What is Scientific Thinking and How Does it Develop?

Deanna Kuhn

Teachers College Columbia University

In U. Goswami (Ed.),
Handbook of Childhood Cognitive Development (Blackwell)
(2nd ed., 2010)

Author address:

Box 119
Teachers College Columbia University
New York NY 10027
dk100@columbia.edu

What does it mean to think scientifically? We might label a preschooler's curious question, a high school student's answer on a physics exam, and scientists' progress in mapping the human genome as instances of scientific thinking. But if we are to classify such disparate phenomena under a single heading, it is essential that we specify what it is that they have in common. Alternatively, we might define scientific thinking narrowly, as a specific reasoning strategy (such as the control of variables strategy that has dominated research on the development of scientific thinking), or as the thinking characteristic of a narrow population (scientific thinking is what scientists do). But to do so is to seriously limit the interest and significance the phenomenon holds. This chapter begins, then, with an attempt to define scientific thinking in an inclusive way that encompasses not only the preceding examples, but numerous other instances of thinking, including many not typically associated with science.

WHAT IS SCIENTIFIC THINKING?

Scientific thinking as knowledge seeking

Is scientific thinking of any relevance outside of science? In this chapter I answer this question with an emphatic *yes* and portray scientific thinking as a human activity engaged in by most people, rather than a rarefied few. As such, it connects to other forms of thinking studied by cognitive psychologists, such as inference and problem-solving. In particular, I highlight its connection to argumentive thinking (Kuhn, 1991) and characterize its goals and purposes as more closely aligned with argument than with experimentation (Kuhn, 1993; Lehrer, Schauble, & Petrosino, 2001). Scientific thinking is most often social in nature, rather than a phenomenon that occurs only inside people's head. A group of people may rely jointly on scientific thinking in pursuing their goals.

To fully appreciate scientific thinking, it must be situated in a developmental framework, with a goal of identifying both its origins and endpoints. These endpoints are more general than the practices and standards of professional science. The most skilled, highly developed thinking that we identify here is essential to science, but not specific to it.

The definition of scientific thinking adopted in this chapter is *knowledge seeking.* This definition encompasses any instance of purposeful thinking that has the objective of enhancing the seeker's knowledge. One consequence that follows from this definition is that scientific thinking is something people *do*, not something they *have*. The latter we will refer to as *scientific understanding*. When conditions are favorable, the process of scientific thinking may lead to scientific understanding as its product. Indeed, it is the desire for scientific understanding -- for explanation -- that drives the process of scientific thinking.

Scientific thinking and scientific understanding

The distinction between scientific thinking and scientific understanding is an important one, since there has arisen in recent years an extensive literature on children's developing understandings in the domains of physics, biology, and psychology (see Gelman & Kalish, 2006, for review). From their earliest years, children construct implicit theories that enable them to make sense of and organize their experience. These early theories are most often incorrect, as well as incomplete. In a process that has come to be referred to as *conceptual change*, these theories are revised as new evidence is encountered bearing on them. Knowledge acquisition, then, is not the accumulation of isolated bits of knowledge, but, rather, this process of conceptual change.

In contrast to the sizable body of knowledge that has accrued regarding the content of children's evolving theories within specific domains, less is known about the process by means of which theory revision is accomplished. It is this process that is the concern of the present chapter. How is theory revision possible, is there a single process by means of which it occurs, and where does scientific thinking come into this picture? From an applied, educational perspective, as well as a theoretical one, the process of theory revision assumes particular significance. Enhanced understandings of scientific phenomena are certainly a goal of science education. But it is the *ability to advance these understandings* that depends on scientific thinking and is at least as important as an educational goal.

On the grounds that there is no rigid dividing line between informal and formal theories (Kuhn & Pearsall, 2000), we refer here to any cognitive representation of the way things are, no matter how simple, implicit, or fragmentary, as a theory, rather than reserve the latter term for theories meeting various formal criteria that might be invoked (Brewer & Samarapungavan, 1991; Wellman & Gelman, 1998). We can claim, then, that in the early years of life, theories and theory revision are common, as children seek to make sense of a widening array of experience. This early theory revision shares two important attributes with scientific thinking. First, both involve the coordination of theory and evidence -- a characterization of scientific thinking common to most contemporary accounts of it (Bullock, Sodian, & Koerber, in press; Klahr, 2000; Klahr, Fay, & Dunbar, 1993; Klahr & Simon, 1999; Koslowski, 1996; Kuhn, 1989; Kuhn, Amsel, & O'Loughlin, 1988; Lehrer & Schauble, 2006; Zimmerman, 2000, 2007). Second, both can lead to enhanced understanding. There is one important difference, however, between the two. Unlike scientific thinking, early theory revision occurs implicitly and effortlessly, without conscious awareness or intent. Young children think *with* their theories, rather than about them. In the course of so doing they may revise these theories, but they are not aware that they are doing so.

The modern view of scientific thinking as theory-evidence coordination, note, can be contrasted to the pioneering work on scientific thinking by Inhelder and Piaget (1958). Despite the centrality of meaning making in much of Piaget's writing, in this work Inhelder and Piaget conceptualized scientific reasoning strategies largely as logic-driven devices to be applied independent of any context of understanding of the phenomena being investigated. In the modern view, in contrast, theories are integral to knowledge seeking at every phase of the process, a view consonant with modern philosophy of science (Kitcher, 1993).

Knowledge seeking as the intentional coordination of theory and evidence

It is the intention to seek knowledge that transforms implicit theory revision into scientific thinking. Theory revision becomes something one *does*, rather than something that happens to one outside of conscious awareness. To seek knowledge is to acknowledge that one's existing knowledge is incomplete, possibly incorrect -- that there is something new to know. The process of theory-evidence coordination accordingly becomes explicit and intentional. Newly available evidence is examined with regard to its implications for a theory, with awareness that the theory is susceptible to revision.

The coordination of theory and evidence entailed in scientific thinking may yield either of two broad categories of outcomes -- congruence or discrepancy. In the first case, the new evidence that is encountered is entirely compatible with existing theories, and no new understanding results. A new instance is simply absorbed into existing understanding. In the second, more interesting case, some discrepancy between theory and evidence exists and relations between the two need to be constructed. It is possible that the discrepancy will go unrecognized, because the theory, the new evidence, or both have not been adequately represented in a manner that allows relations between them to be constructed. In this case, a likely outcome is that the evidence is ignored or distorted to allow assimilation to existing theoretical understanding. If we decide to include this as a case of scientific thinking at all, it can only be labeled as faulty scientific thinking, since one's existing understandings have been exposed to no test. No knowledge seeking occurs, nor is the possibility of new knowledge even allowed.

Alternatively, in a process we can refer to as "data reading" (Kuhn & Katz, in press), a mental representation of discrepant evidence may be formed -- a representation distinct from the theory -- and its implications for the theory identified. Such cases may vary vastly in the complexity of thinking involved, but they have in common encoding and representation of the evidence distinct from the theory, which is also explicitly represented as an object of cognition, and contemplation of its implications for the theory. It is important to note that the outcome of this process remains open. It is not necessary that the theory be revised in light of the evidence, nor certainly that theory be ignored in favor of evidence, which is a misunderstanding of what is meant by theory-evidence coordination. The criterion is only that the evidence be represented in its own right and its implications for the theory contemplated. Skilled scientific thinking always entails the coordination of theories and evidence, but coordination cannot occur unless the two are encoded and represented as distinguishable entities.

We turn now to tracing the developmental origins of these capacities and then go on to examine them in their more sophisticated forms. Note that none of the processes identified above restricts scientific thinking to traditional scientific content. We are tracing, then, the development of a broad way of thinking and acquiring knowledge about the world, rather than an ability to reason about “scientific” phenomena narrowly conceived.

DEVELOPMENTAL ORIGINS OF SCIENTIFIC THINKING

A now sizeable literature on children’s theory of mind (Flavell, 1999; Wellman, 1988, this volume) affords insight into the origins of scientific thinking because it identifies the earliest forms of a child’s thinking about thinking. Thinking about thinking is not delayed until adolescence, as Inhelder and Piaget’s (1958) account of formal operations might suggest. Rather, it is identifiable in the early forms of awareness preschool children display regarding their own and others’ thinking. By age 3, they show some awareness of their own thinking processes and distinguish thinking about an object from perceiving it (Flavell, Green, & Flavell, 1995). They also begin to use mental-state concepts such as desire and intention in describing their own and others’ behavior.

Differentiating claims from evidence

By at least age 4, however, a child comes to understand that mental representations, as products of the human mind, do not necessarily duplicate external reality. Before children achieve this concept of false belief, they show unwillingness to attribute to another person a belief that they themselves know to be false (Perner, 1991). Children of this young age hold a naïve epistemological theory of beliefs as mental copies of reality. Mental representations are confined to a single reality defined by what the individual takes to be true. The world is thus a simple one of objects and events that we can characterize for ourselves and others. There are no inaccurate renderings of events.

At this level of mental development, the evaluation of falsifiable claims that is central to science cannot occur (Kuhn, Cheney, & Weinstock, 2000). The early theory-of-mind achievement that occurs at least by age 4 --in which assertions come to be understood as generating from human minds and are recognized as potentially discrepant from an external reality to which they can be compared -- is thus a milestone of foundational status in the development of scientific thinking. Assertions become susceptible to evaluation vis-à-vis the reality from which they are now distinguished. The complexity of claims that a 4-year-old is able to evaluate as potentially false is extremely limited. A child of this age is capable of little more, really, than determining whether a claim regarding some physical state of affairs does or does not correspond to a reality the child can directly observe. Yet, this differentiation of assertion and evidence sets the stage for the coordinations between more complex theoretical claims and forms of evidence that are more readily recognizable as scientific thinking.

A related development during this preschool period is the ability to recognize indeterminacy, that is, to recognize situations in which two or more alternative reality states are possible and it is not known which is true, and to discriminate these indeterminate situations from determinate ones. Fay and Klahr (1996), and before them Pieraut-Le Bonniec, 1980), report development in this respect beginning in early childhood (but continuing through adolescence), as do Sodian, Zaitchik, and Carey (1991). Sodian et al. found that by age 7, children were able to choose a determinate over an indeterminate test to find out if a mouse was large or small by placing food in a box overnight. The indeterminate option was a box with a large opening (able to accommodate a large or small mouse) and the determinate option a box with a small opening (big enough for only the small mouse). In choosing the latter, 7 year olds also show some rudimentary skill in investigative strategy, an aspect of inquiry we discuss at length later.

An early competency that is less compelling as an origin of scientific thinking is identification of correspondences between theory and data (Ruffman, Perner, Olson, & Doherty,1993). Connecting the two does not imply their differentiation, as Ruffman et al. claim, based on findings that 5-7-year-olds make inferences from evidence (e.g., dolls who choose red food over green food) to theory (the dolls prefer red food to green), and vice versa. Instead, theory and evidence fit together into a coherent depiction of a state of affairs. In neither the Ruffman et al. nor the Sodian et al. studies, however, is there reason to assume that the child recognizes the differing epistemological status of theory and evidence. (See Kuhn & Pearsall, 2000, for further discussion of these studies.)

Identifying evidence as a source of knowledge

Once assertions are differentiated from evidence that bears on their truth value, it becomes possible for evidence to be appreciated as a source of support for a theory and for relations between evidence and theory to be constructed. To appreciate the epistemological status of evidence, one must be sensitive to the issue of how one knows -- to the sources of one's knowledge. Several researchers have reported increasing sensitivity to the sources of knowledge during the preschool years, for example in distinguishing imagining from perceiving (Woolley & Bruell, 1996), seeing from being told (Gopnik & Graf, 1988), and something just learned from something known for a long time (Taylor, Esbensen, & Bennett, 1994).

In a study of 4-6-year-olds, Pearsall and I (Kuhn & Pearsall, 2000) investigated specifically whether children of this age were sensitive to evidence as a source of knowledge to support the truth of a claim, distinguishable from theory that enhances plausibility of the claim. Participants were shown a sequence of pictures in which, for example, two runners compete in a race. Certain cues suggest a theory as to why one will win; for example, one has fancy running shoes and the other does not. The final picture in the sequence provides evidence of the outcome -- one runner holds a trophy and exhibits a wide grin. When asked to indicate the outcome and to justify this knowledge, 4-year-olds show a fragile distinction between the two kinds of justification -- “How do you know?” and ‘Why is it so?’ -- in other words, the evidence for the claim (the outcome cue in this case) versus an explanation for it (the initial theory-generating cue). Rather, the two merge into a single representation of what happened, and the child tends to choose as evidence of what happened the cue having greater explanatory value as to why it happened. Thus, children often answered the “How do you know [he won]?” question, not with evidence (“He's holding the trophy”) but with a theory of why this state of affairs makes sense (“Because he has fast sneakers”). A follow-up probe, “How can you be sure this is what happened?” elicited a shift from theory-based to evidence-based responses in some cases, but, even with this prompt, 4-year-olds gave evidence-based responses on average to less than a third of the items. At age 6, confusions between theory and evidence still occurred, but children of this age were correct a majority of the time. A group of adults, in contrast, made no errors.

Development of theory-evidence coordination skill as a continuing process

By the end of the preschool years, when children have begun to show an appreciation of the role of evidence in supporting a falsifiable claim, do they confront further challenges in coordinating theories and evidence? The research on older children and adolescents that we turn to now contains substantial evidence of difficulties in this respect, with degree of difficulty influenced by the number and level of complexity of the theoretical alternatives, as well as complexity of the evidence. Thus, as Klahr (2000) similarly concludes, coordination of theory and evidence is not a discrete skill that emerges at a single point in cognitive development. Rather, it must be achieved at successively greater levels of complexity, over an extended period of development. This is especially so if it is to keep pace with increasingly complex models of scientific understanding that are encountered with increasing age. In evaluating such models, requisite skills are invoked: What data support or contradict this piece of the model? How can we test whether particular segments of the model are correct? In such contexts, even able adults' limitations in coordinating theory and evidence become evident. The range and variability in the scientific thinking skills of adults is in fact striking (Kuhn et al., 1988, 1995; Kuhn & Pease, 2006).

PHASES OF SCIENTIFIC THINKING: INQUIRY, ANALYSIS, INFERENCE, AND ARGUMENT

Preschool children, we noted, are able to coordinate a simple event claim and evidence regarding its truth, e.g., they can verify whether the claim that candy is in the pencil box is true or false. More complex claims, however, which begin to assume greater similarity to genuine theories, cause difficulty among school-age children. One such form of rudimentary theory is the imposition of a categorization scheme on a set of instances. Categorization constitutes a theory, in stipulating that some instances are identical to others but different from a third set with respect to some defining attribute(s). Lehrer and Romberg (1996) describe the conceptual obstacles young school-age children encounter in representing theory and data as they engage in such seemingly simple tasks as categorizing classmates' favorite activities and representing their findings. Another series of studies shows only gradually developing skills in children's making appropriate inductive inferences regarding category definition based on a sample of exemplars (Lo, Sides, Rozelle, & Osherson, 2002; Rhodes, Gelman, & Brickman, 2008). We turn now to this coordination process in the more complex forms characteristic of scientific thinking.

As Klahr (2000) notes, very few studies of scientific thinking encompass the entire cycle of scientific investigation, a cycle I characterize here as consisting of four major phases: inquiry, analysis, inference, and argument. A number of researchers have confined their studies to only a portion of the cycle, most often the evaluation of evidence (Amsel & Brock, 1996; Klaczynski, 2000; Koslowski, 1996; Masnick & Morris, 2008), a research design that links the study of scientific reasoning to research on inductive causal inference (Gopnik & Schultz, 2007; Koslowski, this volume). Of studies in which participants acquire their own data, many studies, following the lead of Inhelder and Piaget (1958), have focused their attention on the control of variables strategy (in which a focal variable is manipulated to assess its effect, while all other variables are held constant), as an isolated cognitive strategy divorced from a context of the theoretical meaning of the phenomena being investigated or the goals of the investigations conducted. In the remainder of this chapter, as well as focusing on research that examines strategies in a context of theoretical understanding, we focus on more recent studies that encompass the entire cycle of inquiry, analysis, inference, and argument. These studies offer a picture of how the strategies associated with each phase of scientific investigation are situated within a context of all the others and how they influence one another.

The microgenetic method

We also focus in this chapter on *microgenetic* research (Kuhn & Phelps, 1982; Kuhn, 1995; Siegler & Crowley, 1991; Siegler, 2006), that is, studies in which an individual engages in the same essential task over multiple sessions, allowing the researcher to observe a dynamic process of change in the strategies that are applied to the task. Participants in microgenetic studies are observed in the process of acquiring new knowledge over time. Knowledge acquisition is best conceptualized as a process of theory-evidence coordination, rather than an accumulation of facts (Kuhn, 2000). A major finding from microgenetic research has been that an individual applies a range of alternative strategies in knowledge-acquisition tasks. The selection of strategies chosen for application evolves over time, toward more frequent use of more developmentally advanced strategies. The theory-evidence coordination process of concern to us here, then, while itself dynamic, is likely to undergo modifications in its own nature as it is applied over time. Microgenetic change can thus be observed at two levels: Knowledge (or understanding) changes, but so do the strategies by means of which this knowledge is acquired. Indeed, the latter is a primary thesis of this chapter: the process of theory-evidence coordination shows developmental change. The microgenetic method offers insight into how this change occurs.

The studies by Klahr and his associates (Klahr, 2000; Klahr, Fay, & Dunbar, 1993: Klahr, Triona, & Williams, 2007; Masnick & Klahr, 2003) have followed children and adults asked to conduct scientific investigations, for example of the function of a particular key in controlling the behavior of an electronic robot toy, or, in another version, the behavior of a dancer who performs various movements in a computer simulation. To do this, individuals need to coordinate hypotheses about this function with data they generate, or, in Klahr's (2000) terminology, to coordinate searches of an hypothesis space and an experiment space. Consistent with the findings reported in this chapter, Klahr and his associates find younger children less able to meet this challenge than are older children or adults.

My own microgenetic studies (Kuhn & Phelps, 1982; Kuhn, Schauble, & Garcia-Mila, 1992; Kuhn, Garcia-Mila, Zohar, & Andersen, 1995; Kuhn, Black, Keselman, & Kaplan, 2000; Kuhn & Pease, 2008), as well as studies by Schauble (1990, 1996), Echevarria, (2003), and Penner and Klahr (1996), address what we have regarded as a prototypical form of scientific inquiry -- the situation in which a number of variables have potential causal connections to an outcome and the investigative task is choose instances for examination and on this basis to identify causal and noncausal variables, with the goals of predicting and explaining variations in outcome. Considered here in their simplest, most generic form, these are common objectives of professional scientists engaged in authentic scientific inquiry.

Following our initial assessment of their own theories regarding the presence and direction of causal effects and the mechanisms underlying them, participants in our studies engage in repeated investigative cycles (within a session and across multiple sessions) in which they identify a question, select instances for examination, analyze and make comparisons, and draw conclusions. They also make predictions regarding outcomes and justify these predictions, allowing us to compare implicit causal theories regarding effects of the variables with the earlier voiced explicit theories regarding these effects. We have conducted these studies in a variety of physical and social domains involving, for example, the speed of cars travelling around a computerized racetrack, the speed of toy boats travelling down a makeshift canal, the variables influencing the popularity of children's TV programs, the variables affecting children's school achievement, the variables affecting a teacher-aide's performance in the classroom, and the variables influencing several kinds of natural disasters -- floods, earthquakes, and avalanches.

The illustrations in this chapter are drawn from preadolescent boys' investigations of a single domain (earthquakes), to facilitate comparison and to highlight differences in performance. The earthquake problem is presented as a computer simulation in which five dichotomous features have potential causal effects on the risk of earthquake (portrayed on a "risk meter" with four gradations from lowest to highest risk). Two of the features --type of bedrock (igneous or sedimentary) and speed of S waves (slow or fast) in fact have no effect on outcome, while the other three -- water quality (good or poor), radon gas levels (light or heavy), and snake activity (high or low) -- have simple additive effects. (A version of the problem can be examined at educationforthinking.org.)

The inquiry phase

We begin with an excerpt from the investigations of 10-year-old Brad, who does not see the goal of the task as analysis. In identifying the second instance he wishes to examine, he commented:

> Last time , the [sedimentary] rock was like white. This one [igneous] is sort of like not. It looks like it's going to just blow up any second. This [sedimentary] one looks like it's okay. [So which one do you want to choose to investigate?] Sedimentary [Why?] Because last time I chose sedimentary as well and it seemed to work out pretty good. The igneous looks like it's about to explode any second.

Brad's primary objective, it appears, is to achieve a "good" outcome, rather than to understand the role of the different features in producing different kinds of outcomes. Another approach common among students of Brad's age is to have no other goal than to "experiment," to "try different stuff and see what happens," with no particular intention or organization shaping their investigations. These students, we find, rarely go on to make any informative comparisons in the analysis phase.

The *inquiry* phase of scientific investigation (figure 1) is a crucial one in which the goals of the activity are formulated, the questions to be asked identified, and the remaining phases thereby shaped (see left side of figure 1, which lists the tasks that characterize the inquiry phase). The ovals in the upper center of figure 1 portray the meta-task and metastrategic knowledge associated with this phase.

The most fundamental challenge of the inquiry phase is to recognize that the data base I have the opportunity to access yields information that bears on the theories I hold -- a recognition that eludes many young investigators. The issue is not how heavily such data are weighed relative to preexisting theories, but simply to recognize that these data stand independently of and *speak to* a claim being made. Once the relevance of the data in this respect is recognized, questions can be formulated of a form that is productive in connecting data and theory.

The various strategies that can be observed in response to the tasks of the inquiry phase are portrayed on the right side of figure 1. Here (in contrast to the left side of figure 1, where objectives are compatible), there appears a set of competing strategies which overlap in their usage and are of varying degrees of adequacy (with more adequate strategies appearing further down in the figure). At the lowest level, a strategy for some individuals (or for a particular individual some of the time) may be the simple one of activity, i.e., choosing instances and generating outcomes. Later, after the phenomenon has been observed a number of times, the dominant strategy may become one of producing the most desirable or interesting outcome, as Brad illustrates. The major developmental shift is one from strategies of activity to genuine inquiry, which in its most rudimentary appearance takes the form of "What is making a difference?" or "What will enable me to predict outcomes?" In more advanced forms, inquiry becomes focused on the specific features in terms of which there is variability, and, ultimately, on the effect of a specific feature, "Does X make a difference?"

Analysis and inference phases

The *analysis* phase of scientific inquiry is depicted in figure 2. To engage in productive analysis (left side of figure 2), some segment of the data base must be accessed, attended to, processed, and represented as such, i.e., as evidence to which one's theory can be related, and these data must be operated on (through comparison and pattern detection), in order to reach the third phase, which yields the product of these operations -- *inference*. The strategies that can be observed being applied to this task reflect the struggle to coordinate theories and evidence. As seen on the right side of figure 2, theory predominates in the lower-level strategies, and only with the gradually more advanced strategies does evidence acquire the power to influence theory.

In moving from the analysis to the *inference* phase, we move from procedural strategies to declarative claims. As shown on the left side of figure 3, the inference phase involves inhibiting claims that are not justified, as well as making those that are. The inferential processes that may be applied to this task (right side of figure 3) range in adequacy from no processing of the evidence and no conscious awareness of one's theories (so-called "theories in action") to the skilled coordination of theory and evidence, which entails understanding the implications of evidence as supporting or disconfirming one's theories.

In contrast to Brad, 11-year-old Tom exhibits a more advanced level of investigation in which he sets out to identify effects of individual features. Two characteristics, however, limit the effectiveness of Tom's investigations. First, he believes he can find out the effects of all features at one time and hence does not focus his inquiry on any particular feature. Second, his investigations are theory-dominated to the undesirable extent that the evidence he generates he does not mentally represent in a form that is distinct from his theories.

In response to the first instance he chose to examine, Tom noted the outcome of highest risk level, but, contrary to Brad, he regarded this result favorably and commented:

> I'm feeling really good about this. [Why?] Like I said before on everything. The water quality being poor. Obviously the earthquake would contaminate the water in some way. The S-waves would go fast because logically thinking even big earthquakes happen pretty quickly. Gas, I figured it'd be kind of hard to breathe in an earthquake. Like I said before about the snakes, in the '86 earthquake, dogs started howling before it happened.

Tom, then, appeared quite ready to interpret multiple variables as causally implicated in an outcome, based on a single co-occurrence of one level of the variable and an outcome. We have referred to the mental model of causality underlying this stance as a *co-occurrence* model (Kuhn et al., 2000). Typically, the single-instance inferences deriving from this model are theory-laden in the sense that the empirical observation is seen not so much as a test of, or even evidence bearing on, the theory, as it is simply an “illustration” of the theory in operation. Tom is “feeling good” because his theories have, from his perspective, withstood the test of empirical verification. In reality, of course, they have not been tested at all.

In choosing a second instance to observe, Tom used a strategy that is also common:

> I’m going to do everything the opposite of what I did before. [Why?] Because I want to see if there’s risk or no risk involved.

Tom went on to declare, however, “Actually, I’m going to mix it up kind of,” and after several alternations back and forth he ended up changing bedrock from igneous (chosen for the first instance he examined) to sedimentary, water quality from poor to good, and S-wave rate from fast to slow. Gas levels he left unchanged at heavy, and snake activity unchanged at high.

Tom then observed the outcome fall to medium-high risk and made this interpretation:

> Bedrock makes no difference. No actually it makes a difference. Snake activity makes a difference. And water quality. . . hmm, yeah it brought it down. That was probably half the reason that lowered [the risk] down, with the bedrock and the S-wave. Bedrock made a difference I think because it lowered it down because … well, it [sedimentary rock] seemed less threatening, so I figured it lowered it down. Snake activity, like I said before, animals act up before disasters happen.

Note the predominant role that theory plays in Tom’s implicating bedrock as affecting the outcome. Snake activity, note, Tom implicated as causal even though it remained at an unchanged high level in both of the instances he has observed to this point. The interviewer probed Tom about this feature:

> [Suppose someone disagreed and said that snake activity makes no difference. What would you tell them? Could you tell them that you found out here that it did make a difference?] Well, if you did it low, probably everything’s normal because the snakes wouldn’t be acting up in some odd way.

Thus, Tom’s claim rests on evidence (regarding low snake activity) that he does not in fact have. Tom and the interviewer go on to have a similar exchange regarding gas level, which also has remained at the same (heavy) level across the two instances, and this time, Tom did not even make reference to evidence:

> The gas makes a difference because the heavier it is, the harder it would be to breathe. [Suppose someone disagreed and said that gas level makes no difference. How could you show them that you’re right? Did you see anything here that shows you that it does make a difference?] Well, I think it makes a difference because … let me summarize this up. When it’s heavy there are more things in the air to clog up your lungs.

Finally, based on the two instances available, Tom again implicated water quality, which has covaried with outcome, as causal. He changed his mind about S-wave rates, however, which also covaried with outcome, now claiming this feature to be noncausal:

> Water quality makes a big difference. If it’s good it wouldn’t be contaminated by an earthquake, which also brought [risk] down. And the S-waves, they’re going slowly, always moving. So they don’t really make a difference.

These excerpts from Tom’s investigative activity suggest that when data are not represented in their own right distinct from theory, the potential for scientific analysis remains limited. It should be emphasized again, however, that the scientific thinking tasks described here are *not* ones that ask individuals to cast aside their own beliefs about the world in favor of some arbitary new information. Rather, they assess the ability to access and represent new evidence and to appreciate the relation it bears to different theoretical claims. Skilled scientific thinking always entails the coordination of theories and evidence, and this coordination requires that the two be clearly distinguished. Someone could say, "This is what this evidence implies for these theories, although other sources of support I have for some of these theories lead me to maintain belief in them in the face of your disconfirming evidence." This individual would do perfectly well in our tasks. More troubling are those whose beliefs *are* influenced by the evidence but who remain metacognitively unaware that this has happened and, more broadly, of why they claim what they do.

Mental models of causality and their implications for scientific investigation

Mark, also age 11, does better than Tom in representing data separately from his theories and drawing on these data as a basis for his inferences. In other respects, however, his approach is like Tom's. Mark implicates features as causal based on a single co-occurrence of variable level and outcome. In choosing an instance for observation, he intended, "to try to find out about everything," and in choosing a second instance, he decided to "do the opposite of each one." Mark saw risk level drop (from instance 1 to 2) from medium-high to low risk. In interpreting the second outcome, he implicated four of the five varying features as causal (with the justification that they covaried with outcome) and yet dismissed the fifth (for which evidence was identical) on the basis of his theory that it doesn't matter.

The performance of both Mark and Tom is consistent with the interpretation of their causal analysis and inference as based on the co-occurrence mental model. Both boys falsely include as causal a variable that either co-occurs with outcome in a single instance or covaries with outcome over two instances. Mark also shows an even more interesting inferential error, which (following Inhelder & Piaget, 1958) we have called *false exclusion* (in contrast to the *false inclusion* errors just noted). In choosing a third instance for examination, Mark changed some features and left others the same and observed a low-risk outcome. Following causal inferences for several features, Mark made two noncausal inferences, using false exclusion to justify each. Water quality, he said, made no difference because

> before [instance 1] it was good and had medium-high risk. This time it's good and has low risk. [What does that tell you?] It probably doesn't matter.

The implication is that another feature has produced variation showing that feature's causal power in affecting the outcome, and the feature in question can therefore be discounted. Mark's inference regarding snake activity was identical in form. Both of these features, note, he had earlier implicated as causal, illustrating the vacillation in claims that our microgenetic studies have shown to be common.

Both false exclusion and false inclusion are consistent with a co-occurrence criterion for inferring causality. The co-occurrence of a level of one variable and an outcome is sufficient to explain that outcome. The potential causal influence of a second variable, therefore, need not be treated as additive. Instead, it can be invoked as a different explanation for a later outcome, or it can be discounted because the first feature explains the outcome (false exclusion, if the discounted variable has not been varied). Accordingly, then, the co-occurrence mental model treats causal influences as neither consistent nor additive.

Computing the consistent effects of multiple variables on an outcome rests on a different, more advanced model of causality. Identification of an individual effect ("Does X make a difference?") is only one step in explaining the causal structure of a domain. The broader task is to identify the effect of each of the varying features, and then -- a part of the task that has received little attention -- considering their joint effects on outcome. Doing so is of course the only way to achieve the goal of accurate prediction of outcomes. It requires that a different mental model of causality replace the co-occurrence model, one in which multiple causes operate individually in a consistent fashion, simultaneously and additively producing an outcome. (Interactive effects require a further level of understanding.)

In our research, we have observed an association between the goal of identifying effects of individual features and use of controlled comparison as an analysis strategy (Kuhn et al., 2000). Arguably this is so because both rest on the mature mental model of causality in which multiple individual variables additively influence an outcome. In the absence of this model, one's task goal is unlikely to be identification of the effect of each of the individual variables. Accordingly, neither attribute of the controlled comparison strategy will be compelling. The "comparison" attribute is not compelling, given it entails comparing the outcomes associated with different levels of an individual variable to assess its effect. And the "controlled" attribute is even less compelling, since it is the individual effects of other variables that need to be controlled.

The immature mental model of causality underlying Tom's and Mark's performance, then, limits adoption of either the goals or strategies that make for effective scientific investigation. Unsurprisingly, neither Mark's nor Tom's investigations led to judgments of any greater than chance correctness. Mark, for example, concluded (after examining 4 instances) that all features except water quality are causal. He was thus wrong about 3 of the 5 features. Moreover, when asked how sure he was that he had found out which features were and weren't making a difference, on a 1-10 scale, Mark rated his certainty as "9."

The performance of 12-year-old Robbie can be contrasted to that of Tom and Mark. Robbie's approach initially does not look that different. He chose as the second instance "the opposite of what I did last time." When asked for his inferences, however, he initially implicated S-wave rates, but then said:

> Well, I should … I can do a test to find out actually.

Robbie then said:

> I am going to keep everything the same as last time and just change the igneous to sedimentary to see if it alters the thing.

In response to the interviewer's question, "Why are you keeping the others the same?" Robbie responded:

> If you alter one thing and it's different, that means it has to be the difference. So the type of bedrock does not make a difference. [How do you know?] Because it [the outcome] didn't change. If it had changed, it would mean that it mattered.

Robbie proceeded in an identical manner to assess effects of the remaining features and was able to explain his strategy explicitly:

> I'm doing the same thing as last time. I'll keep everything the same except for gas level, which I am changing to the opposite, light.

After satisfying himself that he had discovered which features did or did not make a difference, Robbie went on to the next phase of the activity, in which he is asked to make predictions about outcomes and then to indicate (as an assessment of implicit causal judgments) which features had influenced the prediction. For each of his predictions, Robbie implicated the same three features. The interviewer asked, ""Would it always be these three for every prediction, or would it be different for some predictions?", to which Robbie replied:

> It would always be these three for all predictions, because [the other two] didn't matter. It was only these three that actually mattered.

With this awareness of what and how he knows, Robbie would be well equipped to defend his claims in discussion with others. This is an important achievement, since a final *argument* phase of scientific thinking consists of debate of the claims that are the product of the earlier phases, in a framework of alternatives and evidence (see figure 4). Again, a range of strategies can be identified, strategies that an individual draws on with varying degrees of probability. Given sustained exercise, these argumentive strategies undergo development (Kuhn, 1991; Kuhn, Shaw, & Felton, 1997; Felton, 2004; Kuhn & Udell, 2003; Kuhn, Goh, Iordanou, & Shaenfield, 2008; Udell, 2007). The products of the argumentive process are revised and strengthened *claims*, strengthened in the sense of being better supported as an outcome of the argumentive process.

The progression from absence of an analysis goal on the part of 10-year-old Brad to 12-year-old Robbie's explicit awareness of the effective analysis strategies he used should not be taken as implying an orderly age-related progression in the development of scientific thinking skills. To the contrary, the norm is wide inter-individual variability. As often as 12-year-olds like Robbie, we see children of similar age and even adults, reported in our earlier research (Kuhn et al., 1995), who over 7 or 8 weekly sessions examine a database that provides no support for a theory and yet maintain to the end not just the correctness of their theory, but, more importantly, that the data they have examined *shows* that it is correct.

Recognizing the centrality of argument to scientific thinking extends scientific thinking beyond traditional science and into the realm of everyday thinking. People typically hold entrenched beliefs, supported by contextually rich representations and sometimes significant affect. This richness may facilitate thinking about such topics, but it may also make it more difficult to think well about them. This was the finding of a comparison of inquiry processes in social science vs. physical science domains (Kuhn et al., 1995). Thus, while it would be hard to contest that "valid experimentation strategies support the development of more accurate and complete knowledge" (Schauble, 1996, p. 118), it is less clear that rich knowledge necessarily enhances the selection of valid experimentation strategies.

Coordinating effects of multiple variables – an overlooked aspect of scientific thinking

Tom and Mark, we noted, were constrained by their satisfaction with a single factor as sufficient in explanatory power to discourage investigation of additional factors acting on the outcome. In further work (Kuhn & Dean, 2004; Kuhn, Iordanou, Pease, & Wirkala, 2008; Kuhn, Pease, & Wirkala, in press), we have in fact found that coordination of the effects of multiple factors on an outcome is a skill in its own right that is by no means implied by the ability to identify the effect of each individual feature. This skill is called upon when individuals are asked to predict outcomes on the basis of multiple features that they have identified as causal and to explain the basis for their predictions. Typical among preadolescents (and even many adults) is inconsistent causal attribution across consecutive predictions, for example implicating variable A as the sole basis for an outcome prediction for one case and variable B as the basis for the prediction in the next case. Moreover, they typically implicate fewer variables (and very often only one) as having affected the prediction than they earlier identified as causal variables.

These behaviors cannot be attributed to failure to maintain mental representations of the multiple effective variables. Even when we provided pictorial memory aides depicting the various effects, these response patterns persisted (Kuhn et al., 2008). Performance does improve with practice over time (Kuhn et al., in press), but the weaknesses are persistent and do not disappear entirely. In particular, we observed a common conceptual error that is crucial in scientific reasoning, a confusion between the levels of a variable and the variable itself. Thus, a frequent response to the question of which variables had affected a prediction judgment was, for example, "I considered the snake activity, because it's high and that increases risk, but I didn't consider any of the others because they were all low so they wouldn't matter." What the student in this case does not recognize is that, unlike the noncausal variables, the other two causal variables of course had to be considered, or she would not have been able to categorize their levels as ones associated with lower risk. They could not be ignored.

Students of this age, appear to have at best a fragile concept of what a variable is, without which it is difficult to reason explicitly or with precision about the effect of one variable on another. In particular, the concept that under consistent conditions a variable operates in a consistent way across occasions is fundamental to science, and yet it is a concept that children appear to only gradually acquire and one therefore that cannot be assumed to be in place. Equally fundamental to science is the assumption that to be adequately explained most events require that a confluence of multiple causes be invoked. In the absence of this assumption, scientific thinking is severely constrained. Variables and multiple causation are the bread and butter of science. Despite their often being taken for granted in the design of science curricula, our studies suggest that neither of these assumptions is easily come by.

THE ROLE OF META-LEVEL PROCESSES IN SCIENTIFIC THINKING

Fully as important as the inter-individual variability portrayed by the preceding examples is the intra-individual variability that microgenetic studies have found to be the norm: Individuals typically have available a range of different strategies of differing levels of advancement and effectiveness (Kuhn, 1995; Siegler, 1996). Development consists of shifts in the frequencies with which different strategies are chosen for application. To explain development, we therefore need to turn to a meta-level of functioning (Kuhn, 2000, 2001; Kuhn & Pease, 2009) -- the level at which strategies are selected and their use monitored. The meta-level involves knowing about knowing. Strong meta-level processes afford Robbie the certainty that he has drawn correct conclusions, while insufficiently developed meta-level processes are implicated in Tom and Mark's false certainty that their inferences are correct.

Figure 4 places the strategies and phases of knowledge seeking in the context of meta-level processes that regulate them. On the left side is the *procedural* meta-level that selects knowledge-seeking strategies to apply, in relation to task goals, and manages and monitors their application. Feedback from this application is directed back to the meta-level. This feedback leads to enhanced awareness of the task goal and the extent to which it is being met by different strategies, as well as enhanced awareness of the strategies themselves (in particular, increased recognition of the power and the limitations associated with each). These enhancements at the meta-level lead to revised strategy selection and hence changes in the distribution of strategies observed at the performance level. In a continuous process, this modified usage in turn feeds back to enhanced understanding at the meta-level, eventually getting the individual to the performance goal of consistent use of the more powerful strategies (Kuhn, 2000).

A notable feature of this model is that it accounts for the common finding that efforts to induce change directly at the performance level have only limited success, reflected in failures to transfer outside a specific context. As the figure 4 model predicts, if nothing has been done to influence the meta-level, new behavior will quickly disappear once the instructional context is withdrawn and individuals resume meta-level management of their own behavior. This limitation applies to many of the studies that have undertaken to improve scientific thinking simply by teaching strategies ("do this"), and, if meta-level understanding is addressed at all, by assessing children's knowledge that this is what they should do. The meta-level understanding that is critical, in contrast, is *why* this is what to do and why other strategies are less effective or wrong. In one of the most meticulously designed training studies, for example, Chen and Klahr (1999) explained to 2nd-4th graders that some (confounded) comparisons were bad comparisons while other (unconfounded) comparisons were good comparisons because just one feature changed. In posttests in new contexts, many children were able to choose a good comparison over a bad one and to justify it as good because only one feature changed. Indicative of their fragile meta-level knowledge, however, was the continued mixture of correct and incorrect strategies shown by a majority of children in conducting their own investigations.

The rights side of figure 4 depicts *declarative* meta-level understanding regarding what it means to know something. Epistemological understanding regarding knowledge and knowing is a crucial underpinning of scientific thinking. What is science and scientific knowledge? Most children bring an *absolutist* understanding to their study of science (Hofer & Pintrich, 1997; Kuhn, et al., 2000; Smith, Maclin, Houghton, & Hennessey, 2000): Scientific knowledge is an accumulating set of certain facts. By adolescence, most have made the radical shift to a *multiplist* epistemology that embraces an awareness of the uncertain, subjective nature of knowing. This awareness initially assumes such proportions, however, that it overpowers any objective standard that could serve as a basis for evaluating conflicting claims. Because claims are subjective opinions freely chosen by their holders and everyone has a right to their opinion, all opinions are equally right. By adulthood, many, though by no means all, people have reintegrated the objective dimension of knowing and espouse the *evaluativist* understanding that some claims are superior to others to the extent they are better supported by argument and evidence. Only at this level is the coordination of theories and evidence that marks authentic science fully understood. If facts can be readily ascertained with certainty, as the absolutist understands, or if any claim is as valid as any other, as the multiplist understands, scientific inquiry has little purpose.

Values are the final component that figures importantly in figure 4. Epistemic understanding informs intellectual values (Kuhn & Park, 2005), in connection with each of the four knowledge seeking phases; value in turn affect disposition to action. Meta-level procedural knowing is necessary if one is to be able to apply knowing strategies effectively, but it is the intellectual values depicted in figure 4 that determine whether one regards knowledge seeking as worthwhile and is therefore disposed to engage in it. Our earlier definition of scientific thinking as knowledge seeking thus accords values a central place in conceptions of scientific thinking.

SCIENTIFIC THINKING AS ARGUMENT

Meta-level understanding is a crucial part of what needs to develop in scientific thinking. Fortunately, like performance, it shows improvement over time, when thinking is exercised, and correspondences are apparent in the improvements that occur at the two levels (Kuhn & Pearsall, 1998). Returning scientific thinking to its real-life social context is one approach to strengthening the meta-level components of scientific thinking. When students find themselves having to justify claims and strategies to one another, normally implicit meta-level cognitive processes become externalized, making them more available. Social scaffolding, then, may assist a less-able collaborator to monitor and manage strategic operations in a way that he or she cannot yet do alone, as in this example from two girls working on the problem of what variables affect the speed that model boats travel down an improvised canal (Kuhn, 2000):

S: We found out about the weight.
N: No, about the boat size, that's all.
S: Oh, the boat size.
N: Just talk about the boat size.

Peer collaboration can be highly variable, however, in its form and effects. The interchange in table 1 (which occurs just after a second instance of evidence has been observed), comes from a single segment of a session with the earthquake problem in which 10-year-old Brad is working together with 11-year-old Tod (Kuhn, unpublished). In discussing the gas level feature, Tod supports and strengthens Brad's theory-based claim by drawing on evidence. With respect to the snake activity and water quality features, the boys disagree and we see Tod vacillate between endorsement and rejection of Brad's incorrect inference strategies. In one case, Tod ends up succumbing to Brad's inferior reasoning. In the other, he does not and the disagreement stands. It is clear, nonetheless, that both boys' scientific thinking has been exercised by the exchange.

The excerpt in table 1 brings individual reasoning strategies into the richer context of social discourse or argument. Increasingly, contributors to both the cognitive development and science education fields have emphasized scientific thinking as a form of discourse (Berland & Reiser, 2009; Bricker & Bell, 2008; Duschl, 2008; Duschl, Schweingruber, & Shouse, 2007; Erduran, Simon, & Osborne, 2004; Garcia-Mila & Andersen, 2007; Iordanou, 2009; Lehrer, Schauble, & Lucas, 2008; Osborne, 2004; Zimmerman, 2007). This is of course the richest and most authentic context in which to examine scientific thinking, as long as the mistake is not made of regarding these discourse forms as exclusive to science. Scientific discourse asks, most importantly, "How do you know?" or "What is the support for your statement?" When children participate in discourse that poses these questions, they acquire the skills and values that lead them to pose the same questions to themselves (Olson & Astington, 1993). Although central to science, this critical development extends far beyond the borders of traditional scientific disciplines.

EDUCATING SCIENTIFIC THINKING AND THINKERS

Science education does not necessarily involve scientific thinking. In the kinds of learning experiences that are commonplace in much of science education, information may be presented or a phenomenon demonstrated, with the questions the new information is intended to answer either left unclear or externally imposed. Students may, in such cases, respond in routinized ways that avoid scientific thinking entirely.

In addition to what they may undertake to teach children about science, science educators hope that the educational activities they design will develop the scientific thinking skills that have been the subject of the present chapter. Much educational practice in preschool and early elementary years rests on the idea of young children's "natural curiosity." Practices that encourage children to ask questions, to observe, and to express their ideas in response to teachers' questions have been accepted as sufficient components to define good "constructivist" teaching practice. It becomes clear, however, by the middle elementary school years, that these practices do not by themselves constitute an adequate instructional model. Palincsar and Magnusson (2001) note ". . . the impossibility that children will come to meaningful understandings of the nature of scientific thinking simply through the process of interacting with materials and phenomena." Video-based teacher training material of a constructivist bent commonly features a teacher asking a bright-eyed, appealing youngster, "What do you think, Tommy?," making a minimal acknowledgment "("Okay, good."), and then turning to the next child with the same query. The richness of inquiry teaching and learning depends on the teacher's *doing something* with that child's response, in a way that leaves the child with a richer, more elaborated conceptual representation than the child had previously. Such conceptual representations encompass far more than specific content, extending, for example, to understandings of what kinds of questions are worth asking and why. To develop these instructional skills, teachers need to understand what the child is bringing to the instructional situation and exactly what kinds of process skills are in the process of developing (Lehrer et al., 2008; Kuhn, 2005; Kuhn & Pease, 2008).

With respect to the process skills of investigation and inference that lie at the heart of authentic scientific thinking, there is a divergence of opinion as to the most productive instructional methods. Klahr and colleagues (Chen & Klahr, 1999; Klahr & Nigam, 2004), have focused their efforts on single-session direct instruction, specifically of the control-of-variables strategy, whereas others have engaged children in the practice of scientific inquiry over longer periods of time (Kuhn & Pease, 2008; Lehrer, Schauble, & Lucas, 2008). (See educationforthinking.org for examples of the Kuhn & Pease software-based curriculum.) When evaluation is extended over time, a study by Strand-Cary & Klahr (2008) and a study by Dean and Kuhn (2007) nonetheless show very similar results. Direct instruction with respect to the control-of-variables strategy confers a temporary benefit. This benefit recedes over time and in transfer assessments, however, unless, as in the Dean and Kuhn (2007) study, it is accompanied by sustained practice with problems requiring the strategy. Moreover, in a direct comparison, Dean and Kuhn (2007) found, a group engaged in practice alone performed as well after several months as the group who in addition had initially received direct instruction.

Most needed now are studies examining the mechanisms by means of which thinking improves as it is practiced. From an educational perspective, the importance of doing so is undeniable, for it has become clear that most of the thinking skills, as well as dispositions, examined in this chapter develop only in environments conducive to them (Bullock et al., in press; Lehrer et al., 2008). As highlighted in this chapter through its emphasis on meta-level understanding, perhaps most important for teachers to convey to children about science is less the *what* or the *how* but the *why* , including ultimately, why inquiry and analysis are worth the effort they entail. These values, as shown in figure 4, are supported by epistemological understanding of what scientific knowing entails (Kuhn & Park, 2005). It is here that the variance emerges with respect to whether learned skills will be used.

If we can clearly identify what the cognitive skills are and how they develop, we are in the best position to learn how to promote understanding of their value. Thus, science educators need to base their efforts on a sound understanding of the entire complex of skills and meta-skills that have the potential to develop during the childhood and adolescent years. Educators who are informed developmentalists stand to bring the strengths of both traditions to the challenge that science education poses.

References

Amsel, E., & Brock, S. (1996). The development of evidence evaluation skills*. Cognitive Development, 11,* 523-550.

Brewer, W., & Samarapungavan, A. (1991). Children's theories vs. scientific theories: Differences in reasoning or differences in knowledge. In R. Hoffman & D. Palermo (Eds.), *Cognition and the symbolic processes* (pp. 209-232). Hillsdale NJ: Erlbaum.

Bullock, M., Sodian, B., & Koerber, S. (in press). Doing experiments and understanding science: Development of scientific reasoning from childhood to adulthood. In W. Schneider & M. Bullock (Eds.), *Human development from earl childhood to early adulthood: Findings from a 20-year longitudinal study*. Mahwah NJ: Erlbaum.

Chen, Z., & Klahr, D. (1999). All other things being equal: Children's acquisition of the control of variables strategy*. Child Development, 70,* 1098-1120.

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

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

Erduran, S., Simon, S., & Osborne, J. (2004). TAPing into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse. *Science Education, 88,* 915-933.

Fay, A., & Klahr, D. (1996). Knowing about guessing and guessing about knowing: Preschoolers' understanding of indeterminacy. *Child Development. 67*, 689-716.

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 skills. *Discourse Processes,32,* 135-153.

Flavell, J. (1999). Cognitive development: Children's knowledge about the mind*. Annual Review of Psychology.* Palo Alto CA: Annual Reviews.

Flavell, J., Green, F., & Flavell, E. (1995). Young children's knowledge about thinking*. Society for Research in Child Development Monographs, 60* (1), (Serial no. 243).

Garcia-Mila, M. & Andersen, C. (2007). Cognitive foundations of learning argumentation. In S. Erduran & M. P. Jimenez-Aleixandre (Eds.), *Argumentation in science education: Perspectives from classroom-based research.* New York: Springer.

Gelman, S., & Kalish, C. (2006). Conceptual development. In D. Kuhn & R. Siegler (Eds.), (W. Damon & R. Lerner, Series eds.), *Handbook of Child Psychology: Vol. 2. Cognition, Perception, and Language.* (6th edition). Hoboken NJ: Wiley.

Gopnik, A., & Graf, P. (1988). Knowing how you know: Young children's ability to identify and remember the sources of their beliefs. *Child Development, 59,* 1366-1371.

Gopnik, A., & Schultz, L. (2007). *Causal learning: Psychology, philosophy, and computation.* New York: Oxford University Press.

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.

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

Kitcher, P. (1993). *The advancement of science.* New York: Oxford University Press.

Klaczynski, P. (2000). Motivated scientific reasoning biases, epistemological beliefs, and theory polarization: A two-process approach to adolescent cognition. *Child Development,* 71, 1347-1366.

Klahr, D. (2000). *Exploring science: The cognition and development of discovery processes.* Cambridge MA: MIT Press.

Klahr, D., Fay, A., & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study*. Cognitive Psychology, 25,* 111-146.

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science, 15*, 661–667.

Klahr, D., & Simon, H. (1999). Studies of scientific discovery: Complementary approaches and convergent findings. *Psychological Bulletin, 125,* 524-543.

Klahr, D., Triona, L., & Williams, C. (2007). Hands on what? The relative effectiveness of physical vs. virtual materials in an engineering design project by middle school children. *Journal of Research in Science Teaching, 44,* 183-203.

Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning.* Cambridge MA: MIT Press.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96,* 674-689.

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

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

Kuhn, D. (1995). Microgenetic study of change: What has it told us? *Psychological Science, 6,* 133-139.

Kuhn, D. (2000). Why development does (and doesn't) occur: Evidence from the domain of inductive reasoning. In R. Siegler & J. McClelland (Eds.), *Mechanisms of cognitive development: Neural and behavioral perspectives.* Mahwah NJ: Erlbaum.

Kuhn, D. (2001). How do people know? *Psychological Science,* 12, 1-8.

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

Kuhn, D., Amsel, E., & O'Loughlin, M. (1988)*. The development of scientific thinking skills.* Orlando FL: Academic Press.

Kuhn, D., Black, J., Keselman, A., & Kaplan, D. (2000). The development of cognitive skills that support inquiry learning. *Cognition and Instruction, 18, 495-523.*

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

Kuhn, D., Garcia-Mila, M., Zohar, A., & Andersen, C. (1995). *Strategies of knowledge acquisition. Society for Research in Child Development Monographs, 60* (4), Serial no. 245.

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., Iordanou, K., Pease, M., & Wirkala, C. (2008). Beyond control of variables: What needs to develop to achieve skilled scientific thinking? *Cognitive Development, 23, 435-451.* [Special issue, The Development of Scientific Thinking, B. Sodian & M. Bullock, eds.]

Kuhn, D., & Katz, J. (in press). Are self-explanations always beneficial? *Journal of Experimental Child Psychology.*

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

Kuhn, D., & Pearsall, S. (1998). Relations between metastrategic knowledge and strategic performance*. Cognitive Development, 13,* 227-247.

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

Kuhn, D., & Pease, M. (2006). Do children and adults learn differently? *Journal of Cognition and Development, 7,* 279-293.

Kuhn, D., & Pease, M. (2008). What needs to develop in the development of inquiry skills? *Cognition and Instruction, 26,* 512-559.

Kuhn, D., & Pease, M. (2009). The dual components of developing strategy use: Production and inhibition. In H.S. Waters & W. Schneider (Eds.),*Metacognition, Strategy Use, and Instruction.* New York: Guilford Press.

Kuhn, D., Pease, M., & Wirkala, C. (in press). Coordinating effects of multiple variables: A skill fundamental to causal and scientific reasoning. *Journal of Experimental Child Psychology.*

Kuhn, D., & Phelps, E. (1982). The development of problem-solving strategies. In H. Reese (Ed.), *Advances in child development and behavior,* Vol. 17. New York: Academic Press

Kuhn, D., Schauble, L., & Garcia-Mila, M. (1992). Cross-domain development of scientific reasoning*. Cognition and Instruction, 9,* 285-32.

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

Lehrer, R., & Romberg, T. (1996). Exploring children's data modeling*. Cognition and Instruction, 14,* 69-108.

Lehrer, R., & Schauble, L. (2006). Scientific thinking and scientific literacy: Supporting development in learning contexts. In K. A. Renninger & I. Sigel (Eds.), *Handbook of child psychology. Vol. 4.* Hoboken, NJ: Wiley.

Lehrer, R., Schauble, L., & Lucas, D. (2008). Supporting development of the epistemology of inquiry. *Cognitive Development, 23,* 512-529. [Special issue, The Development of Scientific Thinking, B. Sodian & M. Bullock, eds.]

Lehrer, R., Schauble, L., & Petrosino, A. (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: Erlbaum.

Lo, Y., Sides, A., Rozelle, J., & Osherson, D. (2002). Evidential diversity and premise probability in young children's inductive judgment. *Cogntive Science, 26,* 181-206.

Masnick, A., & Morris, B. (2008). Investigating the development of data evaluation: The role of data characteristics. *Child Development, 79,* 1032-1048.

Olson, D., & Astington, J. (1993). Thinking about thinking: Learning how to take statement and hold beliefs. *Educational Psychologist, 28*, 7-23.

Palincsar, A., & Magnusson, S. (2001). The interplay of first-hand and second-hand investigations to model and support the development of scientific knowledge and reasoning. In S. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress.* Mahwah NJ: Erlbaum.

Penner, D., & Klahr, D. (1996). The interaction of domain-specific knowledge and domain-general discovery strategies: A study with sinking objects. *Child Development, 67*, 2709-2727.

Perner, J. (1991). *Understanding the representational mind.* Cambridge MA: MIT Press.

Pieraut-Le Bonniec, G. (1980). *The development of modal reasoning.* New York: Academic Press.

Rhodes, M., Brickman, D., & Gelman, S. (2008). Sample diversity and premise typicality in inductive reasoning: Evidence for developmental change. *Cognition, 108,* 543-556.

Ruffman, T., Perner, J., Olson, D., & Doherty, M. (1993). Reflecting on scientific thinking: Children's understanding of the hypothesis-evidence relation. *Child Development, 64,* 1617-1636.

Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology, 49,* 31-57.

Schauble, L. (1996). The development of scientiifc reasoning in knowledge-rich contexts. *Developmental Psychology, 32,* 102-119.

Siegler, R. (1996). *Emerging minds: The process of change in children's thinking.* New York: Oxford University Press.

Siegler, R., & Crowley, K. (1991). The microgenetic method: A direct means for studying cognitive development. *American Psychologist, 46, 606-620.*

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

Sodian, B., Zaitchik, D., & Carey, S. (1991). Young children's differentiation of hypothethical beliefs from evidence. *Child Development, 62,* 753-766.

Strand-Cary, M., & Kllahr, D. (2008). Developing elementary science skills: Instructional effectiveness and path independence. *Cognitive Development, 23, 435-451.* [Special issue, The Development of Scientific Thinking, B. Sodian & M. Bullock, eds.]

Taylor, M., Esbensen, B., & Bennett, R. (1994). Children's understanding of knowledge acquisition: The tendency for children to report they have always known what they have just learned. *Child Development, 65,* 1581-1604.

Woolley, J., & Bruell, M. (1996). Young children's awareness of the origins of their mental representations. *Developmental Psychology, 32,* 335-346.

iZimmerman, C. (2000). The development of scientific reasoning skills*. Developmental Review, 20,* 99-149.

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

Table 1

Excerpt of Discussion of Brad and Tod in the Earthquake Problem

GAS LEVEL
B [Brad]: Gas level makes a difference because if an earthquake is coming the level of oxygen would decrease I think because the earthquake is taking up all the oxygen.
T[Tod]: Can I say something? We also know that by trial and error. Last time we had heavy gas and got low-medium risk. Since we changed the gas to low, we found out gas did make a difference.

SNAKE ACTIVITY
T: Brad, before you move on, how do you know snake activity makes a difference?
B: It makes no difference.
T: Okay, okay.
I[Interviewer]: Well, let’s talk about this a minute. One of you says it makes a difference, the other that it doesn’t. Did you find that out, Brad, by looking at these cases?
B: Yes.
I: How did you know?
B Well, last time I kept snake activity the same and I only changed two things, and I believe those were the ones that made a difference. Now I’m just gonna keep snake activity the same, because it does not make any difference.
I: Tod, do you disagree?
T: No. I was asking him the same question; how did he know?
I: Well, he’s told you how he knows; what do you think of his answer?
T: I think it’s correct, because he just made some changes [itemizes] … and all those changes made it go to low. But if he’d changed snake activity, it may have made a difference and it may have not.
I: But he didn’t change snake activity.
T: Right, he didn’t and that’s why he said that snake activity didn’t make a difference.
I: So, do we know then that snake activity makes no difference?
B: No, not positively, but it’s a good estimate.
T: If we changed it to heavy next time, and keep all of these [other features the same], we may find out if it makes a difference.
I: But I thought before you were both telling me you’d already found out it does not.
B: That’s my good estimate.
T: I agree. But just to make sure, change it to heavy. If risk went back to low, we’d know it would make a difference.
I: And what do we know now about snake activity?
T: We are pretty sure it makes no difference.
I: Why is that?
T: Because he made changes in everything else and he kept snake activity the same and it went to the lowest.

WATER QUALITY
B: Water quality made a difference because last time I kept it good. I think it made a difference because the water would be sinking down if there was an earthquake coming.
I: Do the records show whether it made a difference?
B: Yes. Because last time it was good and this time it was good too. I think it should be good because I got the lowest risk and last time I got medium-low risk. I changed two [others] and it got me down to the lowest risk. So I think the water quality should be good.
I: And how do you know water quality makes a difference?
B: Because last time it was good and this time it was good and [both times] we got the lowest risk.
I: Tod, what do you think?

T: Last time it was good and this time good. I say it wouldn’t really make a difference. They were both the same and how would you see that it made a difference if they were both the same, if they were both good?
I: Brad, what do you think?
B: It makes a difference, because I kept it good and I got a medium-low risk; this time I got a [even] lower risk.
I: So what does that tell you again?
B: Two of my answers were wrong last time. And I think I changed them to the right answers this time.
I: And so does water quality make a difference?
B: It makes a difference.
I: Tod, what do you think?
T: We don’t know if it does.

Figure titles

Figure 1. The inquiry phase
Figure 2. The analysis phase
Figure 3. The inference phase
Figure 4. The role of meta-level operators in scientific thinking (from Kuhn, D., 2001 How do people know? *Psychological Science*).