JOURNAL OF RESEARCH IN SCIENCE TEACHING

Relations among Three Aspects of First-Year College Students' Epistemologies of Science

Carol L. Smith,[1] Laura Wenk[2]

[1]*Department of Psychology, University of Massachusetts, Boston, MA 02125*

[2]*School of Cognitive Science, Hampshire College, 893 West Street, Amherst, MA 01002*

Received 14 July 2004; Accepted 19 April 2005

Abstract: At the start of their first semester, 35 college freshmen were given an interview probing (a) their differentiation of scientists' ideas from evidence, and hypotheses from theories; (b) their understanding of the inherent uncertainty of scientific knowledge; and (c) their reasoning about scientific controversies. The most common responses were in terms of an epistemology in which scientists' ideas and evidence are differentiated, and theories are understood as tested hypotheses (Level 2 in our system based on Carey, Evans, Honda, Jay, & Unger, 1989), although students varied in how consistently they differentiated theories and evidence across all questions. Responses in which theories are understood as broader explanatory frameworks guiding hypothesis testing (Level 3) were virtually nonexistent, but some students gave responses that showed awareness of processes of interpreting and reinterpreting patterns of results (Level 2.5). Responses across the three parts of the interview were significantly related. Consistently differentiating scientists' ideas from evidence was strongly related to appreciating the inherent uncertainty of scientific knowledge and with having a deeper understanding of the reasons for scientific controversies and how to resolve them. © 2006 Wiley Periodicals, Inc. J Res Sci Teach

Reform efforts in science education at the K–12 and college levels have stressed the importance of developing students' epistemological understandings of science (American Association for the Advancement of Science, 1993; National Research Council, 1996; National Science Foundation, 1996). At the same time, there is concern that most college students and preservice science teachers have at best a very naïve understanding of knowledge construction in science as following an atheoretical "scientific method" (Abd-El-Khalick & Lederman, 2000; Lederman, 1992; Windschitl, 2004). Developing more sophisticated epistemological understandings may be important for diverse reasons: to help create better science students, science teachers, scientists, and even more informed citizens capable of reasoning critically about science

Contract grant sponsor: NSF; Contract grant number: REC-99 80519; Contract grant sponsor: Howard Hughes Medical Institute; Contract grant number: HHMI 71100–503803.
Correspondence to: C.L. Smith; E-mail: Carol.Smith@umb.edu
DOI 10.1002/tea.20113
Published online in Wiley InterScience (www.interscience.wiley.com).

issues. More specifically, if one's epistemological understandings in a given field function as a kind of metacognitive control structure (Hofer, 2004), then they would guide one's goals, reasoning, and sense making in situations in which they are activated. For example, they could guide how one structures firsthand inquiry (Sandoval, 2003; Windschitl, 2004), reasons about scientific controversies (Bell & Linn, 2002; Driver, Leach, Millar, & Scott, 1996; Sadler & Zeidler, 2004), conducts searches for reliable information on science topics (Hofer, 2004), approaches learning difficult science content (Hammer, 1994; Songer & Linn, 1991), and designs classroom inquiry experiences for students (Chinn & Malhotra, 2002).

Although there is broad agreement that developing students' epistemological understandings may be important as an educational outcome, there is less consensus about how these understandings should be conceptualized and assessed. One issue is the extent to which epistemological views are ''tacit'' versus ''articulated.'' Other issues concern what ''grain size'' it is useful to use in describing these understandings, the ways that context affects them, the extent to which even the novice's thinking (as well as the expert's) is coherent and organized, and the extent to which epistemological understandings are domain general versus domain specific. At present, researchers have conceptualized student epistemological understandings as a domain-general developmental structure (King & Kitchener, 1994), as domain-general beliefs (Schommer, 1990), as domain-specific theories (Hofer, 2004; Hofer & Pintrich, 1997), and as highly situation-specific resources (Hammer & Elby, 2002). We believe that debate about these issues can be productive at this early point in charting out this complex terrain, especially if those with competing views take care to articulate their positions clearly, generate testable predictions from their views, and identify the important phenomena about student thinking and reasoning that any theory of epistemological reasoning needs to account for. Further, it might even be the case that each tradition is trying to understand an important ''part of the elephant'' and that we may ultimately have more comprehensive theories that integrate diverse views.

The initial studies of epistemological development among college students concerned students' personal epistemologies rather than their epistemologies of science and were conceptualized as a domain-general developmental structure (Baxter Magolda, 1992; King & Kitchener, 1994, Perry, 1970). Work in this area has been influenced by William Perry's (1970) pioneering longitudinal studies of Harvard University students, with subsequent researchers attempting to further develop and refine his basic coding scheme. These schemes consider the extent of student recognition of the inherent uncertainty of knowledge, the role of viewpoint and knower in the knowing process, and the need to make commitments in the face of uncertainty and to identify some standards of argument that transcend individual perspectives. Researchers have identified a progression of student views of increasing complexity from (a) the naïve view that knowledge is unproblematic, known by authorities with a high degree of certainty to (b) transitional radical relativist views in which all knowledge is uncertain and everyone uses evidence to support his or her own opinion to (c) more sophisticated views of knowledge as uncertain, affected by the interpretative perspectives of the knower, but in which knowledge claims can nonetheless be evaluated by standards that take account of and transcend individual frameworks. According to the longitudinal studies mentioned earlier, college freshmen typically still believe in the certainty of knowledge, although some are beginning to embrace radical relativism (Baxter Magolda, 1992; King & Kitchener, 1994; Perry, 1970). These studies paint a picture of limited progress in epistemological development during college, in which sophisticated epistemological positions still elude most college graduates (Baxter Magolda, 1992; Hofer & Pintrich, 1997; King & Kitchener, 1994).

One limitation of this approach is that it is silent with respect to issues of important domain-specific differences in epistemological views or how particular kinds of educational experiences affect epistemological development. Thus, other researchers have begun to characterize student and expert epistemologies in science (Abd-El-Khalick & Lederman, 2000; Carey & Smith, 1993; Driver et al., 1996), math (Lampert, 1990; Schoenfeld, 1992), and history (Lee & Ashby, 2000). They have argued that students' experiences with particular domains affect their epistemological views about those domains and hence that students could be more sophisticated in reasoning about some domains than others. (see Boix-Mansilla, 2001, for evidence of differences in high-school students' epistemological sophistication in history vs. science as a function of their differential experiences with deep, firsthand inquiry in those domains). In light of this work, Hofer and Pintrich (1997) proposed that it might be more useful to conceptualize epistemologies as a progression of domain-specific intuitive theories, with succeeding theories characterized by different interrelated beliefs about what knowledge is and how it is justified in particular domains.

The framework informing the current work is that epistemological stances are organized as domain-specific intuitive meta-theories in which history, science, and math are seen as distinct domains.[1] Unlike Hofer and Pintrich (1997), however, we assume that student thinking is coherent in the sense that key epistemological *concepts* tend to be interdefined within a domain (e.g., the notions of theory, hypothesis, and experiment are within the domain of science), not that students have uniform or highly codified specific *epistemological beliefs*.[2] We also assume that different domains use different epistemological concepts and standards; hence, there is not a simple one-to-one correspondence between the epistemological distinctions used across domains. For example, what counts as "proof" or "evidence" is different in math, science, history, and literary analysis. Contrasting disciplines vary in how they conceptualize their main goals (e.g., explanation vs. interpretation, explication, or axiomatization) and in what methods they use (e.g., designing experiments vs. analyzing reliability or motives of sources vs. interpreting texts and symbolism). Further, disciplines that emphasize explanatory goals vary in their views of what makes a compelling explanation (e.g., appeal to a sense of mechanism vs. a compelling narrative account of how something came to be in a historical sense). Thus, it is important to identify and describe the underlying concepts that students use to organize their epistemological thinking in a given domain and to examine the ways those concepts change as they develop expertise in that domain. Conceptual change can include differentiating two ideas that were previously conflated in one concept, coalescing into one concept two ideas that were initially thought to be radically different kinds, or changing what is a core or peripheral feature of a concept.

What conceptual changes might underlie the development of a more informed epistemology of science? Carey and Smith (1993) argued that as students develop a more sophisticated epistemology of science, they must make fundamental changes in their concepts of scientific ideas and evidence, theories, and hypotheses. These changes, of course, do not occur in isolation but occur as they develop increasing knowledge of deep explanatory (and often highly counter-intuitive) theories that have been created in science and of the complex methods that have been employed both in generating and evaluating these theories. Further, changes in the extent to which they differentiate and interrelate the notions of theories, hypotheses, and evidence have widespread implications for how they organize their thinking about the overall goals, purposes, and activities of science.

More specifically, students' initial framework epistemology of science (dubbed Level 1 in Carey & Smith's, 1993, system) makes no differentiation between scientists' ideas, activities, or experimental results. Consequently, students have no appreciation of the role of scientists' ideas in guiding their activities/experiments, of experimental results (or other data) as providing *evidence* for ideas, or of any uncertainty in scientific knowledge. Scientists simply make (local) observations of what happens, do tests, find out what works or how to do something correctly, and amass a collection of true beliefs about the world. Interview studies have suggested that many seventh graders are solidly Level 1 in their thinking (Carey et al., 1989; Carey & Smith, 1993).

A more sophisticated epistemology of science (Level 2) makes a clear differentiation between the ideas scientists are developing about the world and their experimental results or data that they have collected in the service of testing or evaluating those ideas. At Level 2, students now view scientists as fundamentally concerned with understanding how things work or why things happen, not just with knowing what happens or how to do things. Furthermore, they now view scientists as designing experiments, using careful measurements, or making particular observations in the service of testing their initial ideas. Thus, two new notions that emerge at this level are ideas of explanation and hypothesis testing, both of which support making a fundamental differentiation between scientists' ideas and results (An explanation provides *an account* of data. A hypothesis is evaluated *in light of* data.) Thus, science proceeds through a process of testing ideas, and scientific knowledge consists of a collection of well-tested hypotheses. Students at Level 2, however, make no distinction between scientists' overarching theories and the specific hypotheses they test in experiments (or by other means) and hence have no awareness of the role of scientists' theories in guiding inquiry. Rather, they see hypotheses as suggested (more immediately) by the observations that scientists have made and see scientists' theories simply as their well-tested hypotheses. Interview studies have suggested that these Level 2 ideas are often becoming more salient among high-school students (Honda, 1994, 1996; Sandoval & Morrison, 2003).

Movement to a still more sophisticated epistemology (Level 3) involves making a further differentiation between scientists' framework theories and more specific hypotheses, a level that was not directly observed in interviews with students but which corresponds to ideas about science in historical and philosophic studies (e.g., Kuhn, 1970; Lakatos, 1978). At Level 3, a theory is a coherent network of interrelated concepts (or causal principles) that informs all aspects of inquiry—the questions raised, the methods used, and the formation of specific testable hypotheses. Scientists are seen as able to employ a framework theory to generate a multitude of rival hypotheses. They develop and flesh out a framework theory by working to distinguish among and elaborate on these hypotheses. Experimental results (or other kinds of high-quality data) not only provide evidence for and against specific hypotheses but also provide support (more indirectly) for or against the guiding framework theory. In this way, the process of hypothesis testing may ultimately lead to results that challenge the framework theory, but scientists will only abandon an existing theory if there is a more viable alternative. In this sense, the process of evaluating hypotheses is not only constrained by available data but also available theories. Thus, even well-supported theories may be revised or changed, although well-supported theoretical frameworks are less likely to change than more fledgling frameworks.

Table 1 provides a brief description of some of the concepts that inform the three framework epistemologies. Our work builds on these analyses and the elaborations provided in Smith, Maclin, Houghton, and Hennessey (2000) that include the intermediate constructions of Levels 1.5 and 2.5. Note that these epistemologies form a developmental sequence in the sense that the succeeding epistemologies are more complex and hierarchically integrated than the previous ones. For example, at Level 2, one does not abandon the goal of making observations but now sees that observations are often made in the service of generating or evaluating hypotheses. Similarly at Level 3, one does not abandon the goal of hypothesis testing but now realizes that hypothesis testing serves the larger goal of theory construction and evaluation. Note also that Level 3 is a framework theory—rather than a fully fleshed out epistemological theory—in that it does not take a stance on the ultimate ontological status of theoretical entities, about which there has been much philosophical debate. Thus, one could craft equally sophisticated "realist" versus "idealist" views of the status of theoretical entities within this framework theory.

Table 1
Description of some of the core concepts in three epistemologies of science

Level	Key Differentiations	Nature of Knowledge	Acquisition Processes	Certainty/ Uncertainty
1	No differentiation of ideas and evidence	True beliefs about what happens and what works	Making observations, doing tests, finding answers	Certain knowledge
2	Simple differentiation of ideas and evidence	Well-tested hypotheses Generalizations about how or why something works	Simple hypothesis testing	Transitional (not fully specified)
3	Differentiation among framework theories, specific hypotheses, and evidence	Well-tested coherent theories (explanatory frameworks)	Cycles of hypothesis testing that indirectly test theories	Theory-based uncertainty

Note. Based on Carey and Smith (1993).

Other researchers have provided somewhat similar characterizations of progressions in students' epistemological views in science, although they have not worked from an intuitive meta-theories perspective and have not assumed that novice views are especially coherent. For instance, based on their extensive investigations of the thinking of elementary- and secondary-school students, Driver and colleagues (1996) developed a framework that distinguishes among three forms of reasoning: phenomenon-based reasoning, relation-based reasoning, and model-based reasoning. Similar to the Carey and Smith framework (1993), these three forms of reasoning have implications for the ways students conceive of the goals of inquiry and the extent to which explanation and description are differentiated. For example, in phenomenon-based reasoning, there is no distinction between explanation and description. In relation-based reasoning, there is a simple distinction, but explanation is seen as still emerging from the same categories as the data. In contrast, in model-based reasoning, explanations are seen as based on theoretical entities that are not part of the simple observation categories. Driver and colleagues stressed that there are contexts in which each form of reasoning is valid, and that students can have different views (or images of science) in different tasks; however, phenomenon-based reasoning was the modal pattern among elementary-school students, and relation-based reasoning was the modal pattern among secondary students. Model-based reasoning was rarely observed in their sample at any age.

Lederman, Abd-El-Khalick, Bell, and Schwartz (2002), working primarily with college students and preservice teachers, made a twofold distinction between more naïve and informed epistemological views, which they characterize in terms of seven interrelated facets. Their characterization of more sophisticated views is similar to Level 3 in Carey and Smith's (1993) system and model-based reasoning in Driver et al.'s (1996) system. Perhaps because Lederman et al. were working primarily with college students and teachers, they did not distinguish among different forms of naïve reasoning (e.g., Levels 1 and 2 in Carey and Smith's system, or phenomenon-based and relation-based reasoning in Driver et al.'s system). They have developed and refined the Views of Nature of Science Questionnaire, which they have carefully validated in multiple ways (Lederman et al., 2002). Their data on college undergraduates indicates that a majority have ''naïve'' views on all seven facets (Abd-El-Khalick & Lederman, 2000). Their scoring system requires consistency (or at least lack of disconfirming evidence) on a given facet to be credited with sophisticated understanding; however, they noted that students with ''naïve'' understandings can be fragmented or inconsistent in their reasoning, although explicit data on the extent of these inconsistencies has not been presented.

At present, although there is much data that students can look more or less sophisticated in different contexts depending upon the amount of scaffolding supplied (Bell & Linn, 2002), it is not always clear that the strategies students use in those situations depend on the students' own epistemological conceptualizations of the problem. Thus, there is remarkably little systematic data that speak to the issue of the coherence or lack of coherence of students' thinking about epistemological issues in situations where they are called on to structure the situation. The goal of this study is to see if there might, in fact, be the kind of coherence as would be assumed by an intuitive meta-theories perspective in these more unstructured situations.

Approaches to Assessment

How might students' science epistemologies be assessed? The present study seeks to assess epistemology using three different types of probes (each with a scoring system that focuses on scoring the level of articulation of underlying *concepts* rather than simple agreement with specific *beliefs*) and to explore the relations among these three assessments. Consistency among multiple probes is one important way to assess the validity of epistemology of science as a coherent construct. That is, if it is a coherent or intuitive theory, then the various aspects of epistemology are facets of a system of thinking, and one would expect interrelationships among them. Of course, this assumption is open to disconfirmation if the different facets are independently scored as in this study. One could find that there was little within-student consistency across different probes.

The first assessment is a modified Nature of Science interview first developed by Carey et al. (1989) in which students are asked explicit questions about the goals of science, nature of hypotheses and theories, and so forth. This is a decontextualized assessment, and students are not given any examples but rather are asked to provide the meaning of key terms and their own examples. Student answers are scored for differentiation of ideas from evidence and theories from hypotheses. The existing instrument has been used with younger children, but has not been used with college students. The extension of this instrument to the college level allows for further articulation of the coding categories (especially for the higher levels) and a fuller developmental analysis.

Our second assessment probes student views by asking them to respond to a provocative statement that expresses a fairly low-level epistemological position focusing on ideas of truth and certainty of knowledge. It also is a decontextualized assessment, but probes students' conception of scientific truth—an aspect not directly probed in Carey et al.'s (1989) Nature of Science interview, but one that is central to progressions from knowledge unproblematic to knowledge problematic perspectives. We wanted a separate, more direct probe and coding system to help further describe students' changing comfort with and conceptualization of uncertainty, especially at the intermediate epistemological levels. The interview probe is inspired by a question used by Belenky, Clinchy, Goldberger, and Tarule (1986), but adapted for science.

The third assessment used in this study presents interviewees with a specific controversy and asks them to reason about it. Students are purposely not given much information about the controversy so that they have to use their epistemological ideas to structure the situation, but it also is more contextualized than the first two sections of the interview. In developing measures in this area, we were inspired by some prior work by King and Kitchener (1994), but adapted it to science and foregrounded students' explanations about the sources of disagreement. Because we wanted to know whether students understood the power of empirical scientific research to test competing explanations, we added a probe about how scientists could resolve the controversy and specifically asked whether they believe the controversy could be resolved through experimentation. Students are asked to describe an experiment that could help scientists resolve the controversy. Our work thus complements the work of other researchers who have presented students with more structured scenarios that involve students in reasoning about specific data, to probe their understanding of scientific controversies (e.g., Driver et al., 1996; Sadler & Zeidler, 2004).

This last portion of the interview, where students are asked to reason about a scientific controversy, acts in an interesting way as a validation of our more decontextualized measures. If the decontextualized measures assess the salience of certain epistemological ideas for students, and these underlying epistemological ideas are part of a coordinated system of thinking and not just isolated pieces of verbal knowledge, then all three measures should be highly interrelated. That is, the salience of certain epistemological ideas for students should affect how they structure their thinking about a controversy, especially when they have limited information about that controversy.

Current Hypotheses

Based on previous research using the Nature of Science or similar interviews with high-school students (e.g., Driver et al., 1996; Honda, 1994, 1996; Sandoval & Morrison, 2003), we hypothesized that most college freshmen would have made a simple differentiation between scientists' ideas and evidence, but would not have made a deeper differentiation between scientists' theories and hypotheses wherein theories are seen as explanatory frameworks guiding hypothesis testing. That is, students would generally show Level 2, but not Level 3, insights in the Nature of Science Interview.

We further hypothesized that the conceptual reorganizations that occur as students come to clearly differentiate ideas and evidence both support and are supported by changes in students' ways of thinking about the reasons for scientific uncertainty. More specifically, coming to appreciate the inherent uncertainty of scientific knowledge should go hand in hand with having achieved a fairly consistent differentiation between ideas and evidence (Level 2 epistemology). Initially, this would be a simple inductive uncertainty (in keeping with students' more limited conceptions of hypotheses as constrained by data rather than theories at Level 2) rather than a deeper interpretive or full-blown theory-based uncertainty (associated with students' greater appreciation of the role of theories in constraining inquiry at Levels 2.5 and 3). In making this hypothesis, we propose that it is important to give a more nuanced analysis of student thinking about uncertainty than has traditionally been provided in the scientific epistemology literature. Previous work has often assumed that students' initial differentiation between ideas and evidence does not undermine their belief in the certainty of scientific knowledge and that awareness of uncertainty comes primarily with a sophisticated epistemological stance (Carey & Smith, 1993; Driver et al., 1996); however, we believe that there should be important changes in understanding of uncertainty at each epistemological position, and that consistent with the analysis provided in the personal epistemology literature, an intermediate stage akin to "radical relativism" should bridge between more naïve and sophisticated views in science as well.

Finally, we predicted that the way students' reason about specific controversies (in the absence of detailed knowledge about those controversies) would be related to their epistemological views. For example, students with simple Level 2 understandings of hypotheses and theories should have more simplistic views of controversy as stemming from sampling and generalization problems rather than from interpretive problems or theory-based differences. In addition, they should expect that controversies are resolvable primarily by gathering more data, without awareness of the strengths and limitations of different research designs. Carey and Smith (1993) suggested that developing a more adequate epistemology of science (which includes understanding there can be alternative interpretations of the same pattern of data) goes hand in hand with gaining important science process skills (e.g., understanding the logic of the design of controlled experiments).

Method

Participants

Thirty-five college freshmen were given an extensive, three-part interview during the first few weeks of the fall semester. The entire interview was completed in one session, averaging about 45 min. Each interview was tape recorded, transcribed, and blinded for data analysis. Eighteen of the students were from a private alternative liberal arts college (College). Seventeen of the students were from a large public university (University).

College enrolls approximately 1,200 students, including an average incoming class of 300 first-year students, 12% of whom are students of color and 3% of whom are international students. Students' average high-school grade point average is 3.36. The college does not require SAT scores for admission. Approximately 84% of applicants reported their SAT scores. Of those students admitted, the middle 50% SAT scores ranged from 600 to 700 on Verbal and from 540 to 640 on Math. Fifty-six percent of the alumni go on to graduate studies.

The University enrolls approximately 18,000 undergraduates, including 3,300 first-year students of whom 18% are students of color. Students' average weighted high-school grade point average is 3.42. On average, first-year students ranked in the top quarter of their high-school class and had combined SAT scores of 1123. Fifty-three percent of graduating seniors reported plans to attend graduate school within the next 2 years after graduation.

Equal numbers of male and female students were selected randomly (using random numbers) from course rosters stratified by sex. All University students were selected from survey courses for students intending to major in science. Since the first year courses at College are mixed in terms of students' intended majors, the sample of interviewees from College contains both science and nonscience majors.

Assessment Instruments

Part 1 used a slightly modified form of the Nature of Science interview (Carey et al., 1989; Smith et al., 2000). It probed student ideas about the goals of science, the nature of scientific questions, experiments, hypotheses, and theories, the role of ideas in scientists' work, the relation between hypotheses and theories, and whether and when scientists might change their hypotheses and theories. One way it was elaborated from previous versions of the interview was to probe more fully for students' conceptions of the relation between hypotheses and theories, including asking students to provide specific examples to justify their general points.

This interview has been previously given to middle- and high-school students and has an established coding system for responses to each question. Level 1 responses use concepts that do not require the differentiation of ideas and evidence. For example, students talk of the goal of science as "making discoveries" or "doing experiments," give examples of scientific questions as factual questions about what happens or how to do something, and describe the purpose of experiments as finding answers or what works. Level 2 responses reveal a simple differentiation between scientist's ideas and evidence, using the new concepts of explanation or hypothesis testing. For example, students now see the goal of science as either testing one's ideas or explaining why things happen or how things work. They give examples of scientific questions that show concern with explanation or with finding evidence for their ideas, and they now describe experiments as ways to test ideas or hypotheses. Level 3 responses reveal a differentiation between a theory (as a coherent explanatory framework) and the specific hypotheses that one formulates and investigates (in the service of developing the theory). For example, the larger goal in science is now seen as building coherent theories of ever greater explanatory power, and involves cycles of careful hypothesis testing to flesh out the theoretical framework. Experiments are devised as direct tests of specific hypotheses that indirectly provide tests of the theories themselves. Theories although well tested, are still always open to revision based on their explanatory power compared to their rivals. At Levels 1.5 and 2.5, students are beginning to make the relevant differentiations for the subsequent level, but do not yet have fully articulated understandings (see Appendix B for a more complete description of all the levels and coding categories associated with each).

Prior to undertaking this study, we piloted the revised interview with college juniors and seniors, and elaborated on the coding system in light of their responses. We developed coding grids for each question (identifying common ideas expressed by students and categorizing them as at a particular level based on a conceptual analysis of their fit with different epistemological positions) and then coded a small subsample of the present cohort. We finalized our coding grids based on scoring that subsample before coding the remaining students. Each of the questions was thus scored both qualitatively (for the specific idea expressed) and quantitatively using the five-level category system (Levels 1, 1.5, 2, 2.5, and 3).

Part 2 probed student understanding of the uncertainty of scientific knowledge. Students were given the statement:

> Everyday, in more and more areas of science, the right answer is known. In areas where the right answer is known I look to experts to tell me what is right. In areas where no right answer is known, I think anyone's opinion is as good as another's.

They were then asked whether they agreed or disagreed with the statement, and the reasons for their agreement or disagreement. They also were asked where they go when they have questions about a scientific issue, what they do if they find disagreement among sources, and what, if anything, makes one answer better than another.

Coding grids for these questions were developed based on pilot data with other college students and a small subsample of the present cohort, but independently of any knowledge of what students had said in the Nature of Science Interview. More specifically, we described the specific ideas that were expressed in response to these probes and then categorized different responses as reflecting the different levels based on a conceptual analysis of how well that idea fit with a given level. For example, some students said simply that they thought that science knew the right answers with no further elaboration, qualification, or doubt. This was scored as consistent with a Level 1 epistemology because if ideas and evidence are undifferentiated, scientific knowledge is simply the facts, the right answers, the way things are. In contrast, other students expressed the belief that all scientific knowledge was fundamentally uncertain because we could always get disconfirming evidence at a later date. This was scored as consistent with a Level 2 epistemology in which there is a simple differentiation between ideas and evidence. That is, if students realize that scientists are concerned with making generalizations that go beyond the data, they should realize that data provide evidence for, but does not prove, those generalizations (a kind of "inductive" uncertainty). A Level 3 epistemology, in which there is a distinction between the general concepts of the theory and the specific ideas or hypotheses formulated using that theory, would allow for recognition of a deeper, ‘‘theory-based’’ uncertainty. Scientific knowledge is theory relative in the sense that there can always be another theory that might provide a more satisfactory or deeper account of all the data.

Student responses in Part 2 were scored quantitatively (for Levels 1, 1.5, 2, 2.5, and 3) and qualitatively (for specific idea at that level) on two scoring grids: (a) one for their ideas about the certainty/uncertainty of scientific knowledge and (b) one for ideas about how knowledge is justified in cases where experts disagree.

Part 3 of the interview presented students with a specific scientific controversy. Two controversies were used with half the students receiving one about the efficacy and safety of fluoridation of water in preventing tooth decay and the other half one about the efficacy and safety of using Echinacea in preventing colds. Students were asked why the scientists might disagree, how they might try to resolve the controversy, and what makes the controversy difficult to resolve. We developed these situations, borrowing on previous work by King and Kitchener (1994), in an attempt to examine students’ understanding of scientific controversy and their methods of justifying decisions in science.

Again, we developed coding grids for these questions based on a conceptual analysis of the themes expressed in pilot data and with a small subsample of the present group, independent of knowing particular students’ responses on Parts 1 and 2. Student responses were scored quantitatively (for level) and qualitatively (for specific idea) on two separate grids by analyzing whether student answers embodied Level 1, 1.5, 2, 2.5, or 3 concepts. One grid concerned the reasons for the controversy; the second grid concerned the way the controversy might be resolved.

Appendix A provides a complete script for the three parts of the interview. Appendix B provides a more detailed description of the coding levels used for each part of the interview.[3]

Two coders independently coded all interviews, which were blinded for course, professor, and institution. Both coders were highly experienced in using the five-level category system for coding data in the Nature of Science Interview, based on their experience analyzing such data from middle-school students. Hence, they already shared an understanding of the important conceptual distinctions embodied at the different levels. Each of the three parts was scored without knowledge of what had been said by that participant on the other part. In addition, the coders had not given any of the interviews. Coders used grids for each question to check off all the distinct ideas that occurred, with the final-level score representing the level of the participant’s highest clear idea on that grid. The two coders had very similar mean scores and ranges, suggesting they had similar ways of applying the coding system. Further, their average scores for individuals (on a given part) were highly correlated: .88 (Part 1), .89 (Part 2), and .84 (for Part 3). Absolute agreement between coders in each assigned level score ranged from 73% for Parts 1 and 2 to 84% for Part 3, with virtually all disagreements being between adjacent levels. Final-level scores for each participant on a given grid reflected the average of the two coders’ scores.

Results

Table 2 shows the means and range of scores across the three parts of the interview. In keeping with our expectations, college freshmen averaged at (or slightly below) a Level 2 epistemology; however, individual student averages ranged broadly from a Level 1 to a Level 2.5 epistemology, and student responses on the three parts were related in significant and interesting ways. In the following sections, we highlight what students said in each part of the interview as well as some of the key relations among the parts.

Table 2
Average scores across the three parts of the interview

Epistemology Questions	Average Score M	Average Score Range
Part 1: Nature of Science	1.81	1.38–2.23
(a) Questions 1a–3b	1.68	1.17–2.04
(b) Questions 4a–7b	1.92	1.50–2.39
Part 2: Uncertainty	1.80	1.00–2.75
Part 3: Controversy	2.00	1.50–2.50

Part 1. Nature of Science Interview

The Nature of Science Interview probes the extent to which students are aware of the guiding role of ideas in scientist's work by differentiating (a) the scientist's ideas from the evidence and (b) the scientist's broader theories from specific hypotheses. The first part of the interview (Questions 1a–3b) is less scaffolded as it asks about goals, questions, and experiments, but does not bring up the ideas of hypotheses and theories. The second part of the interview (Questions 4a–7b) more directly probes for students' conceptions of hypotheses and theories and the extent to which they differentiate and interrelate them.

All but 1 of the students answered at least one question in the Nature of Science interview from the perspective of a simple Level 2 epistemology in which ideas and evidence are differentiated; however, students varied in how consistently they adopted this perspective, often slipping into a simpler Level 1 or Level 1.5 way of talking on a given question, especially on the first half of the interview. A minority of students had some Level 2.5 insights, especially on the more scaffolded questions at the end of the Nature of Science interview.

Table 2 shows the overall differences in student scores between the two halves of the interview. Students had a higher mean score on the second half of the interview in which they were asked the more scaffolded questions about hypotheses and theories (1.68 vs. 1.92, $p < .0001$, Wilcoxon signed-ranks test, two-tailed). In general, student scores for individual questions on the first half of the interview ranged from simple Level 1 to simple Level 2. In contrast, student scores for the second half of the interview ranged from Level 1.5 to Level 2.5. Thus, simple Level 1 responses virtually disappeared on the second half of the interview while Level 2.5 responses emerged in the second half of the interview. Table 3 shows the distribution of responses for seven key questions throughout the interview (three from the first half of the interview; four from the second).

Overall Consistency of Responses. Clearly, there was some variability in students' level of responding from question to question, reflecting in part the degree of scaffolding provided by the question. Nonetheless, by averaging student scores across all 13 questions in the Nature of Science Interview, we can get one index of the consistency with which students were differentiating between theory and evidence throughout the interview. When we did this, we found that 10 students had average scores between 1.38 and 1.67, with most of their responses at Level 1 or Level 1.5. For these students, clearly differentiating theory and evidence was not the norm. In contrast, 12 students had average scores between 1.92 and 2.23, with most of their responses at Level 2 or Level 2.5. These students thus provided strong evidence of consistently differentiating between theory and evidence throughout the interview. Finally, 13 students had average scores between 1.71 and 1.87, and were more evenly split between Level 1.5 and 2 responding.

Table 3
Distribution of scores to selected questions across the Nature of Science Interview

Level	Goals (1a)	Questions (2a)	Experiments[a] (3a)	Hypotheses (4a)	Theories (5)	Hypothesis/ Theory Relation (6b)	Change (7a)
1	8	1	6	0	0	1	3
1.5	8	8	10	11	8	10	5
2	19	26	18	23	23	16	23
2.5	0	0	0	1	4	8	4
3	0	0	0	0	0	0	0

(Columns 1a–7a: Selected Question)

[a]One student did not give a codable response on Question 3a; hence, *n* is 34 rather than 35 for this question.

Student Conceptualizations of Hypotheses and Theories. Exactly how did students conceptualize hypotheses and theories on these more directed probes? No student had a simple Level 1 response in which hypotheses/theories are entirely conflated with an experimental procedure (how to make something happen) or an experimental result (what happens). Instead, all had at least the Level 1.5 idea that hypotheses were ‘‘educated guesses’’ about what will happen in a certain situation or experiment. In this way, they were all making some differentiation between the scientist’s ideas and results, as they were aware that these guesses could turn out to be right or wrong; however, students varied in terms of how consistently and clearly they connected their ideas about hypotheses and theories to the Level 2 ideas of explanation or hypothesis testing, whether they saw theories as broader in scope than hypotheses and whether they were aware that scientists’ theories could affect hypothesis testing.

At Level 1.5, hypotheses and theories were either both guesses or both things that could be right or wrong rather than intrinsically defined and related through a process of hypothesis testing. In this sense, ‘‘making guesses’’ might be regarded as an optional step in the scientific method that they had learned about—one that could be omitted without serious consequences—rather than an intrinsic part of the process that guided the design of experiments and the development of theories. In addition, theories and hypotheses were considered in black-and-white terms as being right or wrong rather than as being supported or challenged by lines of evidence. A few students reasoned from this perspective fairly consistently across Questions 4 to 7. For example:

> CHS: (*What is a hypothesis?*) It is an educated guess. (*Can you tell a little bit more about it?*) Well they know small stuff which helps them decide what would be the best way to go about something. But it’s just a guess, they don’t know if it’s actually going to work. (*Example?*) . . . If you’re trying to decide how a chemical is going to react . . . , you can make a guess on that. . . . (*Does a hypotheses influence the experiments that he or she does?*) It shouldn’t if you’re a good scientist . . . things are going to happen because of science, not because some person wants it to . . . . But I suppose if you didn’t have the hypothesis, you wouldn’t know how to go about doing something.
> DIK: (*What is a hypothesis?*) A hypothesis leads into an experiment of science. (*What is a theory?*) I don’t know. I know it’s based on a scientific hypothesis. Maybe that it states the hypothesis. (*Are hypotheses and theories related?*) I think they are. I think a theory is based on if the hypothesis is true. (*Could a scientist’s theory affect his or her specific hypothesis in an experiment?*) I don’t think so.

In contrast, at Level 2, students’ very definitions of hypotheses and theories are now intrinsically connected to their ideas about hypothesis testing or explanation. Indeed, the most common Level 2 response was to see hypotheses as ideas that are to be tested (evaluated) in experiments and to view theories as well-tested hypotheses. Students now have a stronger idea about the central role of hypothesis formation in science: Forming a hypothesis is central to designing an experiment; however, in this view the differentiation between hypotheses and theories is very limited, as hypotheses and theories differ not in content or scope but only in degree of testing and certainty. Hypotheses are initial guesses; theories are more certain because they have been tested, even though scientists might still later find out the theories are wrong. Further, this view allows students to see only a simple one-way relation between hypotheses and theories (i.e., hypotheses lead to theories) rather than the more complex, bidirectional relations (i.e., theories constrain hypotheses; hypotheses provide indirect support for theories). Consequently, when directly probed in Question 6b about whether a scientist's theory could affect his or her hypothesis tested in a particular experiment, these students resisted and argued that it in fact works the other way: Hypotheses lead to theories through a process of testing. They also cautioned that your theory *should not* influence your hypothesis—that shows bad form and a lack of objectivity. The majority of students argued from this position for Questions 4 to 7, which is well exemplified with the following two protocols:

> WAM: (*What is a hypothesis?*) It's basically a theory that you come up with but before you do an experiment . . . and it basically states what you're trying to prove by doing the experiment. (*Example?*) You'd say that the growth rate of the grass varies . . . that sunlight in one of the causes of the rate change. (*What is a theory?*) In the American language, a theory would just be kind of like a hypothesis but like in the scientific world a theory is something that'd I'd say is almost a law but this thing where it can be disproven. (*Are hypotheses & theories related?*) Yeah. Because once you prove a hypothesis and you prove it many times and people agree upon it, it would become a theory if it wasn't a law. (*Could a scientist's theory affect his or her specific hypothesis in an experiment?*) When you're doing the experiment, you're trying to prove the hypothesis and the theory. I mean, in a certain sense, it's the same thing.
>
> JOS: (*What is a hypothesis?*) A scientific hypothesis is when you come up with a statement like "I believe that lilacs will grow better in light than in darkness." That's the statement that you base your entire experiment around. And that's what they're testing in the experiment. They're testing to see if their hypothesis is correct. And after enough experiments, that hypothesis becomes a theorem. (*What is a theory?*) That is a hypothesis that has been tested in many experiments, and it's constantly been proven correct so it becomes a theory. It's basically just a well-tested hypothesis. (*Example?*) I'll go back to the lilacs. They would say "It's my hypothesis that lilacs will grow better in light than in darkness" and they'll test it and the experiment will dictate that the lilacs in the dark die and the lilacs in the light thrive. So they'll test it again and again and again and as they continue to get this response. (*Are hypotheses & theories related?*) In a sense, since a hypothesis is an early form of a theory . . . (*Could a scientist's theory affect his or her specific hypothesis in an experiment?*) I'm not really sure about that because a theory usually comes from the hypothesis, so I would think the theory would come after, so the hypothesis wouldn't be affected.

Finally, a few students were beginning to make a deeper differentiation between a theory and a hypothesis on Questions 4 to 7, in which a theory is seen as broader in scope than a hypothesis. For some, insight about the broader scope of theories first emerged on the interview questions: "What is a hypothesis?" and "What is a theory?" (Questions 4 & 5). For most, however, a statement of the broader nature of theories only came through on the most probed question: "Does a scientist's theory influence the hypothesis he or she tests in a particular experiment?'' (Question 6b). Prior to that time, they primarily talked of theories simply as well-tested hypotheses. The view they articulated was of a theory as a collection of tested hypotheses that provided the *general background* for testing subsequent hypotheses, falling short of the Level 3 understanding of a theory as a *coherent explanatory framework* that constrains the very concepts used in specific hypotheses. The following 2 students exemplify this Level 2.5 perspective:

> HIK: (*What is a hypothesis?*) It is a statement that can be . . . not proven, but attempted to be proved, at least, through data. (*Example?*) Things will fall to the earth because of gravity. And you can get more specific as you get in the relationship to the density of the earth and things like that. And you can get speeds in there and make it sort of more testable. (*What is a theory?*) It's a hypothesis that has been proven to a reasonable certainty . . . . Like covalent bonding and molecules . . . . It's a theory on how things work. But there might be some other weird mechanism operating that they don't know about. So they just cop out a little bit and say ''Well, we're not absolutely sure.'' (*Are theories and hypotheses related?*) Yeah, the hypothesis needs to be tested and tested before it becomes a theory. (*Can the theory influence the hypothesis they test in a specific experiment?*) Yes, because [scientists] rely on the theories to give them background information . . . . To say . . . if I take these theories as true, what else do I want to figure out? So that helps create the hypothesis. (*Example?*) Well we know that things fall to the earth. There's gravity. But we're not sure about anything more. So then the next hypothesis would be ''Okay, is it related to the size of the object, the density of the object—what's going on?''

> MOM: (*What is a scientific hypothesis?*) A prediction that is stated in a manner such that it can be proven, or disproven . . . . (*Example?*) If you are studying how different colored light affects plant growth, you could say: If I expose plants to only green light they will grow better . . . . (*Does a scientist's hypothesis influence the experiments that he or she does?*) Yes, in the fact that the experiments that you're doing should be attempting to prove your hypothesis . . . . (*What is a scientific theory?*) A theory is kind of like a hypothesis, but on a broader scale, usually not as provable because it's so broad in general. For instance, Einstein, with his theory of relativity . . . . Darwin's theory of natural selection . . . . (*How formed?*) Theories require a lot more . . . . You have to form a logic based on several hypotheses, kind of like you're making a large group of assumptions that all do things . . . because they're so broad they tend to be a little abstract. (*Are theories and hypotheses related?*) Yes, I think a lot of it just comes down to the scale, but I think a theory can be based on several hypotheses. (*Does a scientist's theory influence the hypotheses that he or she tests in a particular experiment?*) Yes, I think so. (*How?*) I think that if I was a scientist and I had come up with a theory . . . . I might be forming hypotheses that would fit in nicely with it . . . . It's kind of like you're devoted to your theory.

Overall, 8 students were scored as Level 2.5 on Question 6b. Almost all (7 of 8) had at least one other Level 2.5 response in the Nature of Science interview; however, no student articulated a full Level 3 understanding of a theory as a guiding and constraining conceptual framework. Indeed, no student even articulated a consistent Level 2.5 view across the entire Nature of Science interview.

Part 2. General Probe about Scientific Truth

Part 2 of the interview probed whether students believed science has the right answers and how students decide what to believe in cases where the right answer is not known. At issue was whether they saw scientific knowledge as fundamentally certain or uncertain, and if uncertain, the

Table 4
Distribution of responses on the four grids in Parts 2–3 of the Interview

Level	Part 2		Part 3	
	Certainty/Uncertainty of Knowledge (Grid 8)	How Decide if Disagreement (Grid 9)	Reasons for Controversy (Grid 10)	How Resolve Controversy (Grid 11)
1	12	3	0	0
1.5	8	10	6	10
2	10	17	25	17
2.5	5	5	4	8
3	0	0	0	0

deeper reasons for uncertainty (inductive uncertainty, interpretative uncertainty, or theory-based uncertainty). Also at issue was whether they appealed to firsthand evidence (rather than simply the beliefs of experts) in resolving disagreements, and if so, the degree of sophistication they showed in reasoning about evidence. For example, did they focus (in an uncritical way) simply on the existence or amount of firsthand evidence? Were they beginning to be aware of the strength of evidence based on a more detailed understanding of experimental design? Finally, were they also considering the strength of the larger theoretical perspective that was informing the design of specific experiments?

Table 4 shows the distribution of responses on the two grids for this part of the interview: Grid 8 (Certainty of Knowledge) and Grid 9 (How Decide if Disagreement). Responses ranged from Level 1 to Level 2.5. Tables 5 and 6 give examples of responses at each level.

In response to the probe about whether science knew the right answers (Certainty/Uncertainty of Knowledge, Grid 8), 34% of the students thought that there were definite right answers in science (Level 1); another 23% thought that there were many definite answers in science, although there were some things that were opinions (and hence could not be right or wrong) or some things about which we might be mistaken (Level 1.5). In contrast, 43% of the students (15 of 35) made a general statement that there were not absolutely right answers in science. Among the latter group, about two thirds gave simpler Level 2 reasons for uncertainty, consistent with an idea of "inductive uncertainty." That is, they argued scientists have often been mistaken about things in the past, such as when they thought that the earth was flat, so they could continue to be mistaken about things in the future; or scientists can always find disconfirming evidence at a later date. The other third gave deeper Level 2.5 reasons for uncertainty, consistent with the idea of interpretive uncertainty, arguing that there could always be other possible interpretations or explanations of a specific pattern of data, even if some specific facts have so much support that they are likely to be true (a view of uncertainty that pervades Levels 2.5 and 3). No student, however, articulated the full-blown, theory-based view of Level 3 or talked of the fundamental uncertainties of scientific theories.

In response to the probe about what they do if they find disagreement among experts and is there anything that makes one answer better than another (How Decide if Disagreement, Grid 9), 38% of the students gave Level 1 or 1.5 responses in which they either had no basis for justification (Level 1) or deferred to more knowledgeable sources or to what the majority of scientists believed (Level 1.5). Thus, at best, they focused on the knowledge and beliefs of the experts rather than examining the evidence that supported those knowledge or beliefs. In contrast, the other 62% of students (22 of 35) indicated their commitment to resolving disagreements through consideration of the scientific evidence (Level 2 or 2.5). Of these, most ($n = 17$) focused on fairly simple considerations of evidence as in "Do the experiment yourself," "Read the research articles" (but with no discussion of what they would look for in reading the articles), or "See which position has

Table 5
Certainty/uncertainty of scientific knowledge (Grid 8): Descriptions of coding levels with examples of student responses

Level	Description	Examples of Student Response
1	Temporary uncertainty: Knowledge is certain, but some things might not yet be known.	(*Do you think science knows the right answers*?) Not all of them, but they are finding most of them out, I think. (*Do you agree with this statement*?) Yes I think I do. (*Can you say why*?) Well, every day, scientists are finding out other things they didn't know before, and they're coming up with solutions to things. And if a right answer is not known, all you can do is guess, till it's right or close to it.
1.5	Partial uncertainty: Knowledge is certain in some areas; uncertain in others	(*Do you think science knows the right answers*?) Sometimes but because several times scientific theories have been changed it causes us to wonder whether or not scientists really know what they're talking about. (*Do you think science knows the right answers*?) No not really. I think they have really good ideas to what they think is true. Maybe in some cases they have the right answers but I guess it depends on the questions. There are some questions that can have right answers and there are questions that can't.
2.0	Inductive uncertainty: All knowledge is uncertain because can be always disproved in the future.	(*Do you think science knows the right answers*?) That's a very good question. No, not entirely. Because, I mean, in the past, in the 1600s, or whatever, it may have been, science thought that the world was flat. And that was our science. So I guess scientists kind of like, the way we did it, we just kind of understand things. So to us it may be right, but in the future it could be proven wrong. (*Do you think that science knows the right answers*?) I don't think that science ever will know the right answer, I think that it's the best known answer at the time, which is always open to challenge, and if somebody finds something that's more accurate then that would be suddenly the right answer at the moment.
2.5	Interpretive uncertainty: All knowledge is uncertain because there can be many different explanations consistent with the data.	(*Do you agree with this statement*?) Well, it's kind of a little unscientific because they're talking about the right and wrong answers. Because science . . . never sort of claims to be the absolute right answer, or anything like that . . . . From a scientist's point of view it'd be really bad to just take the expert's opinion for verbatim and not just question it at all, because scientists like to question things. (*Do you think that science knows the right answers*?) I'd like to think so for the most part. But I mean it's not an absolute right answer, but it's a really good way of interpreting the world. (*Do you think science knows the right answers*?) I think science probably has a good idea but as far as concrete right answer I don't know if anyone really has the concrete right answer to anything because it's always open to interpretation. And each individual views things differently and has their own ideas as to how things work.

the most experiments in favor or the 'most proof' or 'most reasons.' " Only a few ($n = 5$) espoused more sophisticated Level 2.5 positions, in which they acknowledged the need to read the primary source material itself to look for design flaws, to consider the quality of the evidence and underlying argument, and/or even consider ways they might develop a new position that synthesizes the two opposing points of view.

Table 6
How to decide if disagreement (Grid 9): Description of coding levels with examples of student responses

Level	Description	Examples of Student Response
1	No real basis for justification: Go with whatever feels right or what authorities say	(*What makes one answer better than another*?) Just whatever seems right. (*And that is based on*?) What I know already, what other people tell me. (*What do you do when you find disagreement*?) Go back to the teacher for help. (*Is there anything that makes one answer better than another*?) Sometimes the book could be old, so the teacher might know something that's newer; he might have the right answer.
1.5	Resolve based on more knowledgeable source, more reasons, majority opinion	(*What makes one answer better than another*?) The experience and the knowledge that it's based on. (*Is there anything that makes one answer better than another*?) Yeah I think if you have more reasons for why you have an opinion that makes it better.
2	Resolve based on firsthand evidence or data from studies, but with no consideration of the quality of the evidence	(*Is there anything that makes one answer better for you*?) If I can carry out the tests myself, obviously that would make—prove the argument to me. But I'm not a scientist, so it's not likely that I'm going to test everything to find out. But if I know that several scientists have carried out the experiment usually I'll accept the theory. (*What makes one answer better than another for you*?) What has the most proof behind it . . . The evidence behind it, I guess. Like I'd ask the teacher, all right, why do you think this or that? And why do you think these people are wrong? And just weigh it out.
2.5	Resolve based on analysis of the quality of data; be open to integrating diverse viewpoints	(*Is there anything that makes one answer better than another . . .* ?) Things that would make one better than another would be if in one of them there was flawed scientific methods used or if there were errors within the experiments . . . . Or if you can find ways in which the experiment was skewed one way to create results. Or statistically they used certain slightly incorrect or less appropriate models or tests. (*What do you do when find disagreement*?) Pay attention to both. Be open about what each one says . . . because even though they may be conflicting it doesn't necessarily mean they can't work together.

In general, students' ideas about the certainty/uncertainty of knowledge were related to their beliefs about how to deal with disagreement (see Table 7). Being aware of the importance of evidence in resolving disputes (Levels 2 and 2.5) was more common among students who viewed scientific knowledge as uncertain (Levels 2 and 2.5) than among those who viewed scientific knowledge as certain (Levels 1 and 1.5), a relation that approached significance (Fisher exact test, $p = .07$, one-sided). Indeed, the majority of students who did not focus on evidence at all in resolving disputes believed in the certainty of scientific knowledge. Further, there were differences between simple Level 2 and Level 2.5 responders. Those who believed in inductive uncertainty (Level 2) focused (at best) on the *existence* or *amount* of firsthand evidence (Level 2) in evaluating competing views while those who believed in interpretative uncertainty (Level 2.5) focused more on evaluating the *quality* of evidence and strength of argument or trying to synthesize competing points of view (Level 2.5) (Fisher exact test, $p = .004$, one-sided).

In sum, Part 2 of the interview highlighted that there was a range of ways in which students think about uncertainty in science. At Level 1, where scientific knowledge is conceptualized as

Table 7
Relation between students' beliefs about certainty and ways of resolving disagreements

	How Decide When Sources Disagree		
Certainty of Knowledge	Based on Majority Opinion or Amount of Knowledge (Level 1/1.5)	Based on Amount of Firsthand Evidence (Experimental Results) (Level 2)	Based on Quality of Evidence; Be Open to Reshaping Views (Level 2.5)
Some Knowledge Certain (Level 1/1.5)	10	9	1
Inductive Uncertainty (Level 2)	2	8	0
Interpretive Uncertainty (Level 2.5)	1	0	4

factual, students acknowledge temporary uncertainty in the sense that there are some answers that might not be known yet but in principle everything is knowable. At Level 1.5, students are beginning to consider that some claims in science might be uncertain, although they still consider that most are knowable with certainty. At Level 2, as students move to see that science is concerned with forming deeper generalizations about what happens and why, and that such generalizations necessarily go ''beyond the information given,'' students now consider knowledge claims in science to be fundamentally uncertain—a kind of inductive uncertainty because one can always find some disconfirming evidence at a later date. At Level 2.5, students are more aware that science involves thinking about how best to interpret or explain a pattern of data. Given that there is always more than one way to interpret a pattern of data, there must be inherent uncertainty about what is the best interpretation or explanation. Significantly, only a few freshmen were beginning to think about such deeper interpretive uncertainty in science. It was more common to embrace a belief in temporary, partial, or inductive uncertainty.

Part 3. Reasoning about a Specific Scientific Controversy

Part 3 of the interview presented students with a specific controversy about whether fluoridation of water prevents tooth decay and/or is toxic or whether Echinacea prevents colds and/or causes cancer.[4] They were told that some studies indicated that fluoridation/Echinacea prevented tooth decay/colds while others did not. They were then asked how this was possible (Reasons for Controversy, Grid 10) and how they might go about resolving the disagreement (How Resolve Controversy, Grid 11). In probing their ideas for resolving the controversy, they were specifically asked to describe what a particular experiment might look like. At issue was whether students would appreciate the deeper interpretive uncertainties of scientific studies and understand how to design studies that help to distinguish among conflicting interpretations. For example, would they realize that it is possible that fluoridation or Echinacea was not effective at all, despite some positive studies, due to some confounding variable in these studies, and propose a more controlled experimental design? Such a position would be consistent with a Level 2.5 epistemological stance. Alternatively, students might focus on more superficial inductive uncertainties: Many variables affect the results, therefore, it is hard to know for sure; the treatment probably worked for some people or subgroups, but not for others. The challenge, therefore, is simply to get more positive evidence that it works, not to distinguish among rival interpretations—an approach more consistent with a Level 2 epistemological stance.

Table 8
Reasons for controversy (Grid 10): Descriptions of coding levels with examples of student responses

Level	Description	Examples of Student Responses
1.5	Not enough data: can't determine trend	I suppose their experiences have been inconclusive. They hadn't been able to set a very steady trend.
	Different people/plants: used different things (no elaboration of specific variables)	It could be something within the people. Some people may react differently to Echinacea. The studies were different. They could have done something differently; they might have researched a different group of people.
2	Different Variables: Scientists have looked at different things or used different methods (different, specific variables can affect results)	Maybe they've done their studies in areas where there are other variables, because some people are going to be more prone to tooth decay than others. Just depending on the amount of people and kind of people in their study . . . Also how much fluoride you are putting in the water and what ages. Because they're looking at different populations . . . .They would work with different groups of people. I don't think that their experiments are entirely the same. And there's also genetic dispositions to tooth decay. I don't think that everyone's doing the exact same research on the exact same thing . . . some disagreement is bound to occur . . . .
2.5	Different Interpretations: May be entirely mistaken or faulty interpretation because of failure to control for variables	It can't be that for some scientists it worked, and for some it didn't, or that it does prevent colds for some people and not others—that's not a logical conclusion. The conclusion one could reach is that is has to do with how you gathered the data. Definition of preventing a cold (One scientist might be seeing if it lessens the symptoms; others scientists might be looking for complete prevention.) I think it basically comes down to how you look at it. (*Why disagree about whether causes cancer?*) It's very difficult to isolate one environmental factor as causing a disease in people, at least. If you're talking laboratory animals you could put them in an environment exposing them to nothing but Echinacea, but with people you kind of have to let them out. . .and they're exposed to other environmental factors, it's really tough to say whether one particular thing caused the cancer.

Student responses varied from Level 1.5 to Level 2.5, with a modal response of Level 2 for each probe (see Table 4). Tables 8 and 9 give descriptions and examples of student responses at each level for both coding grids.

All students acknowledged that a controversy could exist. In this sense, they went beyond simple Level 1 responses in which they would have failed to engage with the question, denied that such a controversy could exist, or asserted their own viewpoint; however, a few showed only a fairly minimal understanding of the reasons for the controversy, locating the reasons for the controversy in terms of simple limitations in sample size (Level 1.5). They expected variability in the data because people are different and respond differently to things (or Echinacea plants are not all alike and might not all have the same effects). Scientists might not have studied enough people (or plants) to detect a clear trend. These responses were scored as Level 1.5 because they expected the controversy would be resolvable rather simply by testing more people or gathering more data. Thus, there was as yet no appreciation of the need to design or carefully craft research studies beyond having a large sample size.

The majority of answers were from a more sophisticated Level 2 epistemology, in which they acknowledged that the scientists might have looked at different things or used somewhat different

Table 9
How resolve controversy (Grid 11): Descriptions of coding levels with examples of student responses

Level	Description	Examples of Student Responses
1.5	Do more tests with larger sample	Carry out tests and take the majority . . . .The larger number of people you do the experiment on the more accurate your result will be.
		More studies . . . .It needs to have a large sample group.
2	Decrease the variables: test similar people; use similar methods	Have standardized tests, maybe test larger groups of people, people maybe in a closer area or closer age groups. . .I would suggest finding people with very similar, or somewhat similar backgrounds and similar diets too . . .. I would just try to decrease the variables as much as possible.
	Switch experiments/swap ideas	I think it would be cool if they ran each other's experiments.
		They should try and work together to try to figure out what the real thing is, and if they're going to work apart, then hopefully they try and think of the other person's point of view, so they can get different ideas and backgrounds for conducting different experiments.
	Do simple experiment (no control of other variables)	Just get a group that takes the medicine and a group that doesn't. Check in with them every once in a while to see how their health is and if then they can make a conclusion out of it . . ..
2.5	Do controlled experiment (vary treatment holding other factors constant, or including placebo controls)	Um I guess it would probably be sort of . . . little spots of dividing up the country, basically, and sort of people who drink fluoridated water and who don't and . . . you'd have to quarantine all people off because they couldn't go and get someone else's water or whatever, or filter their own water . . . and sort of feed them just what you want to feed them so they're not eating too much sugar, or they're all eating the same things so it's not a diet issue. And then feed them all the same water . . . .Oh yeah, you'd also have to mandate teeth brushing or something . . . In short, keep everything the same except for the fluoridation or not. So then you'd see if it a effected their rate of tooth decay.

methods. Thus, the limitation was no longer simply in the amount of data but some aspects of scientist-crafted experimental design. In addition, they were no longer simply stating the platitude that ''people are different and respond differently to things'' but were able to articulate some of the variables that might have affected the results (e.g., whether participants brushed their teeth, what their genetic predispositions were, what kinds of food they ate, what the composition of their water was like, etc.); however, they tended to think of these other variables as extra factors that act to modify the effectiveness of the treatment rather than as completely rival causal hypotheses. That is, they did not fully question that the treatment might work for some—the real question was to identify for whom and under what conditions.

In keeping with the progress made in thinking about the reasons for the controversy, students also had more sophisticated (Level 2) ideas about how to resolve the controversy. There were three main kinds of responses that we scored as Level 2. Some called for experimental designs that ''decreased the variables.'' If part of the problem is that scientists have studied different populations and used different methods, then they should make sure that they use the same methods or study similar people. Others called for the scientists to try to ''swap'' experiments. That is, each group of scientists would try to do the other's experiments to see if they could get each other's results. Still others called for a simple experimental design: Give the treatment to some but not others and compare results. Each of these proposals represents an important step forward beyond simply testing more people: It *is* important for scientists to decrease the variables, to be aware of other testing methods, and to include no-treatment comparison groups. The chief limitation of these responses is that they focused on only *one* of these improvements in isolation rather than integrating them. Hence, they stopped short of proposing more sophisticated experimental designs that actually can *distinguish* among rival causal hypotheses.

Only a few students systematically engaged with the controversy in deeper (Level 2.5) ways. Level 2.5 responses are more sophisticated in that they show awareness that scientists could have been *entirely mistaken* in their interpretations of their experimental results due to their failure to consider or control for rival interpretations. Thus, they no longer simply assume that the treatment is effective at least for a subgroup but realize that it potentially might not be effective at all—rival variables might even explain away all positive results. Further, they propose more sophisticated experimental designs to test the scientist's hypotheses. Most typically, these more sophisticated designs call for controlled experimentation in which one simultaneously holds other variables constant while varying whether one gets a given treatment or includes comparisons between treatment and placebo controls. In several other cases, it involves looking for converging evidence from multiple lines of investigation (e.g., combining more naturalistic observational approaches with more experimental, lab-based approaches).

No student thought that the controversy might stem from theory-based differences (Level 3), and only a couple considered doing experiments about the underlying causal mechanism as a way of resolving the controversy.

Table 10 shows that there was a clear relation between student ideas about the reasons for the controversy and how it should be resolved: The majority of students were scored as reasoning at the same level on both issues, and no student showed a discrepancy greater than half a level. Although the expected values in some cells are too small to do an overall chi-square analysis for the table, two follow-up Fisher exact tests confirmed the significance of the main relations. More specifically, students with Level 1.5 reasons for controversy (e.g., not enough data) were much more likely to give simple Level 1.5 ways to resolve the controversy (e.g., test more people) than were students with higher Level 2 or 2.5 reasons for controversy (Fisher exact test, $p = .043$, one-sided.). In addition, students with sophisticated Level 2.5 reasons for controversy (e.g., mistaken interpretations) were much more likely to give sophisticated Level 2.5 ways of resolving the controversy based on controlled experimentation than students with simple Level 2 reasons for controversy (Fisher exact test, $p = .003$, one-sided).

Table 10
The relation between student ideas about reasons for controversy and how resolved

	How Controversy Resolved		
Reasons for Controversy	Test More People (Level 1.5)	Do Similar Experiments; Make Simple Comparisons (Level 2)	Do Controlled Experiments; Test in Multiple Ways (Level 2.5)
Not enough data; different people (Level 1.5)	4	2	0
Looked at different variables (Level 2)	6	15	4
Mistaken/conflicting interpretations (Level 2.5)	0	0	4

Overall, it was striking how few college freshmen considered that there could have been major interpretive errors made by the scientists and showed an understanding of the role of controlled experimentation in resolving these problems of interpretation (Level 2.5); however, such results make sense in light of the limited epistemological understandings displayed by these students in Parts 1 and 2 of the interview. In the next section, we consider how students' responses across various parts of the interview were related.

Relations among the Three Parts of the Interview

Students' scores on all main parts of the interview were significantly interrelated (Table 11). The best predictor of student scores on Parts 2 and 3 was students' average score on Questions 4 to 7 of the Nature of Science interview, the questions that directly probed their conceptions of hypotheses, theories, and their relation (Indeed, their scores on the Nature of Science 4–7 were just as good and sometimes an even better predictor than their overall Nature of Science scores.) Figure 1 shows the scatter plot of the relation between student average scores on Nature of Science Questions 4 to 7 and their average scores on Parts 2 and 3 combined. The overall correlation between student scores on Nature of Science Questions 4 to 7 and their scores on Parts 2 and 3 combined was .80.

As predicted, there were strong relations between students' level of differentiation of scientists' ideas and evidence on the Nature of Science interview and their awareness of the uncertainty of scientific knowledge in Part 2 (see Table 12). No student with an average Nature of Science score less than 1.7 (predominately Level 1 and 1.5 responders) regarded scientific knowledge as fundamentally uncertain in Part 2 of the interview (Level 2 or 2.5) whereas almost all of those with average Nature of Science scores greater than 1.9 (consistent Level 2 responders) did so. Those with average Nature of Science scores between 1.7 and 1.9 were more variable, although the majority still regarded scientific knowledge as at least partially certain. Overall, there was a significant association between being a consistent Level 2 responder on the Nature of Science interview (Nature of Science scores > 1.9) and being aware of the uncertainty of scientific knowledge in Part 2 of the interview (Fisher exact test, $p = .001$, one-sided).

Similarly, there were clear relations between how students reasoned about a specific controversy (in Part 3) and their differentiation of evidence/hypotheses/theories (in Part 1) and their understanding of the uncertainty of scientific knowledge (in Part 2). For example, all students who systematically engaged with the controversy in deeper ways (e.g., by showing awareness that the scientists' may have made interpretive errors) had been aware of the uncertainty of scientific knowledge in Part 2. Further, all of these students had consistently differentiated ideas and evidence in the Nature of Science Interview (Part 1); they also had at least one and often several Level 2.5 scores on this part as well.

Table 11
Correlation between different parts of the epistemology interview

Part of Interview	1a	1b	2	3
1a. Questions 1a–3b	—	$.60^{***}$	$.61^{***}$	.24
1b. Questions 4a–7b			$.72^{***}$	$.66^{***}$
2. Part 2: Uncertainty				$.48^{**}$
3. Part 3: Controversy				—

$^{**}p < .01$, two-tailed, Pearson correlation.
$^{***}p < .001$, two-tailed, Pearson correlation.

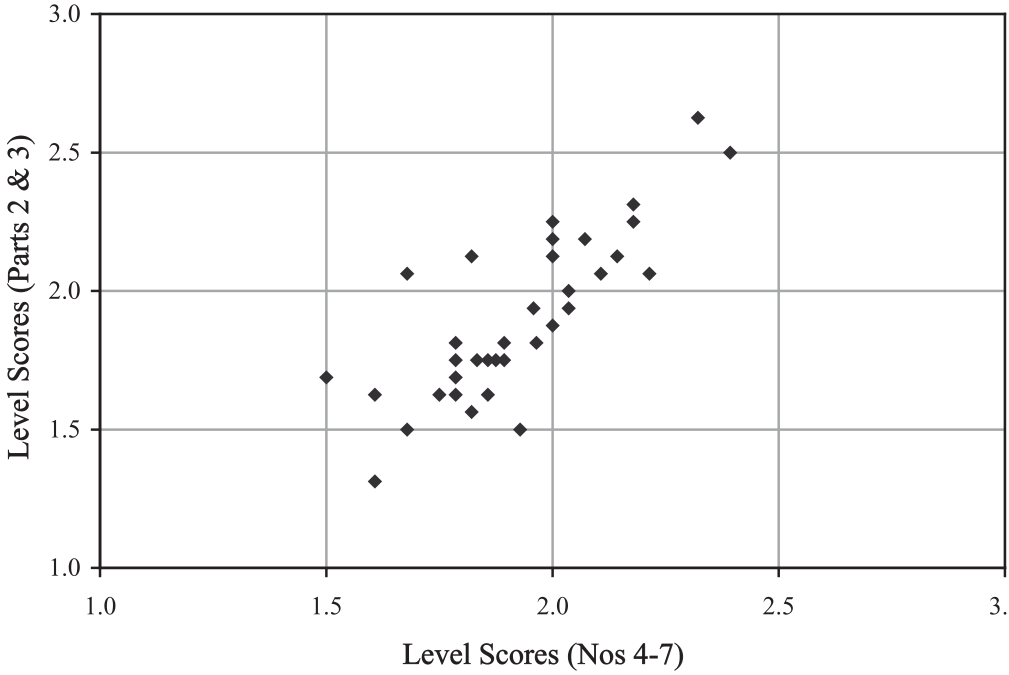

Figure 1. The relation between level scores on Questions 4 to 7 of the Nature of Science Interview (Part 1) and the level scores on uncertainty and controversy (Parts 2 & 3).

Comparison of the Two Samples

Table 13 compares student responses across all parts of the interview for the two samples: students at the small, liberal arts college and students at the large public university. As can be seen, the mean scores for students at the college were consistently higher across all parts of the interview; however, the differences tended to be small (i.e., less than one fifth of a level) for all parts of the interview except for Part 2, where the difference was a full half level. In addition, there was a broad (and generally similar) range of student responses at both colleges.

To test that the strong pattern of correlations among epistemological measures obtained in our previous analyses was not an artifact of the differences in means between samples, we examined the pattern of correlations among students at each school separately. This analysis confirmed that

Table 12
Relation between average Nature of Science score in Part 1 and awareness of uncertainty in Part 2

	Awareness of Uncertainty in Part 2			
Average Score	Certain Knowledge Level 1	Some Certain; Some Uncertain Level 1.5	Inductive Uncertainty Level 2	Interpretive Uncertainty Level 2.5
<1.70	8	2	0	0
1.70–1.90	4	4	4	1
> 1.90	0	2	6	4

Table 13
Comparison of average scores and range of scores for the two samples

Epistemology Questions	College		University		Comparison of Means	
	M	Range	*M*	Range	Difference	*p*
Part 1: Nature of Science	1.89	1.62–2.23	1.72	1.38–2.12	.17	.01
(a) Questions 1a–3b	1.77	1.33–2.04	1.58	1.17–2.04	.19	.02
(b) Questions 4a–7b	2.00	1.68–2.39	1.85	1.50–2.21	.15	.03
Part 2: Uncertainty	2.06	1.50–2.75	1.54	1.00–2.00	.52	<.001
Part 3: Controversy	2.04	1.50–2.50	1.96	1.50–2.50	.08	n.s.

the pattern of correlations was highly robust for students at each school and quite similar to the correlations we obtained in the analysis of the full sample (see Table 14). As in that analysis, Nature of Science scores on Questions 4 to 7 were especially highly correlated with scores in both Parts 2 and 3. Further, the overall correlation between Nature of Science Questions 4 to 7 scores with scores on Parts 2 and 3 combined was .78 for the College sample and .76 for the University sample.

Discussion

In this study, we characterized students' science epistemologies when entering college to assess whether there was some coherence in students' thinking about these issues, as the intuitive meta-theories perspective would suggest, and to obtain baseline data for a longitudinal study we will be conducting with these and other cohorts of students as they move through college. Our interviews probed (a) students' differentiation of scientists' ideas from evidence and hypotheses from theories, (b) their understanding of the inherent uncertainty of scientific knowledge, and (c) their reasoning about scientific controversies. These aspects have previously been studied separately and typically have been scored with noncomparable rubrics. Thus, our work required not only the creation of additional tasks but also the revision of existing, or development de novo of, scoring rubrics that are coordinated in their assessment of the concepts underlying different epistemological stances. The results reported here paint a picture of entering college students' thinking about science that is not only in keeping with the literature on any one aspect but also very rich in its description of their thinking across the three different dimensions of scientific epistemology. What is more, the intercorrelations among the three aspects of epistemology

Table 14
Correlations between Part 1 (Nature of Science Questions 4–7) and Parts 2 and 3 of the Interview for the full sample and each school

Part 1 Sample	Other Part of Interview	
	Part 2	Part 3
Questions		
4–7 Full sample	$.72^{***}$	$.66^{***}$
4–7 University	$.70^{**}$	$.61^{**}$
4–7 College	$.66^{**}$	$.69^{**}$

$^{**}p < .01$, two-tailed, Pearson correlation.
$^{***}p < .001$, two-tailed, Pearson correlation.

provide some validation of our measures and of the existence of different coordinated systems of concepts for thinking about science.

Key Findings on Individual Aspects of First-Year College Students' Epistemologies of Science

The Differentiation of Scientists' Ideas from Evidence and Theories from Hypotheses. Previous work has suggested that from middle school to high school, students make progress in differentiating theories from evidence (Carey & Smith, 1993; Driver et al., 1996; Honda, 1994, 1996; Kuhn, 1991; Sandoval & Morrison, 2003). Our data are consistent with that suggestion, although they underscore that for many college freshmen that differentiation is still incomplete. That is, although the modal response for most questions was in terms of Level 2 epistemology in which scientist's ideas and evidence are differentiated, only 34% of the college freshmen in our sample were *consistently* giving Level 2 responses across the entire Nature of Science Interview (i.e., average Nature of Science scores of 1.9 or greater). The remaining students were either primarily Level 1/1.5 responders (i.e., average Nature of Science scores = 1.3–1.7) or split between Level 1.5 and 2 responding (i.e., average Nature of Science scores = 1.7–1.9).

Furthermore, our data suggest that although college freshmen are beginning to differentiate scientists' ideas from evidence, most do not yet differentiate between scientists' theories and hypotheses by conceptualizing theories as the broader framework that shapes the formation of more specific testable hypotheses. Indeed, only 23% of the students acknowledged that theories *could* influence the specific hypotheses scientists test in a particular experiment (in our most scaffolded Question 6b) and provided an explicit example that showed they at least thought of the theory as broader in scope than the hypothesis. The remaining students either totally conflated theory and hypothesis (i.e., the theory *is* the hypothesis that one is testing in the experiment), argued that the influence goes in the reverse direction (i.e., hypotheses lead to theories, not vice versa), or misunderstood the question at an even lower level (e.g., by arguing that experiments influence theories). Because our revised Nature of Science interview probed this specific issue much more clearly and directly than previous versions of the Nature of Science interview (used by Sandoval, 2003, and Honda, 1994, 1996), our study provides the clearest evidence to date that students do *not* conceptualize theories as frameworks (or even collections of background knowledge) that guide all aspects of scientific inquiry. Instead, theories are thought of as well-tested hypotheses—a finding consistent with the recent results reported by Sandoval and Morrison (2003) for high-school students. Thus, although science instruction at the middle- and high-school levels may be helping students to differentiate theories and evidence, science instruction does not appear to be helping them to differentiate hypotheses from theories. This conclusion also fits with the prior findings of Abd-El-Khalick and Lederman (2000) and Windschitl (2004), both of whom, working with college students or preservice teachers, argued that students did not appreciate the theory guided nature of inquiry.

Uncertainty of Scientific Knowledge. Previous work has been less consistent in the conclusions reached about students' understandings of the inherent uncertainty of scientific knowledge. On one hand, some studies have found that high-school students treat scientific controversies as resolvable with enough diligence and effort (Driver et al., 1996) and that college students consider science as different from other disciplines because it is "definite, correct, or 'proven true.'" (Abd-El-Khalick & Lederman, 2000). Both findings have been interpreted as evidence that students conceive of scientific knowledge from a more naïve absolutist perspective as certain and true. On the other hand, other studies have found that even high-school students have a greater appreciation of the tentative nature of scientific knowledge than has typically been realized. For example, Lederman and O'Malley (1990) noted that high-school students are aware that scientific theories can change; further, they often use the word ''proof'' not to mean absolute certainty but as synonym for ''support;'' finally, they are frequently aware that scientists can look at the same evidence differently.

One limitation of these prior studies is that the data are typically scored in dichotomous ways (i.e., students have a naïve vs. informed view, or view knowledge as absolute vs. tentative) rather than in terms of a progression of views. Another limitation is that they have not distinguished whether students see uncertainty, change, and revision as accidental or essential properties of scientific knowledge or the underlying reasons for uncertainty and change. In a conceptually based scoring system, however, attention to such details is crucial. Simply asserting that theories *can* change is very different from making a principled assertion that theories are *always* subject to revision and change or that all scientific knowledge is fundamentally uncertain. Similarly, asserting that one can always change a theory in light of new data is different from realizing that one can always change a theory in light of a new idea that is more compelling or that seems to provide a better account of the data.

Our work thus provides further perspective on this issue of students' awareness of the uncertainty of scientific knowledge not only by probing students' awareness of uncertainty in a novel way (i.e., by probing their reactions to a general statement that asserts that science knows the ''right answers'') but also by developing a more nuanced scoring system that is sensitive to whether students are making universal and principled assertions and to the nature of their reasoning used to support those assertions. Interestingly, we found this probe served as a valuable lightning rod for eliciting dramatically different student views about the nature of scientific knowledge, producing the widest spread in student views of any question throughout the interview. These views ranged from asserting that science has (or can have) the right answers to all questions (34% of the students), to views that assert science has the right answers to only some questions with other questions deemed to be unanswerable in principle (23% of students), to assertions that scientific knowledge is essentially uncertain (43% of students). Within the latter group, however, most students gave simple inductive reasons for scientific uncertainty rather than deeper interpretive or theory-based reasons for uncertainty.

This progression of views in thinking about certainty in science has striking parallels to the progression of views described in the general epistemology literature (for a review, see Hofer & Pintrich, 1997). In that literature, ''radical relativism'' bridges between absolutist and more nuanced views that acknowledge a role for the knower's perspective. Similarly, in thinking about science, awareness of rampant inductive uncertainty may bridge between viewing scientific knowledge as factual and certain to viewing scientific knowledge as founded in rigorously tested, but ever revisable, theories.

Reasoning about and Resolving Scientific Controversies. Finally, our work has documented both strengths and limitations in the way first-year college students reason about scientific controversies. College students have enough intellectual sophistication to engage with the issue of scientific controversies. They are not surprised by the existence of scientific controversies or the fact that two groups of scientists who have studied the same problem can come to different conclusions. Further, they do not trivialize the controversies by simply asserting that one scientist must be right and the other wrong or that they occur simply because of mistakes, errors, or deceit on the part of scientists. Rather, they are aware of some of the difficulties in arriving at generalizations about the effects of fluoride or Echinacea, in large part because of their awareness that people are different and respond differently to things.

Interestingly, this awareness seems to be a "two-edged sword." On one hand, it alerts the students to the dangers of "overgeneralization." On the other hand, it undermines their expectation that they will find any strong generalizations. Indeed, they often conclude that both scientists must be (at least partially) right because of the differences in their methods or sample. That is, students typically assume that each treatment genuinely worked for some subgroups but not for others because there were both positive and negative empirical results. Strikingly, only a few considered that there may have been major interpretive errors due to the failure to control for other relevant variables and thus that the treatments may not have "worked" for any subgroups at all. No students considered that the scientists might have operated under different theoretical frameworks, and so defined the problem and the variables to be tested entirely differently.

Complementing their views on the underlying reasons for controversy were their views on how the controversies could be resolved. In keeping with students' finding the difficulty to be one of generalization rather than interpretation, the main methods for resolving the controversies involved either: (a) testing larger, more diverse samples; (b) testing more homogeneous samples with similar methods; or (c) simple treatment/no treatment comparisons (but with no attempt to control other relevant variables). Only a minority of students described how one needs to both manipulate the presumed causal variable in experiments while controlling other possible causal variables to sort out issues of proper interpretation. No students considered the need to have detailed studies of the specific causal mechanism by which each treatment worked as part of resolving the controversy.

In general, our findings complement the prior work of Driver et al. (1996), who provided small groups of 16-year-olds with background information about two scientific controversies (e.g., the 1920s' controversy about the hypothesis of continental drift and the current controversy about the safety of food irradiation) and asked them to think about the reasons for the controversies and how they might be resolved. Like us, they found that students were not at all surprised at the existence of scientific controversies; they had an abundance of reasons (both empirical and social) that such controversies might exist. For example, scientists needed more or better (more conclusive) evidence; disagreements among individuals were "natural" in the face of inconclusive evidence because they might look at different evidence or because of personal rivalries. Like us, they also found little mention of disagreement as involving genuinely different interpretations of results or as stemming from different underlying theories of mechanism (even though the background information for the continental drift controversy specifically included information about the absence of any mechanism for continental drift as being one reason for the dispute in the 1920s).

Our work extends this previous work by making some finer distinctions among students' understanding of the reasons for scientific controversies and in relating this understanding to their emerging domain-specific knowledge of experimental design. In particular, we distinguished (a) students who simply expected diverse data from (b) those who were beginning to articulate different variables that might affect the results or how well a treatment worked from (c) those who understood that the scientists could have been entirely mistaken in their interpretations because of uncontrolled confounding variables. We found that students who simply expected diverse data typically proposed gathering more data from a larger sample to resolve the controversy. Those who were focusing on specific variables affecting the results were more sophisticated, but stopped short of articulating a fully controlled experimental design. They typically proposed either (a) testing more similar people (to reduce the variables) or (b) making a simple treatment/no treatment comparison (with no control of other variables). Significantly in light of the fact that elements of controlled experimental design have been taught from middle school on, only a small number of students recognized that there could have been genuinely conflicting interpretations and proposed a full-scale, controlled experimental design as part of the process of resolving the disagreement.

College Students' Epistemologies of Science: Interrelations among Key Aspects

All three measures of epistemological understanding in science were highly intercorrelated, providing some validation of this construct. Specifically, students' level of idea/evidence differentiation (in Part 1) was strongly related to both awareness of uncertainty in science (Part 2) and reasoning about specific controversies (Part 3); student's awareness of uncertainty in science (Part 2) also was related to their reasoning about specific controversies (Part 3). In general, the best predictor of their reasoning about scientific controversies was their level of idea/evidence differentiation.

We believe these findings are important for two reasons. First, as mentioned earlier, it provides some validation of the assumption that students have an interrelated set of ideas that constrains their thinking about epistemological issues in science. In particular, making a consistent differentiation between ideas and evidence was associated with having an appreciation of the general uncertainty of scientific knowledge, with having an appreciation of the deeper interpretive issues in scientific controversies, and with understanding some of the domain-specific methods scientists have developed for resolving these controversies. The interrelation of these aspects may account for both the slowness of change in students' thinking as well as their overall resistance to change. That is, if students' ideas on one aspect of epistemology are coordinated or supported by ideas about other aspects, a great deal of reorganization and reconceptualization would need to occur to move to a more sophisticated stance.

Second, the fact that students' answers to both the decontextualized Nature of Science Interview and the General Probe about Scientific Truth were not only strongly related to each other but also to their reasoning about a specific science controversy supports the idea that these epistemological understandings are not simply isolated bits of ''verbal'' knowledge but instead may function more actively as a metacognitive control structure (Hofer, 2004) that can guide student reasoning about ambiguous situations. That is, coming to verbalize the interrelated ways that theorizing, hypothesizing, and designing experiments help scientists in their quest to understand how the world works helps consolidate a new way of thinking about scientific inquiry and heightens their awareness of the deeper reasons for controversies and how to resolve them. Of course, we do not believe the verbal knowledge demonstrated in the Nature of Science interview comes either from memorizing simple verbal definitions or from unreflective engagement with scientific inquiry. Rather, as researchers such as Abd-El-Khalick and Lederman (2000) so eloquently argued based on their teaching studies in classes that analyze historical texts, meaningful lessons about scientific epistemology may come from the combined reflection on, analysis of, and discussion of real episodes of scientific inquiry. Further, reflective student discourse during extended episodes of classroom inquiry also has been shown to promote epistemological growth (Smith et al., 2000). Currently, we are following college students who are actively practicing science, critically evaluating current primary literature, designing their own experiments, analyzing data, and making decisions about the extent to which they can generalize from their findings beyond their own experimental conditions. Given the richness of their experiences, one can imagine a host of issues arising, such as the logic of the prospective clinical trial in medicine or considering Popper's refutation criterion in the abstract. We suspect that such practices might prove to be very important in the development of a sophisticated understanding of controversy.

One final caveat is in order about the significance of these interrelations for an intuitive meta-theories perspective on students' epistemological stances in science. Within the science education community, there have been debates about the amount of coherence that exists in students' epistemological thinking between those with knowledge-in-pieces viewpoints versus intuitive-theories viewpoints. Those with knowledge-in-pieces views have tended to emphasize the inconsistencies in students' reasoning across diverse contexts (e.g., Hammer & Elby, 2002; Sandoval & Morrison, 2003) while those in the intuitive-theories camp have emphasized the patterns of coherence (e.g., Smith et al., 2000). Of course, resolving the debate does not simply depend on data about student consistencies or inconsistencies. For example, even if students do have coherent epistemological viewpoints, there could be reasons to expect inconsistency (e.g., knowledge limitations could prevent the application of concepts in certain contexts). Conversely, consistency in student reasoning does not automatically favor the intuitive-theories viewpoint (a) in the absence of a detailed conceptual analysis of the contrasting ways that concepts cohere within different theories or (b) if alternative, more compelling reasons for those intercorrelations can be found (e.g., general differences in intelligence, critical thinking, or verbal-reasoning ability that affect the pace and ultimate level of development). Thus, we note the reason we think our data argue for some underlying coherence in students' epistemological thinking is not simply because there were high correlations across epistemological measures but rather because the pattern of observed relations is consistent with our specific conceptual analysis (e.g., consistently differentiating theory and evidence was strongly related to having an understanding of inductive and interpretive uncertainty).

Ultimately, we believe that an adequate account of epistemological development will need to integrate the insights from both perspectives. In fact, we believe there are both piecemeal and systemic aspects to epistemological development and that resources for change come from both within and outside one's initial epistemology in a specific domain. Epistemological concepts develop not only in interaction with each other (i.e., one's concept of theory has implications for one's concept of evidence and truth or proof) but also in response to real-world situations that one is trying to understand. Part of conceptual knowledge is not just interrelations among conceptual elements but knowledge of how and when to apply those ideas to the specific situations. Thus, in arguing for an intuitive-theories perspective on epistemological development, we are not denying the existence of local and contextual knowledge or of importance of resources outside the theory but rather reminding researchers why—in their awareness of the intricacies and complexities of knowledge acquisition—they should not overlook the systemic aspect of the conceptual element. It may be important to understand this aspect not only to explain "resistance to change" but also to understand the complex processes by which students make new inductions from experience.

Implications for Instruction

The impact of different pedagogies on epistemological development is not yet well understood. In general, most agree that the impact of individual courses is limited, although detectable progress is made, especially when students are explicitly engaged in active reflection such as through journal writing (Schwartz, Lederman, & Crawford, 2004) or in discussing the nature of science at a meta-conceptual level (Abd-El-Khalick & Lederman, 2000, Carey et al., 1989; Honda, 1994, 1996). Additionally, more progress is made when interventions are sustained over longer periods of time (Smith et al., 2000). Although Schwartz et al. (2004) found that engagement with inquiry alone does not affect students' understanding of the nature of science, involvement in courses with robust arrays of inquiry activities and opportunities for students to synthesize their ideas might help students understand scientific inquiry and increase their propensity to try to make sense of the controversy inherent in the practice (Bell & Linn, 2002). Inquiry, then, might pave the way for just the type of active reflection and metaconceptual discussion that have been reported to affect epistemological understanding. Wenk (2000) found that first-year college students in inquiry-based courses cited designing/running/analyzing data from their own experiments and justifying their positions using primary literature (especially when confronted with studies whose results are contradictory) as affecting their views of the certainty of scientific knowledge and their understanding of the role of evidence in justification.

It might be especially important to involve students in complex cycles of designing experiments or analyzing the experiments of others to evaluate competing explanations of phenomena. Of course, science courses in different disciplines do not lend themselves to the same kinds of learning opportunities nor do they have exactly the same epistemological issues. For example, it is common to collect and work with real data in an introductory physics course, but it would be difficult to provide students with accessible primary literature (Therefore, it is unlikely that students would design experiments in the service of inquiry into current physics debates.). The epistemological issues for physics students center more on engaging and refining their everyday intuitions through attempts to build coherent mental models that consistently apply to a broad range of everyday situations (Hammer, 1994; Redish, 1994; Redish, Saul, & Steinberg, 1998). In contrast, in a course on evolutionary biology, students might not perform experimental studies, but might look instead at the relationships among evidence from the fossil record, morphology, and molecular genetics to help them understand current accounts of evolutionary change. The epistemological issues these students face might be examining how the results of testing many smaller hypotheses contribute to the support of the framework theory, and how theory guides further inquiry (from the development of new hypotheses to the methods for testing them and the mode of analysis of data).

As our longitudinal study progresses, we will describe the nature of any activities that seem to influence students' abilities (a) to differentiate theories, hypotheses, and evidence; (b) to understand the uncertainty of knowledge claims; and/or (c) to resolve scientific controversy. We will be interested in whether we replicate the findings of this study with future cohorts and whether we find that these different aspects change in a coordinated fashion in our longitudinal study. At this point, it is our hope that this baseline data and our findings about the relationships among the three aspects of students' epistemological understandings will help college faculty gain a richer understanding of the ways their students' ideas about epistemological issues form an interconnected system and will shed light on why traditional instruction often fails to promote epistemological development.

For example, given the finding that most college students are at least beginning to differentiate scientists' ideas and evidence, it might be a relatively simple task for faculty to convince students that we do not know everything in science with certainty just by pointing out the kinds of questions they and their colleagues are currently researching. Similarly, it might be relatively easy to convince students that one has to be careful about generalizing from a limited sample. In fact, students seem too quick to overlook conflicting results as actually challenging a theory by falling back on the easy idea that "it worked for some, but not for others." Thus, the harder task might actually be to get them to take generalizations seriously and to expect that mechanisms work for a broad sample of individuals. It also might be difficult to help students understand that scientific ideas and theories are interpretive extrapolations from observations and, as such, are subject to imprecision, error, and doubt even if the observations were done most carefully and systematically on a large sample. And it might be easy to demonstrate that our ideas about what is correct in science change over time by lecturing on the historical unfolding of a particular concept or theory in science. But it might be much more difficult, and one would need a different tactic, to help students understand the role of scientists' theory-bound interpretations in changing ideas in science. Overall, we found that very few students had developed an understanding of a theory as a larger framework that powerfully guides all aspects of scientific inquiry, including hypothesis testing. Instead, they thought of theories as end products of scientific inquiry.

In general, the empirical evidence, the conceptual change model of epistemological growth, and the picture of a complex system of thinking about science that our own data support all point to the fact that epistemological progress should be slow. Yet, one cannot ignore the potential retarding effect of an entrenched instructional system of lecture, textbook readings, and recitation on the students' epistemological development. In the subsequent years of our study, we will examine how students' epistemologies of science develop, or fail to, under different kinds of programs of study. Half of our participants are enrolled in more traditional lecture/lab courses and the other half in inquiry-based courses that emphasize engagement in authentic inquiry. By "authentic," we mean an inquiry process that mirrors the practice of scientists (Roth, 1995). Having authentic inquiry embedded in a course allows for explicit discussion about the ways scientists think when confronted with a question, and thereby engages students in meta-level processing. These inquiry-based courses use activities in which students are asked to critique the primary literature and reflect on the meaning of scientific endeavors, identifying underlying assumptions, examining the limitations of the methodologies, and discussing alternative explanations of the data. It is our hope that data from our longitudinal study will shed light on the kinds of science experiences that challenge students' thinking, causing them to become dissatisfied with their current conceptions, and the types that support them in making, what can be uncomfortable, shifts in their views about science.

Notes

[1]Of course, exactly what counts as an epistemological domain for a student is a function of their knowledge and experience; however, we assume that because schooling is organized around distinct subjects (e.g., history, science, math, literature) that are informed by different disciplines, students have experience in figuring out how the intellectual game is played in each area and in understanding certain ideas that are part of the intellectual discourse of each subject (e.g., the ideas of postulates, theorems, and proofs within mathematics; the ideas of an account, source, and artifact in history, etc.). Of course, with further experience and specialization, students may make further distinctions within these broad disciplines—for example, contrasting epistemological issues in physics versus biology, or even between molecular and evolutionary biology (given, for example, the greater importance of historical reasoning and argument in the latter).

[2]In this sense, intuitive theories function as a framework theory that guides the construction of more specific (and even rival) beliefs within that framework. Thus, the intuitive-theories perspective allows one to make a distinction between underlying epistemological concepts and specific epistemological beliefs—a distinction that is not made by some other perspectives for studying epistemological development (for a discussion of this issue, see Carey, 1999). For example, the statements "Scientific knowledge is certain" and "Scientific knowledge is uncertain" are specific epistemological beliefs (i.e., assertions that can be regarded as either true or false); however, the meaning of any belief is informed by one's underlying concepts. Thus, two people may agree with the statement "Scientific knowledge is certain," yet have different concepts of "scientific knowledge" and "certainty." Similarly, two people may have opposite beliefs about the truth value of the statement, yet have similar underlying concepts of knowledge and certainty. For this reason, we do not assess epistemological stances via simple epistemological beliefs questionnaires in which students only circle whether they agree or disagree with given statements. (see critiques of these methods by Elby & Hammer, 2001, and Lederman, Abd-El-Khalick, Bell, & Schwartz, 2002). Rather, we consider the reasoning used in thinking about a topic or reacting to a statement in coding

student responses. At the same time, we acknowledge that changes in one's underlying epistemological concepts (e.g., one's concepts of knowledge, truth, and certainty) often lead to changes in which specific epistemological beliefs are embraced, as conceptual change supports belief change as well.

[3]A more complete coding manual and set of grids for all questions are available upon request.

[4]There were no differences in average student responses for the two controversies. The average score for the Echinacea controversy was 1.99, and for Water Fluoridation, 2.01.

Appendix A. Interview Scripts

Part 1: Nature of Science Interview

1. a. What do you think the goal of science is?
   b. What sorts of things do scientists do that help them reach those goals?
2. a. What sorts of questions do scientists try to answer? Can you give me a specific example of a question that a scientist would ask?
   b. What would a scientist do to answer that question? Would they do experiments? Can you give me an example of an experiment they might do?
3. a. What is an experiment and why do scientists do them?
   b. In general, how do scientists decide what experiment to do?
4. a. What is a scientific hypothesis? Can you give a specific example of a scientific hypothesis?
   b. Does a scientist's hypothesis influence the experiments (s)he does?
5. What is a scientific theory? Can you give a specific example of a scientific theory?
6. a. Are theories and hypotheses related? If yes, how?
   b. Does a scientist's theory influence the hypotheses he or she tests in a particular experiment? If yes? How? Can you give an example of that? If no, why not?
7. a. Do scientists ever change their hypotheses or theories? When would they do that and why?
   b. Do you think it is easier to change a hypothesis or a theory? Why?

Part 2: General Probe about Scientific Truth

I will ask you to read a statement made by another student while I read it out loud. Then I'll ask you to comment on it, and then I'll ask you some questions about it. This student said: "Everyday, in more and more areas of science the right answer is known. In areas where the right answer is known, I look to experts to tell me what is right. In areas where no right answer is known, I think anyone's opinion is as good as another's."

a. Do you agree with this statement? Why or why not?
b. Do you think science knows the right answers? Why or why not?
c. Where do you go when you have questions about a scientific issue?
d. What do you do when you find disagreement among sources?
e. Do you agree with this person who says that when there are no right answers anybody's opinion is as good as another's?
f. Is there anything that makes one answer better than another?

Part 3: Probe about a Specific Scientific Controversy

I will be presenting a situation and asking you some questions about it. I am not concerned with how much information you have about the issues, but how you think about them. I will read a statement out loud while you follow along on a card. After I finish reading the statement, I'll give you a minute to think about the issue and then we will talk about it.

Version A: The Effects of Water Fluoridation. There have been frequent reports about the relationship between water fluoridation and tooth decay. Some studies indicate that fluoridation of drinking water supplies substantially decreases tooth decay in large portions of the population. Other studies, however, indicate that adding fluoride to water supplies does not decrease tooth decay. The issue of whether or not to add fluoride to drinking supplies is controversial for this reason and because there are also disagreements in the literature about the possible toxicity of fluoride.

Version B: The Effects of Echinacea. Preparations of Echinacea root are among the most popular herbal supplements in the United States marketplace. Echinacea is used for preventing and treating the common cold, flu, and upper respiratory tract infections. It's also used to increase general immune system function. Because Echinacea has become so popular, scientists have been conducting tests on the effectiveness of Echinacea. Some of the studies show positive effects of Echinacea in preventing colds, flu, and upper respiratory tract infections. Other studies show no effect. The use of Echinacea for preventing colds and other ailments is controversial for this reason and because there are also disagreements in the literature about possible carcinogenic effects of Echinacea.

a. How can scientists disagree about (whether or not fluoridation of drinking water decreases tooth decay) OR (whether or not taking Echinacea prevents colds and flu)?
b. How can scientists disagree about (whether or not fluoridation poses a health hazard) OR (whether or not taking Echinacea causes cancer)?
c. In a controversy like this, is one answer right and one wrong? If yes: What would make one right and one wrong? If no: Could one be better than the other? What would make it better?
d. In the case of this controversy, how might scientists go about resolving it? If they mention experiments: What might the experiment be in this case? If no mention of experiments: Is it possible they might do an experiment? What would an experiment look like in this case?

Appendix B. Description of Coding Categories and Levels for Each Part

Coding Part 1: Differentiation of Theories, Hypotheses, and Evidence

Level 1. No differentiation between scientist's ideas, procedures, and results. Ideas (and hypotheses/theories) are either conflated with procedures (i.e., how to do experiments) or with results (i.e., what happened). More specifically:

- *Goals of science (1a)*: The purpose of science is to do things, to learn things, or to make discoveries about what happens.
- *Scientific questions (2a)*: Scientists ask procedural questions (i.e., how to do things) or factual questions (i.e., what happens).
- *Nature of experiments (3a)*: Experiments are trying things out to produce a good result or something you do to find out what happens.

- *Nature of hypotheses and theories (4a; 5)*: Students may not know these words; at best, they map onto notion of scientists' ideas which is conflated with experiments (a set of steps for accomplishing something) or experimental results (what you find out), or a vague question, topic, or opinion; no differentiation between hypotheses or theories.
- *Relation between hypotheses and theories (6)*: As there is no differentiation between ideas and experiments, students cannot formulate any coherent relation between them; they may say both are the same (e.g., both procedures, questions, topics, facts).
- *Process by which hypotheses and theories change (7)*: Change occurs when something doesn't work, when you make mistakes, or when you find out you are wrong.

Level 1.5. Beginning awareness of the role of ideas in scientist's work, although the nature of ideas is vague (i.e., students may say scientists are trying to find our how something works but can't give an example of what they mean) and ideas are not yet central to the scientific process.

- *Goals of science (1a)*: Science is concerned with finding out how something works, but students provide no example of what they mean so it is ambiguous whether they are referring to some concrete procedure (how to work something) or underlying mechanism.
- *Scientific questions (2a)*: Scientists ask questions about how something works, but students are unable to provide any specific example of what they mean. Alternatively, they may say scientists ask questions about how one (observable) variable affects another (observable) variable.
- *Nature of experiments (3a)*: Experiments are something scientists do to find out how something works (no further elaboration) or something that involves measuring/ comparing variables (but no articulation of why this is important).
- *Nature of hypotheses and theories (4–5)*: Hypotheses are now thought of as educated guesses, but there is no mention that these ideas are then tested or evaluated by doing experiments. Theories too are thought of as guesses or as simple facts (what happens). If students do attempt some distinction between hypotheses and theories, it is simply the distinction between a guess and a known fact.
- *Relation between hypotheses and theories (6)*: Hypotheses and theories are either undifferentiated (Both are educated guesses.) or if they attempt a distinction, it is between a guess (hypothesis) and an observed outcome or fact (theory).
- *Process by which hypotheses and theories change (7)*: Change occurs by doing experiments, accumulating facts. Students may consider that hypotheses can be changed (because just guesses) whereas theories cannot (because known facts, proved.)

Level 2. Initial differentiation between scientist's ideas and experimental results using either new concepts of (a) explanation (Scientists have ideas about how to explain their results.) or (b) hypothesis testing (Scientists evaluate their ideas through their fit with experimental results.). These two new ideas are used separately rather than coordinated. Ideas are now more central to inquiry. Theories, however, are simply well-tested hypotheses; there is no differentiation between hypotheses and theories in terms of content or scope.

- *Goals of science (1a)*: Science is concerned with either understanding how something works (with an appropriate specific example) or it is defined as a process of testing one's ideas via systematic observation and experiment.
- *Scientific questions (2a)*: Scientists ask questions about how something works, or why something happens. These are no longer mere slogans: Students can generate a host of specific examples.

- *Nature of experiments (3a)*: Experiments are defined as tests of hypotheses. Scientists decide what experiment to do based on their hypothesis.
- *Nature of hypotheses and theories (4–5)*: Hypotheses are typically defined as ideas that are tested (evaluated) by doing experiments, and theories are well-tested hypotheses; however, students make no distinction between the scope of hypotheses and theories, just the degree of evidence or testing. Occasionally, students may focus on explanation (rather than testing) in the definition of hypotheses and theories.
- *Relation between hypotheses and theories (6)*: Typically theories are well-tested hypotheses. When asked if a scientist's theory can affect his hypothesis, students generally say "No it's the reverse: The scientist's hypothesis affects his theory!"
- *Process by which hypotheses and theories change (7)*: Change occurs through explicit hypothesis-testing process. Students don't spontaneously make a distinction between ease of changing a hypothesis and theory, but when probed may acknowledge it is easier to change a hypothesis than a well-tested theory.

Level 2.5. Deepening understanding of the conjectural/explanatory nature of hypotheses/theories and that the same pattern of results can be interpreted in more than one way. Beginning recognition that theories may be broader in scope than hypotheses and can affect hypothesis testing; however, such a theory-guided effect is not yet central to inquiry.

- *Goals of science (1a)*: Science is concerned with testing ideas about how something works; there may be acknowledgment that these ideas are always open to revision.
- *Scientific questions (2a)*: Scientists ask a series of questions to find out how something works; there also may be acknowledgment of the need to formulate questions in a way that makes them testable.
- *Nature of experiments (3a)*: Experiments are defined as either tests of competing explanatory hypotheses or as the controlled designs that allow the test of competing hypotheses.
- *Nature of hypotheses and theories (4–5)*: Hypotheses are typically defined as explanatory ideas that are tested by doing experiments, and theories are a well-tested set of hypotheses.
- *Relation between hypotheses and theories (6)*: With probing, students acknowledge that one's theories can provide the background that affects the formulation of a specific hypothesis. Hence, there is some beginning differentiation between hypotheses and theories: Theories are not just well-tested hypotheses, but broader in scope than hypotheses.
- *Process by which hypotheses and theories change (7)*: Change occurs through a thoughtful process of testing ideas and considering them against alternatives. Thus, change involves both new ideas and data. There also may be some appreciation that it may be hard to change one's ideas (not because of stubbornness but because you use those ideas to interpret data) or that one's ideas are always open to deep revision in light of future evidence.

Level 3. Differentiation between scientist's theory as general explanatory framework and the specific hypotheses that are tested within the theory. Understanding of the way hypothesis testing indirectly tests the larger theory and of the theory-relative nature of all knowledge.

- *Goals of science (1a)*: Science is concerned with developing theories about the world with greater explanatory scope; these theories are developed through a process of testing specific hypotheses that are consistent with one's theory.

- *Scientific questions (2a)*: Scientists' explanatory framework influences how they ask a question; students may appreciate the nested nature of scientific questions (e.g., a more general why question (about mechanism) requires that scientists pursue a series of implications of that mechanism; implications must be formulated in ways that are testable).
- *Nature of experiments (3a)*: Experiments are designed not only to directly test a specific hypothesis but also to indirectly test one's larger theory.
- *Nature of hypotheses and theories (4–5)*: Theories are defined as an interrelated set of concepts and causal principles that are used to explain phenomena in a broad domain that have been carefully tested. Hypotheses are more specific ideas about how some particular phenomena within the domain might work.
- *Relation between hypotheses and theories (6)*: Specific hypotheses are formulated using the concepts and explanatory principles of theories. In this sense, theories constrain all aspects of hypothesis formulation and data interpretation and evaluation.
- *Process by which hypotheses and theories change (7)*: Theory change is harder than hypothesis change because theories provide the framework within which hypotheses are tested. Thus, theory change involves new explanatory ideas (i.e., new concepts that are incommensurable with one's initial theory) as well as new data.

Coding Part 2: Uncertainty & Justification of Scientific Truths

Level 1. Knowledge is certain, although there are things that authorities do not yet know. When answers are not known, one can believe anything one wants.

Level 1.5. Knowledge is certain in some areas, but not others. In cases of uncertainty, believe a more knowledgeable source or go with majority opinion.

Level 2. Knowledge is uncertain, as scientists can always find some exception to their generalizations (a kind of inductive uncertainty). One tries to understand the reasons for the controversy and uses amount of evidence to resolve controversy.

Level 2.5. Knowledge is uncertain because there can always be alternative explanations of a phenomena (interpretive uncertainty). Justification is based on quality of evidence and ability to integrate diverse evidence.

Level 3. Knowledge is inherently uncertain because of the theory-guided nature of inquiry. Justification is based on both the strength of evidence and one's theory.

Coding Part 3: Reasons for Controversy and How to Resolve

Level 1. Controversy exists because authorities do not yet have the correct information, have made a mistake or an error, or do not yet know. Resolve by getting more information or checking for mistakes.

Level 1.5. Controversy exists because there is not yet enough data or because people are different and respond differently to things. Resolve by testing large sample to make sure treatment really works.

Level 2. Controversy exists because scientists have different data/results from studies and have looked at different things. Both scientists are probably right: The treatment may have really worked for some, but not others, due to the operation of specific variables that modify its effectiveness. Resolve by having scientists work together, reducing the variables when testing, or by having a simple comparison group (but no control of other variables).

Level 2.5. Controversy exists because scientists may have failed to control for other variables and thus may have been entirely mistaken in their interpretations of data. They also may have defined important variables differently. Resolve by doing an experiment that controls for other variables, distinguishes between contrasting interpretations, or provides converging lines of evidence.

Level 3. Controversy exists because scientists may have rival theories of mechanism of treatment. Resolve by sustained experimentation to evaluate the competing theories.

References

Abd-El-Khalick, F. & Lederman, N.G. (2000). The influence of history of science courses on students' views of nature of science. Journal of Research in Science Teaching, 37, 1057–1095.

American Association for the Advancement of Science. (1993). Benchmarks for science literacy: A Project 2061 report. New York: Oxford University Press.

Baxter Magolda, M.B. (1992). Knowing and reasoning in college: Gender related reasoning in intellectual development. San Francisco: Jossey-Bass.

Belenky, M.F., Clinchy, B.M., Goldberger, N.R., & Tarule, J.M. (1986). Women's ways of knowing: The development of self, voice, and mind. New York: Basic Books.

Bell, P. & Linn, M.C. (2002). Beliefs about science: How does science instruction contribute? In B. Hofer & P. Pintrich (Eds.), Personal epistemology: The psychology of beliefs about knowledge and knowing (pp. 321–346). Mahwah, NJ: Erlbaum.

Boix-Mansilla, V. (2001). The pursuit of understanding: A study of exemplary high school students' conceptions of knowledge validation in science and history. Unpublished doctoral dissertation, Harvard University, Cambridge, MA.

Carey, S. (1999). Sources of conceptual change. In E. Scholnick, K. Nelson, S. Gelman, & P. Miller (Eds.), Conceptual development: Piaget's legacy (pp. 293–326). Mahwah, NJ: Erlbaum.

Carey, S., Evans, R., Honda, M., Jay, E., & Unger, C. (1989). "An experiment is when you try it and see if it works:" A study of grade 7 students' understanding of the construction of scientific knowledge. International Journal of Science Education, 11, 514–529.

Carey, S. & Smith, C. (1993). On understanding the nature of scientific knowledge. Educational Psychologist, 28, 235–251.

Chinn, C.A. & Malhotra, B.A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. Science Education, 86, 175–219.

Driver, R., Leach, J., Millar, R., & Scott, P. (1996). Young people's images of science. Buckingham, United Kingdom: Open University Press.

Elby, A. & Hammer, D. (2001). On the substance of a sophisticated epistemology. Science Education, 85, 554–567.

Hammer, D. (1994). Epistemological beliefs in introductory physics. Cognition and Instruction, 12, 151–183.

Hammer, D. & Elby, A. (2002). On the form of a personal epistemology. In B. Hofer & P.R. Pintrich (Eds.), Personal epistemology: The psychology of beliefs about knowledge and knowing (pp. 169–190). Mahwah, NJ: Erlbaum.

Hofer, B. (2004). Epistemological understanding as a metacognitive process: Thinking aloud during online searching. Educational Psychology, 39, 43–55.

Hofer, B. & Pintrich, P. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. Review of Educational Research, 67, 88–140.

Honda, M. (1994). Linguistic inquiry in the science classroom: "It is science, but it's not like a science problem in a book." Cambridge, MA: MIT Working Papers in Linguistics.

Honda, M. (1996, February). Developing an epistemology of science through linguistic inquiry. Paper presented at the annual meeting of the American Association for the Advancement of Science/AMSIE (Annual Meeting and Science Innovation Exposition), Baltimore.

King, P. & Kitchener, K. (1994). Developing reflective judgment: Understanding and promoting intellectual growth and critical thinking in adolescents and adults. San Francisco: Jossey-Bass.

Kuhn, D. (1991). The skills of argument. New York: Cambridge University Press.

Kuhn, T. (1970). The structure of scientific revolutions. Chicago: University of Chicago Press.

Lakatos, I. (1978). The methodology of scientific research programmes. Cambridge, England: Cambridge University Press.

Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. American Educational Research Journal, 27, 29–63.

Lederman, N. (1992). Students' and teachers' conceptions of the nature of science: A review of research. Journal of Research in Science Teaching, 29, 331–359.

Lederman, N.G., Abd-El-Khalick, F., Bell, R.L., & Schwartz, R.S. (2002). Views of Nature of Science Questionnaire: Toward valid and meaningful assessment of learner's conceptions of nature of science. Journal of Research in Science Teaching, 39, 497–521.

Lederman, N. & O'Malley, M. (1990). Students' perceptions of tentativeness in science: Development, use, and sources of change. Science Education, 74, 225–239.

Lee, P. & Ashby, R. (2000). Progression in historical understanding among students ages 7–12. In P. Stearns, P. Seixas, & S. Wineburg (Eds.), Knowing, teaching and learning history: National and international perspectives (pp. 199–222). New York: New York University Press.

National Research Council. (1996). National science education standards. Washington, DC: National Academy Press.

National Science Foundation. (1996). Shaping the future: New expectations for undergraduate education in science, mathematics, engineering, and technology. Washington, DC: Advisory Committee to the Directorate for Education and Human Resources.

Perry, W. (1970). Forms of intellectual and ethical development in the college years: A scheme. New York: Holt, Rinehart & Winston.

Redish, E.F. (1994). The implications of cognitive studies for teaching physics. American Journal of Physics, 62, 796–803.

Redish, E.G., Saul, J.M., & Steinberg, R.N. (1998). Student expectations in introductory physics. American Journal of Physics, 66, 212–224.

Roth, W.M. (1995). Authentic school science. Dordrecht, The Netherlands: Kluwer.

Sadler, T. & Zeidler, D. (2004). Student conceptualizations of the nature of science in response to a socioscientific issue. International Journal of Science Education, 19, 387–409.

Sandoval, W. (2003). Conceptual and epistemic aspects of students' scientific explanations. Journal of the Learning Sciences, 12, 5–51.

Sandoval, W.A. & Morrison, K. (2003). High school students' ideas about theories and theory change after a biological inquiry unit. Journal of Research in Science Teaching, 40, 369–392.

Schoenfeld, A.H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D.A. Grouws (Ed.), Handbook of research on mathematics teaching and learning (pp. 334–370). New York: Macmillan.

Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. Journal of Educational Psychology, 82, 498–504.

Schwartz, R., Lederman, N., & Crawford, B. (2004). Developing views of nature of science in an authentic context: An explicit approach to bridging the gap between nature of science and scientific inquiry. Science Education, published online May 12, www.interscience.wiley.com

Smith, C., Maclin, D., Houghton, C., & Hennessey, M.G. (2000). Sixth graders' epistemologies of science: The impact of school science experiences on epistemological development. Cognition and Instruction, 18, 349–422.

Songer, N.B. & Linn, M.C. (1991). How do students' views of science influence knowledge integration? Journal of Research in Science Teaching, 28, 761–784.

Wenk, L. (2000). Improving science learning: Inquiry-based and traditional first-year college science curricula. Dissertation Abstracts International, 61, 3885A. (University Microfilms No. AAT99–88852)

Windschitl, M. (2004). Folk theories of "inquiry:" How preservice teachers reproduce the discourse and practices of an atheoretical scientific method. Journal of Research in Science Teaching, 41, 481–512.