removed, or whether these early concepts are simply subordinated to their "scientific" counterparts over the course of instruction. Indeed, there is some evidence (Clement, 1987; McDermott, 1979) that even physics experts show lingering signs of these robust misconceptions. Certainly, experts make wide use of erroneous substance-like models of heat, light, and electricity, although they know the limitations of these models, as well as when the models should be abandoned. If indeed the naive conceptions are not actually removed or replaced, then the process of conceptual change is seen as one of

Design Overview

In order to test these predictions, a training study was designed in which one group of subjects received special, focused training on the nature of the *equilibration process* ontology (step 1, above), and then an instructional text concerning electric current that was completely free from suggestions of a *material substance* ontology (step 2, above; e.g., it made no use of the famous water analogy for instruction of electric circuits). A control group received no training in the *equilibration process* ontology (performing a control task instead), and then read the same instructional text as the experimental group. The question of greatest interest is whether the experimental group demonstrated conceptual change, as measured by a shift in their ontological associations for the concept of electric current. All subjects performed a pre-test consisting of eight qualitative physics problems (concerned with electric current), where they were asked to verbally explain their reasoning about the problems. The problem solutions, as well as the verbal explanation data were used to obtain measures of subjects' initial ontological commitments. These pre-test measures were compared to similar post-test measures in order to test the hypothesis that direct instruction concerning the *equilibration process* ontology would facilitate conceptual change.

An essential feature of the design is that both the experimental and the control groups received exactly the same text for learning about the target concept of electric current. The two groups differed only in that the experimental group received prior instruction about several properties of the *equilibration process* ontology, where this training included no mention of electricity whatsoever, nor any foreshadowing of its application to electricity concepts. The sole intent of the ontology training was to provide the students with some knowledge of the *equilibration process* ontology so that they could have a chance of making the correct association when subsequently learning about an *equilibration process* concept (electric current).

Assessment of conceptual change was performed according to the method developed by Slotta et al (1995), where verbal explanation data is analyzed for its content of two specific sets of conceptual attributes which were determined a priori to indicate ontologies of *material substance* or *equilibration process,* respectively. If a subject talked about electric current as if it had attributes of a material substance, this was taken to reflect an underlying conception of electric

---

learning a new conception alongside a robust old one. Nevertheless, the problem remains the same: How do we prevent new instruction from being assimilated into the framework of the robust incorrect old one?

current as a *material substance*[6]. Similarly, the use of verbal predicates reflecting ontological attributes of *equilibration process* is taken to reflect the presence of a *process* association. It was predicted that the experimental group would show a transition from the pre-test, where they explained problems in terms of a *substance* ontology, to the post-test, where they drew upon more *process* predicates in their explanations. In addition to this analysis of ontological commitment, the pre- and post-tests were also scored according to subjects' accuracy of response. These tests consisted of conceptual physics problems designed to be sensitive to the existence (or removal) of substance-based misconceptions of electric current.

Session 1

The study consisted of two sessions (see Table 1), each lasting between one and two hours (subjects were self-paced, and varied in reading speed). In the first session, university students with no science background were asked to complete a pre-test (all materials are described below) of their qualitative knowledge of simple electric circuits. They were then asked to sit at a computer and work through the *Equilibration Process* Training Module, which consisted of a carefully written text, and several accompanying simulations. Throughout the training module, subjects were occasionally interrupted by computer-presented explanation prompts that were meant to assure some level of attention was paid to the content[7] . Subjects were also aware that a post-test would be administered after the session, covering the content material. This training post-test also helped assure attention to content, but most importantly, provided some means to determine which subjects comprehended the training and which did not. After all, we would not predict much of an effect from unsuccessful training.

Insert Table 1 about here

---

[6] Note: the use of materialistic words or phrases is not sufficient evidence of a material substance conception. The subject is required to use these words or phrases in such a way that s/he predicates the concept with them meaningfully. So the subject's explanation won't necessarily be scored as "materialistic" if she uses the word "moves", whereas if she used the phrase "the electric current moves ___", this would be coded as evidence of a material substance conception.

[7] There is a growing body of evidence that the act of explaining new conceptual material to yourself or another can facilitate learning (Chi, Bassok, Lewis, Reimann & Glaser, 1989; Chi, de Leeuw, Chiu & LaVancher, 1994).

A second group of control subjects spent the first session working through a control module, where they read a completely different text from the computer. This control text was selected from an existing published science text (Hewitt, 1987) so that it was roughly matched to the training module text, both in topic (ideal gases, temperature, and liquid diffusion) and level of difficulty. While reading this control text, subjects were occasionally interrupted by computer-presented explanation prompts. At the end of the session, all control subjects received the control post-test, which consisted of qualitative questions concerning the definition and properties of the material described in the control text. Subjects were aware of this test at the beginning of the session, so that it provided some motivation for them to attend to the material. Additionally, it provided some means of assessing the extent to which subjects were able to learn the material in the control text.

Session 2

All subjects received the same materials and procedure in Session 2, which consisted of a physics text concerning electricity and electric circuits. The subjects read this electricity text from a 3-ring binder in the form of photocopied text, as it appeared in the published textbook from which it was drawn (Hewitt, 1987). Experimental subjects were instructed to try to apply what they learned in Session 1 to what they would read in the present session, while control subjects were simply instructed to try to learn the material as well as they could. In the course of reading through this text, subjects (both control and experimental) encountered occasional explanation prompts, which helped them reflect on their understanding and assured that attention was paid to the content. After completing the electricity text, all subjects received the post-test, which was identical to the pre-test. Finally, subjects were asked to complete an exit survey in which they provided information concerning their high school achievement (g.p.a., and SAT scores), university g.p.a., and some qualitative feedback about their perceptions of the study.

Method

Subjects

Subjects were 24 university undergraduate students (15 female, 9 male) recruited from the University of Pittsburgh and paid for their participation. Subjects had no university-level science background, nor any formal training in electricity. Table 2 provides a profile of the mean SAT scores and grade point averages for subjects in the experimental and control conditions. The table is not complete, because these figures were determined from an exit survey that some subjects could not accurately complete. The "n-value" listed below each figure in the table represents the number of subjects who responded confidently to that item, from which these means were computed (out of a total possible of 12). While these figures were obtained informally, they provide a qualitative sense of the subjects who participated in the study. Note that the subjects in the control group had a slightly higher scores on four of the five measures.

Insert Table 2 about here

Design

12 Subjects were assigned randomly to each group (experimental and control). There were both within-subject and between-subject aspects to this design. Within-subject aspects were concerned with pre-post test differences, and between-subject aspects were concerned with experimental-control differences. Two sorts of dependent measures were obtained from the data: pre-post test gains and verbal predication measures. All materials are described here.

Materials

Many different materials were used in this study including: pre- and post-tests of concepts of electricity; the *equilibration process* ontology training module (interface, simulations, and most importantly, the training module text); the control module; the training module post-test; the control module post test; the physics learning text (in topics of electricity); and the physics learning post-test.

Pre-test (and post-test)

Pre-test and post-test items were identical, consisting of eight conceptual physics problems, each of which involved predicting the behavior of a simple electric circuit. The problems were based on the most successful items (i.e., the most highly diagnostic items) from the materials of Slotta et al. (1995), which made use of circuits with multiple bulbs, connected either in series or parallel. For example, in Figure 2 the subjects are asked to predict whether the bulbs will illuminate at the same time, or with slight time differences once the switch in the circuit is closed. After selecting a response from the multiple choices, subjects were asked to verbally explain the behavior of the circuit.

Insert Figure 2 about here

While using the same items in pre- and post-tests did introduce an aspect of familiarity with the post-test items, subjects did not appear hesitant to either change their responses or keep them the same. That is, subjects appeared to consider each item in the post-test afresh, even if they recognized it from the pre-test; this was apparent in subject interviews and recorded explanations. The advantage of using the same items on both tests is that it allowed a contrast of the verbal predication measures, which might change unpredictably with a novel set of post-test problems. If a subject uses different patterns of verbal predication in explaining the same problem two different times, we can infer more strongly that the subject's conception of the problem has changed. In contrast, different patterns of verbal predication in the explanations of two different problems could simply be a consequence of different surface features within the problems.

The test items were designed so that different responses corresponded roughly with different mental models of electric current. For example, in choosing response "a" in the problem shown in Figure 2 (the light bulbs closer to the battery will illuminate slightly before those farther away), subjects are choosing a response consistent with the substance-like conception of electricity as a fluid which flows through hose-like wires. Slotta et al. (1995) observed that novices who chose this response also explained the problem using patterns of language very similar to those they used in explaining the problem's material substance isomorph (which consisted of a hose with multiple sprinklers in series). Building on the previous work of Slotta et al. (1995), it was possible to design a set of multiple choice problems where wrong answers would likely reflect underlying misconceptions.

Equilibration Process training module

The training module consists of a computerized instructional module which presents textual material to be read by the student at her own pace, where this textual material periodically refers to one of several running simulations on the top portion of the screen. Figure 3 displays a screen capture of the interface used in the training module. Notice the buttons which can be selected with the mouse to move between pages of text, as well as to interact with the simulations. The "Simulation" button was not selectable if the subject was not reading a portion of the text that required interaction with the running simulations. Simulations continued to run at all times, even when the Simulation button was disabled. Thus, during the portion of the training module that dealt with properties of Air Expansion (discussed below), the animated air molecules displayed in Figure 3 continued to bounce around the inside of the piston in the upper portion of the computer display. This provided the subject with a sense of the ongoing nature of *equilibration process*.

Insert Figure 3 about here

The purpose of the training module text was to communicate four attributes of the ontological category of *equilibration processes* in such a way that the subjects could understand and even apply the content of the text. This was done by focusing on two distinct examples of *equilibration processes*, Air Expansion and Liquid Diffusion, both of which are quite distinct from electric current. The text pointed out the four “special qualities” of these concepts, noting that they pertain to a whole class of difficult science concepts, which were referred to as "Equilibration Processes”.

The four ontological attributes used to describe air expansion and liquid diffusion were also ones that were determined by Slotta et al. (1995) to be most relevant to the *equilibration process* concept of electric current. It is possible that there are additional attributes which characterize *equilibration processes*, but it was important to train subjects on easily recognizable attributes that apply to the concepts in the training module text, as well as to the transfer concept (electric current). In the order they are presented within the training module, these four attributes are:

1. Equilibration Processes have no clear <u>cause-and-effect</u> explanation.
2. Equilibration Processes involve a system of <u>interacting</u> components seeking equilibrium amongst several constraints.
3. In an Equilibration Process, certain constraints behave as they do because they are actually the <u>combined effect</u> of many smaller processes occurring <u>simultaneously and independently</u> within the system.
4. Equilibration Processes have <u>no beginning or ending</u>, even if they arrive at an equilibrium position.

Throughout the text, general properties of the *equilibration processes* category were explained first (e.g., “Equilibration Processes have no beginning or ending”), followed by an explanation and simulation of how this property applied to each example (e.g., “The molecules in the air cylinder will continue bouncing around, even after the piston has reached its equilibrium height”). After all four attributes had been presented in the context of an example, each of the attributes was then reviewed once more in the abstract. In this way, it was hoped that subjects could learn the ontological attributes in a somewhat abstracted sense, by seeing both the general definition and the specific instantiation. After completing the tutorial on Air Expansion, the control module text reviewed the four attributes as they apply to a second concept - that of Liquid Diffusion. This second example, together with its accompanying simulation, was designed to provide the student with a second distinct instance of an *equilibration process*.

As the reader progressed through the training module, each example was illustrated by an animated simulation. For the concept of Air Expansion, the simulation (shown in Figure 3) consisted of a cylinder-piston system (a rectangle with a moveable "ceiling") with moving air molecules (circles) that collide with the walls of the cylinder and with the piston. When more molecules of air are pumped into the system (by an animated pump that injects more circles into the cylinder), the piston is seen to rise. The first attribute ("no clear cause-and effect explanation") was illustrated by showing students a faulty model that would have provided a clear causal account of the piston's rising: marbles (packed circles) were arranged within the cylinder so tightly that they forced against one another; newly added marbles had no room, and thus forced the upper marbles against the piston, which then rose. The text pointed out that no such clear chain of cause and effect exists to explain the rising of a piston in a cylinder full of air, and that this special quality is common to all *equilibration processes*. Each of the four attributes was then discussed in turn, defining the system (attribute number 2) and its equilibrium, the constraints on this process (attribute 3), and the fact that it never arrives at an end-point, but continuously approaches the equilibrium state (attribute 4).

Training text explanation prompts

Throughout the ontology training, subjects were occasionally asked to verbally respond to questions that appeared as part of the computer dialog (all explanations were tape recorded for possible later analysis). These questions were designed so that they simply prompted the subject to explain an important part of the text that she may not have completely understood (e.g., "can you name all of the forces acting on the piston?"; "Why does the piston eventually begin to fall?") These "explain questions" did not require subjects to draw any major new inferences, although they were designed to ensure a complete understanding of the material.

Training post-test

All experimental subjects were told at the beginning of the training module that, when finished, they would be asked several questions to determine their comprehension of the text. The first two questions of this test simply asked subjects to recall the four basic properties as they applied to the two examples used in the training module (air expansion and liquid diffusion). The final four items were related to a transfer problem (predator-prey populations), which was described in some detail. Subjects were then requested to apply each of the four properties of *equilibration processes* to their understanding of this new example. The four properties were listed overtly, so that the problem was solely one of applying them to the new example. The purpose of the training post-test was both to assure that subjects attended to the material, as well as to provide some measure of how well the material was actually understood. After all, one would not predict any positive effect from a training session where the subject did not comprehend the training content. Thus, the results from this training post-test actually permitted a contrast between those experimental subjects who understood the material well and those who did not.

Control module

The control group of subjects received a different training module, administered on the same computer interface (minus the simulations), so as to control for the training medium (i.e., computer-based). This text was also concerned with air expansion and liquid diffusion, so as to control for training topic, although its focus was more broadly related to basic concepts of solids, liquids and gases, and their behavior. The control text was drawn directly from a popular conceptual physics text (Hewitt, 1987), as well as a second text book (Towle, 1989) for the pages relating to liquid diffusion.

Control text explanation prompts

Throughout the control task, subjects were occasionally asked to verbally respond to various questions which appeared as part of the computer dialog (all explanations were tape recorded for possible later analysis). These questions were designed so that they simply prompted the subject to explain an important part of the text that she may not have completely understood (e.g., Why does water have a higher specific heat than sand?) All of these "explain questions" were drawn directly from the same book as the text itself (Hewitt, 1987), as the author provides several "Questions" highlighted within the text and at the end of the chapter.

Control text post-test

All Control group subjects were told at the beginning of the control module that, when finished, they would be asked several short questions to determine their comprehension of the text.

The purpose of these questions was both to ensure that the subjects attended to the material, as well as to provide some measure of how well the material was actually understood. All of these test items were drawn directly from the "Think and Explain" questions at the end of the Hewitt (1987) chapters.

Electricity text

This text provided the learning materials concerned with electric current, and was drawn directly from Hewitt's (1987) Conceptual Physics, chapters 33-35. By removing several tangential sections (e.g., one concerning Van de Graf generators and one concerning the difference between ac and dc current), as well as the questions and answers provided by Hewitt, it was possible to condense the text into approximately fifteen full-length pages which covered the basic theory of voltages, current, resistance, and simple resistive circuits. This length of text was manageable for subjects in a single 60-90 minute session (depending on reading speed). As presented by Hewitt, the text included many references to the water analogy of electric current, which is a very common instructional analogy used in the teaching of electric circuits. Because of obvious theoretical concerns, these references could not be included in any text presented to subjects in this study, and were systematically removed (removed text totaled less than 5 percent of the text retained). This text was read by the subjects in the form of photocopied paper, in a 3-ring binder, with approximately one paragraph of text per page, and occasional explanation prompts interspersed within the pages.

Electricity text explanation prompts

Throughout the electricity text, subjects were occasionally asked to verbally respond to a prompt for explanation of the material (all explanations were tape recorded for possible later analysis). These questions were designed so that they targeted those aspects of electric current that correspond to the ontological attributes instructed in the training module. For example, at one point in the text, the author talks about the actual speed of an electron through the wire as being "slower than a snail's pace", and explains that individual electrons do not actually flow through the wires, but instead that a net statistical drift is imposed on all the electrons. At this point, it was advisable to insert an explanation prompt which asked the subject to explain this in her own words.

Procedure

Session 1

All subjects began Session 1 by receiving instructions about the general course of the two-session study. They were then given the pre-test -- a short multiple choice test concerning several simple electric circuits where they chose their response, and then explained their answer to an interviewer. The interviewer tried to make sure that the subject explained the entire problem, as in providing a causal account of the problem, rather than a simple justification for the response chosen. Subjects were prompted for detailed explanations. After completing the pre-test, they were left alone at the computer to proceed through either the training module or the control module. All subjects were tape recorded during this session, in order to capture their responses to the occasional explanation prompts. The interviewer was in the next room during all sessions, so that any problems or questions could be immediately addressed. Upon finishing this computer session, subjects received a short post-test which measured their comprehension of the material from this session.

Session 2

In Session 2, all subjects (experimentals and controls) read through the Electricity Text, and were instructed that they would receive a short test at the end of the session (the post-test). In addition, the experimental subjects were told that they would be reading about another example of an *equilibration process*, and that they should try to remember and apply what they learned in Session 1 as they read the Session 2 material. Control subjects were simply informed that they should try their best to understand the material, that there would be occasional explanation prompts during the reading, and a post-test afterwards. At the end of Session 2, each subject met with the interviewer, who administered the post-test, prompting for detailed explanations where appropriate.

Results

Contrast of pre- and post-test scores

Choosing the correct answer on either the pre- or the post-test (e.g., saying that all the bulbs in Figure 2 illuminate at exactly the same time) does not necessarily imply possession of a scientific conception. For example, our visual impressions from the lighting of Christmas lights or multiple lights in a room are those of simultaneity. Thus, everyday experiences could lead some students to "know the right answer," without possessing the conceptualization required for a good explanation. But choosing the wrong answer (e.g., saying that the closer bulbs illuminate earlier, or glow brighter than those farther away) is more likely an indication of some misconception, and such responses can be analyzed for an effect of the *equilibration process* training. Furthermore, the small number of items on these tests (reflecting our focus on the explanation results) makes it less likely that we would measure significant differences in post-test scores between the experimental and control conditions. In spite of these two unfavorable factors (subjects obtaining the correct responses based on their perceptual experiences, and the small number of test items), a contrast of pre- and post-test scores revealed significant gains for the experimental group. (See Figure 4) Experimental subjects showed pre-post test gains of 29% compared to the control group's gain of only 9%. This difference was significant, with F(1,22) = 6.765, p = .0163. The difference in pre-test scores suggested by the Figure is not significant, and the conditions of administering the pre-test were strictly uniform between experimentals and controls.

Insert Figure 4 about here

Additionally, the improvement of an experimental subject's test scores depended on how well she understood the training material. Splitting the training group into high and low scorers on the training post-test (which measured how well they remembered and applied the training), Figure 5 shows a breakdown of the Figure 4 results into 3 groups: high-training; low-training; and controls. Clearly, those who scored highest on the training post-test were more likely to achieve gains during the second session of the study. The interaction of test score with training group was highly significant, with F (2, 21) = 13.847, p = 0.0001.

Insert Figure 5 about here

Inadvertently, when we split the Experimental subjects into high and low training successes, the pre-test scores of the "high training" experimental subjects appear more closely matched with those of the control group. While there are no significant differences between any of the groups' pre-test scores, the visual impression from the Figure suggests that the gains of the "high training" experimental group were even more pronounced.

Furthermore, this result of greater gains by the experimental group is not simply a matter of the experimental subjects being brighter or more motivated, thereby doing better on all aspects of the study (training, physics learning, post-test), since the high training group actually had lower pre-test scores on four of the five profile measures (see Table 2 again). However, a more specific way of analyzing this follows. A conglomerate ranking was constructed from the exit survey data, based on high school grade point average, verbal and math SAT test scores, and university grade point average[8]. While most of the high-ranking experimental subjects did become part of the "high training" group (i.e., the group who scored best on the training post-test), there were some exceptions. Ranking alone was not as successful in predicting improved test scores as was the training post test ; this interaction (with F (2, 21) = 3.457; p = .051) is not as great as that shown in Figure 5. Thus, even for lower ranking students, as long as they understood the training material, they achieved some test score improvement. Note that the training materials were difficult, requiring subjects to learn about a new type of concept -- one involving a "system" (a difficult concept in itself) that seeks equilibrium amongst "constraints". The training post-test was very difficult,[9] requiring subjects to transfer what they had learned to a new example of *equilibration processes* (predator-prey populations). This difficult training test was quite effective in assessing whether or not a subject had comprehended the training material. Figure 5 shows that subject improvement in problem solving depends directly on how well the training was understood.

Verbal Predication in pre-test and post-test explanations

The qualitative reasoning gains described above show that training in a specific *Process* ontology can indeed facilitate learning of certain physics concepts (as reflected by gains in qualitative reasoning about electricity). Yet such result does not guarantee that subjects have undergone conceptual change unless we can reliably assess the ontology of their underlying conceptualizations. The predicate analysis of verbal explanations developed by Slotta et al. (1995) permits such an assessment. Verbal predication within a problem explanation is analyzed for the presence of 6 different attributes from either a *process* (used here as a shortened reference for *equilibration process*) or a *substance* conception of electric current. The selection of attributes was informed by previous work of Slotta et al. (1995), who studied spontaneous explanations made by novices and experts

---

[8] This score was obtained from incomplete data on all measures listed. For each measure (e.g., verbal SAT), all subjects who provided that measure were rank ordered. Thus, if nine out of the twelve experimental subjects had provided their verbal SAT scores, they would be ranked as 1/9, 2/9, etc. Such numerical rankings were obtained for each of the four measures, and were then averaged for each subject across all the values available. The highest possible ranking would thus be 1.0 (if a student had been ranked first in all of the measures she provided). 17 subjects provided all four measures; 3 subjects provided only three measures; 3 subjects provided only two measures, and 1 subjects provided only one of the measures. Other rank scores were explored to see if any significant changes occurred within the data (e.g., scores based solely on SAT scores), with no noticeable differences. The score developed and used here is successful in ranking subjects with respect to one another, based on four separate measures, and tolerating incompleteness in the survey data.

[9] This difficulty clearly accounts for why the higher ranking subjects were more likely to succeed in the training.

in a set of very similar problems (in some cases, identical). The presence of any predicate from the list of *process* or *substance* attributes (known as the “basis sets” for the coding) was interpreted as evidence of the corresponding ontological association.

The six most commonly used attributes for descriptions of electric current were chosen from the Slotta et al. (1995) novice explanations as a basis set for the *substance* predicates in the present analysis[10]: Moves (M), is Supplied (S), can be Quantified (Q), comes to Rest (R), can be Absorbed (A), and can be Consumed (C). Similarly, the six attributes used most commonly by experts in their descriptions of electric current were chosen as a basis set for the process predicates in the present analysis: System-Wide (SW), Movement Process (MP), Uniform State (US), Equilibrium State (EQ), Simultaneity (SIM), and Independence (IND). These same attributes had also been used to guide the development of the training module: the four trained attributes of *Equilibration Processes* were: System-Wide, Equilibrium-Seeking, Simultaneous and Independent Processes, and Ongoing Process. Thus, the four attributes used in the training module are represented in the basis set for the *process* attribute coding.

Given a complete coding of all subjects' explanations (a complete coding means simply to search and code each problem explanation for the presence of all six attributes in each of the two basis sets), we can quantitatively address such questions as, (1) To what extent do subjects attribute the concept of electric current with substance-like qualities versus process-like qualities? (2) Is a subject's choice of attributes affected by the *equilibration process* category training (i.e., is there conceptual change)? and (3) Do subjects who scored highly on the training post-test show more conceptual change than those who did not, as measured by increases in process predication or decreases in substance predication? There are a variety of ways to address such questions, once a complete coding of all subjects' explanation data has been performed, but we must first develop a valid measure from the "raw" coded explanation data.

Slotta et al (1995) employed a single conglomerate measure of "substance-like predication" (or, similarly, "process-like predication") which was obtained by summing all occurrences of the *substance* (or *process*) predicates for a given explanation, and then dividing this number by the total number of coded predicates for that problem. This essentially provided a ratio of how many of the total coded predicates were *substance* (or *process*) predicates, and affords a means of normalizing for differences in length of problem explanations (both within and between subjects).

---

[10] As shown earlier in Figure 1, novices in the Slotta et al. (1995) study relied almost exclusively on substance predicates in their explanations.

However, it has the limitation of resulting in many measures of either 1 or 0, as most explanations that include *substance* predication do not also include *process* predication. Additionally, information about the number of distinct *process* or *substance* predicates found in a single explanation is lost. That is, if there were 4 substance predicates and 0 process predicates, this would result in a measure of 1.0, which ignores the important information concerning how many distinct substance predicates contributed to that 4 -- were they 4 instance of the "Moves" predicate, or perhaps 1 each of the "Contains" and "Blocks" predicates and 2 "Moves" predicates?

Another measure that gives a bit more weight to explanations where several different predicates are uttered is to simply tally whether or not each of the six attributes is represented in the explanation, ignoring multiple occurrences, to obtain a maximum value of 6 for any problem. This measure is sensitive to information about the distribution of individual predicates from a specific basis set, although it loses some information about the relative presence of substance vs. process predicates within an explanation. This loss of information is not highly dangerous, however, since Slotta et al. (1995) found that explanations containing one type of predicate rarely contained the other. Slotta et al. (1995) applied both types of measure to their data, and found the same basic effects either way. The present research thus makes use of the second type of measure, and derives a single count of how many predicates from each basis set is present in each explanation. The resulting counts are used below in quantitative analyses to address questions about ontological associations in the experimental and control groups. The test of the reliability of our coding measures is presented in Appendix A.

Comparisons of pre-test to post-test explanation data

When the verbal explanation protocols were coded for the presence of all attributes[11] in both basis sets (a total of 12 attributes in all -- six from each basis set), it was found (Figure 6) that experimental subjects (the dotted lines) used dramatically lower levels of *substance* predication in the post-test, accompanied by a corresponding increase in levels of *process* predication. Meanwhile, control subjects (the solid lines) showed little change from pre-test to post-test in their use of either *substance* predicates (still predominant) or *process* predicates (still scarce). Both control and experimental groups relied almost entirely on *substance* predicates (the squares in the Figure) in explaining their pretest solutions, replicating Slotta et al. (1995).

---

[11] Explanation protocols are actually coded for the presence of verbal predicates that reflect attributes in the basis sets. A variety of verbal predicates would reflect the "Moves" attribute, for instance: goes; comes; travels; shoots; etc. Slotta et al. (1995) provide a detailed discussion of this analysis, including some commentary on the challenge of drawing valid inferences from patterns of verbal predication.

Analysis of post-test explanations revealed the hypothesized conceptual change in the experimental group, who relied greatly on *process* predicates (the triangles), and very seldom drew upon the *substance* predicates (closely resembling the experts in the Slotta et al. study). Both the increase in *process* predication (F(1,10) = 31.04, p = 0.0002) and the decrease in *substance* predication (F(1,10) =20.17, p = .0012) were significant. Control subjects showed no such transition in their preference of conceptual attributes, with no significant differences in level of *process* or *substance* predication. The similarity between the expert-novice graphs of Slotta et. al (Figure 1) , and the experimental-control graphs of Figure 6 is striking.

Insert Figure 6 about here

Examples of pre-test explanations

The experimental and control subjects relied almost exclusively on substance predicates in their pre-test explanations, replicating Slotta et al. (1995). There were no significant differences between groups in the degree to which substance or process predicates were applied to the concept of electric current. Pre-test explanations tended to treat electric current as a substance that emerges from the battery (once a switch in the circuit is closed) and progresses around the circuit, gradually diminishing in size or strength as it is consumed by each successive bulb in the circuit, until finally its remainder drains back into the battery. Table 3 provides four representative explanations of the problem shown in Figure 2 -- two from experimental subjects and two from controls. Overall, pre-test explanations in this study closely resembled the novice explanations reported by Slotta et al. (1995)

Insert Table 3 about here

Examples of post-test explanations

Subjects in the training condition showed significant changes in the way they explained post-test problems such as that of Figure 2. These changes were consistent with the hypothesized conceptual change: away from a *substance*-based, and towards a *process* view. Table 4 provides four representative explanations of the "ten-bulbs problem" of Figure 1, taken from the same four subjects displayed in Table 3. Notice that the explanations of the experimental subjects now refer to a system-wide process occurring throughout the circuit, and involving simultaneous activity at all ten bulbs. The control subjects often answered this problem correctly, since it was treated explicitly within the physics text. Still, this only highlights the importance of appropriate assessment of the underlying conceptions, such as the predicate analysis. Control subject explanations typically retained their substance-based flavor, even though they often become more sophisticated and logically sound, as in the case of subject C-12 (shown in Table 4).

Insert Table 4 about here

Just as the high-trained experimental subjects (those who scored higher than the median on the training post-test) showed more gains in the problem solving measure than their low-trained counterparts, they also showed a greater effect in these measures of conceptual change. Figure 7 shows that successful training was indeed a requirement for conceptual change, with the high scoring subjects responsible for nearly all the gains of the experimental group. The interaction suggested by Figure 7 -- between Training Split (high, low, control) and decrease in *substance* predication -- is significant (F(2, 21) = 7.57, p = 0.003, as is the interaction between Training Split and increase in *process* predication (F(2, 20) = 35.89 , p = 0.0001). The low-scoring experimental group did show a reduction in *substance* predication and an increase in *process* predication compared to the control group, but significantly less so than the high-scoring training group. In general, all apparent differences between high and low scorers are significant at least to p=0.05. This result shows a connection between the effectiveness of the training (how well a subject did on the training post-test) and the conceptual change experienced in subsequent learning from a physics text.

Insert Figure 7 about here

Discussion

Overview

The purpose of this research is to provide empirical support for a theoretical account of conceptual change in learning complex science concepts (Chi, 1992; 1993; 1997; in press-b; Chi & Slotta, 1993; Chi, Slotta and deLeeuw, 1994; Ferrari & Chi, in press; Slotta & Chi, 1996; Slotta, Joram & Chi, 1995;). The theoretical account basically specified an ontological boundary between two explanatory frameworks: the students' misconceptions and the scientific conceptions. In other words, once a student associates a concept with a particular ontology, she will try to understand all subsequent instruction in terms of that association, which will be troublesome to the extent that the instructor pursues a different ontology. Thus, certain concepts are traditionally more difficult for students to learn than others because students are more likely to associate them with an incorrect ontology. Chi (1992, 1993, 1997) argues that students are fundamentally inclined to conceive of concepts such as heat and electric current as a kind of *material substance*. This may result from a variety of different causes: materialistic biases in language, such as in the heat example above; the dominance of the material substance ontology in our conceptual knowledge, such that it becomes a "default" for novel concepts (i.e., most of our experience is with material substances and their observed behavior); or the paucity of examples from alternative ontologies (as perhaps in the case of the *dynamic systems* ontology). Whatever the origin of this bias towards a *material substance* ontology, the challenge of teaching certain physics concepts apparently involves convincing students to either relinquish their initial ontological associations, or at least to learn an entirely distinct conceptualization consistent with the target ontology.

Chi's theory was specified at a level of explicitness that affords a prediction about training: That training of a metaframework (the ontology of *dynamic system or equilibration process)* would favorably influence the effectiveness of standard physics instruction from text in one particular topic (electric current). The logical implication of this prediction is that we could improve the effectiveness of physics texts by leaving them completely unchanged and simply adding an apparently unrelated conceptual module into the curriculum stream. As unlikely as this prediction seems, it was essentially supported by the results discussed above. Not only did subjects begin talking and thinking of electric current in fundamentally different terms than their control counterparts, but they actually improved over controls in their ability to solve conceptual problems. Thus, the present findings are unique in that the content of the training module has no relation to the to-be-learned concepts in electricity other than the theorized ontological connection. Indeed, the actual physics text (i.e., about electricity concepts) was completely unchanged from that received by the control group, so that it must have been the subjects' acquisition of a conceptual framework that was affected by the training. Thus, a seemingly tangential training about an ontological category has yielded dramatic results in terms of qualitative reasoning (problem solving and explanations) in another domain (in electricity concepts) that reflects "far transfer" or deep conceptual change.

Taken together, the results from this study present compelling support for our main prediction that it is possible to facilitate conceptual change in a difficult physics concept by first providing training in the concept's target ontology, followed by normal instruction in the topic (in this case, from a text, with all misleading ontological references removed). Figure 5 suggests that subjects who received and understood the training learned enough from the subsequent electricity text that they significantly improved in both the predictions of the problem solutions and their explanations (on the post-test).

Our analysis of pre-test explanations serves to replicate Slotta et al. (1995) with regard to their observations about novice conceptions of electric current. It also supports a suggestion offered by Slotta et al. (1995) that the expert-novice differences they observed could provide a means of assessing conceptual change. While this suggestion was intuitively clear, a possible objection was that experts used different language in their explanations for some other reason (e.g., they were older or smarter). While such objections were countered by the observation that experts did not differ from novices in their explanations of the Substance problems, the argument for an account of conceptual change was still somewhat indirect. In the present research, however, novice conceptions are actually observed to change. The pattern of means displayed in Figure 6 is strikingly similar to that reported by Slotta et al. (1995, and shown in Figure 1), in support of the arguments offered in that paper. The separation of the experimental subjects according to training effectiveness (shown in Figure 7) provides even further support for the argument that *equilibrium process* ontology training can directly mediate conceptual change.

Implications for Instruction

The direct application of any basic cognitive research to real-world instruction is never a straight-forward process. It involves discerning the relevant basic findings, and determining their significance to the broad issues of instruction in any particular domain (which includes many issues outside the cognitive realm). In the present case, the most important theoretical idea is that different types of learning processes may be involved in coming to understand different types of concepts. Put succinctly: Before designing instruction for certain science concepts, we must first determine whether students are required to undergo radical or gradual conceptual change. In many topics, students' preconceptions will require substantial revision in the course of instruction, but never a complete ontological shift. For example, students' initial conceptions of certain biology concepts, such as those involved in the circulatory system, may be incorrect, but still in the correct ontological category (e.g., the "heart" may be misconceived as a source of blood, and not as a pump, but it is still thought of as a material object, and not as a process or an abstract idea). In these cases, even though students' initial incorrect ideas need major reconceptualization (Chi, in press-a), it is not of the kind that crosses ontological boundary so that the present training approach would not be recommended. The important insight here is that educators may need to first recognize whether or not serious revision of the nature of the concept is required.

In the instruction of concepts such as electric current, heat, light, or force, and perhaps others involving system equilibration, such as diffusion, supply and demand, or natural selection, to name a few, the initial conceptualizations held by students may be so far removed from the conventional scientific view that they involve differences of an ontological nature. In these cases, the present research would be applicable. First of all, it would suggest that teachers not try to "bridge the gap" between students' misconceptions and the target instructional material, since there is no tenable pathway between distinct ontological conceptions. Students who understand "force" as a property of an object cannot come gradually to shift this conception until it is thought of as a process of interaction between two objects. Indeed, their learning may actually be hindered if they are required to relate instruction to such misconceptions. Rather, it is proposed here that no treatment whatsoever of students' misconceptions be entertained[12], and that instruction stress the basic ontological characteristics of the concept. Thus, student misconceptions are not ignored by instruction; rather, they are addressed by carefully avoiding any language, analogies, or phenomena which might reinforce them, and by explicitly drawing attention to fundamental (ontological) aspects of the concepts, in order to help students formulate completely novel conceptions which adhere more closely to the accepted scientific view.

A second, perhaps more revolutionary implication of this research is that students might profit from seemingly tangential or unrelated training in ontological properties. In the present case, a short module which highlighted the properties of *equilibration processes* was found to positively affect subjects' ability to learn electricity concepts -- even though the training module made no mention of electricity! Thus, instruction could benefit from first establishing some knowledge of a concept's ontological type before trying to establish any knowledge of the concept itself. This is far from a radical notion in instruction. Foreign language students first learn about the nature of the

---

[12] Note: such avoidance of students' alternative conceptions is only suggested in these special cases where they have made an ontological error in their initial conceptualizations. All other cases of instruction will certainly profit from the process of demanding that the student reconcile existing conceptions with new phenomena and principles, as occurs in the process of self-explanation (see Chi, in press-a)

conjugation type "pluperfect" or "subjunctive" before they are ever taught a specific verb conjugation in those tenses; and math students are instructed about the nature of vectors or tensors before they attempt to learn any specific operations with these new types of entities. This research identified that such a situation exists within the physics domain, and that instruction is aided by placing some precursory focus on the special ontological nature of certain concepts before they are taught explicitly.

The situation in physics instruction is exacerbated (compared to the linguistics and mathematics examples listed above) by the fact that students enter into instruction with existing knowledge of these concepts that is mistaken in terms of ontology. Further, these ontologies are prominently and continually reinforced by everyday language, terminology, cartoons, etc. Thus, students must first be enabled to learn about the new ontology (*equilibrium process*) and then carefully guided to apprehend the more relevant conceptualization. In the case of the present research, this was achieved by means of the *process* training module, followed by instruction in which all reference to the *material substance* ontology (e.g., the famous water analogy of electric circuits) was carefully avoided. So the implication for instruction in these topics is much more than to simply "start afresh" with no attention paid to preconceptions. Teachers and curriculum designers must first discern whether a concept is likely to have been ontologically misplaced by the student, and then proceed with a two-phased approach: first, train the student in the target ontology, which amounts to providing some knowledge of the relevant attributes of concepts of this type; second, provide instruction which relates the concept to these attributes while completely avoiding any connection with the misconceived ontology.

References

Ausubel, D. P. (1968). Educational Psychology: A cognitive view. New York: Holt, Rinehart, and Winston.

Chi, M.T.H., Bassok, M., Lewis, M., Reimann, P. &. Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. Cognitive Science, 13, 145-182.

Carey, S. (1985). Conceptual Change in Childhood. MIT Press.

Champagne, A. B., Klopfer, L. E. & Gunstone, R. F. (1982). Cognitive research and the design of science instruction. Educational Psychologist, 17, (1), 31-53.

Chi, M. T. H. (1992). Conceptual change within and across ontological categories: Examples from learning and discovery in science. In R. 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. (1993). Barriers to conceptual change in learning science concepts: A theoretical conjecture. In W. Kintsch (Ed.), Proceedings of the Fifteenth Annual Cognitive Science Society Conference (pp. 312-317). Hillsdale, NJ: Erlbaum.

Chi, M. T. H. (1997). Creativity: Shifting across ontological categories flexibly. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), Creative thought: An investigation of conceptual structures and processes (pp.209-234). Washington, DC: American Psychological Association.

Chi, M.T.H. (in press-a). Self-explaining: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), Advances in Instructional Psychology,Vol.5. Hillsdale, NJ: Erlbaum.

Chi, M.T.H. (in press-b). Understanding of complex, abstract, and dynamic concepts. In Encyclopedia of Psychology. APA and Oxford University Press.

Chi, M. T. H., de Leeuw, N. A., Chiu, M.H. & LaVancher, C. (1994) Eliciting self-explanations improves understanding. Cognitive Science, 18, 145-182.

Chi, M. T. H. & Slotta, J. D. (1993). The ontological coherence of intuitive physics. Commentary on A. diSessa's “Toward an epistemology of physics.” Cognition and Instruction, 10, 249-260.

Chi, M. T. H., Slotta, J. D., & deLeeuw, N. A. (1994). From things to processes: Toward a theory of conceptual change. In S. Vosniadou (Ed.), special issue of Learning and Instruction.

Chi, M. T. H. & VanLehn, K. A. (1991). The content of physics self-explanations. Journal of the Learning Sciences, 1, 69-105.

Chinn, C. A. & Brewer, W. F (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. Review of Educational Research, 63, 1-49.

Clement, J. (1987). Overcoming students' misconceptions in physics: The role of anchoring intuitions and analogical validity. In J Novak (Ed.), Proceedings of the second international seminar on misconceptions and educational strategies in science and mathematics, Vol. 3. (pp. 84-97). Ithaca, NY: Cornell University.

diSessa, A. A. (1988). Knowledge in pieces. In G. Forman & P. Pufal (Eds.), Constructivism in the Computer Age (pp. 49-70). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Doran, R. L. (1972). Misconceptions of selected science concepts held by elementary school students. Journal of Research in Science Teaching, 9, 127-137.

Driver, R. & Easley, J. (1978). Pupils and paradigms: A review of literature related to concept development in adolescent students. Studies in Science Education, 5, 61-84.

Driver, R. & Erickson, G. (1983). Theories -in-action: Some theoretical and empirical issues in the study of students' conceptual frameworks in science. Studies in Science Education, 10, 37-60.

Ferrari, M. & Chi, M. T. H. (in press). The nature of naive explanations of natural selection. In International Journal of Science Education.

Hewitt, P. G. (1987). Conceptual physics: A high school physics program. Menlo Park, CA: Addison-Wesley Publishing Company, Inc**.**

Joshua, S. & Dupin, J. J. (1987). Taking into account student conceptions in instructional strategy: an example in physics. Cognition and Instruction, 4 (2), 117-135.

Keil, F. C. (1981). Constraints on knowledge and cognitive development. Psychological Review, 88, 197-227.

Keil, F. C. (1987). Conceptual development and category structure. In U. Neisser (Ed.), Concepts and conceptual development. Cambridge, MA: Cambridge University Press.

Keil, F. C. (1989). Concepts, kinds, and cognitive development. Cambridge, MA: MIT Press.

King, W. H. (1961). Symposium: Studies of children's scientific concepts and interests. Journal of Educational Psychology, 31, 1-20.

Kuethe, L. J. (1963). Science concepts: A study of "sophisticated" errors. Science Education, 47, 361-364.

McCloskey, M. (1983). Naive Theories of Motion. In D. Gentner & A. L. Stevens (Eds.) Mental Models (pp. 299-324). Hillsdale, NJ: Erlbaum

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

Osborne, R. J. & Wittrock, M. (1983). Learning science: A generative process. Science Education, 67, 489-508.

Pfundt, H. & Duit, R. (1988). Bibliography: Students' Alternative Frameworks and Science Education (2nd ed.). Kiel, FGR: Institute for Science Education.

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

Reiner, M., Slotta, J. D., Chi, M. T. H., and Resnick, L. B. (in press). Naive physics reasoning: A commitment to substance-based conceptions. In Cognition and Instruction.

Slotta, J. D. and Chi, M. T. H. (1996). Understanding constraint-based processes: A precursor to conceptual change in physics. In G.W. Cottrell (Ed.), Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society (pp. 306-311). Hillsdale, NJ: Erlbaum.

Slotta, J. D., Chi, M. T. H. & Joram, E. (1995). Assessing students' misclassifications of physics concepts: An ontological basis for conceptual change. Cognition and Instruction, 13, (3), 373-400.

Tasker, R. & Osborne, R. (1985). Science teaching and science learning. In R.Osborne & P. Freyberg (Eds.), Learning in science. The implications of children’s science (pp.15-27). Auckland: Heinemann

Thagard, P. (1989). Explanatory coherence. Behavioral and Brain Sciences, 12, (3), 435-502.

Towle, A. (1989). Modern Biology. New York: Holt, Rinehart and Winston.

Viennot (1979). Spontaneous reasoning in elementary dynamics. European Journal of Science Education, 1, 205-221.

Vosniadou, S. & Brewer, W. F. (1992). Mental models of the Earth: A study of conceptual change in childhood. Cognitive Science, 24, 535-585.

Appendix A

Reliability of coding measures

While Slotta et al. (1995) had already demonstrated the reliability of coding for such predicates, it remains an important element of the method. An independent coder was provided with written training (approximately 6 double-space pages) and discussion (approximately 30 minutes) concerning the 6 attributes from each of the two basis sets (12 coding items in all). She then coded two complete subject protocols (pre- and post-test explanations, amounting to 16 problem explanations for each subject), and was given feedback on her coding. At this point, she was provided with four complete subjects -- two control and two experimental, with no knowledge of either the subject's group (control or experimental) or condition (pre- or post-test).

The secondary coder's blind coding of these protocols was compared in detail with the primary coding. That is, each explanation protocol was examined line-by line, and every occasion of coding was compared with the original coding to check for agreement. This sort of comparison is time consuming and complicated by the fact that each coder should be checked for agreement with the other. That is, we must determine not only the proportion of the original coder's items found by the secondary coder, but also the proportion of the secondary coder's items that are present in the original coding. Even if the secondary coder finds 90 percent of the original coder's items, it is still important to determine how many items she found that the primary coder did not. Figure 8 shows a detailed bar graph of this initial reliability comparison. Each bar in the figure represents the proportion of a coder's codes that was noted by the other coder; in order to be confident in the reliability of coding, this measure should be equally high for both coders.

Insert Figure A1 about here

Overall, the agreement was quite high, and of approximately equal proportion between coders. That is, each coder noted approximately 90 percent of the other's items; an item was only counted as agreed upon if the same portion of protocol was assigned the exact same code by each coder. It is important to recall that both *substance* and *process* codes were included in these comparisons, and that the primary coding found that *process* predicates only occurred with any frequency in the experimental subjects' post-test explanations. The two coders discussed their differences, and the second coder was provided with the protocols from an additional 8 subjects, resulting in a total of 12 subjects, or 50 percent of all protocol data. These 16 subject protocols (pre- and post-tests for each of the 8 subjects) were again coded blind to condition and group. As the detailed comparison shown in Figure 8 had demonstrated substantial agreement in specific coding, the data from all 12 subjects of the secondary coder were simply analyzed in the same method that was applied to the primary coding. In this way, it was possible to demonstrate that the secondary coding resulted in the same effects reported in the Results section above. Figure 9 displays a comparison of results from the primary and secondary codings for these 12 subjects; there are no significant differences between these two sets of results.

Insert Figure A2 about here
