Cognitive Development 23 (2008) 435–451

Contents lists available at ScienceDirect

Cognitive Development

Beyond control of variables: What needs to develop to achieve skilled scientific thinking?

Deanna Kuhn*, Kalypso Iordanou, Maria Pease, Clarice Wirkala

Columbia University, Teachers College, Department of Human Development, New York, NY 10027, United States

ARTICLE INFO

Keywords:
Scientific thinking
Causal reasoning
Multivariable prediction
Epistemological understanding
Argumentation

ABSTRACT

We identify three aspects of scientific thinking beyond the control-of-variables strategy that we claim are essential for students to master as a foundation for skilled scientific thinking. The first is strategic and involves the ability to coordinate effects of multiple causal influences on an outcome. The second is a mature understanding of the epistemological foundations of science, recognizing scientific knowledge as constructed by humans rather than simply discovered in the world. The third is the ability to engage in skilled argumentation in the scientific domain, with an appreciation of argumentation as entailing the coordination of theory and evidence. We present new empirical data with respect to the first two of these competencies, supporting the claim that they are not well developed by early adolescence and warrant attention and provision of effective kinds of scaffolding.

Science educators who write about the development of scientific thinking skills emphasize the extent and complexity of what needs to develop if students are to become effective science learners (Duschl, 2008; Fortus et al., 2006; Kuhn & Pease, 2008)—a complexity and extent clearly reflected in K-12 science curriculum standards (National Research Council, 1996, 2007). A long-standing tradition among developmental psychologists who study scientific reasoning, however, has been to focus attention on a single reasoning strategy, the control-of-variables strategy, featured by Inhelder and Piaget (1958) in their now classic volume (for early reviews, see Neimark, 1975, or Keating, 1980; for contemporary ones Zimmerman, 2007, or Kuhn, 2002.). Despite this volume of research, controversy continues regarding how this strategy develops and how educators can best support its development, and continuing investigation is warranted. In the present article, however, we turn our attention to

* Corresponding author.
E-mail address: dk100@columbia.edu (D. Kuhn).

0885-2014/$ – see front matter
doi:10.1016/j.cogdev.2008.09.006

another question, one that cognitive and developmental psychologists would be expected to have played a major role in addressing but thus far largely have not: What else, beyond control of variables, is involved in the development of skilled scientific thinking? The conduct of a controlled experiment is certainly fundamental to science and a skill students need to acquire and understand the importance of. But it is hardly all there is to be concerned about, or the task of science educators would be much more straightforward than it is. What else does or has the potential to develop in the realm of scientific thinking?

In this article, we examine three broad areas of competence, one having to do with skill in experimentation, one with skill in argumentation, and a third with epistemological understanding of the nature of science. Science educators have done a better job than psychologists in paying attention to the latter two, but much more research is needed to obtain the empirical evidence that educators need regarding the development of argumentation skills and epistemological understanding within the domain of science. We do not undertake any extended discussion here regarding the relations among the three kinds of competencies, nor their relative importance, except to emphasize the critical role of all three. A case could be made for epistemological understanding of the scientific enterprise as the superstructure that gives meaning to the respective skills, but we agree with Duschl (2008) that the interplay among them is complex, with each informing the other.

Nor do we undertake exhaustive treatment of any of the three. Instead, we present three of our own recent studies, one falling within each of the categories, as illustrative of the kinds of questions that might be asked within each of the three domains. We begin with one in the domain of experimentation, then turn to one in the domain of epistemology, and conclude with only brief mention of one in the domain of argumentation, due to limitations of space.

1. Scientific thinking as multivariable reasoning

Scientific phenomena routinely comprise multiple interacting variables whose influences on one another must be identified. Control of variables is clearly an essential strategy—the influences of other variables must be held constant so that the effect of a focal variable can be identified. But even realizing this aim itself entails more than mastering the control-of-variables technique. Recognizing the value of identifying, and examining the effect of, a single variable at a time, for example – in other words, identifying an appropriate question as the object of investigation – poses its own challenge and contributes significantly to success (Kuhn & Dean, 2005).

Once the control-of-variables strategy is successfully executed and the effect of an individual variable identified, research on the control-of-variables strategy has little more to say. In the real world of science and scientific thinking, however, multiple variables continue to be a major presence. Most commonly, multiple variables co-exist, many of which may influence a particular outcome of interest. Often, then, once the individual effects of each have been ascertained, the task that confronts a scientist (or engineer) is to take all of the relevant effects into account with the objective of predicting how they will jointly affect an outcome. This is clearly a task in multivariable scientific reasoning, but unlike control of variables, one that has been given very little attention. Coordination of the effects of multiple variables or features, in contrast, has long been a topic in the study of categorization (Gluck, Shohamy, & Myers, 2002; Minda & Ross, 2004).

Both control of variables (COV) and multivariable analysis and prediction can be regarded in an analysis of variance (ANOVA) framework. Among the assumptions that are part of this framework is first the assumption that causes have consistent effects under the same conditions, and second that multiple effects may operate jointly on an outcome, in either additive or interactive fashion. Kuhn and Dean (2004) found that preadolescents, as well as some adults, perform poorly on such a multivariable prediction (MVP) task. In addition to typically faulty predictions, individuals often make inconsistent causal attributions across consecutive predictions, for example implicating variable A as causal (when asked, following a prediction, to indicate which variables influenced the prediction) for one prediction and variable B as causal for the next prediction. Furthermore, they often fail to implicate as many variables as influencing their predictions as they had earlier identified as causal when asked to make explicit judgments of the causal roles of each variable (in a multivariable context in which a COV method must be used to identify causal and noncausal effects). For almost half the sample, the median number of variables implicated as influencing an outcome prediction (and hence at least implicitly judged causal) in the context of the MVP task was one (of a possible five). Both of these patterns clearly violate the analysis of variance assumptions of consistency and additivity of effects. Kuhn and Dean (2004) thus characterized these patterns as reflecting an immature mental model of multivariable causality. This study, as well as related work (Dean & Kuhn, 2007; Kuhn & Dean, 2005; Kuhn & Pease, 2008; Zohar, 1995), make it clear that children and even many adults cannot readily conceptualize additive main effects, much less interaction effects.

It is this multivariable scientific reasoning task that we report on here. In a study reported by Kuhn (2007), the possibility was considered and rejected that difficulties with the MVP task are an epiphenomenon of immature scientific method skill (specifically, control of variables), such that once individuals understand how to analyze multivariable constellations into component individual effects, they should have no trouble taking into consideration and aggregating these individual effects so as to perform the MVP task. In a test of this hypothesis, the COV strategy was induced using an engagement and practice method over time with a fourth-grade sample. Although the intervention was successful in inducing the COV strategy among a majority of participants, successful participants continued to do poorly on an MVP task, failing to take into account the effects of all of the causal variables they had identified and frequently violating the consistency principle (i.e., that a factor that produces an effect can be expected to produce the same effect on other occasions, given similar conditions).

In the study reported here, we investigate another potential source of difficulty with the task. In the Kuhn (2007) study with fourth graders, all participants classified as successful in employing the COV strategy had by the end of the intervention period identified the three causal and two noncausal effects operating in the multivariable context that they explored. Conceivably, however, it may have represented a cognitive overload for them to keep in mind and then take into account the operation of all three causal factors in the MVP task. To investigate this possibility, we worked with slightly older students (sixth graders) and, additionally, made available to each of them a chart that clearly identified, pictorially and with supporting text, the three causal and two noncausal effects operating in the problem. The task, then, was simply to apply this information to the problem of predicting outcomes based on the status of the three effective variables. The question we ask is whether providing this support, and thus eliminating any cognitive demand of knowing, remembering, or keeping in mind the relevant effects, the difficulties in executing the MVP task would be eliminated.

1.1. Method

Participants were 91 sixth-grade students attending an urban public middle school that requires an entrance examination for admission and is therefore moderately selective. The students were functioning academically at or near grade level. Roughly two thirds were ethnic minorities (largely Hispanic, Asian, or African-American). Gender representation was approximately equal.

Two thirds of the students were randomly paired with a classmate, creating 15 pairs. Pairs were instructed to discuss with their partner and reach a joint decision. The remaining 31 students worked alone. The task, which involved which of five dichotomous variables influenced the risk of avalanches, was administered via laptop computer to each pair or individual. As shown in Fig. 1, students needed to make a prediction for the case given (left segment of the screen) and indicate which variables had figured in their prediction (right segment of the screen). A total of three such cases were presented and predictions elicited in one iteration of the program. The data we examine consist of students' predictions given in response to the computer probe (left side of screen, Fig. 1) and responses to the question of which variables had entered into their prediction (right side of screen). The cases presented to students were randomly generated by the program. Students received no feedback (and hence no true causal structure was depicted by the software).

To assist in their task, a set of five large colored cards, one for each feature, was left for reference at each table of 4–6 students, after the set was introduced to the class as a whole. For causal features, the card showed, for example, in large type, the printed name "cloud cover" and a sketch followed by the text "Makes a Difference." In addition, one variable level, "light," appeared in text and sketch on one side of the card with a "+" sign next to it, while the alternate variable level, "heavy," appeared on the other side with a "−" next to it, providing the information as to which level was associated with higher

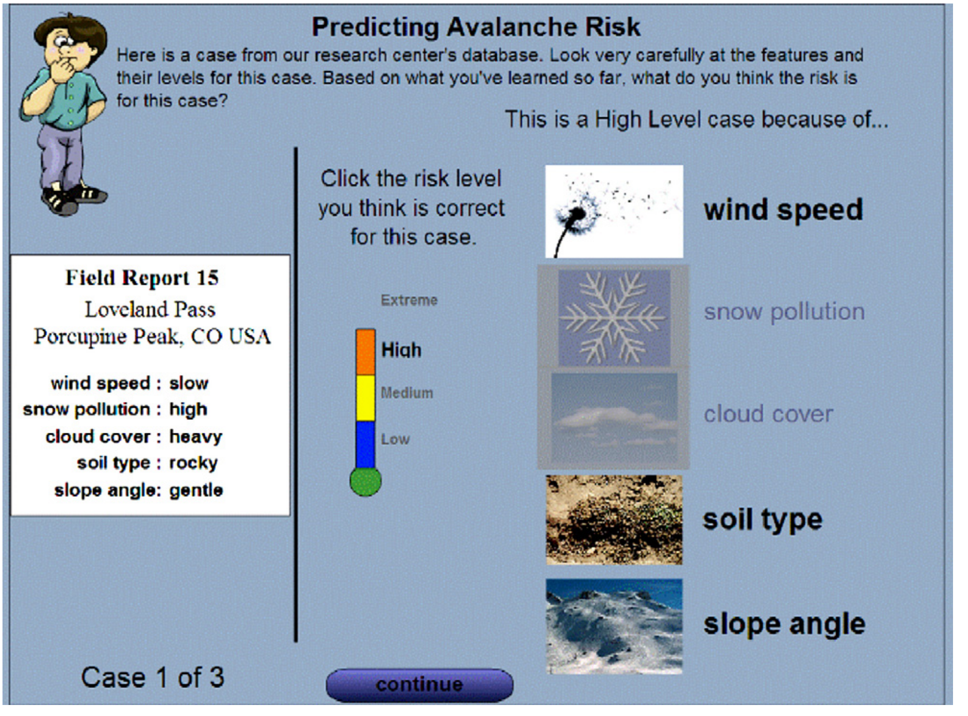

Fig. 1. Predicting avalanche risk.

risk. For noncausal features, the card showed the printed name and sketch of the feature followed by the text "Makes No Difference."

Students in addition completed a "reflection sheet" that promoted their reflection on the three cases together. On the sheet they recorded what the variable levels were for each of the cases, their predictions, and justifications for each prediction.

1.2. Results

Students made predictive judgments about three cases. For each case, from 0 to 5 variables could be implicated as having played a role in the outcome. Our primary interest was in the degree of consistency of these causal attributions over the three cases. Inconsistency for a particular variable is defined as not consistently implicating the variable as either causal or noncausal (i.e., implicating the variable as causal with respect to one or two of the three cases and noncausal with respect to the other one or two).

At each session, inconsistency can be shown for any number from 0 to 5 of the five variables identified in the task. As summarized in Table 1, compared to the three causal variables, students were more likely to be consistent in not invoking either of the two noncausal variables as having played a role in the outcome (by highlighting them on the right side of the screen; see Fig. 1). Almost half of both the individual students and pairs achieved this consistency. Thus, students were relatively more successful in recognizing and then consistently disregarding the two noncausal variables.

Consistency is much lower, however, with respect to the variables invoked as having a causal effect. Only a small minority (again of both individuals and pairs) show complete consistency across the three cases, and from a third to half show no consistency, invoking a different set of variables as having contributed to the outcome in each of the three cases. (These attributions, note, need not be correct to be counted as consistent; a student, for example, might fail to attribute causality to one of the causal

Table 1
Students' consistency in a multivariable prediction task.

	Consistent on all variables	Consistent on some variables	Consistent on no variables	Total
Solitary condition				
3 causal variables	5	16	10	31 students
2 noncausal variables	14	10	7	31 students
Pair condition				
3 causal variables	5	11	14	30 pairs
2 noncausal variables	14	11	5	30 pairs

Note: Entries are numbers of students or student pairs.

Table 2
Number of variables students implicate as causal (averaged across three cases).

	Number of variables			
	3.00 or more	Between 2.00 and 2.99	Between 1.00 and 1.99	Below 1.00
Solitary condition	9	10	9	3
Pair condition	11	9	9	1

Note: Entries are numbers of students or student pairs.

variables across all three cases and be categorized as consistent for this variable.) As seen in Table 1, performance across conditions is very similar.

In Table 2 are shown for each condition how many variables students tended to implicate as causal. As seen there, a majority of students in both conditions implicated as having influenced their prediction on average fewer than the three variables they had been instructed "make a difference." Only about a third of students in each condition implicate at least three (or occasionally more) variables. The remaining majority implicate fewer than three. In other words, then, in making individual predictions the majority tend not to take into account, or at least acknowledge that they are taking into account, all of the variables that need to be considered as the basis for an accurate prediction.[1] Again, performance across conditions is very similar.

The pair Francesca and Angela (pseudonyms) provide a typical example. For the first prediction, they implicate a single variable, snow pollution, as influencing their prediction, elaborating "Because it shows the snow pollution is high; snow is what causes an avalanche." For the second prediction, they again implicate a single variable but this time it is slope angle, and they explain, "Slope angle is an important part of how snow falls." For the third prediction, they again implicate a single variable, this time wind speed. "We chose the wind speed because it affects how fast the snow falls."

A second example, the pair Jason and Rosa, makes it clear that thinking in terms of a single causal factor as responsible for an effect is not the only cause of their difficulties (and inconsistency across predictions). Jason and Rosa implicate two variables to account for their first prediction, three variables to account for the second, and four of the five variables to account for the third. They show the same vacillation as do Francesca and Angela, however, in which of the variables they implicate as having influenced each prediction. For the first prediction, they implicate soil type, slope angle and wind speed, but then note, "The wind speed didn't factor in because it wasn't fast enough to make an impact." For the

[1] Note our data reflect the features students *say* they considered in making each prediction judgment. The ones they in fact took into account may have been different (and indeed the data suggest this was so, since a frequent error was to deny taking into account a variable when its level was associated with low risk, although it had to have been considered to determine this). This approach on our part reflects our interest in students' explicit understandings regarding how multiple variables additively affect an outcome. Prediction error, nonetheless, provides an index of how successful they were in performing the neeessary integrations of effects, whether or not they were aware of how they were doing so. Interestingly, in earlier work (Anderson, 1991; Capon & Kuhn, 1980) involving implicit judgments (assessed by computing ANOVAs for each individual to reveal which features had influenced judgments), results are similar to those found here: fewer features influence judgments than those the individual claims do so.

second prediction, they implicate wind speed and cloud cover, but then explain, similarly, “Since the cloud cover was light, it wouldn’t make any difference.” For their final prediction, they again implicate wind speed and cloud cover, but this time they add slope angle as having influenced their prediction. Hence soil type is implicated only once and slope angle and cloud cover twice. Only wind speed is implicated consistently across all three cases, although, as noted, in the case of the first prediction they vacillate regarding its relevance.

1.3. Discussion

In the social psychology literature, discounting (of a second causal variable if the first one is perceived as explaining the effect) has been a topic of much research, carrying with it an implication that other potentially explanatory variables have been considered and rejected, and discounting has been noted to appear in children by age 8 or 9 (Sedlak & Kurtz, 1981). For many individuals at least, it may be more accurate to characterize the inference process as based on a model in which single factors often suffice to explain outcomes and need not remain consistent across instances. How should these severe limitations on scientific thinking be explained?

We know that children younger than the age of those in this study can under certain conditions appropriately integrate information from at least two sources in an additive fashion (Anderson, 1991; Dixon & Tuccillo, 2001; Wilkening, 1982), so failures to do so in the present causal context cannot readily be attributed to processing limitations. In the present case, they needed to integrate the effects of three variables, these individual effects having been explicitly identified. If the problem were one of processing overload, they might simply have consistently ignored one of the variables and focused on integrating the other two. The results, however, do not support such a model.

The present results also fail to support the explanation that the difficulty lies in availability of the knowledge regarding the individual variables. Providing an external aid to represent their effects did not eliminate the coordination challenge. What, then, does account for difficulty with the MVP task? The task poses two challenges, corresponding to the two assumptions of the ANOVA model underlying multivariable causality. First, all variables that affect the outcome must be taken into account, rather than only those that happen to be the momentary focus of attention. The second challenge posed by the MVP task, however, corresponding to the other ANOVA assumption, is more difficult to accomplish, and that is a revision of one’s mental model of multivariable causality to incorporate consistency across occasions as a constraint. If one is operating under a model in which no (or inadequate) constraints exist requiring that causal variables have to operate consistently across different occasions, the factors that would lead to imposing such constraints are not obvious.

In current work, we are examining microgenetically the progress that students make in this task over time, with repeated engagement (Kuhn, Pease, & Wirkala, 2008). In particular we are examining a conceptual difficulty students appear to have with the distinction between variables and individual levels of those variables. We are also continuing to examine the potential benefit that working in a social (paired) context may confer, over a longer period of time. At a broader level, we see as notable regarding the challenges students confront in this area the fact that they encompass not only procedural skills but also conceptual understanding that is meta-level or epistemological in nature, e.g., the question of whether causal effects must operate consistently.

Most kinds of scientific thinking, we have suggested, encompass both procedural and epistemological aspects. In the next section, we turn to a study that examines the epistemological underpinnings of scientific understanding. In a final section, we consider argumentation in science, where both procedural and epistemological aspects figure importantly.

2. Scientific thinking as understanding the nature of science

To engage in scientific study in a way that is at all authentic and more than superficial, science students must come to understand that scientific knowledge is constructed by humans, not simply discovered in the world (Sandoval, 2005). By this criterion, researchers have shown science students at the intermediate-school and even high-school level to do poorly, with many continuing to regard science from an absolutist stance, as the accumulation of certain facts (Leach, Driver, Millar, & Scott, 1997; Smith, Maclin, Houghton, & Hennessey, 2000). Progression beyond the absolutist belief in certain knowledge has been examined in both scientific and non-scientific domains, the latter most often in an area of research that has come to be known as "personal epistemology" (Hofer & Pintrich, 1997). Studies that have compared levels of epistemological understanding across a science and non-science domain have shown some but far from perfect correspondence (Buehl & Alexander, 2005; Buehl, Alexander, & Murphy, 2002). It is not entirely clear what conclusions to draw from such findings, however, and, as Muis, Bendixen, and Haerle (2006) note in their review, the studies reported have been subject to variable interpretation, supporting both domain-generality and domain-specificity across scientific and social topics.

The question is an important one in understanding what it is that needs to develop in the realm of epistemological understanding. In the work presented here, rather than the quantitative approach of assessing degree of correspondence between allegedly parallel problems across domains, we undertake a more qualitative and conceptual approach, asking first of all what kinds of understandings need to develop in science and non-science domains with respect to the nature of knowledge and knowing. In other words, we ask the qualitative question of *how* the developmental challenges and progression in the different domains are different, rather than only the quantitative question of the degree to which performance is parallel across them.

Three-year-olds make the now well-known error of unwillingness to attribute to another a belief they know to be false. Less often noted about the development of false-belief understanding by age 4 or 5 is its foundational status in the development of scientific thinking. Three criteria must be met if one is to engage in the coordination of theory and evidence, the essence of scientific thinking (Kuhn, 2002): The theoretical claim must be recognized as falsifiable, evidence must be recognized as the means of falsification, and theoretical claim and evidence must be recognized as distinct epistemological categories—evidence must be distinguishable from the theory itself and bear on its correctness. The 3-year-old *realists* just described meet none of these criteria. Thus, when asked for justification for a knowledge claim (that an event occurred), preschoolers are likely to respond with a theory as to why occurrence of the event is plausible rather than with evidence that it did in fact occur (Kuhn & Pearsall, 2000).

Once it is recognized that assertions are produced by human minds and need not necessarily correspond to reality, assertions become susceptible to evaluation vis-à-vis the reality from which they are now distinguished. Here the potential for scientific thinking emerges. The products of knowing, however, for a time still remain more firmly attached to the known object than to the knower. Hence, while inadequate or incorrect information can produce false beliefs, these errors are easily correctable by reference to an external reality—the known object. To be wrong is simply to be misinformed, mistaken, in a way that is readily correctable once the appropriate information is revealed (Carpendale & Chandler, 1996). At this *absolutist* level of epistemological understanding, knowledge is thus regarded as an accumulating set of certain facts.

Further progress in epistemological understanding can be characterized as an extended task of coordinating the subjective and the objective elements of knowing (Kuhn, Cheney, & Weinstock, 2000). At the realist and absolutist levels, the objective dominates. It is at the next levels that the progression in scientific and non-scientific domains appears to diverge. In the personal epistemology realm, there is general agreement that further development proceeds toward transition to a broad level of *multiplism* or relativism, followed by, in at least some individuals, an evaluativist level (Hofer & Pintrich, 1997, 2002; Moshman, 2008). The discovery that reasonable people – even experts – disagree is a likely source of recognizing the uncertain, subjective aspect of knowing. This recognition initially assumes such proportions, however, that it eclipses recognition of any objective standard that could serve as a basis for evaluating conflicting claims. At this level, knowledge consists not of facts but of opinions, freely chosen by their holders as personal possessions and accordingly not open to challenge. Knowledge is now clearly seen as emanating from the knower, rather than the known, but at the significant cost of any discriminability among competing knowledge claims. By adulthood, many, though by no means all, adolescents will have reintegrated the objective dimension of knowing and achieved the understanding that while everyone has a right to their opinion, some opinions are in fact more right than others. Knowledge at this *evaluativist* level consists of judgments, which require support in a framework of alternatives, evidence, and argument.

In progressing beyond an absolutist position, the challenges in a scientific domain are related but somewhat different. The entry of human interpretation into what was previously regarded as direct perception of a single reality is to be understood in positive terms—science comes to be appreciated as a human construction and interpretation as an essential resource for knowing. The “filter” that human minds represent empowers scientific endeavor. Human rational construction and interpretation (with the inevitable subjective element that accompanies human interpretation) are recognized as essential to science. They entail the construction of alternative possibilities (multiple representations of truth, or theories) that need to be coordinated with empirical evidence, in an ongoing process that constitutes scientific work. Scientific knowledge accordingly comes to be understood as evolving and subject to revision.

In the social domain, in contrast, when human subjectivity first impinges on the absolutist realm of objective fact, it is typically regarded in negative terms, as the intrusion of human “bias”—a necessary evil to be recognized and accommodated. The absolutist often looks for a way to get beyond this bias to uncover the “true facts.” But the danger is one of a permanent stall in the multiplist’s embrace of radical relativism, with the evil of subjectivity seen as overpowering the quest for any knowledge beyond subjective opinion (a position rarely seen in the domain of science). In the social domain, then, the major challenge is to come to terms with the concern that human interpretation plays an unmanageable, overpowering role, while in the science domain, the major challenge is to recognize that human interpretation plays any role at all.

2.1. Method

2.1.1. Design

We report here a study in which we presented to sixth graders (as well as teachers) the three scenarios shown in Tables 3–5, one in the history domain and the other two in the science domain, and asked the same key questions in each case regarding the certainty of knowledge: 1. Could anyone ever be certain about what happened in the Fifth Livian War? [about why dinosaurs became extinct/about the effect of eating fish on health]? 2. What would help us become more certain?

The available research (see Muis et al., 2006, for review) leads to a prediction of greater willingness to relinquish the absolutist idea of certain knowledge in the social domain and greater resistance to doing so in the science domain. Our purpose was to identify a problem in the science domain that would facilitate willingness to loosen the absolutist’s commitment to knowledge

Table 3
The Livia problem.

North and South Livia are two small countries that existed in the 1800s in Asia. There occurred a series of conflicts between the two countries, termed the Livian Wars. Here are two brief accounts of the Fifth Livian War that took place in 1878.
A brief account of the Fifth Livian War by J. Abman, National Historian of North Livia
On 19th July 1878, during a national ceremony in North Livia to honor one of their national leaders, ceremonies were interrupted by a sneak attack from South Livia. Thus began the Fifth Livian war. Because the North Livians were caught by surprise, they were unprepared at first and the South Livians won a few early battles. Then North Livia began to win. But before the North Livians could reach a final victory, a neighboring large country intervened to stop further bloodshed.
Despite their early setbacks, the later sweeping victories of the North Livians showed they would have won had the fighting continued. As a result of this war, South Livians finally recognized anything they gained from North Livians would have to be worked out through peaceful negotiations. So ended the Livian Wars.
A brief account of the Fifth Livian War by N. Ivan, National Historian of South Livia
In the last war, North Livia had beaten South Livia, taken some of its land, and refused to leave. South Livia could no longer accept this situation and spent much money to strengthen its military. On 20th July 1878, the Fifth Livian war began. The war took place with rapid, dramatic victories for South Livia, resulting in great celebration. After these dramatic victories, the South Livians suffered some minor losses. But then a neighboring large country intervened to stop further bloodshed.
Despite their later setbacks, South Livia’s victory seemed assured because of its position of strength. As a result of this war, South Livians felt a new self-respect. They had felt embarrassed by their previous defeats, but now they had proven they were the equals of the North Livians. Because South Livians had achieved military respect, they were willing to work out differences through peaceful negotiations, thus ending the Livian Wars.

Table 4
The Dinosaur problem.

Dinosaurs dominated the Earth for nearly 150 million years. Dinosaurs disappeared at the end of the Cretaceous Period, about 65 million years ago. There are different views about why dinosaurs disappeared. Recently a new finding was reported – a layer rich in Iridium near the geological layers of the Cretaceous Period.
According to scientist Luis Alvarez, this finding supports his view that dinosaurs died out when the earth was hit by a meteorite. (Meteorites contain a lot of Iridium.) The collision left enormous amount of dust in the air that blocked the sunlight, resulting in a long dark winter that caused plants to die. Dinosaurs died from starvation and the very cold climate.
According to scientist Norman MacLeod, this finding supports his view that dinosaurs died out because of the difficult climate conditions caused by a series of giant volcanic eruptions from deep in the earth. (Large quantities of Iridium are found at the earth's core.) The volcanic eruptions filled the air with poison gas. This caused a Greenhouse effect, which raised the Earth's temperatures. Dinosaurs died from the poison gas and very hot temperatures.

Table 5
The Fish problem.

[From the New York Times, Wednesday 18 October, 2006 page F5, Washington]
A report about the risks and benefits of eating seafood, released by the **Harvard School of Public Health**, said eating fish reduces the risk of heart attack death by 36% and total death by 17%.
A similar report released simultaneously by the **National Institutes of Health** concluded that there is only enough evidence to say that eating fish, especially fatty fish like salmon and mackerel, "may" reduce the risk of heart disease. . .
Dr. D. Mozaffarian, one author of the **Harvard study** said, "Seafood is likely the single most important food one can consume for good health."
Dr. Marion Nestle, a professor of nutrition, public health and food safety at New York University remains unconvinced. "Those of us who have been in nutrition for a long time have seen miracle foods come and go: vitamin E for heart disease, beta carotene to prevent cancer; now it's fish."
Dr. J. M. Ordovas of the **National Institute of Health** panel and a professor of nutrition at Tufts. . . said the 36% figure "is based on circumstantial evidence that does not provide definite proof."
Dr. Mozaffarian responded, "It's the best evidence we have."

as objective fact (and thereby foster developmental advance), to a degree that would equal or exceed that observed in the social domain. Because the Dinosaur phenomena are situated in a pre-human era, the possibility of the direct observation by humans that an absolutist conception requires is precluded. Moreover, because direct observation is impossible, the potential negative influence of human subjectivity (the "biased" observation that is a dominant concern in the social domain) is minimized. In the Dinosaur problem, then, the only way in which knowledge can advance is by means of the positive, constructive role of human theorizing and its coordination with various forms of indirect evidence. The Fish problem remains in a science domain but reintroduces both of the factors that the Dinosaur problem seeks to minimize—first, the availability of direct observation and, second, the prominence of human subjectivity (and potential "bias") not typically highlighted in a science domain. We thus predicted that the Dinosaur problem would facilitate epistemological understanding beyond an absolutist level, to a level at least equal to and perhaps beyond that observed in the non-scientific (Livia) problem, but that the Fish problem would not have this facilitative effect.

2.1.2. Participants

Participants were 43 teachers and 35 sixth graders attending an urban independent school. The students were functioning academically at or near grade level. Roughly one half were ethnic minorities (largely Hispanic, Asian, or African-American). Gender representation was approximately equal.

The 43 teachers were a convenience sample solicited from a variety of schools, backgrounds and kinds of teaching experience. Three quarters were female and slightly over half held advanced degrees. Roughly two thirds taught at the high school level and the remainder at the elementary level. Only five

Table 6
Levels of epistemological understanding on Dinosaur and Livia problems.

Level	Livia problem	Dinosaur problem
Absolutist		
Certainty empirically possible via direct observation	Talk to anyone who was around at that time. *A neutral eyewitness to the event, such as a neighboring country.*	A series of carbon dating and seeing the pattern in global warming could prove it.
Certainty not possible due to absence of observers, but theoretically would be if an observer had been present	*If we could find direct evidence to prove facts about the war.*	An invention that would travel us back in time to see what really happened.
Multiplist		
Certainty not possible due to the subjective nature of knowing	All history is someone's opinion, point of view, or side of a story. *Both parties have their own viewpoint, making each account "truthful" for them.*	I don't really know how to become more certain about how dinosaurs became extinct. I don't think there's much more scientists can do.
Evaluativist		
Certainty not possible but approachable through investigation and interpretation of evidence	Look at more accounts, noting similarities and differences. *Look up records from the countries in that area.*	Trying to find artifacts that will give us clues of what happened. *Comparing findings and valid information may help narrow down some reasons.*

Note: Teachers' responses in italics.

teachers were specialists in science. Teaching experience ranged from 1 to 39 years with a median of 10 years.

2.1.3. Procedure

All participants were asked to respond in writing to the three problems in Tables 3–5. Order of presentation of the three was counterbalanced across participants.

A portion of responses (12%) from both samples was used as a basis for developing a coding scheme that captured the range of responses observed. After devising the initial scheme, coders coded another 10% of responses to insure that the coding scheme was able to capture all responses. Reliability of classification (into the categories to be introduced below) was assessed on the remainder of the sample. Average percentage agreement between two coders was 89% for the Livia problem, 80% for the Dinosaur problem and 83% for the Fish problem. Disagreements were resolved through discussion.

2.2. Results

As no differences were observed as a function of order of administration of the three problems, data were collapsed across orders and this variable was not examined further. We begin with an overview and examples of the kinds of responses we observed to the Livia problem and Dinosaur problem, postponing examination of responses to the Fish problem. Responses were classified into the four categories shown in Table 6, which includes examples of each, drawn from both sixth graders and teachers (the latter appearing in italics). Those in the category presented in the first row responded that certainty was possible and achieved via direct observation or examination of data. Those in the category presented in the second row were similar in their reliance on direct observation as the path to knowing, but they responded that certainty was not possible due to the impracticability of such observation in this situation. Respondents in the two remaining categories also claimed certainty to be impossible, in the multiplist category in the Livia problem because of the subjectivity of human knowing and in the Dinosaur problem because no means of knowing can be envisioned. In the evaluativist category, in contrast, respondents also reject certainty, but investigation, analysis, and interpretation are all embraced as ways of knowing.

Table 7
Percentages of sixth-grade students showing different levels of epistemological understanding in response to Dinosaur and Livia problems.

Level	Livia problem (%)	Dinosaur problem (%)
Absolutist		
Certainty empirically possible	46	03
Certainty only theoretically possible	20	31
Total absolutist	66	34
Multiplist	09	09
Evaluativist	26	57

Table 8
Levels of epistemological understanding on the Fish problem.

Level	Fish Problem
Absolutist	
Certainty empirically possible via direct observation	If you actually run a test on someone.
Certainty only theoretically possible, due to practical limitations	*The only way seems to test people who ONLY eat fish and this is near impossible.*
Multiplist	
Certainty not possible due to the subjectivity associated with human phenomena	*The effect of eating fish on health is likely true for some people and perhaps not true for others.* *Other factors such as lifestyle play a role.*
Evaluativist	
Certainty not possible but approachable through investigation and interpretation of evidence	Test on many different people that have heart disease, and if it reduces their chance, you can be more certain. *A longer period of testings with large control and test groups from different areas.*

2.2.1. *Sixth graders' performance*

As seen in Table 7, sixth graders' responses to the Livia problem show the largest proportion of students of this age to be at an absolutist level of understanding, consistent with expectation (Hofer & Pintrich, 1997, 2002). In the Dinosaur problem, however, despite its scientific content and despite the expectation from previous research noted earlier of a greater clinging to absolutism in the science domain, the proportion of students at the absolutist level drops by almost half, and the proportion classified at the evaluativist level is more than twice what it is for the Livia problem.

Individual patterns across the two problems are consistent with these group differences. Of those classified as absolutist on the Livia problem, 57% achieve the multiplist or evaluativist level on the Dinosaur problem; of those classified as multiplist 67% achieve the evaluativist level on the Dinosaur problem. Only two students of 35 (6% of the sample) drop from the evaluativist to a less advanced level when encountering the Dinosaur problem.

Students' responses to the Dinosaur problem can be considered further in relation to their responses to the second problem in the science domain, the Fish problem. Responses to this problem were classified into the same four categories as the other two problems. Examples appear in Table 8. A total of 71% of sixth graders' responses were classified in the absolutist category, with 68% reporting certainty empirically possible and 3% certainty only theoretically possible. None fell into the multiplist category and 21% into the evaluativist category. A comparison with the data in Table 7 shows that these percentages are less advanced than those for the Dinosaur problem and in fact revert to the level of those for the Livia problem.

2.2.2. *Teachers' performance*

Results for teachers appear in Table 9. Here, it is seen, teachers overall are more likely to be in the multiplist or evaluativist categories, and less likely to be in the absolutist category, relative to sixth graders, consistent with the assumption that it is the former modes of thinking that are more developmentally advanced. The superiority of performance on the Dinosaur problem over that on the

Table 9
Percentages of teachers showing different levels of epistemological understanding in response to Livia, Dinosaur and Fish problems.

Level	Livia problem (%)	Dinosaur problem (%)	Fish problem (%)
Absolutist			
Certainty empirically possible	23	00	33
Certainty only theoretically possible	21	21	00
Total absolutist	44	21	33
Multiplist	21	09	16
Evaluativist	35	70	51

Livia problem seen in the younger sample, nonetheless, is replicated. Teachers are twice as likely to be classified in the evaluativist category on the Dinosaur problem as they are on the Livia problem.[2] In contrast to the pattern observed among the sixth graders, however, performance of teachers on the Fish problem is intermediate between their performance on the other two problems (Table 9).

2.3. Discussion

The patterns of performance of the two samples across the three problems support our theoretical analysis of the common and unique challenges that developing epistemological understanding poses in the social and science domains. The Fish problem remains in a science domain but reintroduces both of the factors that the Dinosaur problem sought to minimize—the availability of direct observation and the prominence of human subjectivity (and potential “bias”) not typically characteristic of the science domain. If these are indeed the challenges that are confronted in developing mature epistemological understanding, then reintroducing them in the Fish problem should cause levels of understanding to revert to levels no higher than those exhibited on the Livia problem. Alternatively, if it is something else about the human social context of the Livia problem that creates its challenge, and/or something else about the Dinosaur problem that reduces this challenge, then the superior performance seen on the Dinosaur problem might be maintained on the Fish problem.

Results for the student sample support our analysis—performance on the Fish problem is inferior to that on the Dinosaur problem and reverts to its level on the Livia problem. For the teachers, in contrast, who are as a whole already at a more advanced level than the students, this prediction is only partially confirmed. Their level of performance on the Fish problem on average reverts to a level midway between that on the Livia and Dinosaur problems. A likely explanation is that their further progress toward an evaluativist level of understanding, relative to the students, makes the teachers more aware of the role of human knowledge, enabling them to see its relevance in both the Fish and Dinosaur contexts. Nonetheless, the potential for direct observation in the Fish problem is seductive and pulls some teachers back down toward absolutist thinking (as it does many sixth graders).

What are the broader implications of our analysis? In his analysis of student epistemologies of science, Sandoval (2005) characterizes the central questions that such epistemologies must address as these: How do we know what we know? And why do we believe it? He goes on to claim that, “Probably the most important epistemological notion for students to understand is that scientific knowledge is constructed by people, not simply discovered out in the world” (p. 639).

It would be hard to claim that these are not the core questions and challenges that characterize developing epistemological understanding of knowledge in general, not only students’ understanding of science. And yet Sandoval in his review, and many other science educators along with him, treat the literature on personal epistemology as only tangentially related to how students understand the nature of science. A goal of the study presented here is to bring these two lines of investigation together

[2] Individual patterns of performance are again consistent with group trends. Of teachers classified as absolutist on the Livia problem, 79% advance to the multiplist or evaluativist level on the Dinosaur problem; of those classified as multiplist, 56% advance to evaluativist on the Dinosaur problem.

by seeking to identify both the common and the more particular challenges that developing epistemological understanding poses across science and social domains. The common, and broadest, challenge, we have suggested, is the need to allow into one's conception of knowing the human knower as the source and constructor of knowledge. And the further challenge is to do so in a way that does not damage the knowing process but, to the contrary, enables it to flourish.

If so, the analytic task becomes one of fleshing out just how this broad challenge is met in scientific and social domains, in the process identifying in qualitative terms *how* they are different, not simply whether or to what degree they are different (the latter the goal of a strictly empirical approach). The kind of knowledge being contemplated stands to influence conceptions of the nature and source of this knowledge, certainly. But this specificity, we have seen here, applies not just across traditional knowledge domains but within domains as well. The phenomena involved in our Dinosaur and Fish problems would be widely accepted as falling within a physical science domain. Yet within this domain, the more subtle nuances that distinguish the Dinosaur and Fish problems affect the reasoning they elicit. These nuances, it is also worth noting, extend not only to problem content but also to the specific kinds of questions that are asked. Students of the age examined here may abandon absolutism to the extent of acknowledging that two views "can both be right" (Kuhn et al., 2000). Yet, as we saw here, they are highly likely to cling to absolutism when asked about the certainty of knowledge.

This specificity of content and context by no means dictates a retreat to radical domain-specificity. To the contrary, a common developmental progression can be identified, with a number of different and more specific challenges that may be encountered along the way. Different problem content and contexts emphasize these challenges to a greater or lesser extent. Analyzing these challenges within a framework of the more global intellectual development that is taking place keeps them from appearing as fragmentary bits of understanding to be mastered individually.

This way of analyzing the issue, we believe, has educational implications, for science education certainly but for other subjects as well. It has by now been widely recognized that epistemological understanding has both broad and specific implications for academic performance (Buehl & Alexander, 2005; Kuhn, 1991, 2005; Mason & Boscolo, 2004) as well as performance outside of academic contexts (Kuhn, Weinstock, & Flaton, 1994; Moshman, 2005). Although only a few studies have undertaken to advance participants' epistemological understanding as a goal (Khishfe & Abd-El-Khalick, 2002; Sandoval & Morrison, 2003; Smith et al., 2000), a broadly held view has been that dense engagement in knowledge-building activities is the most likely path to effecting such advances. Sandoval (2005) emphasizes engagement in authentic scientific inquiry as a means of developing epistemological understanding in science. Studies by Kuhn and colleagues and others (Felton & Kuhn, 2001; Kuhn, 1993; Kuhn & Udell, 2003, 2007; Mason & Scirica, 2006) have emphasized engagement in dialogic argument as a vehicle for coming to recognize and value its role in knowledge construction, the topic we turn to in the next section. By engaging in argument one comes to see the point of it.

The data presented here, and in particular the better performance that both students and teachers displayed on the Dinosaur problem, relative to the other two problems, demonstrates that epistemological understanding is indeed nuanced. The implication, we suggest, is the fruitfulness of closely examining these nuances, both as a means of better understanding what needs to develop and as a means of identifying how best to support students in confronting these developmental challenges. The view has by now been widely expressed that the development of epistemological understanding is critical to young people's academic progress in science (Leach et al., 1997; Metz, 2004; Sandoval, 2005), to their academic progress more broadly (Buehl & Alexander, 2005; Mason & Boscolo, 2004; Mason & Scirica, 2006), and to their intellectual productivity outside of and beyond school (Kuhn, 2005; Kuhn & Park, 2005). It perhaps comes as no great surprise that teachers do not routinely show the advanced levels of epistemological understanding we would like to see (Olafson & Schraw, 2006; Windschul & Thompson, 2006). Such findings, while troubling, only underscore the developmental challenge that is at stake as we seek ways to support students in meeting it.

3. Scientific thinking as argumentation

Argumentation has by now become explicitly recognized as an essential aspect of scientific thinking (Duschl & Osborne, 2002; Erduran & Jimenez-Aleixandre, 2008; Kuhn, 1993; Kelly, Druker, & Chen, 1998; Lehrer, Schauble, & Petrosino, 2001). Students who gain an epistemological understanding of science as involving coordination of evidence with evolving theories constructed by human knowers come to recognize argument as central to the process by which science advances. They need to both be aware of its relevance and become skilled in its execution if they are to engage deeply in science. In the space remaining here, we describe briefly a recent study by Iordanou (submitted for publication) examining this process.

The study follows a line of work devoted to fostering the development of argumentation skills in early adolescents based on engagement and practice in argumentive discourse (Felton, 2004; Kuhn, Shaw, & Felton, 1997; Kuhn & Udell, 2003; Kuhn, Goh, Iordanou, & Shaenfield, 2008; Udell, 2007). A particular feature of the study described here is its use of instant-messaging computer software as the medium of discourse, modeled on the successful use of this method by Kuhn, Goh, et al. (2008). Several studies by other researchers (Andriessen, 2006; Bell & Linn, 2000; Clark, Sampson, Weinberger, & Erkens, 2007) suggest that this medium is a fruitful one for scaffolding argumentation in science domains. In contrast to these studies, however, ours involves no software-based scaffolding of argument construction and evaluation, beyond the instant-messaging software itself. The further question the study investigates is that of transfer of argumentation skills across scientific and non-scientific domains. Do skills developed in the scientific domain transfer to non-scientific domains and vice versa?

A potential problem with applying in a science domain the dialogic methods used in work to foster development of argument skills is that young students are widely regarded as lacking sufficient knowledge about science topics to engage in productive debate. In social domains, in contrast, people of all ages are likely to consider themselves “experts” or, at the very least, to feel entitled to hold and defend strong positions on a topic. In the study described here, we undertake to address this challenge by providing students with a constrained knowledge base (a set of “possibly relevant facts”) that is equated across students and can serve as a basis for their argumentation.

In the Iordanou study, 40 sixth graders engaged in instant-messaging (IM) electronic discourse on a controversial topic; for half, the topic was a choice between two rival theories to account for the extinction of dinosaurs (the science topic) and for the other half the topic was whether parents should be allowed to home school their children (the social topic). A sheet containing a list of several “Possibly Relevant Facts,” each a sentence or two in length, was provided, for either the science or social topic depending on condition, and remained available throughout the series of IM discussions. Over 13 sessions, students collaborated with a partner in conducting arguments via IM with a succession of pairs of classmates who held an opposing view on the topic. They were instructed to collaborate with their partner in formulating what to say and to try to persuade the opposing pair to their position if possible. In addition each pair engaged in reflective activities (using transcriptions of the dialogs that were provided to them), in which they identified their own and the opposing pair’s arguments and counterarguments and considered how they could be strengthened or addressed.

The results indicate that argumentation skills in scientific domains are amenable to development in the same way as are skills in social domains, using a computer-based method centered on engagement, practice and reflection. Compared to an additional group of students in a control (non-intervention) condition, from initial to final assessment students exhibited increased frequency of usage of skilled (counterargument and rebuttal) argument strategies and decreased frequency of less advanced (simple exposition, ignoring the other’s position) strategies. The intervention proved also to be successful in producing transfer of argument skills across domains in both directions. However, a difference in the magnitude of transfer was observed, with only participants in the science topic condition able to transfer their achievements in argumentation skill to the non-intervention (social) topic to the same degree that these skills were mastered in the science topic. Intervention gains in the social-topic condition transferred less well to the science topic.

At least part of the explanation for the transfer observed may lie at the meta-level. Kuhn, Goh, et al. (2008), based on a microgenetic examination of developing meta-level discourse during argumentation, maintain that reflection on the argumentation process co-exists with argumentation skills that operate at the procedural level. As meta-level understanding of argumentation develops, it supports the execution of argument skills at the procedural level, an outcome that extends across domains of application. In current work we are investigating gains in both epistemological understanding and meta-level strategic understanding as products of the intervention, advances that in turn play a role in strategic gain.

Results of this study have implications for science education. They show that students' limited argument skills in the science domain reported in several studies (Driver, Newton, & Osborne, 2000; Solomon, 1992) are not due to constraints imposed by the nature of the science domain itself. The findings also indicate that direct attention to the development of argumentation skill (and its associated epistemological understanding) within science domains is warranted. Argument skill in the science domain is amenable to the same development as argument skill in the social domain has been shown to be, but specific engagement and practice within the science domain is required for optimum development of such skill. We lack the space here to delve into the critical issues that loom large in the field of science education today. But the policy recommendation supported by the present findings – engagement and practice in argumentation within the context of authentic science topics – is consistent with much recent thought in the field (National Research Council, 2007).

4. Conclusion

We have by no means provided in this article an exhaustive account of what needs to or might develop in the realm of scientific thinking. By sampling from the three broad areas of inquiry, argument, and epistemology, however, we convey an idea of the necessarily wide scope and detail required of a map detailing the development of scientific thinking. It also becomes clear that scientific thinking occurs and develops at multiple levels – the strategic, the metastrategic, the epistemological – and these levels themselves encompass multiple strands. Both strategies and metastrategies, for example, are entailed in inquiry, design of an investigation, data analysis, and argumentation.

It follows that a broad map of this nature ought to inform science education. Educators debate what it is we should try to teach elementary and middle school children about science, and neither scholars nor classroom teachers are certain of the answers. Although much of the terrain is not yet fully charted, a map of the sort we refer to here stands to play an important role in informing their deliberations.

References

Anderson, N. (1991). *Contributions to information integration theory: Vol. III. Development*. Hillsdale, NJ: Erlbaum.

Andriessen, J. (2006). Collaboration in computer conferencing. In A. O'Donnell, C. Hmelo-Silver, & G. Erkens (Eds.), *Collaborative learning, reasoning, and technology*. Mahwah, NJ: Erlbaum.

Bell, P., & Linn, M. (2000). Scientific arguments as learning artifacts: Designing for learning from the Web with KIE. *International Journal of Science Education*, *22*, 797–817.

Buehl, M., & Alexander, P. (2005). Motivation and performance differences in students' domain-specific epistemological belief profiles. *American Educational Research Journal*, *42*, 697–726.

Buehl, M., Alexander, P., & Murphy, P. (2002). Beliefs about schooled knowledge: Domain general or domain specific? *Contemporary Educational Psychology*, *27*, 415–449.

Capon, N., & Kuhn, D. (1980). A developmental study of consumer information-processing strategies. *Journal of Consumer Research*, *7*(3), 225–233.

Carpendale, J. I., & Chandler, M. J. (1996). On the distinction between false belief understanding and subscribing to an interpretive theory of mind. *Child Development*, *67*, 1686–1706.

Clark, D., Sampson, V., Weinberger, A., & Erkens, G. (2007). Analytic frameworks for assessing dialogic argumentation in online learning environments. *Educational Psychology Review*, *10*, 343–374.

Dean, D., & Kuhn, D. (2007). Direct instruction vs. discovery: The long view. *Science Education*, *91*, 384–397.

Dixon, J., & Tuccillo, F. (2001). Generating initial models for reasoning. *Journal of Experimental Child Psychology*, *78*, 178–212.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, *84*, 287–312.

Duschl, R. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of Research in Education*, *32*, 268–291.

Duschl, R., & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. *Studies in Science Education*, *38*, 39–72.

Erduran, S., & Jimenez-Aleixandre, M. P. (Eds.). (2008). *Argumentation in science education*. New York: Springer.

Felton, M. (2004). The development of discourse strategies in adolescent argumentation. *Cognitive Development*, *19*, 35–52.

Felton, M., & Kuhn, D. (2001). The development of argumentive discourse skill. *Discourse Processes*, *32*, 135–153.

In *Paper presented at the annual meeting of the National Association for Research in Science Teaching.*

Gluck, M., Shohamy, D., & Myers, C. (2002). How do people solve the "Weather Prediction" task? Individual variability in strategies for probabilistic category learning. *Learning and Memory*, *9*, 408–418.

Hofer, B. K., & Pintrich, P. R. (Eds.). (2002). *Personal epistemology: The psychology of beliefs about knowledge and knowing*. Madwah, NJ: Erlbaum.

Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation of learning. *Review of Education Research*, *67*(1), 88–140.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.

Iordanou, K. (submitted for publication). *Arguing on the computer in scientific and non-scientific domains*.

Keating, D. (1980). Thinking processes in adolescence. In J. Adelson (Ed.), *Handbook of adolescent psychology*. New York: Wiley.

Kelly, G. J., Druker, S., & Chen, C. (1998). Students' reasoning about electricity: Combining performance assessments with argumentation analysis. *International Journal of Science Education*, *20*, 849–871.

Khishfe, R., & Abd-El-Khalick, F. (2002). Influence of explicit and reflective versus implicit inquiry-oriented instruction on sixth graders' views of nature of science. *Journal of Research in Science Teaching*, *39*(7), 551–578.

Kuhn, D. (2007). Reasoning about multiple variables: Control of variables is not the only challenge. *Science Education*, *91*(5), 710–716.

Kuhn, D. (2005). *Education for thinking*. Cambridge, MA: Harvard University Press.

Kuhn, D. (2002). What is scientific thinking and how does it develop? In U. Goswami (Ed.), *Handbook of childhood cognitive development*. Oxford, England: Blackwell.

Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education*, *77*(3), 319–337.

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

Kuhn, D., Cheney, R., & Weinstock, M. (2000). The development of epistemological understanding. *Cognitive Development*, *15*, 309–328.

Kuhn, D., & Dean, D. (2004). Connecting scientific reasoning and causal inference. *Journal of Cognition and Development*, *5*, 261–288.

Kuhn, D., & Dean, D. (2005). Is developing scientific thinking all about learning to control variables? *Psychological Science*, *16*, 866–870.

Kuhn, D., Goh, W., Iordanou, K., & Shaenfield, D. (2008). Arguing on the computer: A microgenetic study of developing argument skills in a computer-supported environment. *Child Development*, *79*, 1311–1329.

Kuhn, D., & Park, S. (2005). Epistemological understanding and the development of intellectual values. *International Journal of Educational Research*, *43*, 111–124.

Kuhn, D., & Pease, M. (2008). What needs to develop in the development of inquiry skills? *Cognition and Instruction*, *26*, 1–48.

Kuhn, D., Pease, M., & Wirkala, C. (2008). *Coordinating the effects of multiple variables*: *A skill fundamental to causal and scientific reasoning*. Unpublished manuscript, Teachers College Columbia University.

Kuhn, D., & Pearsall, S. (2000). Developmental origins of scientific thinking. *Journal of Cognition and Development*, *1*(1), 113–129.

Kuhn, D., Shaw, V., & Felton, M. (1997). Effects of dyadic interaction on argumentive reasoning. *Cognition and Instruction*, *15*, 287–315.

Kuhn, D., & Udell, W. (2007). Coordinating own and other perspectives in argument. *Thinking and Reasoning*, *13*(2), 90–104.

Kuhn, D., & Udell, W. (2003). The development of argument skills. *Child Development*, *74*(5), 1245–1260.

Kuhn, D., Weinstock, M., & Flaton, R. (1994). How well do jurors reason? Competence dimensions of individual variation in a juror reasoning task. *Psychological Science*, *5*, 289–296.

Leach, J., Driver, R., Millar, R., & Scott, P. (1997). A study of progression in learning about 'the nature of science': Issues of conceptualization and methodology. *International Journal of Science Education*, *19*, 147–166.

Lehrer, R., Schauble, L., & Petrosino, A. J. (2001). Reconsidering the role of experiment in science education. In K. Crowley, C. Schunn, & T. Okada (Eds.), *Designing for science: Implications from everyday, classroom, and professional settings* (pp. 251–277). Mahwah, NJ: Lawrence Erlbaum Associates.

Mason, L., & Boscolo, P. (2004). Role of epistemological understanding and interest in interpreting a controversy and in topic-specific belief change. *Contemporary Educational Psychology*, *29*(2), 103–128.

Mason, L., & Scirica, F. (2006). Prediction of students' argumentation skills about controversial topics by epistemological understanding. *Learning and Instruction*, *16*, 492–509.

Metz, K. (2004). Children's understanding of scientific inquiry: Their conceptualization of uncertainty in investigations of their own design. *Cognition and Instruction*, *22*, 219–290.

Minda, J., & Ross, B. (2004). Learning categories by making predictions: An investigation of indirect category learning. *Memory and Cognition*, *32*(8), 1355–1368.

Moshman, D. (2005). *Adolescent psychological development: Rationality, morality, and identity* (2nd ed.). Mahwah, NJ: Erlbaum.

Moshman, D. (2008). Epistemic development and the perils of Pluto. In M. Shaughnessy, M. Vennman, & C. K. Kennedy (Eds.), *Metacognition: A recent review of research, theory and perspectives* (pp. 161–174). Hauppauge, NY: Nova.

Muis, K., Bendixen, L., & Haerle, F. (2006). Domain-generality and domain-specificity in personal epistemology research: Philosophical and empirical reflections in the development of a theoretical framework. *Educational Psychology Review*, *18*, 3–54.

National Research Council, National Academy of Science. Committee on science learning. (2007). *Taking science to school: Learning and teaching science in grades K*-8. Washington DC: National Academies Press.

National Research Council. (1996). *The National Science Education Standards*. Washington, DC: National Academies Press.

Neimark, E. (1975). Intellectual development during adolescence. In F. Horowitz (Ed.), *Review of child development research*. Chicago: Chicago University Press.

Olafson, L., & Schraw, G. (2006). Teachers' beliefs and practices within and across domains. *International Journal of Educational Research*, *45*, 71–84.

Sandoval, W. A. (2005). Understanding students' practical epistemologies and their influence on learning through inquiry. *Science Education*, *89*(4), 634–656.

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*(4), 369–392.

Sedlak, A., & Kurtz, S. (1981). A review of children's use of causal inference principles. *Child Development*, *52*, 759–784.

Smith, C. L., Maclin, D., Houghton, C., & Hennessey, M. G. (2000). Sixth-grade students' epistemologies of science: The impact of school science experiences on epistemological development. *Cognition & Instruction*, *18*(3), 349–422.

Solomon, J. (1992). The classroom discussion of science-based social issues presented on television: knowledge, attitude and values. *International Journal of Science Education*, *14*(4), 431–444.

Udell, W. (2007). Enhancing adolescent girls' argument skills in reasoning about personal and non-personal decisions. *Cognitive Development*, *22*, 341–352.

Wilkening, F. (1982). Integrating velocity, time, and distance information: A developmental study. *Cognitive Psychology*, *13*, 231–247.

Windschul, M., & Thompson, J. (2006). Transcending simple forms of school science investigation: The impact of preservice instruction on teachers' understandings of model-based inquiry. *American Educational Research Journal*, *43*, 783–835.

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, *27*(2), 172–223.

Zohar, A. (1995). Reasoning about interactions between variables. *Journal of Research in Science Teaching*, *32*, 1039–1063.