WILEY THE STEVENS' HANDBOOK OF EXPERIMENTAL PSYCHOLOGY AND COGNITIVE NEUROSCIENCE, FOURTH EDITION

Development of Scientific Thinking

Journal:	*The Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience, Fourth Edition*
Manuscript ID	STEVENS-0072
Wiley - Manuscript type:	Entry
Date Submitted by the Author:	21-Feb-2016
Complete List of Authors:	Zimmerman, Corinne; Illinois State University Department of Psychology Klahr, David
Keywords:	scientific thinking, hypothesis formation, investigation skills, evidence evaluation, science education
Abstract:	For almost a century, psychologists interested in cognitive development have devised empirical investigations to uncover the trajectory of scientific thinking, and they have explored a variety of methods for enriching children's understanding of scientific procedures and concepts. Topics have ranged from the origins of early childhood curiosity, through the development of everyday understanding of scientific phenomena, to the practices necessary to advance our knowledge about the natural world. In this chapter, we review the empirical research that has informed us about the ways in which the child is -- and is not -- like a scientist, and the ways in which scientific thinking needs to be fostered and supported, via educational, social, and cultural scaffolds. We define scientific thinking by drawing on two relatively distinct lines of inquiry, present a taxonomy to categorize these lines of inquiry and the cognitive skills involved, describe illustrative examples of experimental studies of scientific thinking, and then summarize what has been learned about the similarities and differences between children's scientific thinking and mature scientific thinking.

Development of Scientific Thinking

Corinne Zimmerman
Illinois State University

David Klahr
Carnegie Mellon University

Author Note

Corinne Zimmerman, Department of Psychology, Illinois State University; David Klahr, Department of Psychology, Carnegie Mellon University.

In preparing this review, we have drawn heavily upon work produced through collaborations over the years with several colleagues, including Sharon Carver, Zhe Chen, Steve Croker, Anne Fay, Kevin Dunbar, Jamie Jirout, Amy Masnick, Brian Matlen, Bradley Morris, David Penner, Chris Schunn, Martin Schwichow, and Herbert Simon. This chapter could not have been written without their prior contributions, but we do not hold them responsible for any errors, distortions, or egregious omissions in this chapter.

CHAPTER 7: DEVELOPMENT OF SCIENTIFIC THINKING

Why Study Scientific Thinking?

For almost a century, psychologists interested in cognitive development have devised empirical investigations to uncover the origins and trajectory of scientific thinking and have explored a variety of methods for enriching children's understanding of scientific procedures and concepts. The study of scientific thinking is particularly appealing to cognitive and developmental psychologists not only because of science's cultural value, but also because of the inherent importance and challenge of investigating scientific thinking and the paradox of "the child as scientist." In addition, science educators are interested in the topic because of its obvious relevance for improving science instruction.

Cultural Value

Science and technology have had profound effects on human culture. Scientific thinking has enhanced the ability of human beings to understand, predict, and control the natural forces that shape our world. *Scientific literacy* refers to the skills required by citizens in a scientifically advanced society. Students, citizens, and policy makers need to understand how to investigate, evaluate, and comprehend science content (e.g., climate change, evolution, astronomy, disease), processes (e.g., how to test hypotheses effectively), and products (e.g., from evaluating data about the most effective cancer treatments to the possibility of space colonization). In addition to the "factual" aspects of scientific knowledge and scientific procedures, the well-being of a society depends on a wide-spread appreciation of the value of science (e.g., the necessity of evidence-based decisions about policies, practices, and programs). A report from the National Research Council (NRC; 2010) argues that science is *the* discipline that should be used to convey the skills required for the 21st century workforce, such as non-routine problem solving, adaptability, complex communication skills, self-management, and systems thinking.

Inherent Importance and Difficulty

Although scientific thinking has its roots in "everyday thinking," it is much more complex, highly structured, and refined. There has been a rich mythology about the importance, on one hand, and the intractability, on the other, of studying the psychological processes that lead to scientific discovery. With respect to intractability, Einstein once mused, "I am not sure whether there can be a way of really understanding the miracle of thinking" (Wertheimer, 1959 p. 227). If Einstein's bewilderment were correct then this chapter could conclude right here. However, in a more optimistic and constructive reflection, Einstein also said:

> The whole of science is nothing more than a refinement of everyday thinking. It is for this reason that the critical thinking of the physicist cannot possibly be restricted to the examination of concepts of his own specific field. He cannot proceed without considering critically a much more difficult problem, the problem of analyzing the nature of everyday thinking." (Einstein, 1936, p.59)

In the decades since Einstein had this remarkable insight, cognitive scientists have made substantial advances in understanding the "nature of everyday thinking." At the same time, there has been a rich and active line of research on the cognitive and social processes involved in in scientific inquiry and discovery. Thus, the very difficulty of studying this complex, multilayered, and socially impactful topic provides a strong justification for the endeavor.

The Paradox of the "Child as Scientist"

Both children and scientists are described as "naturally curious" with an inherent and enthusiastic approach to finding out about the world. Research shows that infants and young children have some of the precursors and abilities needed to engage in formal scientific thinking. At the same time, older children and adults struggle with scientific thinking tasks; they can be unsystematic, ignore and misinterpret evidence, try to prove what they already believe to be true, and design uninformative experiments (just to name a few of the documented difficulties). Although U- and inverted-U-shaped curves are common in developmental psychology, there is more to this story than early and late competence. By systematically studying age-appropriate tasks across the lifespan, we have learned a lot about the extent to which the child-as-scientist view is supported, and the extent to which, from pre-school to college, people have deep and well-entrenched misconceptions about both the content and processes of science.

Educational Relevance

The study of basic cognitive processes involved in scientific thinking has obvious and important implications for science education across the curriculum, from pre-Kindergarten through college. For example, Piaget's theory of cognitive development was very influential in the design and development of science curricula from the late 1950's through the end of the last century (Blake & Pope, 2008; Elkind, 1972; Kamii, & DeVries, 1993; Klahr, 2012; Metz, 1995, 1997). And even as the field of cognitive development has distanced itself from much of Piaget's theoretical edifice, many current educational practitioners, from K through college level, continue to base much of their instruction upon Piagetian stages of cognitive development. Science education (in the US in particular) has undergone remarkable change within the past several decades (DeBoer, 2000). Recently, in the US, national organizations such as the National Research Council have been focused on evidence-based efforts to improve science education. These efforts have included an explicit acknowledgement of the basic psychological research on scientific thinking (e.g., NRC, 2000, 2007, 2008). The "Taking Science to School K-8" report (NRC, 2007) has emphasized and summarized several important emergent themes in recent research in the learning sciences and cognitive development. For example, with respect to summarizing what research has told us about how science is learned, the "surprisingly sophisticated" nature of children's thinking is acknowledged, as is their ability to engage in a variety of reasoning processes that represent the precursors to mature scientific thinking. The role of adults, teachers, and learning experiences is emphasized as a way to turn children's "rich but naïve understandings of the natural world" (NRC, 2007, p. 3) into proficient skills needed for engaging in investigation, evidence evaluation, explanation, argumentation, and discourse as science students and as scientifically literate adults. Recommendations about best practices for supporting this transition have been proposed (NRC, 2007, 2008), which acknowledge children's existing and developing capabilities, focus on core knowledge areas and skills, and promote the opportunity for the students to engage in the practices of science.

The plan for this chapter is as follows. We will (1) define scientific thinking by drawing on two relatively distinct lines of inquiry; (2) present a taxonomy to categorize these lines of inquiry and the cognitive skills involved; (3) describe illustrative examples of experimental studies of scientific thinking; and finally (4) summarize what has been learned about the similarities and differences between children's scientific thinking and mature scientific thinking.

What is Scientific Thinking?

Science, as a human endeavor, can be approached as an individual, social, and cultural activity. At the individual level, scientific thinking shares many characteristics with other forms of problem solving and reasoning (Klahr, Matlen & Jirout, 2013; Zimmerman & Croker, 2013). It can be further described as a specific type of *intentional information or knowledge seeking* (Kuhn, 2011). Curiosity emerges early and spontaneously in children (Jirout & Klahr, 2012). However, before this innate curiosity can effectively address scientific issues, it must be refined and shaped by instruction through deliberate activities such as exploring, asking questions, testing hypotheses, engaging in inquiry, and evaluating evidence (Jirout & Zimmerman, 2015; Morris, Croker, Masnick, & Zimmerman, 2012). An additional defining feature of mature scientific thinking involves metacognitive and metastrategic knowledge – the ability to reflect on the process of knowledge acquisition and the changes that result from engaging in scientific activities (Kuhn, 2011). Metacognitive skills are evident in children's emerging theory-of-mind skills. Children must learn where beliefs about the world around them come from, that others may have different beliefs, that beliefs can be more or less certain, and in particular – that beliefs may be formed on the basis of inference or from evidence (e.g., Sodian & Wimmer, 1987). This broad and encompassing definition provides a context in which to summarize a very wide range of investigations, from studies of young children making observations in a school classroom (e.g., Chinn & Malhotra, 2002) to descriptions of a research team in a laboratory discussing the results of a set of experiments (e.g., Dunbar, 1995).

Psychological studies of scientific thinking have taken several forms, including historical accounts and case studies of individual scientists or groups of scientists (see Klahr, 1994; Klahr & Simon, 1999 for review) and computational models of the cognitive processes underlying scientific problem solving and discovery (see Shrager & Langley, 1990 for overview). In this chapter, we focus on psychological studies of participants in simulated discovery contexts. Participants from some characteristic population (e.g., school children, college students, scientists) are presented with problem-solving situations that isolate one or more essential aspects of "real world" science. The "thing-to-be-discovered" can range from something as simple and arbitrary as a "rule" that the experimenter has in mind (e.g., Wason, 1960) to something as complex as the physics of an artificial universe (Mynatt, Doherty & Tweney, 1977) or the mechanisms of genetic inhibition (Dunbar, 1993). The advantage of this approach is that it enables the researcher to exert some experimental control over participants' prior knowledge and complete control over the "state of nature." Most important, this approach enables the researcher to observe the dynamic course of scientific discovery processes in great detail.

Conceptual Thinking in Science

The beginning of psychological research on scientific thinking has been widely attributed to Jean Piaget's meeting with Albert Einstein in 1928. Einstein was curious about the developmental origins of fundamental physical concepts. Piaget began a line of research that included children's understanding of scientific concepts and their skills for investigating the world and credits Einstein for inspiration (Piaget 1946/1969). Piaget investigated children's developing thinking processes about time, speed, distance, number, movement, velocity, living things, people, space, mathematics, logic, morality, physical causality, and psychology. The legacy of his steadfast focus on children's early understanding of causes and effects in the natural world cannot be overstated. Piaget's influence is eloquently summarized by John Flavell (1996), and continues to be reiterated (Klahr, 2012).

Following on in this tradition of studying scientific thinking, researchers have examined the concepts that children and adults hold in the various domains of science, such as biology (e.g., Carey, 1985; Hatano & Inagaki, 2013), chemistry (Calik & Ayas, 2005; Garnett, Garnett, & Hackling, 1995), and physics (e.g., Chi, Feltovich, & Glaser, 1981).Numerous examples of specific concepts within these domains have been studied. For example, conceptual development has been studied in astronomy (Vosnaiadou & Brewer, 1994), gravity (Hood, 1998), genetics (Echevarria, 2003), evolution (Emmons & Kelemen, 2015; Samarapungavan & Wiers, 1997), ecology (Zimmerman & Cuddington, 2007), earth and space concepts (Sackes, 2015), life science concepts (Akerson, Wieland, & Fouad, 2015), buoyancy, current, and bubbles (Tenenbaum, Rappolt-Schlichtmann, & Zanger, 2004).

The focus in this line of research has been to identify and describe the mental models or domain-specific theories that children and adults hold about scientific phenomena and the progression of changes that these models undergo with experience or instruction. Here, scientific thinking is studied by asking participants to use their conceptual knowledge about particular scientific phenomena to answer questions and reason about novel scenarios. Two classic studies illustrate this type of scientific thinking. To probe children's understanding of the shape of the earth, Vosniaduou and Brewer (1992) asked 6- to 11-year-olds factual questions such as "What is the shape of the earth?" and questions that would differentiate children's conceptualizations, such as "What is above/below the earth?" and "Can you fall off the edge [of the earth]?" These responses, along with drawings, uncovered a variety of alternative mental models that varied in internal consistency (e.g., disc earth, hollow sphere, rectangular earth). In the domain of genetics, Clough and Driver (1986) asked early adolescents to reason about situations of genetic inheritance involving the offspring of humans or animals with acquired traits (e.g., a gardener with rough skin, a mouse that had lost its tail). Developmental trends in beliefs about inheritance and accompanying explanations were evident from 12 to 16 years of age.

The vast number of possible concepts that can be investigated makes it difficult to adequately summarize the findings from this literature, because of the domain-specific nature of such research. For example, as of 2009, Reinders Duit, a researcher at University of Kiel, compiled a bibliography of research studies on conceptual change in science with more than 8,400 entries.[1] Consider, for example, a single chapter on children's understanding of physical science concepts. Hadzigeorgiou (2015) reviews studies about children's ideas about matter, heat,

[1] For last update see http://archiv.ipn.uni-kiel.de/stcse/.

temperature, evaporation, condensation, the water cycle, forces, motion, floating, sinking, electricity, and light. Each of these topics can be further unpacked to constituent subcomponents (e.g., electricity concepts include current, voltage, charge, electrons, resistance, and circuits, to name a few).

In the most current science standards in the USA, the NRC (2012) has identified three dimensions that represent “a broad set of expectations for students in science” (p. 1), including *Disciplinary Core Ideas*, *Crosscutting Concepts*, and *Scientific and Engineering Practices*. The core ideas are restricted to the traditional sciences (i.e., physical, life, earth/space). The crosscutting concepts, in contrast, are broadly applicable to several domains of science, and include ideas such as causality, patterns, time, feedback, analogy, and equilibrium. There is a rich literature on what individuals understand about the concept of *causality* across the lifespan, from infancy studies (e.g., Baillargeon, 2004) through childhood (e.g., Gopnik, Sobel, Schulz & Glymour, 2001; Piaget, 1974) and adulthood (for a review see Koslowski & Masnick, 2010). Although less work has been done on the other domain-general crosscutting concepts, there are a few exceptions. Swanson (2015) examined students’ understanding of patterns (as a process or behavior) that can underlie physical, social, or psychological phenomena (e.g., threshold, equilibration, and oscillation) in a middle-school science course called “the Patterns Class.”

Procedural Thinking in Science

In addition to studies addressing what children know about science, there is a substantial literature on how children acquire that knowledge. Here too, Piaget led the way in his decision to study not only what children know about the world at various stages of their development, but also, and perhaps more importantly, the methods and processes that they use to acquire, integrate, and refine this knowledge. This second line of research examines the development of scientific thinking by focusing on the scientific practices of observing, asking questions, conducting experiments, evaluating evidence, constructing models, and generating explanations. In this line of investigation, "scientific thinking is something people *do*, not something they *have*" (Kuhn, 2011, p. 498).

The distinction between the products and processes of scientific thinking is reflected in the often distinct research programs that developmental psychologists have undertaken. It is possible to develop a line of research about children's *knowledge* about various content (e.g., astronomy, biology) without needing to be concerned about their investigation skills (i.e., how they came to know it). For example, Vosiadou and Brewer's (1992) study of children's conceptions about the shape of the earth is a canonical study of scientific knowledge, but it does not address children's *knowledge-producing* activities. Studies of such knowledge-producing investigation skills have utilized two main strategies. One is to reduce the reliance on (or interference by) conceptual knowledge by creating tasks that reduce the role of prior knowledge (e.g., Wason, 1960; Siegler & Liebert, 1975). The other is to examine these knowledge types in a more integrated way, motivated by the fact that concepts and procedures are intertwined in authentic scientific thinking.

A Taxonomy for Categorizing Experimental Studies of Scientific Thinking

Our taxonomy utilizes the two principle features that have been used previously to characterize scientific thinking: domain-specific concepts and domain-general procedures (Klahr, 1994; Klahr, Zimmerman, & Jirout, 2011; Zimmerman, 2000; 2007). In addition, Table 1 extends this taxonomy to incorporate some of the *Scientific and Engineering Practices* that have been identified by the NRC's (2012) national science education standards. The cognitive processes identified by Klahr and Dunbar (1988) are included in the first column. The scientific practices that map on to these cognitive processes are included in the second column. In the third and fourth columns, we distinguish between domain-general and domain-specific knowledge because some studies focus on specific content whereas others use simple or abstract contexts that do not require much (or any!) domain-specific knowledge. For example, a typical domain-specific study might investigate what children understand about the domain of chemistry by asking them to reason about processes such as dissolution and chemical change (Calik & Ayas, 2005). Domain-general studies, in contrast, focus on broadly applicable reasoning processes that can be investigated in arbitrary and abstract forms, such as Bruner's classic concept learning tasks (Bruner, Goodnow, & Austin, 1956) or Wason's famous 2-4-6 task (Wason, 1960).

Some preliminaries are in order. Some studies might fit neatly into a single cell, but for the majority of this literature (described in more detail in the next section), even studies that focus on a particular cognitive process or scientific practice have to traverse more than one cell. For example, in order to examine experimentation skills, participants must set up (or, select from a pre-defined set of choices) an experiment to address a particular hypothesis. Likewise, the evaluation of evidence must be done in some context, and requires a consideration of either the experiment that produced it, or the hypothesis that it is meant to support or refute.

Empirical Investigations of the Development of Scientific Thinking

Klahr and Dunbar's (1988) *Scientific Discovery as Dual Search* (SDDS) model will serve as the general framework for organizing illustrative empirical findings to be discussed (see Table 1). The SDDS framework captures the complexity and the cyclical nature of the process of scientific discovery and includes both inquiry skills and conceptual change (see Klahr, 2000, for a detailed discussion). The top-level categories of the model include the three major cognitive components of scientific discovery: Searching for hypotheses, searching for experiments (or investigations more generally), and evidence evaluation. The studies to be reviewed involve one or more of these three processes. SDDS is an extension of a classic model of problem solving from the field of cognitive science (Newell & Simon, 1972, Simon & Lea, 1974) and explains how people carry out problem solving in varied science contexts, from simulated inquiry to professional scientific practice. The fundamental aspects of SDDS are (a) the concept of two distinct, but closely related, "problem spaces": a space of hypotheses and a space of experiments; and (b) *coordinated* search in these two problem spaces.

Individuals begin inquiry tasks with some existing or intuitive ideas, or perhaps no ideas at all about how particular variables influence an outcome. Given some set of possible variables (i.e., independent variables) and asked to determine their effect on an outcome (i.e., the dependent variable), participants negotiate the process by coordinating search in the set of possible hypotheses and the set of possible experiments. Experiments are conducted to determine the truth status of the current hypothesis, or to decide among a set of competing hypotheses. Experiments may also be conducted to generate enough data to be able to propose a hypothesis (as might be the case when one has little or no prior knowledge). Evidence is then evaluated so that inferences can be made whether a hypothesis is correct or incorrect (or, in some cases, that the evidence is inconclusive). Depending on the complexity of the task, the number of variables, and the amount of time on task, these processes may be repeated several times as an individual negotiates search in the hypothesis and experiment spaces and makes inferences based on the evaluation of self-generated evidence. Factors such as task domain, amount of prior knowledge, and the perceived goal of the task influence how these cognitive processes are deployed.

Searching the Hypothesis Space

Of the three cognitive processes of SDDS, search in the hypothesis space has the most in common with conceptual thinking in science, as it typically involves a search of relevant domain-specific knowledge as represented in the hypothesis space. When one is engaged in inquiry or investigation activities, hypothesis-space search is instantiated in the service of the *scientific practices* of asking questions and developing or using models (NRC, 2012).

Asking questions and curiosity. Asking questions is one of the foundational process skills of scientific practice (NRC, 2012). Older students often believe that the goal of science is to demonstrate what is already known (Kuhn, 2005), or to see if something "works" or to invent things (Carey, Evans, Honda, Jay, & Unger, 1989). However, asking questions for which the answer is *not* yet known is a crucial element of inquiry that students must learn (Kuhn & Dean, 2005). They must learn not only how to ask "good" questions, but also that *question-asking* is a defining feature of science. An essential precursor to asking good questions is *curiosity* (Jirout & Klahr, 2012; Klahr, Zimmerman, & Jirout, 2011). Curiosity's fundamental importance in science education is indicated by its nearly universal inclusion across a variety of highly influential science curricula, educational standards, and assessment goals (AAAS, 1993; National Education Goals Panel, 1995; NAEYC, 2011; NRC, 2000). Curiosity is the desire or motivation to explore and ask questions. Specifically, we define curiosity as the *preferred level of uncertainty* – or the amount of uncertainty that will lead to deliberate question asking or exploratory behavior (Jirout & Klahr, 2012, 2016; Jirout & Zimmerman, 2015).

Simple problem-solving tasks that require question asking have been used for investigating children's ability to recognize specific instances of uncertainty and to evaluate information. *Referential tasks* assess children's general ability to ask categorical questions ("is it an animal?", "does it bark?") that will help them to identify a target from a group of possibilities (e.g., one picture from an array of pictures). When children are given the opportunity to ask a question to figure out which of the objects is hidden in a box before guessing, they are correct on about five of the six trials; if they are told to guess what's in the box without being allowed to ask a question, their accuracy is at chance (Chouinard, 2007). Thus, children can determine which questions to ask to address uncertainty, but they can also use information yielded by the answer to their questions to resolve it.

Research on conceptual thinking in science has used various methodologies (drawings, standardized interviews, reasoning scenarios) to uncover children's understanding of various phenomena. At this intersection with process skills, there are examples of children's question asking in particular domains. For example, Greif, Kemler Nelson, Keil, and Gutierrez (2006) investigated young children's ability to ask domain-specific questions on a structured task. Children were instructed to ask questions about unfamiliar objects and animals, which they were able to do – averaging 26 questions asked across 12 pictures. Many questions were quite general, such as "what is it?" Other questions, however, showed that children recognized and understood that different questions should be asked of the different categories (i.e., objects and animals). Children tended to ask more function questions about the objects (i.e., what can I do with this?), whereas the unfamiliar animals prompted questions about category membership, food choices, and locations. Samarapungavan, Mantzicopoulos, and Patrick (2008) explored whether kindergarteners could generate meaningful questions about biological topics such as structure, function, and growth with respect to living things in general, and about monarch butterflies in particular. Given a supportive inquiry unit, the majority of students were found to be proficient at asking questions that were meaningful and biologically relevant. Similarly, Metz (2004) examined the specific questions that children in the second and fourth grades asked about cricket behavior a supportive inquiry-based curriculum. With appropriate scaffolding, all pairs of students were able to formulate appropriate and relevant research questions. Moreover, these children were able to generate fairly sophisticated and domain-specific categories of researchable questions (e.g., comparison of cricket behavior under different conditions).

Developing and using models. Developing and using models constitute a "signature practice of the sciences" (Quellmalz, Timms, Silberglitt, & Buckley, 2012, p. 366) that is becoming increasingly emphasized in science education and science assessment (Clement, 2000; Lehrer & Schauble, 2000, 2012; NRC, 2012). The use of models to support theory building, argumentation, and explanation is common in science and engineering (Nersessian, 2008) and has been documented in socio-historical analyses of practicing scientists (e.g., Thagard, 2000; Tweney, 2002). Scientists use physical models (e.g., Watson and Crick's model of the structure of DNA), drawings, and schematic representations (e.g., Faraday's sketches of electromagnetic toruses or Darwin's tree of life). In both science and science education, the ability to develop and use models is becoming increasing easier (and thus by extension, easier to be more sophisticated) due to the scaffolding provided by computers and computer simulations. Simulation models can be used to learn about and investigate phenomena that are "too large, too small, too fast, too small, or too dangerous to study in classrooms" (Quellmalz et al., 2012, p. 367).

Schunn and Klahr's (1995, 1996) investigation of participants' explorations of a complex computer microworld (see Klahr, 2000, Chapter 7, for additional details) led them to propose that as tasks become more rich, complex, and authentic, the SDDS model of scientific thinking should include the search of two additional problem spaces. As well as hypotheses and experiments, scientific discovery requires the search of abstract data representations (*aka* "models") and a space of experimental paradigms at a level of abstraction above the specific instantiation of an experiment. The data representation (or model) space is used to select attributes of the data under consideration. This may involve attention to regularities of the data set, or may result from analogy to existing knowledge (e.g., the superficial similarity between an atom and a solar system). Experimental paradigms function as general-purpose templates that can be deployed in multiple specific, but isomorphic, contexts. One example of a simple experimental paradigm that is located in the paradigm space is the concept of ensuring that a variable hypothesized to be critical is manipulated. The SDDS model, and the extension involving a total of four search spaces, represents a domain-general model of the scientific thinking processes involved in scientific discovery.

In science education, there are numerous domain-general and domain-specific examples of the instantiation of such model-based practices. A concept like *variability* is one that can be applied in several different domains (Lehrer & Schuable, 2004). Domain-specific examples are varied and include learning about decomposition in a first-grade classroom (Ero-Tolliver, Lucas, & Schauble, 2013), evolution in elementary school (Keleman, Emmons, Seston, & Ganea, 2014; Lehrer & Schauble, 2012), ecosystems in sixth-grade (Lehrer, Schauble, & Lucas, 2008), and biomechanics of the human elbow in college (Penner, Lehrer, & Schauble, 1998).

Searching the Experiment Space

Given a hypothesis (or a set of competing hypotheses), one usually needs to design an experiment[2], and the design of that experiment can be construed as a problem to be solved, via search in a problem space (Newell & Simon, 1972). A "solution" to problem solving search in the experiment space is an experiment that assesses the truth or falsity of the current hypothesis.

[2] Of course, experimentation is just one of several types of legitimate scientific inquiry processes (see Lehrer, Schauble, & Petrosino, 2001), but here we focus on the substantial body of literature on the development of experimentation skills.

Experimentation's privileged status in science education is indicated by its inclusion in science standards (NGSS Lead States, 2013; NRC, 2012). In the next sections, we first describe research on the developmental precursors of experimentation skills, followed by studies in which participants are engaged in the full cycle of experimentation.

Early experimentation skills. Science education for young children tends to focus on investigation skills such as observing, describing, comparing, and exploring (NAEYC, 2012; NSTA, 2007). Even though, as noted earlier, few contemporary researchers in cognitive development accept the assertions of Piaget's stage theory (e.g., Inhelder & Piaget 1958), it has often been used to justify waiting until adolescence before attempting to teach science process skills (French & Woodring, 2013; Metz, 1995). However, an accumulation of evidence about human learning (e.g., NRC, 2000a) has resulted in a more nuanced story about the development of experimentation and investigation skills, and the extent to which well-designed instruction can accelerate that development (NRC, 2007). The identification of causal factors in the world *via* experimentation involves the coordination of several component processes: identifying and manipulating variables, and observing and measuring outcomes. Not until the later school years, after extended instruction, scaffolding, and practice can children successfully coordinate all of these steps (e.g., Kuhn, Black, Keselman, & Kaplan, 2000). Several studies have examined the precursors of the later ensemble of experimentation skills, however, and we turn to them next.

In her classic study, Tschirgi (1980) presented children and adults with everyday problem-solving situations that involved a positive or negative outcome such as "John baked a cake and it turned out terrible" or "Susan made a paper airplane and it turned out great." The character would propose a hypothesis about a variable that may have caused the outcome (e.g., sweetener type in the cake story). Three options were presented to help the character test the hypothesis. In the vary-one-thing-at-a-time (VOTAT) choice, the proposed variable was changed, but the others were kept the same (i.e., corresponding to a controlled experiment). In the hold-one-thing-at-a-time (HOTAT) choice, the hypothesized variable was kept the same, but the other variables were changed. The change-all (CA) choice consisted of changing all of the variables. All participants were more likely to select the HOTAT strategy when the outcome was *positive*. That is, the presumed causal variable was held constant (consistent with a confounded experiment) to maintain the positive outcome. For a *negative* outcome, the logically correct VOTAT strategy (consistent with a controlled experiment) was chosen more frequently than HOTAT or CA, suggesting that participants were searching for the one variable to change in order to eliminate the negative outcome. Although second- and fourth-graders were more likely to select the CA strategy for the negative outcomes (likely as a way to eliminate all possible offending variables), all participants were influenced by the desire to reproduce good effects and eliminate bad effects by choosing a strategy based on pragmatic outcomes (rather than logical grounds).

Croker and Buchanan (2011) used a task similar to Tschirgi's, but included contexts for which 3.5- to 11-year-olds held strong prior beliefs (e.g., the effect of cola vs. milk on dental health). For all age groups, there was an interaction of prior belief and outcome type. The logically correct VOTAT strategy was more likely to be selected under two conditions: (a) when the outcome was positive (i.e., healthy teeth) and consistent with prior belief, or (b) when the outcome was negative (i.e., unhealthy teeth) and inconsistent with prior belief. Even the youngest children were influenced by the context and the plausibility of the domain-specific content of the situations that they were reasoning about.

Sodian, Zaitchik, and Carey (1991) investigated the extent to which children in the early school years understand that one can use an experiment to test a hypothesis. Their classic *mouse house task* presents children with a situation that pits "finding out" against "producing an effect." Sodian et al. found that children as young as 6 can distinguish between a conclusive and inconclusive experimental test of a simple hypothesis when provided with the two mutually exclusive and exhaustive hypotheses or experiments. Chen and Klahr (2008) used tasks isomorphic to those used by Sodian et al. in a training study aimed at determining the extent to which kindergartners through second graders could distinguish conclusive from inconclusive experimental tests. In addition, they used different types of feedback (implicit, verbal, physical demonstration) to show that children could transfer strategies for selecting/generating an experimental test of a hypothesis over delays up to 24 months. Piekny and Maehler (2013) used the mouse house task with preschoolers (4- and 5-year-olds) and school children (7-, 9-, and 11-year-olds). It was not until age 9 that children scored significantly above chance, and not until age 7 that children showed a recognition of, and justification for, conclusive or inconclusive tests of a hypothesis.

Klahr, Fay, and Dunbar (1993) were specifically interested in how participants tested hypotheses that are plausible or implausible. They provided third- and sixth-grade children and adults with hypotheses to test that were incorrect, but either plausible or implausible. When a hypothesis was plausible, all participants set up experiments to *demonstrate the correctness* of the hypothesis. When given an implausible hypothesis to test, adults and some sixth-graders proposed a plausible *rival hypothesis*, and set up an experiment that would discriminate between the two. The third graders also proposed a plausible rival hypothesis, but got sidetracked in the attempt to demonstrate that the rival plausible hypothesis was correct. Klahr et al. (1993)

identified two useful heuristics that participants used: design experiments that produce informative and interpretable results, and attend to one feature at a time. The third- and sixth-grade children were much less likely than adult participants to restrict the search of possible experiments to those that were informative.

Bullock and Ziegler (1999) collected longitudinal data on participants, starting when they were age 8 and following them through to age 12. They examined the process skills required for experimentation, using separate assessments to tease apart an *understanding* of experimentation from the ability to *produce* controlled experiments. When the children were 8-year-olds, they were able to *recognize* a controlled experimental test. The ability to *produce* a controlled experiment at levels comparable to adults did not occur until the children were in the sixth grade. This study provides additional support for the idea that young children are able to understand the "logic" of experiments long before they are able to produce them.

When task demands are reduced – such as simple story problems or when one can select (rather than produce) an experimental test – even young children show competence with rudimentary science process skills. Children, like adults, are sensitive to the context and the content of what is being reasoned about.

Planning and carrying out investigations. For older children and adults, much of the research on the development of investigation skills involves presenting participants with a multivariable causal system, such as physical apparatus or a computer simulation. The participants' goal is to investigate the system so as to identify the causal and non-causal variables in the system; they propose hypotheses, make predictions, plan and conduct experiments, collect and evaluate evidence, make inferences, and draw conclusions in the form of either new or updated knowledge. For example, Schauble's (1996) participants conducted experiments in hydrodynamics, where the goal was to determine which variables have an effect on boat speed. Participants could vary the depth of the canal, and the size, shape, and weight of the boat

A foundational science process skill is the control-of-variables strategy (CVS), which is a domain-general skill (Chen & Klahr, 1999). The fundamental goal of an experiment is to unambiguously identify causal factors and their effects, and the essential procedure for doing this is to contrast conditions that differ only with the respect to the variable whose causal status is under investigation. *Procedurally*, CVS includes the ability to create experiments in which conditions differ with respect to only a single contrasting variable, as well as the ability to recognize confounded and unconfounded experiments. *Conceptually,* CVS involves the ability to make appropriate inferences from the results of unconfounded experiments (e.g., that only inferences about the causal status of the variable being tested are warranted), as well as an awareness of "the inherent indeterminacy of confounded experiments" (Chen & Klahr, 1999, p. 1098). The conceptual aspects of CVS are relevant for argumentation and reasoning about causality in science and everyday life, as CVS includes an understanding of the invalidity of evidence from confounded experiments (or observations) and the importance of comparing controlled conditions (Kuhn, 2005). Thus, CVS is relevant to broader educational and societal goals, such as inquiry, reasoning skills, and critical thinking.

Mastery of CVS is required for successful inquiry learning as it enables students to conduct their own informative investigations. However, without instruction, students and even adults have poor inquiry skills (e.g., Kuhn, 2007; for review see Zimmerman & Croker, 2013). Siler and Klahr (2012) identified the various "misconceptions" that students have about controlling variables. Typical mistakes include (a) designing experiments that vary the wrong (or "non-target") variable, (b) varying more than one variable, or (c) not varying anything between the contrasted experimental conditions (i.e., overextending the “fairness” idea so both conditions are identical).

A recent meta-analysis of CVS interventions (Schwichow, Croker, Zimmerman, Höffler, & Härtig, 2016) summarized the results of 72 studies. Possible moderators of the overall effect size included design features (e.g., quasi-experimental vs. experimental studies), instructional features (e.g., use of demonstrations), training features (e.g., use of hands-on experiences), and assessment features (e.g., test format). Of the various instructional features coded for, only two were found to be effective: (a) interventions that induced a *cognitive conflict* and (b) teacher *demonstrations* of good experimental design. In this context, a teacher draws attention to a particular (confounded) comparison and asks what conclusions can be drawn about the effect of a particular variable. Cognitive conflict is induced in students by drawing attention to a current experimental procedure or interpretation of data; the teacher attempts to get the student to notice that the comparison is confounded or that the conclusion is invalid or indeterminate (Adey & Shayer, 1990). Interestingly, the cognitive conflict technique is often presented *via* a demonstration by the teacher and so additional research is necessary to disentangle the unique effects of these two instructional techniques (Schwichow et al., 2016). Other instructional techniques that are often presumed to be important, such as the need for “hands on” engagement with experimental materials did not have an impact on student learning of CVS. In a follow-up to the meta-analysis, Schwichow, Zimmerman, Croker, and Härtig (2016) determined that it is important for there to be a match between the way students learn CVS and the test format used to assess the extent to which they have learned it.

Evaluating Evidence

The goal of most experiments[3] is to produce evidence that bears on a hypothesis, and once that evidence is generated it must be interpreted. The final cognitive process and scientific practices we will discuss are those that enable people to *evaluate* and *explain* how evidence relates to the hypothesis that inspired it. Evidence evaluation is the part of the cycle of inquiry aimed at determining whether the result of an experiment (or set of experiments) is sufficient to reject or accept a hypothesis under consideration (or whether the evidence is inconclusive), and to construct possible explanations for how the hypothesis and evidence are related.

Interpreting evidence is always done in the context of prior belief. However, to minimize the effects of (widely varying) prior knowledge, early investigations of the evidence evaluation process used tasks that deemphasized it. Wason's (1960) famous 2-4-6 task was conceived as a "knowledge lean" task to simulate the process of evaluating evidence to test and revise hypotheses. In brief, the participant is given an exemplar (2-4-6) of a general rule for numerical triads and asked to hypothesize the rule that governs the sequence of digits. The cycle includes hypothesis formation, generation of a new triad to test the hypothesis, evidence (feedback that the participant-generated triad is, or is not, an exemplar of the rule), and hypothesis revision; this cycle continues until enough evidence has accumulated to discover the rule. Traditionally, the experimenter chooses a very *general* rule, such as "3 integers in ascending order" but most participants begin the task by (mistakenly) hypothesizing a much *narrower* rule (e.g., "even

[3] We say "most" because, as noted in the SDDS model, there are often cases in which one might perform a "theory free" experiment in the absence of any clearly articulated hypothesis, just to get a "feel" for the nature of the phenomenon. For example, Klahr and Dunbar (1988) report that nearly one-third of the experiments proposed by participants were not accompanied by any clearly articulated hypothesis. Additionally, generating results from hypothesis-free experiments to collect information inductively is one of the characteristics that differentiate *experimenters* from *theorists* (see also Dunbar & Klahr, 1989; Zimmerman, Raghavan, & Sartoris, 2003).

numbers increasing by 2"). Research using this task became a cottage industry[4] in cognitive psychology, with many interpretations and reinterpretations of performance (e.g., Klayman & Ha, 1987). Interestingly, there are few developmental studies using the 2-4-6 task:

> [C]hildren do not know enough about numerical relationships to make the mistakes, so typical in the task with adults, and hence they often hit on the rule immediately. They have not yet learnt to erect their own obstacles against finding it. (Wason & Johnson-Laird, 1972, p. 217)

As will be illustrated, much of the literature on evidence evaluation demonstrates that both children and adults have a tendency to "erect their own obstacles" to knowledge acquisition.

Evaluating patterns of evidence. One method of examining the developmental precursors of skilled evidence evaluation with children involves presenting them with pictorial representations of potential causes and effects. These are often simple representations like those between types of food and health (e.g., Kuhn, Amsel, & O'Loughlin, 1988) or plant treatment (e.g., sun, water) and plant health (Amsel & Brock, 1996). The pictures may represent perfect covariation between cause and effect, partial covariation, or no covariation. This cognitive skill is facilitated by the meta-cognitive ability to make a distinction between a hypothesis and the evidence to support a hypothesis (Kuhn, 2005, 2011).

In their classic study, Ruffman, Perner, Olson and Doherty (1993) presented 4- to 7-year-old children with simple story problems involving one potential cause (e.g., type of food: red or green) and an outcome (tooth loss). A "faked evidence task" was used to determine whether children could form different hypotheses based on varying patterns of evidence. For example, children would be shown that green food perfectly covaries with tooth loss: this situation represents the "real evidence." Next, the evidence was tampered with; anyone who was unaware

[4] This paper has been cited over 1,000 times!

of the original pattern would be led to believe that red food causes tooth loss (i.e., the "faked evidence"). Children were asked to interpret which hypothesis the faked evidence supported. The key advantage of this type of task is that it is diagnostic with respect to whether a child can make a distinction between a hypothesis and a pattern of evidence to support a hypothesis. This task requires children to understand that their own hypothesis would be different from that of a story character who only saw the faked evidence. When considering the responses to *both* the initial hypothesis-evidence task and the faked-evidence task, only the 5- to 7-year-olds performed above chance level. Partial covariation evidence was used to determine if 5- to 7-year-olds could form hypotheses based on *patterns* of evidence. When considering both hypothesis-evidence and faked-evidence questions, only the performance of the 6- and 7-year olds was above chance level performance. Most children understood that veridical vs. faked evidence would lead to different beliefs and that a newly formed hypothesis can be used to generalize to future cases.

Ruffman et al. showed that some of the very basic prerequisite evidence evaluation skills required for scientific thinking are present as early as 6 years of age. In follow-up research, Koerber, Sodian, Thoermer, and Nett (2005) examined the performance of 4- to 6-year-olds on a variety of evidence evaluation tasks to examine whether *existing causal beliefs* influence evidence evaluation in the preschool years. In situations where there are no strong prior beliefs and the outcomes are equally plausible, preschoolers correctly interpreted perfect and partial covariation evidence. Preschoolers had difficulty, however, with evidence that contradicts prior plausible beliefs; this finding is consistent with the performance of both older children and adults on scientific thinking tasks (Zimmerman & Croker, 2013). Although young children demonstrate some of the precursors to more advanced evidence-evaluation skills, they too are susceptible to the influences of prior beliefs and considerations of the plausibility of what is being evaluated.

Beginning with the foundational work of Kuhn et al. (1988), we know that the process of revising and acquiring knowledge on the basis of evidence is highly influenced by the prior knowledge that a participant brings to the task. Evaluating evidence is guided by an assessment of the *plausibility* of a hypothesized cause; we make judgments about the world in ways that "make sense" or are consistent with what we already know about how things work. Plausibility is a known constraint in belief formation and revision (Holland, Holyoak, Nisbett, & Thagard, 1986) and is a domain-general heuristic that is used to guide the choice of which hypotheses to test and which experiments to run (Klahr et al., 1993). Because the strength of existing beliefs and assessments of plausibility are considered when evaluating evidence, children and adults often choose to maintain their prior beliefs rather than changing them to be in line with newly acquired evidence (e.g., Chinn & Brewer, 1998; Chinn & Malhotra, 2002). A common finding is that it is generally more difficult to integrate evidence that disconfirms a prior causal belief (which involves restructuring one's belief system) than it is to integrate evidence that disconfirms a prior non-causal belief (which involves incorporating a newly discovered causal relation). For example, children and adults have robust physics misconceptions about weight, mass, and density, and these misconceptions influence the evaluation of evidence in tasks that involve the motion (e.g., falling, sinking, rolling) of objects. In the case of sinking objects, it is difficult to give up the belief that weight matters, but it is easy to add the belief that shape (sphere vs. cube) speeds up or slows down an object based on first-hand evidence (Penner & Klahr, 1996b). Other research shows pervasive difficulties with revising knowledge on the basis of evidence, even when that evidence is generated and observed directly (rather than being provided by researchers; e.g., Chinn & Malhotra, 2002; Renken & Nunez, 2010).

Analyzing and interpreting data. The NRC (2012) science standards include the scientific practice of analyzing and interpreting *data*. They note that “scientific investigations produce data that must be analyzed in order to derive meaning […] data do not speak for themselves” (p. 51). An inescapable aspect of empirical research is that all measurements in the physical world include some degree of error, and children must learn how to deal with it. Masnick, Klahr, and Morris (2007) describe the challenge for the young scientist:

> A young child eagerly awaits the day when she will pass the 100 cm minimum height requirement for riding on the "thriller" roller coaster at her local amusement park. She regularly measures her height on the large-scale ruler tacked to her closet door. As summer approaches, she asks her parents to measure her every week. A few weeks ago she measured 98 cm, last week 99.5 cm, but today only 99.0 cm. Disappointed and confused, when she gets to school she asks the school nurse to measure her, and is delighted to discover that her height is 100.1 cm. Success at last! But as she anticipates the upcoming annual class excursion to the amusement park, she begins to wonder: what is her real height? And more importantly, what will the measurement at the entrance to the roller coaster reveal? Why are all the measurements different, rather than the same? Because she is a really thoughtful child, she begins to speculate about whether the differences are in the thing being measured (i.e., maybe her height really doesn’t increase monotonically from day to day) or the way it was measured (different people may use different techniques and measurement instruments when determining her height). (p. 3)

Although the processes associated with understanding and interpreting error and data variability draw heavily on mathematical reasoning, and therefore are beyond the scope of this chapter, there are a few studies that capture the intersection of analyzing quantitative data, and identifying sources of error. Masnick and Morris (2008) examined how the characteristics of measurement data, such as sample size and variability within the data set (e.g., magnitude of differences, presence of outliers) influenced conclusions drawn by third- and sixth-graders and adults. Participants were shown data sets with plausible cover stories (e.g., testing new sports equipment), and asked to indicate what conclusions could be drawn and their reasons. Third- and sixth-graders had rudimentary skills in detecting trends, overlapping data points, and the magnitude of differences. Sixth graders had developing ideas about the importance of variability and the presence of outliers for drawing conclusions from data. At all ages, participants were more confident of conclusions based on larger samples of observations.

Masnick, Klahr, and Knowles (2016) explored how adults and children (aged 9-11) responded to (a) variability in the data collected from a series of simple experiments, and (b) the extent to which the data were consistent with their prior hypotheses. Participants conducted experiments in which they generated, recorded, and interpreted data to identify factors that affect the period of a pendulum. In Study 1, several children and most adults used observed evidence to revise their initial understanding, but participants were more likely to change incorrect non-causal beliefs to causal beliefs than the reverse. In Study 2, participants were oriented toward either an "engineering" goal (to produce an effect) or a "science" goal (to discover the causal structure of the domain), and presented with variable data about potentially causal factors. Science goals produced more belief revision than engineering goals. Numerical data, when presented in context and with appropriate structure, can help children and adults re-examine their beliefs, and initiate and support the process of conceptual change and robust scientific thinking.

Constructing explanations. The NRC's (2012) Framework for Science Education emphasizes the importance of scientific theories and explanations: "The goal for students is to construct logically coherent explanations of phenomena that incorporate their current understanding of science, or a model that represents it, and are consistent with the available evidence" (p. 52). Scientific explanations are typically constructed after investigations that produce evidence that is to be evaluated and, ultimately, explained.

Much has been written in the scientific thinking literature about the ability to differentiate between evidence (i.e., data, observation, patterns) and the explanation or theory[5] for that finding. In particular, Kuhn's (1989, 2002, 2005) research has emphasized that mature scientific thinking requires the cognitive and metacognitive skills to differentiate between *evidence* and the theory or *explanation* for that evidence. Kuhn argues that effective coordination of evidence and theory depends on three metacognitive abilities: (a) The ability to encode and represent evidence and theory separately, so that relations between them can be recognized; (b) the ability to treat theories or explanations as independent objects of thought (i.e., rather than a representation of "the way things are"); and (c) the ability to recognize that theories can be false and explanations flawed, and that, having recognized that possibility, to assess the evidence in order to determine whether the theory is true or false. These metacognitive abilities are necessary precursors to sophisticated scientific thinking, and represent of the ways in which children, adults, and professional scientists differ.

As noted previously, children are inclined to notice and respond to causal events in the environment; even infants and young children have been shown to have rudimentary understanding of cause and effect (Bullock & Gelman, 1979; Piaget 1929). Keil's (2006; Keil & Wilson, 2000) work on the nature of explanation in general indicates that children and adults alike have a propensity to generate explanations. We often privilege causal explanations, which are arguably quite important in scientific thinking. Koslowski's (1996, 2012, 2013) research shows that people are good at noticing evidence for the covariation between events in the world,

---

[5] We will use the terms "theory" and "explanation" as largely interchangeable in this discussion and consistent with Kuhn and Pearsall's (2000) approach-neutral usage of *theory* to refer to causal or explanatory claims that could be falsified by empirical evidence.

but there is a tendency to only make causal inferences when the link can be explained with a causal mechanism. Participants consider or generate plausible causal mechanisms to explain the relationship between potential causes and their effects. Similarly, if a plausible causal mechanism exists to explain why a cause and effect *should* be linked, it is difficult to let go of that belief. Therefore, we see across many types of scientific thinking tasks, both children and adults have a strong tendency to maintain beliefs rather than change them based on evidence (e.g., Chinn & Brewer, 1998; Chinn & Malhotra, 2002) because the strength of existing beliefs, assessments of plausibility, casual mechanisms, and alternative causal mechanisms are all potentially salient and brought to bear when reasoning (Koslowski, Marasia, Chelenza, & Dublin, 2008).

Summary: Revisiting the Child as Scientist

In this review, we have illustrated the two main approaches to studying the development of scientific thinking. One line of research has focused on the *content* of science: what children and adults think about various science concepts in the traditionally defined disciplines of science. The second has focused on the *processes* or procedures of science: how children and adults ask questions, solve problems, conduct investigations, and evaluate evidence to revise their explanations about how the world works.

The natural and social worlds comprise the laboratory of both the scientist and the child. Scientific thinking is considered "a hallmark intellectual achievement of the human species" (Feist, 2006, p. ix). As developmental psychologists, however, we are interested in the factors that influence the origins and growth of scientific thinking, from the child in a science classroom through to the practicing scientist. In scientific thinking, two developmental endpoints get emphasized – the child and the scientist. Children have been likened to scientists; scientists are said to have the curiosity of young children. But as developmental research from the past century has shown, there is a lot that goes on in between. As is the case with other academic skills such as reading and mathematical thinking, scientific thinking is highly culturally and educationally mediated. Researchers have come to acknowledge that the cognitive processes and the set of scientific practices that must be coordinated in mature scientific thinking require practice and are developed within a social context and with the aid of cultural tools (Lemke, 2001; Zimmerman & Croker, 2014). Importantly, the ability to reflect metacognitively on the process of knowledge acquisition and change is a hallmark of fully developed scientific thinking (Kuhn, 2005).

Science educators now recognize that students should be exposed to learning experiences that reflect how real science is conducted and communicated and educational reform has been aimed at how best to engage students in developmentally appropriate authentic inquiry and argumentation (NRC, 2007, 2008, 2011, 2012). Unlike other basic cognitive skills (e.g., attention, perception, memory), scientific thinking does not “routinely develop,” (Kuhn & Franklin, 2006, p. 974); that is, scientific thinking does not emerge independently of culture and cultural tools. Metacognitive abilities are necessary precursors to sophisticated scientific thinking, but also represent one of the ways in which children, adults, and professional scientists differ. In order for children to go beyond demonstrating the correctness of their existing beliefs (e.g., Dunbar & Klahr, 1989) it is necessary for meta-level competencies to be developed and practiced. With metacognitive control over the processes involved, children can change what they believe based on evidence and, in doing so, are aware not only that they are changing a belief, but also know *why* they are changing a belief. Thus, sophisticated scientific thinking involves the cognitive processes involved in asking questions, forming and refining hypotheses, conducting investigations, developing models, designing experiments, evaluating evidence, and constructing explanations, as well as a meta-level awareness of when, how, and why one should engage in these practices.

References

Adey, P., & Shayer, M. (1990). Accelerating the development of formal thinking in middle and high school students. *Journal of Research in Science Teaching*, *27*, 267–285.

Akerson, V.L., Weiland, I.S., Fouad, K. (2015). Children's ideas about life science. In K. C. Trundle & M. Sackes (Eds.). *Research in early childhood science education* (pp. 99-123). Dordrecht: Springer.

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

Baillargeon, R. (2004). Infants' reasoning about hidden objects: Evidence for event-general and event-specific expectations. *Developmental Science*, *7*, 391-424.

Blake, B., & Pope, T. (2008). Developmental psychology: Incorporating Piaget's and Vygotsky's theories in classrooms. *Journal of Cross-Disciplinary Perspectives in Education*, *1*, 59–67.

Bruner, J.S., Goodnow, J.J., & Austin, G.A. (1956). *A study of thinking*. London: Chapman & Hall.

Bullock, M., & Gelman, R. (1979). Preschool children's assumptions about cause and effect: Temporal ordering. *Child Development*, 89-96.

Bullock, M., & Ziegler, A. (1999). Scientific reasoning: Developmental and individual differences. In F. E. Weinert & W. Schneider (Eds.), *Individual development from 3 to 12: Findings from the Munich Longitudinal Study* (pp. 38 – 54). Cambridge: Cambridge University Press.

Calik, M., & Ayas, A. (2005). A comparison of level of understanding of eighth-grade students and science student teachers related to selected chemistry concepts. *Journal of Research in Science Teaching*, *42*, 638–667.

Carey, S. (1985). *Conceptual change in childhood.* Cambridge, MA: MIT Press.

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

Carey, S., & Spelke, E. (1994). Domain-specific knowledge and conceptual change. *Mapping the mind: Domain specificity in cognition and culture*, 169-200.

Chen, Z., & Klahr, D. (2008). Remote transfer of scientific-reasoning and problem-solving strategies in children. *Advances in child development and behavior*, *36*, 420.

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

Chi, M.T.H., Feltovich,P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121-152.

Chinn, C. A., & Brewer, W. F. (1998). An empirical test of a taxonomy of responses to anomalous data in science. *Journal of Research in Science Teaching, 35,* 623-654.

Chinn, C. A., & Malhotra, B. A. (2002). Children's responses to anomalous scientific data: How is conceptual change impeded? *Journal of Educational Psychology*, *94*, 327-343.

Chouinard, M. M. (2007). Children's questions: A mechanism for cognitive development. *Monographs of the society for Research in Child Development. 72(1)*. 1-129.

Clement, J. (2000). Model based learning as a key research area for science education. *International Journal of Science Education*, *22*(9), 1041-1053.

Clough, E. E., & Driver, R. (1986). A study of consistency in the use of students' conceptual frameworks across different task contexts. *Science Education*, *70*(4), 473-496.

DeBoer, G. E. (2000). Scientific literacy: Another look at its historical and contemporary meanings and its relationship to science education reform. *Journal of Research in Science Teaching*, *37*, 582-601.

Dunbar, K. (1993). Concept discovery in a scientific domain. *Cognitive Science, 17*, 397-434.

Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 365-395). Cambridge, MA: MIT Press.

Dunbar, K., & Klahr, D. (1989). Developmental differences in scientific discovery processes. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 109-143). Hillsdale, NJ: Lawrence Erlbaum Associates.

Einstein, A. (1936). Physics & Reality. Reprinted in A. Einstein (1950) Out of my later years. New York: Philosophical Library

Elkind, D. (1972). Piaget and science education. *Science and Children*, *10*, 9-12.

Emmons, N. A., & Kelemen, D. (2015). Young children's acceptance of within-species variation: Implications for essentialism and teaching evolution. *Journal of Experimental Child Psychology, 139,* 148 – 160.

Ero-Tolliver, I., Lucas, D., & Schauble, L. (2013). Young children's thinking about decomposition: Early modeling entrees to complex ideas in science. *Research in Science Education*, *43*, 2137-2152.

Feist, G. J. (2006). *The psychology of science and the origins of the scientific mind*. New Haven, CT: Yale University Press.

Flavell, J. (1996) Piaget's Legacy. *Psychological Science*, 7, 201 – 203.

Garnett, P. J., Garnett, P. J., & Hackling, M. W. (1995). Students' alternative conceptions in chemistry: A review of research and implications for teaching and learning. *Studies in Science Education*, *25*, 69-95.

Gelman, S. A., & Kremer, K. E. (1991). Understanding natural cause: Children's explanations of how objects and their properties originate. *Child Development*, *62*, 396-414.

Greca, I. M., & Moreira, M. A. (2000). Mental models, conceptual models, and modelling. *International Journal of Science Education*, *22*, 1-11.

Greif, M. L., Kemler Nelson, D. G., Keil, F. C., Gutierrez, F. (2006). What do children want to know about animals and artifacts? Domain-specific requests for information. *Psychological Science, 17*, 455.

Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, *37*, 620-629.

Hadzigeorgiou, Y. (2015). Young children's ideas about physical science concepts. In K. C. Trundle & M. Sackes (Eds.), *Research in early childhood science education* (pp. 67-97). Dordrecht: Springer.

Hatano, G., & Inagaki, K. (2013). *Young children's thinking about biological world*. CITY, ST: Psychology Press.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). *Induction*. Cambridge, MA: MIT Press.

Hood, B. M. (1998). Gravity does rule for falling events. *Developmental Science*, *1*, 59-63.

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

Jirout, J., & Klahr, D. (2012). Children's scientific curiosity: In search of an operational definition of an elusive concept. *Developmental Review*, *32*, 125-160.

Jirout, J., & Klahr, D. (under review) Questions – and some answers -- about young children's questions.

Jirout, J., & Zimmerman, C. (2015). Development of science process skills in the early childhood years. In K. C. Trundle & M. Sackes (Eds.), *Research in early childhood science education* (pp. 143-165). Dordrecht: Springer.

Kamii, C., & DeVries, R. (1993). Physical knowledge in preschool education: Implications of Piaget's theory. New York, NY: Teachers College Press

Keil, F. C. (2006). Explanation and understanding. *Annual review of psychology*, *57*, 227-254.

Keil, F.C. and Wilson, R.A. (Eds.) (2000). *Explanation and cognition*. Cambridge, MA: MIT Press.

Kelemen, D., Emmons, N., Seston, R. & Ganea, P. (2014). Young children can be taught basic natural selection using a picture storybook intervention. *Psychological Science, 25,* 893-902.

Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis-testing. *Psychological Review, 94*, 211-228.

Klahr, D. (1994). Searching for the cognition in cognitive models of science. *Psycoloquy*, 5(94), scientific-cognition.12.klahr.

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

Klahr, D. (2012). Beyond Piaget: a perspective from studies of children's problem solving abilities. In A. Slater & P. Quinn (Eds.), *Refreshing Developmental Psychology: Beyond the Classic Studies* (pp. 56-70). London: Sage Publications.

Klahr, D., & Carver, S. (1995). Scientific thinking about scientific thinking. In *Monographs of the Society for Research in Child Development*, 60(4), 137-151.

Klahr, D., & Dunbar, K. (1988). Dual search space during scientific reasoning. *Cognitive Science*, *12*, 1-48.

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

Klahr, D., Matlen, B. & Jirout, J. (2013). Children as scientific thinkers. In G. Feist & M. Gorman (Eds.), *Handbook of the Psychology of Science* (pp. 223-248). New York, NY: Springer.

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

Klahr, D., Zimmerman, C., & Jirout, J. (2011). Educational interventions to advance children's scientific thinking. *Science, 333*, 971-975.

Koerber, S., Sodian, B. , Thoermer, C., & Nett, U. (2005). Scientific reasoning in young children: Preschoolers' ability to evaluate covariation evidence. *Swiss Journal of Psychology*, *64*, 141-152.

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

Koslowski, B. (2012) How explanation makes information evidentially relevant. In R. Proctor & J. Capaldi (Eds.), *The psychology of science: Implicit and explicit processes* (pp. 112-136). New York, NY: Oxford University Press.

Koslowski, B. (2013). Scientific reasoning: Explanation, confirmation bias, and scientific practice. In G. Feist and M. Gorman (Eds.), *Handbook of the psychology of science* (pp. 151-192). New York, NY: Springer.

Koslowski, B., Marasia, J., Chelenza, M., & Dublin, R. (2008). Information becomes evidence when an explanation can incorporate it into a causal framework. *Cognitive Development*, *23*, 472-487.

Koslowski, B., & Masnick, A. M. (2010). Causal reasoning. In U. Goswami (Ed.), *Handbook of child cognitive development* (2nd ed.) (pp. 377-393). Oxford: Blackwell Publishers.

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

Kuhn, D. (2007). Reasoning about multiple variables: Control of variables is not the only challenge. *Science Education*, *91*, 710-726.

Kuhn, D. (2011). What is scientific thinking and how does it develop? In U. Goswami (Ed.), *Handbook of childhood cognitive development* (2nd ed., pp. 497-523). Oxford, UK: Wiley-Blackwell.

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 to support inquiry learning. *Cognition and Instruction*, *18*, 495-523.

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

Kuhn, D., & Ho, V. (1980). Self-directed activity and cognitive development. *Journal of Applied Developmental Psychology, 1*, 119-130.

Kuhn, D., & Franklin, S. (2006). The second decade: What develops (and how)? In W. Damon & R. Lerner (Series Eds.), D. Kuhn & R. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (6th ed.). Hoboken, NJ: Wiley.

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

Lehrer, R., & Schauble, L. (2000). Developing model-based reasoning in mathematics and science. *Journal of Applied Developmental Psychology*, *21*, 39-48.

Lehrer, R., & Schauble, L. (2004). Modeling natural variation through distribution. *American Educational Research Journal*, 41, 635-679

Lehrer, R., & Schauble, L. (2012). Seeding evolutionary thinking by engaging children in modeling its foundations. *Science Education*, *96*, 701-724.

Lehrer, R., Schauble, L., & Lucas, D. (2008). Supporting development of the epistemology of inquiry. *Cognitive Development*, *23*, 512-529.

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.

Lemke, J. L. (2001). Articulating Communities: Sociocultural perspectives on science education. *Journal of Research in Science Teaching*, *38*, 296-316.

Masnick, A. M., Klahr, D., & Morris, B. J. (2007). Separating signal from noise: Children's understanding of error and variability in experimental outcomes. In M. Lovett & P. Shaw (Eds.), *Thinking With Data* (pp. 3-26). Mawah, NJ: Erlbaum.

Masnick, A., M., Klahr, D, & Knowles, E. R. (in press). Data-driven belief revision in children and adults. *Journal of Cognition and Development*.

Masnick, A. M., & Morris, B. J. (2008). Investigating the development of data evaluation: The role of data characteristics. *Child Development*, 79(4), 1032–1048.

Metz, K. E. (1995). Reassessment of developmental constraints on children's science instruction. *Review of Educational Research*, *65*(2), 93-127.

Metz, K. E. (1997). On the complex relation between cognitive developmental research and children's science curricula. *Review of Educational research*, *67*(1), 151-163.

Morris, B. J., Croker, S., Masnick, A. M., & Zimmerman, C. (2012). The emergence of scientific reasoning. In H. Kloos, B. J. Morris, and J. L. Amaral (Eds.), *Current topics in children's learning and cognition* (pp. 61-82). Rijeka, Croatia: InTech.

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology*, *29*, 85-95.

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1978). Consequences of confirmation and disconfirmation in a simulated research environment. *The Quarterly Journal of Experimental Psychology*, *30*, 395-406.

NAEYC, 2012. All criteria document. Retrieved from: http://www.naeyc.org/files/academy/file/AllCriteriaDocument.pdf

National Education Goals Panel (NEGP). (1993). *Promises to keep: Creating high standards for American students.* Washington, DC: National Education Goals Panel.

National Education Goals Panel (NEGP). (1995). *The 1995 national education goals report: Building a nation of learners*. Washington, D.C. National Education Goals Panel.

National Research Council. (2000). How people learn: Brain, mind, experience, and school. Washington, DC: National Academies Press.

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

National Research Council (2008). Ready Set Science: Putting research to work in K-8 science classrooms Washington, DC: National Academies Press.

National Research Council (2010). Exploring the Intersection of Science Education and 21st Century Skills: A Workshop Summary. Washington, DC: National Academies Press.

National Research Council. (2012). A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas. Washington, DC: National Academies Press.

National Science Teachers Association. (2007). *NSTA position statement: The integral role of laboratory investigations in science instruction*. Retrieved from: http://www.nsta.org/about/positions/laboratory.aspx. Accessed 18 March 2013.

Nersessian, N. J. (2008). Creating scientific concepts. Cambridge MA: MIT Press.

Newell, A., & Simon, H. A. (1972). *Human Problem Solving*. NY: Prentice Hall.

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

Penner, D. E., & Klahr, D. (1996b). When to trust the data: Further investigations of system error in a scientific reasoning task. *Memory & Cognition, 24*, 655-668.

Penner, D. E., Lehrer, R., & Schauble, L. (1998). From physical models to biomechanics: A design-based modeling approach. *Journal of the Learning Sciences*, *7*(3-4), 429-449.

Piaget, J. (1929). *The child's conception of the world*. London: Kegan Paul

Piaget, J. (1970). *Genetic epistemology*. New York, NY: Columbia University Press.

Piaget, J. (1969). *The child's conception of time* (originally published 1946). New York, NY: Ballatine.

Piekny, J., & Maehler, C. (2013). Scientific reasoning in early and middle childhood: The development of domain-general evidence evaluation, experimentation, and hypothesis generation skills. *British Journal of Developmental Psychology*, *31*, 153-179.

Quellmalz, E. S., Timms, M. J., Silberglitt, M. D., & Buckley, B. C. (2012). Science assessments for all: Integrating science simulations into balanced state science assessment systems. *Journal of Research in Science Teaching*, *49*(3), 363-393.

Raghavan, K., & Glaser, R. (1995). Model–based analysis and reasoning in science: the MARS curriculum. *Science Education*, *79*(1), 37-61.

Renken, M. D., & Nunez, N. (2010). Evidence for improved conclusion accuracy after reading about rather than conducting a belief-inconsistent simple physics experiment. *Applied Cognitive Psychology*, *24*, 792–811.

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

Sackes, M. (2015). Young children's ideas about earth and space science concepts. In K. C. Trundle & M. Sackes (Eds.), *Research in early childhood science education* (pp. 35-65). Dordrecht: Springer.

Samarapungavan, A., & Wiers, R. W. (1997). Children's thoughts on the origin of species: A study of explanatory coherence. *Cognitive Science, 21*, 147-177.

Samarapungavan, A., Mantzicopoulos, P., & Patrick, H. (2008). Learning Science Through Inquiry in Kindergarten. *Science Education*, *92*, 868 - 908.

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

Schunn, C. D. & Klahr, D. (1993). Self vs. other-generated hypotheses in scientific discovery. *Proceedings of Fifteenth Annual Conference of the Cognitive Science Society* (pp. 900-905). Hillsdale, NJ: Erlbaum.

Schunn, C. D., & Klahr, D. (1995). A 4-space model of scientific discovery. In J. D. Moore & J. F. Lehman (Eds.) *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society* (pp. 106-111). Mahwah, NJ: Erlbaum.

Schunn, C. D. & Klahr, D. (1996) Integrated yet different: Logical, empirical, and implementational arguments for a 4-space model of inductive problem solving. In G. Cottrell (Ed.), *Proceedings of the Eighteenth Annual Meeting of the Cognitive Science Society* (pp. 25-26). Mahwah, NJ: Erlbaum.

Schwichow, M., Zimmerman, C., Croker, S., & Härtig, H. (2016). What students learn from hands-on activities. Manuscript being revised for *Journal of Research in Science Teaching*.

Schwichow, M., Croker, S., Zimmerman, C., Höffler, T., & Härtig, H. (2016). The control-of-variables strategy: A meta-analysis. *Developmental Review*.

Shaklee, H., & Paszek, D. (1985). Covariation judgment: Systematic rule use in middle childhood. *Child Development*, *56*, 1229-1240.

Shrager, J., & Langley, P. (1990). Computational models of scientific discovery and theory formation. San Mateo, CA: Morgan Kaufman.

Siegler, R. S., & Liebert, R. M. (1975). Acquisition of formal scientific reasoning by 10- and 13-year-olds: Designing a factorial experiment. *Developmental Psychology*, *11*, 401-402.

Siler, S. & Klahr, D. (2012). Detecting, classifying and remediating children's explicit and implicit misconceptions about experimental design. In R. W. Proctor & E. J. Capaldi (Eds.), *Psychology of science: Implicit and explicit processes* (pp. 137–180). New York: Oxford University Press.

Simon, H. A., & Lea, G. (1974). Problem solving and rule induction: A unified view. In L. W. Gregg (Ed.), *Knowledge and cognition* (pp. 105-128). Hillsdale, NJ: Lawrence Erlbaum.

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

Swanson, H. L. (2015). Leveraging Students' Prior Knowledge in Attaining Deep Structural Understanding of Domain General Models. DISSERTATION, UC Berkeley. http://eprints.cdlib.org/uc/item/4z0914hz

Tschirgi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development*, *51*, 1-10.

Tenenbaum, H. R., Rappolt-Schlichtmann, G., & Zanger, V. (2004). Children's learning about water in a museum and in the classroom. *Early Childhood Research Quarterly*, *19*, 40-58.

Thagard, P. (2000). *How scientists explain disease*. Princeton, NJ: Princeton University Press.

Tweney, R. D. (2002). Epistemic artifacts: Michael Faraday's search for the optical effects of gold. In L. Magnani, N. J. Nersessian (Eds.) *Model-Based Reasoning: Science, Technology, Values* (pp. 287-303). N. Y.: Kluwer Academic/Plenum Publishers.

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

Vosniadou, S., & Brewer, W. F. (1994). Mental models of the day/night cycle. *Cognitive Science*, *18*, 123-183.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly journal of experimental psychology*, *12*, 129-140.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content.* London: B. T. Batsford.

Wertheimer, M. (1945/1959). *Productive thinking*. New York, NY: Harper & Row.

Wu, H. K., & Puntambekar, S. (2012). Pedagogical affordances of multiple external representations in scientific processes. *Journal of Science Education and Technology*, *21*, 754-767.

Zimmerman, 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.

Zimmerman, C., & Croker, S. (2013). Learning science through inquiry. In G. Feist and M. Gorman (Eds.), *Handbook of the psychology of science* (pp. 49-70). New York, NY: Springer.

Zimmerman, C., & Croker, S. (2014). A prospective cognition analysis of scientific thinking and the implications for teaching and learning science. *Journal of Cognitive Education and Psychology*, *13*, 245-257.

Zimmerman, C., & Cuddington, K. (2007). Ambiguous, circular and polysemous: Students' definitions of the "balance of nature" metaphor. *Public Understanding of Science, 16,* 393-406.

Zimmerman, C., Raghavan, K., & Sartoris, M. L. (2003). The impact of the MARS curriculum on students' ability to coordinate theory and evidence. *International Journal of Science Education, 25*, 1247-1271.

Table 1

A Taxonomy for Categorizing Experimental Studies of Scientific Thinking, with Representative Examples of Each Type

Cognitive Processes	**Science Practices**	**Type of Knowledge**	
		Domain Specific	**Domain General**
Forming and Refining Hypotheses (Hypothesis Space Search)	**Asking questions**	**A1** Metz (2004) Samarapungavan et al. (2008)	**D1** Chouinard (2007) Jirout & Klahr (2016)
	Developing and using models	**A2** Vosnidadou & Brewer (1992, 1994) Lehrer & Schauble (2004)	**D2** Raghavan & Glaser (1995) Greca & Moreira (2000) Wu & Puntambekar (2012)
Investigation Skills (Experiment Space Search)	**Planning and carrying out investigations**	**B** Tschirgi (1980) Schwichow et al. (2016)	**E** Siegler & Liebert (1975) Kuhn & Ho (1980)
Evaluating Evidence	**Analyzing and interpreting data/evidence**	**C1** Penner and Klahr (1996b) Masnick et al. (2016)	**F1** Shaklee & Paszek (1985) Masnick & Morris (2008)
	Constructing explanations	**C2** Hatano & Inagaki (2013) Gelman & Kremer (1991) Carey & Spelke (1994)	**F2** Schunn & Klahr (1993) Mynatt et al. (1978) Kelemen et al. (2014)

Note: The cognitive processes categories and knowledge types are adapted from Klahr and Dunbar (1988), Klahr (1994), and Klahr and Carver (1995). The science practices are from the *Framework for K-12 Science Education* (NRC, 2012). Two additional NRC (2012) practices -- “Engaging in argument from evidence” and “Obtaining, evaluating, and communicating information” -- are beyond the scope of this chapter. Similarly, “Mathematical and computational thinking” is a separate research literature despite its obvious connection to authentic scientific practice (but see Chapter 6 of this volume, “Development of Mathematical Thinking”). In each cell, we note one or more typical exemplars, although any given study might traverse several cells; additional work representative of each cell is described in the main text.