

The Development of Scientific Reasoning Skills

Corinne Zimmerman

University of Alberta, Edmonton, Alberta, Canada

The purpose of this article is to provide an introduction to the growing body of research on the development of scientific reasoning skills. The focus is on the reasoning and problem-solving strategies involved in experimentation and evidence evaluation. Research on strategy use in science has undergone considerable development in the last decade. Early research focused on knowledge-lean tasks or on tasks in which subjects were instructed to disregard prior knowledge. Klahr and Dunbar (1988) developed an integrated model of scientific discovery that has served as a framework to study the interaction of conceptual knowledge and the set of cognitive skills used in scientific reasoning. Researchers now take a more integrated approach, examining the development and use of strategies in moderately complex domains in order to examine the conditions under which subjects' theories (or prior knowledge) influence experimentation, evidence evaluation, and belief revision. Recent findings from integrated studies of scientific reasoning have the potential to inform and influence science education and conceptualizations of science as both academic skill and content domain. © 2000 Academic Press

The purpose of this article is to provide a general introduction to the body of research in cognitive and developmental psychology conducted under the labels "scientific reasoning," "scientific discovery," and "scientific thinking." There are three main reasons for reviewing this literature. First, within the last decade, there was a call to develop a distinct "psychology of science" (Gholson, Shadish, Neimeyer, & Houts, 1989; Tweney, Doherty, & Mynatt, 1981). Developmental psychologists, however, had been exploring aspects of scientific thinking long before, and concurrent with, this rallying call (e.g., Brewer & Samarapungavan, 1991; Carey, 1985; Carey, Evans, Honda, Jay, & Unger, 1989; Case, 1974; Dunbar & Klahr, 1989; see Flavell, 1963, for a summary of Piaget's work 1921–1958; Inhelder & Piaget, 1958; Kaiser, McCloskey, & Proffitt, 1986; Karmiloff-Smith & Inhelder, 1974; Kuhn, Amsel, & O'Loughlin, 1988; Piaget, 1970; Siegler, 1978). Moreover,

This research was supported by a Doctoral Fellowship from the Social Sciences and Humanities Research Council of Canada. I thank Gay Bisanz, David Klahr, Barbara Koslowski, and an anonymous reviewer for helpful comments on earlier drafts of the manuscript.

Correspondence and reprint requests should be addressed to Corinne Zimmerman, Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA 15260. E-mail: czimm+@pitt.edu.



science educators also have been interested in children's understanding of both scientific concepts and the scientific enterprise (e.g., Aikenhead, 1989; American Association for the Advancement of Science, 1990; DeBoer, 1991; Eisenhart, Finkel, & Marion, 1996; Glaser, 1984; Miller, 1983; Norris, 1995, 1997; Posner, Strike, Hewson, & Gertzog, 1982; Ryan & Aikenhead, 1992; Science Council of Canada, 1984; Schauble & Glaser, 1990). That is, a "psychology of science" has existed for some time among researchers interested in children's thinking. My first goal is to bridge the psychology of science and developmental literatures by reviewing the work at the intersection of these two areas.¹

Second, I want to highlight the considerable changes that have occurred within the study of scientific reasoning. Early research on scientific thinking concentrated on circumscribed aspects of the scientific discovery process (e.g., Siegler & Liebert, 1975; Wason, 1968). Within the last decade, however, researchers have used *simulated discovery tasks* in moderately complex domains in order to track participants as they explore, propose hypotheses, test hypotheses via experimentation, and acquire new knowledge in the form of revised hypotheses (e.g., Schauble, 1990, 1996). This evolution of the study of scientific thinking has made it an ideal arena in which to study the development of the cognitive skills implicated in reasoning, problem solving, and knowledge acquisition. Furthermore, these newer methods have generated findings that have the potential to inform researchers and teachers in science education.

This intersection with science education provides the third rationale for this review. In a recent *Handbook of child psychology* chapter, Strauss (1998) lamented the lack of communication between cognitive developmentalists and science educators despite the considerable overlap in the goals and interests of both groups. Strauss suggested that one of the reasons for this communication gap is that "Developmentalists often avoid studying the growth of children's understanding of the science concepts that are taught in school" (p. 358). Strauss went on to outline two criteria for the content of future research that would inform both groups of researchers: (a) the content should be from, and have significance for, particular domains of science (e.g., temperature/heat concepts in physics, photosynthesis in biology) and (b) the content should have generality across the domains of science (e.g., evaluating evidence or arguments for theories). Developmentalists, however, have been conducting research in both of these content areas. Strauss (1998) acknowledges some of the psychological literature on concepts and misconceptions

¹ In a recent review, Feist and Gorman (1998) surveyed the "developmental, cognitive, personality, and social psychology of science" (p. 3). They focused on developmental precursors to becoming a professional scientist (e.g., mathematical ability, religious background) and reviewed only a fraction of the growing body of work on the development of scientific reasoning.

in science, but reviews only a handful of studies that exist on the development of domain-general scientific reasoning skills. To bridge this gap, my third goal is to provide a review of this literature for both audiences and to demonstrate how developmental research on scientific reasoning can be used to inform science educators.

The plan of the article is as follows. In the first section, I will provide a general overview of the two main approaches to the study of scientific thinking: one focused on the development of conceptual knowledge in particular scientific domains and a second focused on the reasoning and problem-solving strategies involved in hypothesis generation, experimental design, and evidence evaluation. Both approaches will be introduced to distinguish two different connotations of “scientific reasoning,” but it is the second line of research that is the primary focus of this review. Next, I will introduce Klahr and Dunbar’s (1988) *Scientific discovery as dual search* (SDDS) model, which is a descriptive framework of the cognitive processes involved in scientific discovery. This framework represents an effort to integrate the concept-formation approach with the reasoning and problem-solving approach into a single coherent model.

In the second section, I will review major empirical findings using the SDDS model as a framework. The first subsection will include a brief review of research that has been focused on experimentation skills. In the second subsection, I will review the research on evidence evaluation skills. The last subsection will include a discussion of self-directed experimentation studies. In these integrative investigations of scientific reasoning, participants actively engage in all aspects of the scientific discovery process so that researchers can track the development of conceptual knowledge and reasoning strategies. I will conclude this subsection with some generalizations about development of skills and individual differences.

In the third and final section of the paper, I will provide a general summary around the points outlined above. Based on this review of the skills and strategies of scientific reasoning, I suggest the idea that science may best be taught as both an academic skill and a content domain.

APPROACHES AND FRAMEWORK

The most general goal of scientific investigation is to extend our knowledge of the world. “Science” is a term that has been used to describe both a body of knowledge and the activities that gave rise to that knowledge. In parallel, psychologists have been interested in both the *product*, or individuals’ knowledge about scientific concepts, and the *processes* or activities that foster that knowledge acquisition. Science also involves both the *discovery* of regularities, laws, or generalizations (in the form of hypotheses or theories) and the *confirmation* of those hypotheses (also referred to as justification or verification). That is, there has been interest in both the inductive processes involved in the generation of hypotheses and the deductive processes

used in the testing of hypotheses.² Scientific investigation broadly defined includes numerous procedural and conceptual activities such as asking questions, hypothesizing, designing experiments, using apparatus, observing, measuring, predicting, recording and interpreting data, evaluating evidence, performing statistical calculations, making inferences, and formulating theories or models (Keys, 1994; Schauble, Glaser, Duschl, Schulze, & John, 1995; Slowiaczek, Klayman, Sherman, & Skov, 1992). Because of this complexity, researchers traditionally have limited the scope of their investigations by concentrating on either the conceptual or the procedural aspects of scientific reasoning. That is, focus has been on the acquisition and development of two main types of knowledge, namely, domain-specific knowledge and domain-general strategies.³

The Domain-Specific Approach: Conceptual Knowledge about Science

One approach to studying the development of scientific reasoning has involved investigating the *concepts* that children and adults hold about phenomena in various content domains in science, such as biology (e.g., Carey, 1985; Hatano & Inagaki, 1994; Miller & Bartsch, 1997), evolution (e.g., Samarapungavan & Wiers, 1997), dinosaurs (e.g., Chi & Koeske, 1983), observational astronomy (e.g., Vosniadou & Brewer, 1992), and physics (e.g., Baillargeon, Kotovsky, & Needham, 1995; Hood, 1998; Kaiser et al., 1986; Clement, 1983; diSessa, 1993; Levin, Siegler, & Druyan, 1990; McCloskey, 1983; Pauen, 1996; Spelke, Phillips, & Woodward, 1995). The main focus has been on determining the naive 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

² Note that I intend to suggest only a very loose mapping between *discovery* and induction and between *confirmation* and deduction. The confirmation or verification of hypotheses is often framed as a deductive or conditional argument (e.g., “If my hypothesis is true, then I should observe some pattern of evidence”). Scientific *discovery*, as a process, involves aspects of both inductive and deductive reasoning. As deductive reasoning involves arguments that are *truth-preserving*, scientific discovery could not come about through deductive reasoning alone (Copi, 1986; Giere, 1979). Inductive inferences involve conclusions that are more than the sum of their premises, that is, they are “knowledge expanding” (Giere, 1979, p.37). The growth of scientific knowledge depends upon inductive inference (Holland et al., 1986).

³ This distinction between concepts (domain-specific knowledge) and strategies (domain-general knowledge) is loosely mapped onto the distinction between declarative and procedural knowledge in the memory literature (Kuhn et al., 1995). Although many authors used the term “domain-general knowledge,” I will use the term domain-general *strategies* throughout in order to emphasize the procedural aspects of this type of knowledge. Note that this dichotomy is not a clean one—it is possible to have domain-general knowledge and domain-general strategies, as well as domain-specific knowledge and domain-specific strategies (Klahr, personal communication).

or instruction. Historically, this approach has roots in the pioneering work of Piaget, who was interested in the development of various concepts such as time, number, space, movement, and velocity (e.g., Flavell, 1963; Inhelder & Piaget, 1958; Piaget, 1970). Individuals construct intuitive theories about their experiences with natural and social phenomena. These naive theories may or may not match currently accepted, scientific explanations of those same phenomena (Murphy & Medin, 1985; Vosnaidou & Brewer, 1992). Of interest is the content and structure of these naive theories, possible misconceptions, conceptual change, and explanatory coherence (Samarapungavan & Wiers, 1997; Schauble, 1996; Thagard, 1989).

In the domain-specific approach, a typical scientific reasoning task consists of questions or problems that require participants to use their conceptual knowledge of a particular scientific phenomenon. For example, children were asked to answer questions about the earth such as "Is there an edge to the earth?" and "Can you fall off the edge?" (Vosniadou & Brewer, 1992, p. 553). In the domain of genetic inheritance, participants were asked to reason about issues such as the origins of anatomical and behavioral differences in species by answering questions such as "How does it come about that people have different color of eyes/hair?" and "How did the differences between horses and cows originate?" (Samarapungavan & Wiers, 1997, p.174). In the domain of physics, individuals were instructed to draw the path of a ball as it exits a curved tube (Kaiser et al., 1986) or to predict the trajectory of a falling object (McCloskey, 1983).

In the previous examples, participants were using their current conceptual understanding (or "misconceptions") to generate a solution to the task. They were not required to evaluate evidence, make observations, or conduct experiments to verify their solutions or answers.⁴ As there are several different domains of science, and numerous concepts of interest within each (e.g., within the domain of physics alone, different researchers have studied the concepts of gravity, motion, velocity, balance, heat/temperature, electricity, and force, to name a few), a thorough discussion of these literatures is not appropriate given the scope of this review. Reviews and collections of work on domain-specific concepts can be found in Carey (1985), Gelman (1996), Gentner and Stevens (1983), Hirschfeld and Gelman (1994), Keil (1989), Pfundt and Duit (1988), Sperber, Premack, and Premack (1995), and Wellman and Gelman (1992).

⁴ There are some exceptions to this claim. For example, both children and college students hold the belief that all parts of an object move at the same rate. Levin et al. (1990) used kinesthetic training to counter children's single-object/single-motion intuitions. Children held a rod while walking around a pivot. On separate trials, they held either the far end of the rod or the near end of the rod (by the pivot). Children came to understand that they were required to move much faster when they held the far end of the rod; therefore, different points of a single object can move at different rates.

The Domain-General Approach: Introduction and Background

A second approach that has been taken to understand the development of scientific thinking has involved a focus on domain-general reasoning and problem-solving *strategies* that are involved in the discovery and modification of theories about categorical or causal relationships. These strategies include the general skills implicated in experimental design and evidence evaluation, where the focus is on the cognitive skills and strategies that transcend the particular content domain to which they are being applied. In their review of research on the acquisition of intellectual skills, Voss, Riley, and Carretero (1995) classified scientific reasoning as a “general intellectual skill.” Scientific thinking as defined in this approach involves the application of the methods or principles of scientific inquiry to reasoning or problem-solving situations (Koslowski, 1996). In contrast to the conceptual-development approach, participants are engaged in designing experiments (e.g., Schauble, 1996) or inspecting and evaluating the findings from a fictitious experiment (e.g., Kuhn et al., 1988).

This approach has historical roots in experimental psychology, in the body of research on reasoning and problem solving (e.g., VanLehn, 1989; Wason, 1960; Wason & Johnson-Laird, 1972). The study of scientific reasoning, in many respects, parallels the history of studying problem solving in general. Early approaches to problem solving involved knowledge-lean tasks that take minutes or hours to complete (e.g., the “insight problems” used by the Gestalt psychologists), whereas more recent approaches utilize multistep tasks (e.g., algebraic problems, the Tower of Hanoi) and knowledge-rich tasks in domains such as chess, computer programming, or medical diagnosis (e.g., VanLehn, 1989). Early attempts to understand scientific reasoning focused on simple, knowledge-lean tasks such as the Wason (1968) four-card selection task. The selection task is a test of conditional reasoning that was intended to simulate the process of selecting the best experiments to test or verify a scientific hypothesis (Oaksford & Chater, 1994).

In addition to examining the deductive components of scientific inference, Wason (1960; Wason & Johnson-Laird, 1972) created a knowledge-lean task that simulates the discovery of hypotheses. The 2–4–6 rule discovery task is a laboratory simulation of scientific discovery, in which participants postulate hypotheses, collect relevant experimental evidence, and then evaluate the evidence in order to accept or modify the initial hypothesis. The subject’s task is to determine the rule that governs the sequence of numbers in a triad. The original triad (2–4–6) is provided. Participants form a hypothesis, generate a triad (i.e., the “experiment”), receive feedback from the experimenter (i.e., the evidence), and then modify the hypothesis. The cycle continues until the subject discovers the rule (traditionally, “three numbers in ascending order”).

Initial investigations with adults focused on a what seemed to be a perva-

sive “confirmation bias” that existed, even among scientists (e.g., Mahoney & DeMonbreun, 1977). Confirmation bias refers to the general tendency of participants to propose triads that conform to—and thus if found to be such confirm—the hypothesis under consideration (Baron, 1994; Koslowski & Maqueda, 1993). For example, if the hypothesis under consideration is “numbers increasing by two,” a participant might suggest triads such as 3–5–7 or 10–12–14. This characterization of participants’ normative behavior as incorrect or fallible resulted from comparing it to prescriptions from Popper’s (1959) falsificationist philosophy. The correct way to test a hypothesis, according to this prescription, is to try to falsify it. In the given example, the participant following the falsificationist strategy would suggest triads such as 3–6–9 or 8–10–6 because these triads have the potential to disconfirm the hypothesis under consideration.

Klayman and Ha (1987) provided an insightful reinterpretation of participants’ behavior when they noted that there is a distinction between the *test choice* (i.e., the proposed triad) and the *expected outcome*. It is not the triad itself (e.g., proposing triad 7–9–11) that confirms or disconfirms a hypothesized rule, but the *answer* or feedback that one receives (e.g., the answer “conforms to rule” for 7–9–11 disconfirms the hypothesis “even numbers increasing by two” but provides confirmatory evidence for “numbers increasing by two”). Klayman and Ha referred to the strategy that participants typically use as the *positive test strategy*. The positive test strategy can yield information that can either confirm or disconfirm a hypothesis. Likewise, a *negative test strategy* can yield information that will either confirm or disconfirm a hypothesis. For example, proposing 1–2–3 or 6–4–2 when the current hypothesis is “numbers increasing by two” is a negative test. The result of this test provides the evidence that determines the fate of the hypothesis under consideration, not the test strategy utilized.

Given Klayman and Ha’s (1987) reconceptualization of the task, recent investigations have focused on the different strategies that participants use for generating experimental tests (e.g., positive test, negative test, counterfactual, win–stay lose–shift) and the conditions that influence different strategy use and success (e.g., discovery versus evaluation; under conditions of possible error; abstract versus semantic content; broad versus narrow rules; discovering one versus two rules; listing hypotheses before experimentation) (Farris & Revlin, 1989; Gorman, 1989; Kareev & Halberstadt, 1993; Klayman & Ha, 1989; Penner & Klahr, 1996b; Tukey, 1986; Wharton, Cheng, & Wickens, 1993).

Studies using the 2–4–6 task with children are rare, however. This dearth of developmental studies could be due, in part, to Wason and Johnson-Laird (1972) reporting that children generally discover the rule fairly quickly because they are less familiar with numerical relationships. Adults are more likely to draw upon their conceptual knowledge of relational properties, number facts, and mathematical terms such as odd–even, positive–negative,

rational–irrational, and prime–factorable (Tukey, 1986) to suggest rules such as “the figures have to have an L.C.D.” (lowest common denominator), or “the middle number is the arithmetic mean of the other two” (Wason & Johnson-Laird, 1972, pp. 208, 209). We know very little about developmental differences in strategy use for the 2–4–6 task. Children’s performance on other scientific reasoning tasks will be discussed following a discussion of Klahr and Dunbar’s (1988) framework for scientific discovery.

Integration of Concepts and Strategies: A Framework for the Scientific Discovery Process

The two contrasting approaches outlined represent different conceptualizations about what the development of scientific reasoning involves. In some respects, the different approaches reflect a lack of agreement concerning which type of acquisition (i.e., concepts or strategies) is more important for accounting for developmental differences in scientific reasoning (Dunbar & Klahr, 1989; Klahr, Fay, & Dunbar, 1993; Kuhn et al, 1988). As mentioned previously, however, these two approaches have used different types of tasks that emphasize either conceptual knowledge or experimentation strategies. Moreover, “science” can be characterized as both product and process. Klahr and Dunbar (1988) recognized the importance of both product (concept formation or knowledge acquisition) and process (experimental design and evidence evaluation skills) and have developed an integrated model of the cognitive processes involved in scientific activity. The SDDS framework incorporates domain-general strategies with domain-specific knowledge. Some necessary background information will be followed by a description of the model itself, which delineates the major types of cognitive processes involved in the different phases of scientific discovery.

Background

The SDDS model was influenced by the work and assumptions of Simon and his colleagues (e.g., Newell & Simon, 1972; Simon & Lea, 1974; Langley, Simon, Bradshaw, & Zytkow, 1987). Simon (1973, 1986, 1989) argued that scientific discovery is a problem-solving activity that uses the same information-processing mechanisms that have been identified in other problem-solving contexts. Prior to this, scientific discovery was viewed as the more mysterious, intuitive, or creative aspect of science, whereas the processes involved in the verification or falsification of hypotheses (once discovered) could be studied systematically. For example, Popper (1959) asserted that “there is no such thing as a logical method of having new ideas” (p. 31).

Scientific “reasoning” is the most common label for the research approaches outlined thus far. A careful examination of what is involved in scientific inquiry should reveal that it involves aspects of both problem solving and reasoning (Copi, 1986, Chap. 14; Giere, 1979). Haberlandt (1994)

defines problem solving as “the achievement of a goal within a set of constraints” (p. 365). Evans (1993) defines reasoning as “the central activity in intelligent thinking. It is the process by which knowledge is applied to achieve most of our goals” (p. 561). Definitions of reasoning and problem solving, often indistinguishable, include the idea of *goals* and the idea that direct retrieval of a solution from memory is not possible. The *scientific discovery process* is best conceptualized as including both reasoning and problem-solving skills with the ultimate goal of generating, and then appraising, the tenability of a hypothesis about a causal or categorical relationship.

One of the main generalizations about problem-solving processes is the use of *heuristic searches* (e.g., Hunt, 1994; Newell & Simon, 1972; VanLehn, 1989). A problem solver constructs or develops a representation of the problem situation (i.e., the *problem space*) in order to generate a solution within some set of constraints (Haberlandt, 1994; Simon, 1986). A problem space includes the *initial state*, a set of *operators* that allow movement from one state to another, and a *goal state* or solution. Many information-processing approaches adopt the idea that goal-oriented behavior in general involves a search through problem spaces (Klahr, 1992). In fact, some authors argue that all complex cognition is best viewed as problem solving (e.g., Anderson, 1993; Dunbar, 1998; Klahr & Simon, 1999; Hunt, 1994; VanLehn, 1989).

The SDDS Model

Klahr and Dunbar (1988; Dunbar & Klahr, 1989) conceived of scientific reasoning as problem solving that is characterized as a guided search and information gathering task. The primary goal is to discover a hypothesis or theory that can account for some pattern of observations in a concise or general form (Klahr, 1994). Klahr and Dunbar argued that scientific discovery is accomplished by a dual search process. The search takes place in two related problem spaces—the *hypothesis space* and the *experiment space*. The search process is guided by prior knowledge and previous experimental results.

With respect to searching hypothesis space, Klahr and Dunbar (1988) noted the difference between “evoking” and “inducing” a hypothesis. The key difference is that in some situations, one can use prior knowledge in order to constrain the search of hypothesis space, while in other situations, one must make some observations (via experimentation) before constructing an initial hypothesis. The latter scenario relies more heavily on inductive reasoning, while the former may rely on memory retrieval. One implication of this distinction is that the search through experiment space may or may not be constrained by a hypothesis. Initial search through the space of experiments may be done in the service of generating observations. In order to test a hypothesis, once induced, the search process involves finding an experiment that can discriminate among rival hypotheses. The search through these

two spaces requires different representations of the problem and may require different heuristics for moving about the problem spaces.

The first two cognitive processes of scientific discovery involve a coordinated, heuristic search. The third process of the SDDS model involves *evidence evaluation*. This process was initially described as the decision made on the basis of the cumulative evidence, that is, the decision to accept, reject, or modify the current hypothesis. Initially, Klahr and Dunbar emphasized the "dual search" nature of scientific discovery, while the evidence evaluation process was somewhat neglected in the overall discovery process. In more recent descriptions, Klahr has elaborated upon the evidence evaluation process, indicating that it involves a comparison of results obtained through experimentation with the predictions derived from the current hypothesis (Klahr & Carver, 1995; Klahr et al., 1993). The additional focus on this component of scientific reasoning is the result of a line of work by Kuhn and her colleagues (e.g., Kuhn, Amsel, & O'Loughlin, 1988; Kuhn, 1989). Kuhn has argued that the heart of scientific thinking lies in the skills at differentiating and coordinating theory (or hypotheses) and evidence.

Klahr and Dunbar's original description of the model highlighted the dual search coordination, but updated descriptions acknowledge that scientific discovery tasks depend upon the coordination and integration of all three components (Klahr, 1994; Klahr et al., 1993; Penner & Klahr, 1996a). Furthermore, Klahr grants there are many factors that may influence the amount, type, and strength of evidence needed to make a decision regarding the status of a hypothesis. Klahr et al. also observed that bias (or strength of prior belief) in favor of a particular hypothesis may affect which aspects of evidence are attended to and encoded. Given that the model integrates both concepts and strategies, it is worth noting that a tacit assumption of Klahr and Dunbar's model is that no negative connotation is attached to the idea that people rely on prior knowledge to constrain the search of hypothesis space or to construct an initial hypothesis. Consideration of prior knowledge is treated as scientifically legitimate. The issue of the legitimacy of considering prior knowledge (or theory) when collecting or evaluating evidence is a controversial one and will be discussed in more detail below.

The SDDS framework captures the complexity and the cyclical nature of the process of scientific discovery. The framework incorporates many component processes that previously had been studied in isolation. Table 1 provides a depiction of the cognitive processes associated with the three major components of scientific discovery (columns) and the two knowledge types (rows) involved in the SDDS framework (from Klahr, 1994; Klahr & Carver, 1995). Most investigations of scientific reasoning can be situated along these two dimensions. The SDDS model, as represented in Table 1, can be used as a framework for the review of the empirical investigations in the next section even though much of that work was not conducted with

TABLE 1

Klahr's (1994) Categorization of Types of Foci in Psychological Studies of Scientific Reasoning Processes and Representative Publications

Type of knowledge	Type of cognitive process		
	Hypothesis space search	Experiment space search	Evidence evaluation
Domain-specific	A (Carey, 1985)	B (Tschirgi, 1980)	C (Chi & Koeske, 1983)
Domain-general	D (Bruner et al., 1956, Reception experiments)	E (Siegler & Liebert, 1975)	F (Shaklee & Paszek, 1985)

a strong commitment either to this model or to the information-processing approach.

Summary

Scientific discovery is a complex activity that requires the coordination of several higher level cognitive skills, including heuristic search through problem spaces, inductive reasoning, and deductive logic. The main goal of scientific investigation is the acquisition of knowledge in the form of hypotheses that can serve as generalizations or explanations (i.e., theories). Psychologists have investigated the development of scientific concepts and the development of strategies involved in the discovery and verification of hypotheses. In initial studies of scientific thinking, researchers examined these component processes in isolation or in the absence of meaningful content (e.g., the 2-4-6 task). Klahr and Dunbar (1988) have suggested a framework for thinking about scientific reasoning in a more integrated manner. The SDDS framework is a descriptive account of the processes involved in concept formation and strategy development in the service of scientific discovery. In the next section I will review major empirical findings, beginning with early efforts to study scientific reasoning, in which only particular aspects of scientific discovery were of interest (as represented by the particular cells in Table 1), and ending with a description of more recent investigations that have focused on the integration of the processes and knowledge types represented by the SDDS framework as a whole.

MAJOR EMPIRICAL FINDINGS

Thus far I have described only in very general terms the main approaches to studying scientific reasoning and an attempt to integrate the major components of scientific activity into a single framework. In this section I will

describe the main findings or generalizations that can be made about human performance and development on *simulated discovery tasks*.⁵ Initial attempts to study scientific reasoning began with investigations that followed a “divide-and-conquer” approach by focusing on particular cognitive components, as represented by the cells in Table 1 (Klahr, 1994). The important findings to come out of this component-based approach will be described first. In the first two sections, I will concentrate on studies involving experimentation (cell E) and evidence evaluation (cell F), respectively. As mentioned previously, research that is focused *exclusively* on domain-specific hypotheses (i.e., cell A), exemplified by work on the development of conceptual knowledge in various domains such as biology or physics (e.g., Carey, 1985; McCloskey, 1983), has been reviewed elsewhere (e.g., Wellman & Gelman, 1992). In the third section, I will review investigations that use *self-directed experimentation* tasks. This recent line of research involves simulated discovery tasks that allow researchers to investigate the dynamic interaction between domain-general strategies (i.e., experimentation and evidence evaluation skills) and conceptual knowledge in moderately complex domains. These tasks are superficially similar to the 2–4–6 rule discovery task in that they focus on the three major processes of scientific discovery (cells D, E, and F). The key difference, however, is that these tasks are also focused on the interdependence of strategies with domain-specific knowledge (cells A through F).

Research on Experimentation Skills

Experimentation is an *ill-defined* problem for most children and adults (Schauble & Glaser, 1990). In the terminology of problem solving a *well-defined* problem is one in which the initial state, goal state, and operators are known, as in the case of solving an arithmetic problem or the Tower of Hanoi puzzle (Dunbar, 1998; VanLehn, 1989). This contrasts with ill-defined problems, in which the problem solver may be uncertain of any or all of the components (initial state, operators, goal state). The goal of an experiment is to test a hypothesis or an alternative (Simon, 1989). Although it has been argued that there is no one “scientific method” (e.g., Bauer, 1992; Shamos, 1995; Wolpert, 1993), it can be argued that there are several charac-

⁵ There are other approaches to the study of scientific thinking (Klahr, 1994). These include nonpsychological approaches (e.g., examining political, social, or anthropological forces), psychohistorical accounts of important scientists (e.g., Tweney, 1991), and computational models of scientific discovery (e.g., Langley, Simon, Bradshaw, & Zytkow, 1987). The “psychology of science” subdiscipline is in general concerned with the investigation of scientific behavior, including practicing scientists or individuals (children, lay adults) engaged in simulated discovery tasks. The present paper is focused on the latter group of participants. For a recent review that emphasizes the psychology of the *scientist*, see Feist and Gorman (1998). See Klahr and Simon (1999) and Klahr (2000) for integrative reviews of the major approaches to the study of scientific reasoning.

teristics common to experimentation across content domains. At a minimum, one must recognize that the process of experimentation involves generating observations that will serve as evidence that will be related to hypotheses. Klahr and Dunbar (1988) discussed the "multiple roles of experimentation" with respect to generating evidence. Experimentation can serve to generate observations in order to induce a hypothesis to account for the pattern of data produced (discovery context) or to test the tenability of an existing hypothesis under consideration (confirmation/verification context).

Ideally, experimentation should produce evidence or observations that are *interpretable* in order to make the process of evidence evaluation uncomplicated. One aspect of experimentation skill is to isolate variables in such a way as to rule out competing hypotheses. An alternative hypothesis can take the form of a specific competing hypothesis or the complement of the hypothesis under consideration. In either case, the control of variables and the systematic combinations of variables are particular skills that have been investigated.

Early approaches to examining experimentation skills involved minimizing the role of prior knowledge in order to focus on the strategies that participants used. That is, the goal was to examine the domain-general strategies that apply regardless of the content to which they are applied (i.e., cell E in Table 1). For example, building on the research tradition of Piaget (e.g., Inhelder & Piaget, 1958), Siegler and Liebert (1975) examined the acquisition of experimental design skills by 10- and 13-year-old children. The problem involved determining how to make a electric train run. The train was connected to a set of four switches and the children needed to determine the particular on/off configuration required. The train was in reality controlled by a secret switch so that the discovery of the correct solution was postponed until all 16 combinations were generated. In this task, there was no principled reason why any one of the combinations would be more or less likely. That is, the task involved no domain-specific knowledge that would constrain the hypotheses about which configuration was most likely.

An analogous task is the colorless liquid task used by Inhelder and Piaget (1958) and Kuhn and Phelps (1982). Four different flasks of colorless fluid were presented to 10- and 11-year-old children. The researcher demonstrated that by adding several drops of a fifth fluid, one particular combination of fluids changed color. The children's task was to determine which of the fluids or combinations of fluids was needed to reproduce the demonstration. Again, an individual could not use domain knowledge of the fluids (e.g., color or smell) to identify a likely hypothesis. Therefore, the only way to succeed on the task was to produce the complete combinatorial system. These studies demonstrated that the development of systematic, rule-governed behavior is an important precursor for competence in experimentation involving multiple variables and factorial combinations (Siegler, 1978). An equally important finding from the Siegler and Liebert study was that, in addition to in-

structional condition and age, *record keeping* was a significant mediating factor for success in producing the complete combinatorial solution. The 13-year-olds were more aware of their memory limitations, as most kept records. The 10-year-olds did not anticipate the need for records, but those who relied on memory aids were more likely to produce the complete factorial combination.

Tschirgi (1980) looked at one aspect of hypothesis testing in "natural" problem situations. Story problems were used in which two or three variables were involved in producing either a good or a bad outcome (e.g., baking a good cake, making a paper airplane) and therefore involved some domain knowledge (i.e., cells B and E of Table 1). Adults and children in Grades 2, 4, and 6 were asked to determine which levels of a variable to vary in order to produce a conclusive test of causality. In the cake scenario, for example, there were three variables: type of shortening (butter or margarine), type of sweetener (sugar or honey), and type of flour (white or whole wheat). Participants were told that a story character baked a cake using margarine, honey, and whole wheat flour and believed that the honey was responsible for the (good or bad) outcome. They were then asked how the character could prove this and were given three options to choose from: (a) baking another cake using the same sweetener (i.e., honey), but changing the shortening and flour; (b) using a different sweetener (i.e., sugar), but the same shortening and flour; or (c) changing all the ingredients (i.e., butter, sugar, and white flour).

Tschirgi (1980) found that in familiar, everyday problem situations, the type of outcome influenced the strategy for generating an experiment to produce evidence. In all age groups, participants looked for confirmatory evidence when there was a "positive" outcome. That is, for positive outcomes, they used a "Hold One Thing At a Time" (HOTAT) strategy for manipulating variables (choice *a* above). They selected disconfirmatory evidence when there was a "negative" outcome, using the more valid "Vary One Thing At a Time" (VOTAT) strategy (choice *b* above). The only developmental difference was that the sixth graders and adults (but not second and fourth graders) were aware of the appropriateness of the VOTAT strategy. Tschirgi suggested that the results supported a model of natural inductive logic that develops through everyday problem-solving experience with multivariable situations. That is, individuals base their choice of strategy on empirical foundations (e.g., reproducing positive effects and eliminating negative effects), not logical ones.

Sodian, Zaitchik, and Carey (1991) investigated whether children in the early school years understand the difference between testing a hypothesis and reproducing an effect. The tasks used by Siegler and Liebert (1975), Kuhn and Phelps (1982), and Tschirgi (1980) all involve producing effects and cannot be used to address this issue. Sodian et al. presented children in the first and second grades with a story situation in which two brothers dis-

agree about the size of a mouse in their home. One brother believes the mouse is small, and the other believes it is large. The participants were shown two boxes with different-sized openings (or "mouse houses") that contained food. In the *feed* condition children were asked to select the house that should be used if the brothers wanted to make sure the mouse could eat the food, regardless of its size. In the *find out* condition the children were asked to decide which house should be used to determine which brother is correct about the size of the mouse. Further questions prompted the children for what they could find out with each box (e.g., "Can a big mouse fit in a house with a small opening?"). If a child can distinguish between the goal of testing a hypothesis and generating an effect (i.e., feeding the mouse), then he or she should select different houses under the two conditions.

Over half of the first graders answered the series of questions correctly (with justifications) and 86% of the second graders correctly differentiated between conclusive and inconclusive tests. It is important to point out, however, that the children were provided with the two mutually exclusive and exhaustive hypotheses and, moreover, were provided with two mutually exclusive and exhaustive experiments from which to select (Klahr et al., 1993). In a second experiment, similar results were found with a task in which the story characters were trying to determine whether a pet aardvark had a good or a poor sense of smell. In the aardvark task, however, participants were not presented with a forced choice between a conclusive and an inconclusive test. Spontaneous solutions were generated by about a quarter of the children in both grades. For example, some children suggested the story characters should place some food very far away. If the aardvark has a good sense of smell, then it will find the food. The results support the general idea that children as young as 6 can distinguish between a conclusive and an inconclusive test of a simple hypothesis.

In summary, researchers interested in experimentation skills have focused on the production of factorial combinations and the isolation of variables on tasks in which the role of prior knowledge was minimized. An important precursor for success in producing a combinatorial array in the absence of domain-specific knowledge is systematic or rule-governed behavior, which appears to emerge around age 5. An awareness of memory limitations and of the importance of record keeping appears to emerge between the ages of 10 and 13. With respect to the isolation of variables, there is evidence that the goal of the experiment can affect the strategies selected. When the hypothesis to be tested can be construed as involving a positive or negative outcome, second-, fourth-, and sixth-grade children and adults select valid experimental tests when the outcome is negative, but use a less valid strategy when the outcome is positive. What develops is an awareness of the appropriateness of the VOTAT strategy selected in the case of negative outcomes. When domain knowledge can be used to view the outcome as positive, even adults do not appear to have developed an awareness of the inappropriateness

of the HOTAT strategy. The research reviewed in this section provides evidence that, under conditions in which producing an effect is not at issue, even children in the first grade understand what it means to test a hypothesis by conducting an experiment and, furthermore, that children as young as 6 can differentiate between a conclusive and an inconclusive experiment.

Research on Evidence Evaluation Skills

The evaluation of evidence as bearing on the tenability of a hypothesis has been of central interest in the work of Kuhn and her colleagues (1989, 1991, 1993a,b; Kuhn et al., 1988; Kuhn, Schauble, & Garcia-Mila, 1992; Kuhn, Garcia-Mila, Zohar, & Andersen, 1995). Kuhn (1989) has argued that the defining feature of scientific thinking is the set of skills involved in differentiating and coordinating theory (or hypothesis) and evidence. Fully developed skills include the ability to consciously articulate a theory, to understand the type of evidence that could support or contradict that theory, and to justify the selection of one of competing theories that explain the same phenomenon. The ability to consider alternative hypotheses is an important skill, as evidence may relate to competing hypotheses. Kuhn has asserted that the skills in coordination of theory and evidence are the “most central, essential, and general skills that define scientific thinking” (Kuhn, 1989, p. 674). That is, these skills can be applied across a range of content areas. By defining scientific thinking as the metacognitive control of this coordination process, the thinking skills used in scientific inquiry can be related to other formal and informal thinking skills (e.g., medical, legal, historical, and “everyday” thinking skills) (Kuhn, 1991, 1993a,b).

In the evidence evaluation studies to be reviewed, the evidence provided for participants to evaluate is *covariation* evidence. In the first section, I will provide a general description of what is meant by covariation evidence and generalizations about tasks used in the studies to be reviewed. Then, early studies of rule use in the evaluation of covariation evidence will be described. In the second section I will summarize the landmark work of Kuhn, Amsel, and O’Loughlin (1988). The third section will be devoted to the critical reactions to Kuhn et al.’s methodology along with attempts to improve upon that work. Koslowski’s (1996) criticisms are especially important and will be discussed in a separate section. Koslowski’s criticisms revolve around the assumptions of the primacy of covariation evidence in making inductive causal inferences. This section includes a description of corroborating evidence from related fields. In the last section, I complete the discussion with a general summary and conclusions about the assumptions of evidence evaluation studies.

Covariation Evidence: General Description and Early Research

With respect to determining causal relationships, Hume (1758/1988) identified the covariation of perceptually salient events as one potential *cue* that

TABLE 2
Cells in a 2×2 Contingency Table for Studies
Using Covariation Evidence

	Outcome	
	Present	Absent
Antecedent		
Present	A	B
Absent	C	D

Note. The antecedent is the presumed causal factor.

two events are causally related. Even young children have a tendency to use the covariation of events (antecedent and outcome) as one indicator of causality (e.g., Inhelder & Piaget, 1958; Kelley, 1973; Shultz, Fisher, Pratt, & Rulf, 1986; Shultz & Mendelson, 1975). Covariation between events, however, is a necessary but not a sufficient cue for inferring a causal relationship (Shaklee & Mims, 1981).

In covariation, there are four possible combinations of the presence and the absence of antecedent (or potential cause) and outcome (see Table 2). In the case of perfect covariation, one would find only cases in which both the antecedent and the outcome were present together (cell A) and cases in which they were both absent (cell D). This cell represents instances that confirm a relationship between two events. However, in a noisy and imperfect world, cases exist in which there is a violation of the sufficiency of causal antecedent (cell B). In these cases, the antecedent that is assumed to be causal is present, but the outcome is absent. There are also cases that violate the necessity of the causal antecedent (cell C), in which the outcome occurs in the absence of the assumed cause. Frequencies in cells A and D confirm that a relationship exists between two events, and frequencies in cells B and C provide disconfirming evidence (Shaklee & Mims, 1981). The correct rule for determining covariation between events is the conditional probability strategy, in which one compares $P(A | (A+C))$ with $P(B | (B+D))$. Mathematically, this simply requires a comparison of the frequency ratio in cells $A \div (A+C)$ with $B \div (B+D)$ (Shaklee & Paszek, 1985). If the ratios are the same, then there is no relationship between the antecedent (presumed cause) and the outcome (i.e., in statistical terms, the variables are independent). If there is a difference between these ratios, then the events covary (i.e., a relationship may exist).⁶

⁶ It is not clear *how large* the difference must be in order to conclude that the two events are related (i.e., not independent). When scientists analyze data in the form of contingency tables, it is appropriate to conduct significance tests (e.g., compute a χ^2 statistic and associated probability). This issue is not well articulated in any of the studies using tasks with covariation evidence and requires further empirical and conceptual work.

In evidence evaluation tasks involving covariation of events, participants are provided with data corresponding to the frequencies in the cells of a 2×2 contingency table (i.e., represented in either tabular or pictorial form). The pattern could represent perfect covariation, partial (or imperfect) covariation, or no correlation between the two events. The task may require participants to evaluate a given hypothesis in light of the evidence or to determine which hypothesis the pattern of data support. In either case, the focus is on the inferences that can be made on the basis of the evidence (i.e., in most cases, participants were instructed to disregard prior domain knowledge). Experimental design skills are not of interest.

Early work on covariation detection was conducted by Shaklee and her colleagues (e.g., Shaklee & Mims, 1981; Shaklee & Paszek, 1985; Shaklee, Holt, Elek, & Hall, 1988). Shaklee investigated participants' ability to make judgments based on patterns of covariation evidence. Children in Grades 2 through 8 and adults were presented with a series of 2×2 contingency tables. The data in the table represented two events that may or may not be related (e.g., healthy/sick plant and the presence/absence of bug spray). The task was to determine, given the pattern of evidence, which hypothesis was supported (i.e., if the events are related and the direction of the relationship if any). Participants may use different strategies or rules to judge event covariation, so contingency tables were specifically constructed such that a characteristic pattern of performance would result, depending upon the rule used. This procedure is analogous to the rule-assessment approach described in Siegler (1976, 1981).

Shaklee and her colleagues found a general trend in the rules used to weigh the evidence in the contingency tables. It was predicted that younger children (Grades 2, 3, and 4) would use the frequency reported in cell A to make a judgment and proceed developmentally to a rule in which they compare frequencies in cells A vs. B. However, the cell A strategy was not common. The most sophisticated strategy that participants seemed to use, even as adults, was to compare the *sums-of-diagonals*. The *conditional probability* rule was used only used by a minority of participants, even at the college level. Adults could readily learn this rule, if they were shown how to compare the relevant ratios (see footnote 6). Children in Grades 4 through 8 could be taught to use the sums-of-diagonals rule (Shaklee et al., 1988). In many respects, the task in this form has more to do with mental arithmetic or naive data analysis (i.e., computing χ^2 without a calculator) and less with identification of covariation between events (Holland, Holyoak, Nisbett, & Thagard, 1986). Shaklee's work, however, demonstrated that participants' judgments were rule-governed and that they did consider the information from all four cells but in a less than ideal manner.

The Work of Kuhn, Amsel, and O'Loughlin (1988)

Kuhn et al. (1988) were not concerned with the rules participants used to combine information in the cells. Their primary motivation was to examine

how participants reconcile prior beliefs with covariation evidence presented to them. Kuhn et al. also used simple, everyday contexts, rather than phenomena from specific scientific disciplines. In an initial theory interview, participants' beliefs about the causal status of various variables were ascertained. For example, in Studies 1a and 1b, adults and sixth and ninth graders were questioned about their beliefs concerning the types of foods that make a difference in whether a person caught a cold (35 foods in total). Four variables were selected based on ratings from the initial theory interview: two factors that the participant believed make a difference in catching colds (e.g., type of fruit and type of cereal) and two factors the participant believed do not make a difference (e.g., type of potato and type of condiment). This procedure allowed the evidence to be manipulated such that covariation evidence could be presented which *confirmed* one existing causal theory and one noncausal theory. Likewise, noncovariation evidence was presented that *disconfirmed* one previously held causal theory and one noncausal theory. The specific manipulations, therefore, were tailored for each person in the study.

Kuhn et al.'s (1988) general method involved the presentation of covariation data sequentially and cumulatively. Participants were asked a series of questions about what the evidence shows for each of the four variables. Responses were coded as either *evidence-based* or *theory-based*. To be coded as evidence-based, a participant's response to the probe questions had to make reference to the patterns of covariation or instances of data presented (i.e., the findings of the scientists). For example, if shown a pattern in which type of cake covaried with getting colds, a participant who noted that the sick children ate chocolate cake and the healthy kids ate carrot cake would be coded as having made an evidence-based response. In contrast, theory-based responses made reference to the participant's prior beliefs or theories about why the scientists might have found that particular relationship. In the previous example, a response that chocolate cake has "sugar and a lot of bad stuff in it" or that "less sugar means your blood pressure doesn't go up" (Kuhn, 1989, p. 676) would be coded as theory-based.

Kuhn et al. were also interested in both inclusion inferences (an inference that two variables are causally related) and exclusion inferences (an inference of no relationship between variables). Participants' inferences and justification types could be examined for covariation evidence versus noncovariation evidence and in situations where the prior theory was causal or noncausal. Other variations in the other studies included (a) examining the effects of explicit instruction; (b) use of real objects for evidence (e.g., tennis balls with various features) versus pictorial representations of data; (c) task instructions to relate the evidence to multiple theories instead of a single theory; and (d) a reciprocal version of the task in which the participant generates the pattern of evidence that would support and refute a theory.

Through the series of studies, Kuhn et al. found certain patterns of responding. First, the skills involved in differentiating and coordinating theory

and evidence, and bracketing prior belief while evaluating evidence, show a monotonic developmental trend from middle childhood (Grades 3 and 6) to adolescence (Grade 9) to adulthood. These skills, however, do not develop to an optimum level even among adults. Even adults have a tendency to meld theory and evidence into a single representation of “the way things are.” Second, participants have a variety of strategies for keeping theory and evidence in alignment with one another when they are in fact discrepant. One tendency is to ignore, distort, or selectively attend to evidence that is inconsistent with a favored theory. For example, the protocol from one ninth grader demonstrated that upon repeated instances of covariation between type of breakfast roll and catching colds, he would not acknowledge this relationship: “They just taste different . . . the breakfast roll to me don’t cause so much colds because they have pretty much the same thing inside [i.e., dough]” (Kuhn et al., p. 73, elaboration added).

A second tendency was to adjust a theory to fit the evidence. This practice seems perfectly reasonable, but the bothersome part is that this “strategy” was often outside a participant’s conscious control. Participants were often unaware of the fact that they were modifying a theory. When asked to recall their original beliefs, participants would often report a theory consistent with the evidence that was presented and not the theory as originally stated. An example of this is one ninth grader who did not believe type of condiment (mustard versus ketchup) was causally related to catching colds. With each presentation of an instance of covariation evidence, he acknowledged the evidence and elaborated a theory based on the amounts of ingredients or vitamins and the temperature of the food the condiment was served with (Kuhn et al., p. 83). Kuhn argued that this tendency suggests that the subject’s theory does not exist as an object of cognition. That is, a theory and the evidence for that theory are undifferentiated—they do not exist as separate entities.

Third, there were a variety of errors involved in understanding covariation evidence and its connection to causality. There were also problems in understanding noncovariation. For example, when asked to generate a pattern of evidence that would show that a factor makes no difference in an outcome, participants often produced covariation evidence in the opposite direction of that predicted by their own causal theory.

Criticisms of Kuhn et al. (1988)

Koslowski (1996) considers the work of Kuhn and her colleagues to be a significant contribution to research on scientific thinking, in part because it raises as many questions as it answers. Various authors have criticized the Kuhn et al. research, however, on both methodological and conceptual grounds (e.g., Amsel & Brock, 1996; Koslowski, Okagaki, Lorenz, & Umbach, 1989; Ruffman, Perner, Olson, & Doherty, 1993; Sodian, Zaitchik, & Carey, 1991).

Sodian et al. (1991) first questioned Kuhn et al.'s interpretation that third- and sixth-grade children cannot distinguish between their beliefs (i.e., theories) and the evidence that would confirm or disconfirm those beliefs. Sodian et al. deliberately chose story problems in which children did not hold strong prior beliefs and they used a task that was less complex than those used by Kuhn et al. (1988). In order to demonstrate that children differentiate beliefs and evidence, they selected a task that did not require a judgment about one causal factor while simultaneously ruling out the causal status of three other potential variables (each with two values). As described in the previous section, Sodian et al.'s research demonstrated that even first- and second-grade children can distinguish between the notions of "hypothesis" and "evidence" by selecting or generating a conclusive test of a simple hypothesis.

Ruffman, Perner, Olson, and Doherty (1993) examined children's abilities (aged 4 to 7) to form hypotheses on the basis of covariation evidence. They also used less complex tasks with fewer factors to consider than Kuhn et al. (1988). When only one potentially causal factor (type of food) covaried perfectly with outcome (tooth loss), children as young as 6 years old could form a hypothesis that the factor is causally responsible. Ruffman et al. also ascertained that 6-year-olds were able to form a causal hypothesis on the basis of a pattern of covariation evidence (i.e., imperfect evidence).

In order to rule out the possibility that children were simply describing a state of affairs, Ruffman et al. tested if 4- to 7-year-olds understood the predictive properties of the hypothesis formed on the basis of covariation evidence. Children were asked to evaluate evidence and then form a hypothesis about which characteristics of tennis rackets were responsible for better serves (e.g., racket size, head shape). They were then asked which tennis racket they would buy and how good the next serve would be. The results were consistent with the idea that by age 7, children understood that the newly formed hypothesis could be used to make predictions.

Ruffman et al. deliberately chose factors that were all equally plausible. Correct performance in the Kuhn et al. tasks was defined by considering covariation evidence as more important than the implausible hypothesis it was intended to support. For example, in Studies 3 and 4 of Kuhn et al., adults and third, sixth, and ninth graders were to evaluate evidence to determine the features of tennis balls that resulted in good or poor serves (i.e., color, texture, ridges, and size). Most children and adults do not believe that color is causally related to the quality of a tennis serve. Ruffman et al. argued that revising prior beliefs (e.g., about the causal power of color) is more difficult than forming new theories when prior beliefs do not exist or are not held with conviction. Literature on inductive inference supports this claim (e.g., Holland et al., 1986).

Amsel and Brock (1996) examined whether children (second/third grade, sixth/seventh grade) and adults evaluated covariation evidence independently of prior beliefs. They also used a task that was less complex and

cognitively demanding than Kuhn et al. (1988). Amsel and Brock argued that causal judgments should be assessed independently of the justification for that judgment and that these judgments about the causal status of variables should be assessed on a scale that reflects certainty, rather than a forced choice (i.e., the factor is causal, noncausal, or neither).

Unlike Ruffman et al.'s (1993) criticism about strong prior beliefs, the participants in Amsel and Brock's study were selected only if they did hold strong prior beliefs concerning the variables. That is, participants believed that a relationship exists between the health of plants and the presence/absence of sunshine and that no relationship exists between health of plants and the presence/absence of a charm (represented as a four-leaf clover). Children in second/third grade, sixth/seventh grade, college students, and non-college adults were presented with four data sets to evaluate given by the factorial combination of prior belief (causal or noncausal) by type of contingency data (perfect positive correlation vs. zero correlation). Participants were asked whether giving the plants (sun/no sun) or (charm/no charm) was causally related to whether the plants were healthy or sick and to respond only based on the information given and not what they know about plants.

Standard covariation evidence served as the control (four instances in a 2×2 contingency table), while three conditions involved "missing data." Participants in the control group were presented with four data instances that represented covariation (or noncovariation) between the putative causal factor and the outcome. Participants in the three missing data conditions were shown two additional instances in which either the antecedent was unknown, the outcome was unknown, or both were unknown. Amsel and Brock reasoned that if participants were evaluating the evidence independent of their strongly held prior beliefs, then the judgments in the control and missing data conditions should be the same. That is, participants would simply ignore the evidentially irrelevant missing data. If they were using prior beliefs, however, they might try to explain the missing data by judging the variables as consistent with their prior beliefs. If they were using newly formed beliefs, then judgments would be consistent with the new belief and pattern of evidence (causal with covariation evidence; noncausal with noncovariation).

College adults were most like the "ideal reasoner" (i.e., defined as someone whose causal certainty scores were based solely on the four instances of contingency data). Both groups of children (second/third grade and sixth/seventh grade) judged the causal status of variables consistent with their prior beliefs even when the evidence was disconfirming. The noncollege adults' judgments tended to be in between, leading the authors to suggest that there are differences associated with age and education in making causal judgments independently of prior beliefs. In the missing data conditions, participants did not read into or try to explain the missing data. Rather, the effect was to cause the children of both grade levels, but not the adults, to be less certain about the causal status of the variables on the outcome. There

was an age and education trend for the frequency of evidence-based justifications. When presented with evidence that disconfirmed prior beliefs, children from both grade levels tended to make causal judgments consistent with their prior beliefs. When confronted with confirming evidence, however, both groups of children and adults made similar judgments.

In this section I have outlined the research that was conducted based on criticisms of Kuhn et al.'s (1988) methodology, including issues about task complexity, plausibility of factors, participants' method of responding (e.g., certainty judgments versus forced choice), and data coding (e.g., causal judgments and justifications assessed jointly or separately). In the next section, I discuss Koslowski's (1996) criticisms of Kuhn et al. in particular and of studies employing covariation evidence in general.

Koslowski's (1996) Criticisms

Koslowski (1996) argued that researchers contributing to the psychological literature on scientific reasoning have used tasks in which correct performance has been operationalized as the identification of causal factors from covariation evidence while suppressing prior knowledge and considerations of plausibility. She argued that this overemphasis on covariation evidence has contributed to either an incomplete or a distorted picture of the reasoning abilities of children and adults. In some cases, tasks were so knowledge-lean (e.g., the colorless liquids task) that participants did not have the opportunity to use prior knowledge or explanation, thus contributing to an incomplete picture. Knowledge-rich tasks have been used, but definitions of correct performance required participants to disregard prior knowledge. In this case, a distorted picture has resulted. As in Klahr and Dunbar's (1988) integrated model, Koslowski treats it as legitimate to rely on prior knowledge when gathering and evaluating evidence. Koslowski presented a series of 16 experiments to support her thesis that the principles of scientific inquiry are (and must be) used in conjunction with knowledge about the world (e.g., knowledge of plausibility, causal mechanism, and alternative causes).

The role of causal mechanism. Koslowski questioned the assumptions about the primacy of covariation evidence. One of the main concerns in scientific research is with the discovery of causes. Likewise, in much of the research on scientific reasoning, tasks were employed in which participants reason about causal relationships. Psychologists who study scientific reasoning have been influenced by the philosophy of science, most notably the empiricist tradition which emphasizes the importance of observable events. Hume's strategy of identifying causes by determining the events that covary with an outcome has been very influential. In real scientific practice though, scientists are also concerned with *causal mechanism*, or the process by which a cause can bring about an effect. Koslowski noted that we live in a world full of correlations. It is through a consideration of causal mechanism that we can determine which correlations between perceptually salient events

should be taken seriously and which should be viewed as spurious. For example, it is through the identification of the *Escherichia coli* bacterium that we consider a causal relationship between hamburger consumption and illness or mortality. It is through the absence of a causal mechanism that we do not consider seriously the classic pedagogical example of a correlation between ice cream consumption and violent crime rate.⁷

In the studies by Kuhn et al. (1988) and others (e.g., Amsel & Brock, 1996), correct performance entailed inferring causation from covariation evidence and lack of a causal relationship from noncovariation evidence. Evidence-based justifications are seen as superior to theory-based justifications. In study 4, for example, a ninth grader was asked to generate evidence to show that the color of a tennis ball makes a difference in quality of serve and responded by placing 8 light-colored tennis balls in the "bad serve" basket and 8 dark-colored balls in the "good serve" basket. When asked how this pattern of evidence proves that color makes a difference, the child responds in a way that is coded as theory-based: "These [dark in *Good* basket] are more visible in the air. You could see them better" (Kuhn et al., 1988, p. 170). Participants frequently needed to explain why the patterns of evidence were sensible or plausible. Kuhn asked "Why are they unable simply to acknowledge that the evidence shows covariation without needing first to explain why this is the outcome one should expect?" (p. 678). Kuhn argued that by not trying to make sense of the evidence, participants would have to leave theory and evidence misaligned and therefore need to recognize them as distinct. Koslowski (1996), in contrast, would suggest this tendency demonstrates that participants' naive scientific theories incorporate information about both covariation and causal mechanism. In the case of theories about human or social events, Ahn, Kalish, Medin, and Gelman (1995) also presented evidence demonstrating that college students seek out and prefer information about causal mechanism over covariation when making causal attributions (e.g., determining the causes of an individual's behavior).

Koslowski (1996) presented a series of experiments to support her thesis about the interdependence of theory and evidence in legitimate scientific reasoning. In most of these studies, she demonstrated that participants (sixth graders, ninth graders, adults) do take mechanism into consideration when evaluating evidence in relation to a hypothesis about a causal relationship. In initial studies, Koslowski demonstrated that even children in sixth grade consider more than covariation when making causal judgments (Koslowski & Okagaki, 1986; Koslowski et al., 1989).

In subsequent studies, participants were given problem situations in which a story character is trying to determine if some target factor (e.g., a gasoline additive) is causally related to an effect (e.g., improved gas mileage). They

⁷ We also use this pedagogical example to illustrate the importance of considering additional variables that may be responsible for both outcomes (i.e., high temperatures for this example).

were then shown either perfect covariation between target factor and effect or partial covariation (four of six instances). Perfect correlation was rated as more likely to indicate causation than partial correlation. Participants were then told that a number of plausible mechanisms had been ruled out (e.g., the additive does not burn more efficiently, the additive does not burn more cleanly). When asked to rate again how likely it was that the additive is causally responsible for improved gas mileage, the ratings for both perfect and partial covariation were lower for all age groups.

Koslowski also tried to determine if participants would spontaneously generate information about causal mechanisms when it was not cued by the task (Experiment 16). Participants (sixth grade, ninth grade, adults) were presented with story problems in which a character is trying to answer a question about, for example, whether parents staying in the hospital improves the recovery rate of their children. Participants were asked to describe whatever type of information might be useful for solving the problem. Half of the participants were told that experimental intervention was not possible, while the other half were not restricted in this manner. Almost all participants showed some concern for causal mechanism, including expectations about how the target mechanism would operate. Although the sixth graders were less likely to generate a variety of alternative hypotheses, all age groups proposed appropriate contrastive tests.

In summary, Koslowski argues that sound scientific reasoning requires "bootstrapping," that is, using covariation information and mechanism information *interdependently*. Scientists, she argues, rely on theory or mechanism to decide which of the many covariations in the world are likely to be causal (or merit further study). To demonstrate that people are reasoning in a scientifically legitimate way, one needs to establish that they rely on both covariation and mechanism information and they do so in a way that is judicious. As shown in the previous studies, participants did treat a covarying factor as causal when there was a possible mechanism that could account for how the factor might have brought about the effect and were less likely to do so when mechanism information was absent. Moreover, participants at all age levels showed a concern for causal mechanism even when it was not cued by the task.

Considerations of plausibility. In another study (Experiment 5), participants were asked to rate the likelihood of a possible mechanism to explain covariations that were either plausible or implausible. Participants were also asked to generate their own mechanisms to explain plausible and implausible covariations. When either generating or assessing mechanisms for plausible covariations, all age groups (sixth graders, ninth graders, adults) were comparable. When the covariation was implausible, sixth graders were more likely to generate dubious mechanisms to account for the correlation.

In some situations, scientific progress occurs by taking seemingly implausible correlations seriously (Wolpert, 1993). Similarly, Koslowski argued

that if people rely on covariation and mechanism information in an interdependent and judicious manner, then they should pay attention to implausible correlations (i.e., those with no apparent mechanism) when the implausible correlation occurs often. Koslowski provided an example from medical diagnosis, in which discovering the cause of Kawasaki's syndrome depended upon taking seriously the implausible correlation between the illness and having recently cleaned carpets. In Experiment 9, sixth and ninth graders and adults were presented with an implausible covariation (e.g., improved gas mileage and color of car). Participants rated the causal status of the implausible cause (color) before and after learning about a possible way that the cause could bring about the effect (improved gas mileage). In this example, participants learned that the color of the car affects the driver's alertness (which affects driving quality, which in turn affects gas mileage). At all ages, participants increase their causal ratings after learning about a possible mediating mechanism. In addition to the presence of a possible mechanism, a large number of covariations (four instances or more) was taken to indicate the possibility of a causal relationship for both plausible and implausible covariations (Experiment 10).

Corroborating evidence from other literatures. Research from the literature on conceptual development can be used to corroborate Koslowski's argument that both children and adults hold rich prior theories that include information about covariations and theoretically relevant causal mechanisms. With respect to concepts, a *distinct theoretical domain* is defined as including four properties, including a distinct ontology, causal mechanisms, unobservable constructs, and coherent relations among theoretical constructs (Wellman & Gelman, 1998; cited in Siegler & Thompson, 1998). Although the story situations used in any of the studies reported may or may not be considered an "autonomous domain" like biology (e.g., Solomon, Johnson, Zaitchik, & Carey, 1996), it seems possible that the theories children hold incorporate some of these properties (i.e., whether the naive theory is about diet and health, how to make a good cake, or how tennis ball characteristics affect quality of serve). Brewer and Samarapungavan (1991) argued that children's theories about the sun-moon-earth system embody the same essential characteristics of scientific theories (e.g., they are integrated, consistent, and not perceptually bound and have explanatory power). The key difference is that the child, although taking a rational approach to the sun-moon-earth system and his or her theories about it, lacks the accumulated knowledge and experimental methodology of the institution of science.

Research from the literature on causal reasoning can be used to demonstrate that people use information other than covariation when making causal attributions. White (1988) presented a review of the origins and development of causal processing. Although covariation is common in both physical and social domains there are several other cues-to-causality such as *generative transmission* (i.e., an effect is attributed to a cause if it is capable of generat-

ing the appropriate effect), *intended action* (versus reflexive or accidental causes), *temporal contiguity* (e.g., when cause and effect overlap in time), and *spatial contiguity* (e.g., when cause and effect make spatial contact) (Shultz, Fisher, Pratt, & Rulf, 1986; White, 1988). Use of these cues appears around age 3 and develops in preschoolers by the age of 5. The use of covariation as a cue to causality appears later in development, in part because other cues-to-causality are preferred and used by preschoolers. Also, covariation evidence is difficult to process for several reasons including memory demands (White, 1988), the effort required to weigh the 2×2 instances (i.e., the conditional-probability rule) (Shaklee & Paskek, 1985), and general difficulties in processing negative information (Bassoff, 1985). That is, it is difficult to process the "instances" in which both the antecedent and the outcome are absent. Therefore, the use of covariation evidence is prone to errors and is used imperfectly, even by adults (White, 1988).

There is also evidence to support the idea that individuals consider possible alternative causes when making causal judgments. Schustack and Sternberg (1981) were interested in determining the types of evidence undergraduate adults consider when making inductive causal inferences. The instructions encouraged participants to consider the information presented in contingency tables and prior knowledge about causes involved in the task scenarios. Using regression analysis, Schustack and Sternberg determined that participants considered the information in all four cells of a 2×2 contingency table. In addition to the data, the presence of possible alternative causes that can explain the outcome was taken into consideration when making judgments about causality based on covariation evidence. Cummins (1995; Cummins, Lubart, Alksnis, & Rist, 1991) presented undergraduate participants with arguments about causes and effects couched in conditional statements. The causal statements varied with respect to the number of possible alternative causes or disabling conditions. These two factors had a systematic effect on participants' judgments about the strength of the causal relationship, such that conclusions for arguments that had fewer alternative causes or disabling conditions were found to be more acceptable.

Summary of Evidence Evaluation Research

The research conducted by Koslowski (1996) along with related research on conceptual development and causal reasoning was presented to demonstrate that both children and adults hold rich causal theories about "everyday" and scientific phenomena (Brewer & Samarapungavan, 1991; Murphy & Medin, 1985). These theories incorporate information about covariation between cause and effect, causal mechanism, and possible alternative causes for the same effect. Plausibility is a general constraint on the generation and modification of theories (Holland et al., 1986). Without such constraints, individuals would be overwhelmed by the countless number of possible correlations in a complex world.

One clear message from the initial attempts to study the “domain-general” process of evidence evaluation is that both children and adults typically do not evaluate patterns of evidence in a theoretically empty vacuum (i.e., in terms of Table 1, cells C and F are implicated, not F alone). It seems clear that it is necessary for researchers to reflect on the distinction between performance that reflects a fundamental bias and performance that reflects a consideration of plausibility, causal mechanism, and alternative causes, but that is still scientifically legitimate. Correct performance in the evidence evaluation studies reviewed involved “bracketing” or ignoring theoretical considerations. Koslowski argued that it is scientifically legitimate to attend to theoretical considerations *and* patterns of evidence in scientific reasoning. To do otherwise provides a distorted view of the reasoning abilities of children and adults.⁸ Because of the discrepancy between the assumptions used in early investigations (e.g., Kuhn et al., 1988) and later investigations (e.g., Koslowski, 1996), the picture concerning the development of evidence evaluation skills is somewhat unclear. Early studies support the idea that even adults have difficulty reasoning in a scientifically legitimate manner, whereas later studies demonstrated that even sixth graders reason in accord with the principles of scientific inquiry.

A second message from these studies is that it is important to consider not only the coordination of theory and evidence (Kuhn et al., 1988), but the coordination of theory with evidence evaluation and experimentation skills. Participants in many of Koslowski’s studies were not explicitly designing experiments, but they were reasoning in a manner consistent with someone who understands that confounding factors affect the process of evidence evaluation and judgments about the tenability of a hypothesis. In a recent line of research described next, investigators have addressed the issues of the interdependence of theoretical considerations and experimentation strategies for generating and evaluating evidence along with considerations about the development of these skills.

Self-Directed Experimentation: An Integrated Approach to Scientific Reasoning

In the decade since the appearance of the SDDS framework (Klahr & Dunbar, 1988) there has been a move toward research in which participants take part in all three phases of scientific activity. Interestingly, one of the earliest simulated discovery tasks—Wason’s (1960) 2–4–6 task—included

⁸ For example, the HOTAT strategy is usually described as “inappropriate” and “invalid” but in some contexts, this strategy may be legitimate. For example, in real-world contexts, scientists and engineers cannot make changes one at a time because of time and cost considerations. Therefore, for theoretical reasons, only a few variables are held constant (Klahr, personal communication). In the tasks described here, HOTAT is interpreted as invalid because there are typically a countable number of variables to consider, each with only two or three values.

all components as well. Unlike the rule discovery task though, researchers examine the development of both domain-specific knowledge and domain-general strategies.

The objective is to construct a picture of the reciprocal influences of strategy on knowledge and knowledge on strategy by focusing on a topic in which individuals have some beginning conceptions but then engage in self-directed experimentation to confirm, extend, or change those beginning theories. (Schauble, Glaser, Raghavan, & Reiner, 1991a, p. 203)

Rather than describe each study in detail, I will provide a description of the common features of this approach, including common measures. I will then highlight the main findings with respect to developmental changes and individual differences.

General Features of the Self-Directed Experimentation Studies

In self-directed experimentation studies individuals participate in all phases of the scientific investigation cycle (hypothesis generation, experimentation, evidence evaluation, and hypothesis revision). Participants explore and learn about a multivariable causal system through activities initiated and controlled by the participant. There are two main types of "systems." In the first type of system, participants are involved in a hands-on manipulation of a physical system, such as the BigTrak robot (Dunbar & Klahr, 1989) or the canal task (i.e., a tank with an adjustable floor in which one tests the speed of boats which vary in size, weight, etc.; Kuhn et al., 1992). The second type of system is a computer simulation, such as the Daytona microworld (to discover the factors affecting the speed of race cars) (Schauble, 1990) or Voltaville (a computer-based electric circuit lab) (Schauble et al., 1992). The systems vary in complexity from fairly simple (e.g., Daytona) to moderately complex domains such as hydrostatics (Schauble, 1996) and molecular biology (e.g., Dunbar, 1993; Okada & Simon, 1997).

These systems allow the experimenter to be in control of participants' prior knowledge and the "state of nature" (Klahr, 1994), but in a way that is not arbitrary, like the rule discovery task (Wason, 1960), or artificial, like the task of determining the "physics" of a simulated universe (i.e., describing the motions of particles based on shape and brightness) used by Mynatt, Doherty, and Tweney (1978). One aspect of this control over the system is that some of the variables are consistent with participants' prior beliefs and some are inconsistent. The starting point is the participant's own theory, which presumably is the result of naive conceptions and formal education. Participants hold prior beliefs for both types of systems. For example, people hold strong beliefs that weight is a factor in how fast objects sink (Penner & Klahr, 1996a; Schauble, 1996) and that the color of a race car does not affect speed (Kuhn et al., 1992). By starting with the participants' own theories,

the course of theory revision can be tracked as participants confirm prior conceptions and deal with evidence that disconfirms prior beliefs.

There are two additional features that are common to a subset of the self-directed experimentation studies. The first involves length of time with the task. In some cases, participants work on a problem-solving task for one experimental session, while others involve repeated exposure to the problem-solving environment, often over the course of weeks. The microgenetic method is common to many self-directed experimentation studies because it allows researchers to track the process of change by observing participants as they engage in the task on multiple occasions (Siegler & Crowley, 1991). Researchers can observe both the development of domain knowledge and the individual variability in strategy usage by which that knowledge is acquired (Kuhn, 1995). The microgenetic studies reviewed typically involve observing only children (e.g., fifth and sixth graders in Schauble, 1990) or comparing one group of children (Grade 4) and one group of adults (Kuhn et al., 1995). In more complex domains such as electricity, only adults' performance (nonscience undergraduates) is tracked over a period of weeks (e.g., Schauble et al., 1991a).

Second, in some studies, participants are provided with some type of external memory system, such as a data notebook or record cards to keep track of plans and results, or access to computer files of previous trials. Tweney et al. (1981) noted that many of the tasks used to study scientific reasoning are somewhat artificial, in that real scientific investigations involve *aided* cognition. This aspect is important because it is ecologically valid and the task remains centered on reasoning and problem solving and not memory. However, given the importance of metacognition (e.g., Kuhn, 1989; Kuhn et al., 1995), it is possible to determine at what point children and adults recognize their own memory limitations as they navigate through a complex task.

A final feature, not generally acknowledged, is that these systems provide other cues-to-causation. In the previous section, I outlined Koslowski's (1996) arguments concerning the importance of information other than co-variation. Although causal mechanisms typically are unobservable, other cues-to-causation are present, such as contiguity in time and space, temporal priority, intended action, and generative transmission (e.g., Corrigan & Denton, 1996; Shultz et al., 1986; Sophian & Huber, 1984; White, 1988). Rather than reading data tables (e.g., Shaklee & Mims, 1981) or inspecting pictorial representations of evidence (e.g., Kuhn et al., 1988), participants can observe the antecedent (putative cause) and outcome (effect). Causal (and noncausal) relationships are discovered and confirmed through the generation of evidence via experimentation, which are initiated by the participants' own activities and developing theories. In Schauble's (1996) most recent study, participants' references to causal mechanisms in the domains of hydrostatics and

hydrodynamics were noted (e.g., unobservable forces such as “currents,” “resistance,” “drag,” and “aerodynamics”).

Common Measures

Given the cyclical nature of the discovery process, and the context of multiple sessions, many different types of observations can be made as the participants explore a causal system. There are several measures that are common to this type of study. Typically, there was some measure of *success*, that is, how many participants were successful at determining the function of the RPT button for a programmable robot (e.g., Dunbar & Klahr, 1989) or at determining the causal and noncausal factors in a simulated laboratory (e.g., Schauble, 1990). Knowledge acquisition is often of interest so studies include an assessment of *comprehension* or knowledge gain. For example, in the Voltaville study by Schauble et al. (1992), students' understanding of electricity concepts was measured before and after their exploration sessions.

Researchers recorded participants' *plans* as well as their *predictions* before an individual experiment was carried out. The number of *inferences* the participants made were recorded, along with whether they were inclusion inferences (i.e., that a factor is causal) or exclusion inferences (i.e., that a factor is not causal). Inferences were also coded as being either valid or invalid (i.e., based on sufficient evidence and a controlled design). *Justifications* for inferences were recorded, and they were coded as being evidence-based or theory-based. Other measures of interest were the percentage of experiment-space searched and the proportion of experiments that were conservative (i.e., based on a VOTAT strategy of variable manipulation). Use of data management was also noted (e.g., the percentage of experiments and outcomes recorded).

Developmental Differences in Self-Directed Experimentation Studies

Differences between children and adults were common across studies. I will make some generalized statements about patterns of change across the 14 self-directed experimentation studies that have been published since 1988 (see Table 3 for a summary of characteristics and main findings for each study). Three studies included both children and adults as participants, six reported data from children only (typically Grades 3 through 6), and five used adult participants only. These generalizations must be considered quite tentative because of the small number of studies and the dearth of studies with an adequate cross section of age groups.

Children's performance (third to sixth graders) was characterized by a number of tendencies: to generate uninformative experiments, to make judgments based on inconclusive or insufficient evidence, to vacillate in their judgments, to ignore inconsistent data, to disregard surprising results, to focus on causal factors and ignore noncausal factors, to be influenced by prior

TABLE 3
Self-Directed Experimentation Studies: Characteristics and Main Findings

Study	Participants	Task domain(s)	Unique feature	Main findings
Klahr and Dunbar (1988)	Undergraduates with and without programming experience	BigTrak robot (discovery of a new function)	Development of SDDS ^a model	Two main strategies of discovery: search hypothesis space (theorists) and search experiment space (experimenters); scientific reasoning characterized as problem solving involving integrated search in two problem spaces
Dunbar and Klahr (1989)	Experiment 3: Grades 3-6 (with LOGO training) ^{b,c}	BigTrak robot (discovery of a new function)	Focus on SDDS and developmental differences	Experiment 3: Given the same data, children proposed different hypotheses than adults; subjects had difficulty abandoning current hypothesis (did not search H-space, or use experimental results); subjects did not check if hypothesis was consistent with prior data; experiments were designed to prove current hypothesis rather than discover a correct one
Schauble (1990)	Grade 5/6	Daytona microworld (cars)	Microgenetic study of evolving beliefs and reasoning strategies	Exploratory strategies improved with time; invalid heuristics preserved favored theories; children using valid strategies gained better understanding of microworld structure
Schauble, Glaser, Raghavan, and Reiner (1991a)	Undergraduates (non-science majors)	Voltaville (electric circuits)	Role of conceptual model on strategies of experimentation	Good and poor learners (identified by gain scores) differed in sophistication of conceptual model, goal-directed planning, generating and interpreting evidence, data management, and algebraic competence

Schauble, Klopfer, and Raghavan (1991b)	Grade 5/6	Canal task (hydrodynamics) Spring task (hydrostatics)	Scientist context (goal = understanding), Engineering context (goal = optimization)	Subject's belief about goal important: Science context resulted in broader exploration and more attention to all variables (including noncausal variables) and all possible combinations than in Engineering context; greatest improvement when exploring engineering problem first, followed by science problem
Kuhn, Schauble, and Garcia-Mila (1992)	Experiment 1: Grade 4 Experiment 2: Grade 5/6	Daytona microworld (cars), canal task (boats), sports company research (balls)	Transfer paradigm (across domains)	Developmental change in microgenetic context not specific to a single content domain; coexistence of more and less advanced strategies; domain knowledge alone does not account for development of scientific reasoning; codevelopment with reasoning strategies
Schauble, Glaser, Raghavan, and Reiner (1992)	Undergraduates (non-science majors)	Black boxes task (electric circuits)	Role of conceptual model for understanding a physical system	Appropriate domain knowledge and efficient experimental strategies both required for successful reasoning; hierarchy of causal/conceptual models; each associated with characteristic strategies for evidence generation and interpretation and data management
Dunbar (1993)	Undergraduates with training in biology	Simulated molecular genetics laboratory	Episode from the history of science (discovery of genetic inhibition)	Study 1: Attending to evidence inconsistent with current hypothesis led to solution; Study 2: Subject's goal was to discover <i>two</i> mechanisms (one consistent, one inconsistent with current hypothesis); twice as many subjects found correct solution; goal affects success/errors

TABLE 3—Continued

Study	Participants	Task domain(s)	Unique feature	Main findings
Klahr, Fay, and Dunbar (1993)	Undergraduates, college students, and Grades 3 and 6	BigTrak microworld discovery of a new function)	Initial hypothesis provided (incorrect and plausible or implausible)	Developmental differences in domain-general heuristics for search and coordination of searches in experiment and hypothesis spaces; adults more likely to consider multiple hypotheses; children focus on plausible hypotheses; plausibility affects dual search
Kuhn, Garcia-Mila, Zohar, and Andersen (1995)	Community college students, Grade 4	Daytona (cars), canal task (boats), TV enjoyment task, school achievement task	Transfer paradigm (across physical and social domains)	Strategic progress maintained by both groups when problem content changed midway; social domain lagged behind physical domain; coexistence of valid and invalid strategies for children and adults (i.e., strategy variability not unique to periods of developmental transition)
Penner and Klahr (1996a)	10-, 12-, and 14-year-olds	Sinking objects task	Participants hold strong prior (incorrect) beliefs about role of weight	Prior belief affected initial goal (i.e., to demonstrate effect of weight); all children learned effects of other factors (e.g., material, shape) via experimentation; older children more likely to view experiments as testing hypotheses; younger children experimented without explicit hypotheses

Schauble (1996)	Noncollege adults, Grade 5/6	Canal task (hydrodynamics) Spring task (hydrostatics)	Continuous measures (varying effect sizes and measurement error)	Success required both valid strategic and correct beliefs (bidirectional relation); children and adults referred to causal mechanisms; variability in measurements limited progress in understanding, i.e., hard to distinguish error from small effect size—interpretation depends on theory
Okada and Simon (1997)	Undergraduate science majors	Simulated molecular genetics laboratory	Singles vs. pairs of participants	Pairs more successful in discovery and more active in explanatory activities (e.g., considering alternative hypotheses) than singles; pairs and singles conducted same critical experiments, but pairs exploited the findings; scientific discovery facilitated by collaboration
Chen and Klahr (1999)	Grades 2, 3, and 4	Spring task, slopes task, sinking objects task, paper and pencil tasks (natural and social science domains)	Explicit and implicit training of the control-of-variables (cvs) strategy	Direct instruction, but not implicit probes, improved children's ability to design unconfounded experiments; cvs resulted in informative tests which facilitated conceptual change; ability to transfer learned strategy increased with age

Note. "SDDS, Scientific discovery as dual search.

^bLOGO is a programming language required to operate the BigTrak robot.

^cExperiments 1 and 2 were reported in Klahr and Dunbar (1988); therefore, only the main findings from Experiment 3 (participants in Grades 3–6) are discussed.

belief, to have difficulty disconfirming prior beliefs, and to be unsystematic in recording plans, data, and outcomes (Dunbar & Klahr, 1989; Klahr et al., 1993; Kuhn et al., 1992, 1995; Penner & Klahr, 1996a; Schauble, 1990, 1996; Schauble & Glaser, 1990; Schauble et al., 1991b). In microgenetic studies, though, children in Grades 5 and 6 typically improve in the percentage of valid judgments, valid comparisons, and evidence-based justifications with repeated exposure to the problem-solving environment (Kuhn et al., 1992, 1995; Schauble, 1990, 1996; Schauble et al., 1991b). After several weeks with a task, fifth- and sixth-grade children will start making exclusion inferences and indeterminacy inferences (i.e., that one cannot make a judgment about a confounded comparison) and not focus solely on inclusion or causal inferences (Schauble, 1996). They begin to distinguish between an informative and an uninformative experiment by attending to or controlling other factors.

Adults were also influenced by prior beliefs, but less so than children in Grades 3 through 6. They tend to make both inclusion and exclusion inferences and make more valid comparisons by using the VOTAT strategy. While children would jump to a conclusion after a single experiment, adults typically needed to see the results of several experiments. Rather than ignoring inconsistencies, adults tried to make sense of them. Adults were more likely to consider multiple hypotheses (e.g., Dunbar & Klahr, 1989; Klahr et al., 1993). For children and adults, the ability to consider many alternative hypotheses was a factor contributing to success (Schauble et al., 1991a). What is somewhat problematic, however, was that adults often displayed many of the characteristics of children when the task complexity was changed (Schauble & Glaser, 1990). This finding leaves the question of what exactly is involved in the development of scientific reasoning skills.

Given a novel problem-solving environment, both children and adults display intraindividual variability in strategy usage. That is, multiple strategy usage is not unique to childhood or periods of developmental transition (Kuhn et al., 1995). A robust finding in microgenetic studies is the coexistence of multiple strategies, some more or less advanced than others (Kuhn et al., 1992; Schauble, 1990; Siegler & Crowley, 1991; Siegler & Shipley, 1995). The existence of multiple strategies should not be surprising for scientific reasoning tasks, given the different phases of scientific discovery. However, within the experimentation and evidence evaluation phases, multiple strategies coexist. Developmental transitions do not occur suddenly, that is, participants do not progress from an inefficient or invalid strategy to a more advanced strategy without ever returning to the former. For example, when isolating variables for a particular experiment, individuals typically do not switch abruptly from the less valid strategies of "Hold One Thing At a Time" (HOTAT) and "Change All variables" (CA) to the valid strategy of "Vary One Thing At a Time" (VOTAT). Rather, the repeated exposure to a problem results in participants' dissatisfaction with strategies that do not

result in progress. New strategies begin to develop, but they do not immediately replace less effective strategies.

With respect to inferences based on evidence, valid inferences were defined as inferences of inclusion (i.e., that a variable is causal) or exclusion (i.e., that a variable is not causally related to outcome) that were based on controlled experiments and included both levels of the causal and outcome variables (e.g., Kuhn et al., 1992, 1995; Schauble, 1990, 1996; Schauble et al., 1991b). Even after discovering how to make inferences under these conditions, participants often have difficulty giving up less-advanced inference strategies such as false inclusion and exclusion inferences that are consistent with prior beliefs, are based on a single instance of covariation (or noncovariation) between antecedent and outcome, or are based on one level of the causal factor and one level of the outcome factor (Kuhn et al., 1992; Schauble, 1996). For children and adults, it is more difficult to integrate evidence that disconfirms a prior causal theory than evidence that disconfirms a prior noncausal theory. The former case involves a restructuring of a belief system, while the latter involves incorporating a newly discovered causal relation (Holland et al., 1986; Koslowski, 1996). The experimentation and inference strategies selected depend on the prior theory and whether it is causal or noncausal.

The increasing sophistication of scientific reasoning, whether in children or adults, seems to involve both strategy changes and the development of knowledge. There is a dynamic interaction between the two, that is, the changes in knowledge and strategy “bootstrap” each other: “appropriate knowledge supports the selection of appropriate experimentation strategies, and the systematic and valid experimentation strategies support the development of more accurate and complete knowledge” (Schauble, 1996, p. 118). In one of the few studies comparing children (Grade 4) and adults directly, Kuhn et al. (1995) demonstrated that strategy development followed the same general course in children and adults (but that adults outperformed children) and that there were no developmental constraints on “the time of emergence or consolidation of the skills” involved in scientific reasoning (p. 102). One implication is that researchers can trace the development of these skills in children, adults, and professional scientists (e.g., Dunbar, 1995). True cross-sectional approaches, however, have not yet been pursued and we are thus left with snapshots of behavior of school children and adults.

The development of multiple strategy use in scientific reasoning mirrors development in other academic skills, such as math (e.g., Bisanz & LeFevre, 1990; Siegler & Crowley, 1991; Siegler & Shrager, 1984), reading (e.g., Perfetti, 1992), and spelling (e.g., Varnhagen, 1995; Varnhagen, McCallum, & Burstow, 1997). That is, multiple strategies coexist for solving mathematical problems and sounding out or spelling unfamiliar words. Researchers who study these academic skills have focused on the *set* of strategies that

children use, changes in relative frequency of strategies, changes in effectiveness of execution, and changes in acquisition of new strategies (Siegler, 1995; Siegler & Shipley, 1995).

Individual Differences

Theorists versus experimenters. Simon (1986) noted that individual scientists have different strengths and specializations, but the “most obvious” is the difference between experimentalists and theorists (p.163). Bauer (1992) also noted that despite the great differences among the various scientific disciplines, within each there are individuals who specialize as theorists or experimenters. Klahr and Carver (1995) observed that “in most of the natural sciences, the difference between experimental work and theoretical work is so great as to have individuals who claim to be experts in one but not the other aspect of their discipline” (p. 140).

Klahr and Dunbar (1988) first observed strategy differences between *theorists* and *experimenters* in adults. Individuals who take a theory-driven approach tend to generate hypotheses and then test the predictions of the hypotheses or, as Simon (1986) described, “draw out the implications of the theory for experiments or observations, and gather and analyse data to test the inferences” (p. 163). Experimenters tend to make data-driven discoveries, by generating data and finding the hypothesis that best summarizes or explains that data.

Dunbar and Klahr (1989) and Schauble (1990) also found that children conformed to the description of either theorists or experimenters. In a number of studies, success was correlated with the ability to generate multiple hypotheses (e.g., Schauble et al., 1991a) and, in the Smithtown simulation, the theorists were more successful than the experimenters (Schauble & Glaser, 1990). In the “black boxes task,” because of task constraints, the systematic generation of all combinations does not guarantee success (Schauble et al., 1992). The participant’s goal was to hook up black boxes to a test circuit with light bulbs in order to identify the component inside each of 10 boxes. Participants needed at least some minimal understanding of electric circuits to determine that one of the components needs to be a battery in order to determine the contents of the other black boxes (e.g., resistors, wires). Therefore, experimentation must be, at least in part, theory-driven to succeed on this task. The general characterization of some participants as “theorists”— and that a theory-driven approach can lead to success in discovery contexts—lends support to the idea that inadequate accounts of the development of scientific reasoning will result from studying experimentation or evidence evaluation in the absence of any domain knowledge or under instructions to disregard prior knowledge.

Scientists versus engineers. Research by Tschirgi (1980) initially suggested the possibility that the participant’s goal could affect the choice of experimentation strategy. Recall that participants (adults and children in

Grades 2, 4, and 6) were to decide the types of ingredients (e.g., sweetener: sugar vs. honey) that would produce a good or a bad cake. For good outcomes, a less valid HOTAT strategy was used. For bad outcomes, the more valid VOTAT strategy was used. Schauble (1990) also noted that fifth- and sixth-grade children often behaved as though their goal was to produce the fastest car in the Daytona microworld rather than to determine the causal status of each of the variables.

Schauble et al. (1991b) addressed the issue of goals by providing fifth- and sixth-grade children with an "engineering context" and a "science context." They suggested that some aspects of children's and adults' performance on scientific reasoning tasks could be elucidated by a consideration of what the participant believed the goal of experimentation was. Children worked on the canal task (an investigation in hydrodynamics) and the spring task (an investigation of hydrostatics).

When the children were working as scientists, their goal was to determine which factors made a difference and which ones did not. When the children were working as engineers, their goal was optimization, that is, to produce a desired effect (i.e., the fastest boat in the canal task, and the longest spring length in the springs problem). When working in the science context, the children worked more systematically, by establishing the effect of each variable, alone and in combination. There was an effort to make inclusion inferences (i.e., an inference that a factor is causal) and exclusion inferences (i.e., an inference that a factor is not causal).

In the engineering context, children selected highly contrastive combinations and focused on factors believed to be causal, while overlooking factors believed or demonstrated to be noncausal. Typically, children took a "try-and-see" approach to experimentation while acting as engineers, but took a theory-driven approach to experimentation when acting as scientists. These findings support the idea that researchers need to be aware of what the participant perceives the goal of experimentation to be: optimization or understanding.

Sophistication of causal model. In the Voltaville study, Schauble et al. (1992) examined the change in college students' models of circuits and electrical components (all were nonscience students). An initial task with circuit components, the "black box task," was used to categorize students according to the sophistication of their initial conceptual models. There were four levels of causal models. Over the course of 5 weeks, the students were to conduct experiments in the Voltaville simulation to determine as many laws or principles of electricity as possible.

Based on knowledge gains, college students were also categorized as good or poor learners. All of the good learners began with more sophisticated conceptual models. Good learners generated and stated many alternative hypotheses. Good learners also conducted more controlled experiments (33%) and conducted a broader and deeper search of experiment space. Of-

ten, good learners would test the generality of a causal relationship once discovered and based their generalizations on a sufficient quantity of evidence. Poor learners conducted fewer controlled experiments (11%) and were unsystematic and incomplete in their search of experiment space. Differences were also apparent in the systematicity of data recording. Poor learners made goal-oriented plans for only 15% of their experiments and recorded the results of 52% of their experiments. Good learners recorded goal-oriented plans and experimental results for 46 and 79% of their experiments, respectively.

Schauble et al. (1992) also found differences in conceptual models resulted in different strategies for generating evidence, interpreting evidence, and data management. It is important to highlight that it was not the case that the college students with more sophisticated models simply knew more about electricity concepts. These students did not “know” the answers, but they did learn more. And those who made the greatest knowledge gains spontaneously used more valid experimentation strategies. One of the key differences between the good and poor learners was the ability to use *disconfirming* information. Students with more sophisticated conceptual models attempted to make sense of unexpected or disconfirming evidence. Dunbar (1993) also found that university students who were able to discover the mechanisms of gene control in a simulated molecular biology laboratory were the ones who attempted to make sense of evidence that was inconsistent with a current hypothesis.

Kuhn et al. (1995) found differential performance for physical domains (i.e., microworlds involving cars and boats) and social domains (i.e., determining the factors that make TV shows enjoyable or make a difference in students' school achievement). Performance in the social domains was inferior for both children in the fourth grade and adults (community college students). The percentage of valid inferences was lower than that for the physical domains, participants made very few exclusion inferences (i.e., the focus was on causal inferences), and causal theories were difficult to relinquish, whether they were previously held or formed on the basis of the experimental evidence (often insufficient or generated from uncontrolled comparisons). Kuhn et al. (1995) suggested that adults and fourth graders had a richer and more varied array of existing theories in the social domains and that participants may have had some affective investment in their theories about school achievement and TV enjoyment, but not for their theories about the causal factors involved in the speed of boats or cars.

SUMMARY AND CONCLUSIONS

I outlined three main reasons why a review of the literature on the development of scientific reasoning is timely. First, there has been increasing interest in the “psychology of science,” as evidenced by recent integrative reviews by Feist and Gorman (1998) and Klahr and Simon (1999). The breadth and

scope of those reviews precluded a detailed examination of the literature on the development of scientific reasoning. The present review contributes to these efforts to establish the psychology of science as a thriving subdiscipline.

My second goal was to provide an overview of the major changes that have occurred in the study of scientific reasoning. The empirical study of experimentation and evidence evaluation strategies has undergone considerable development. Early attempts to study domain-general skills implicated in scientific reasoning focused on particular aspects of scientific discovery. The role of prior knowledge was minimized by using knowledge-lean tasks (e.g., Siegler & Liebert, 1975; Wason, 1960) or by instructing participants to disregard their prior knowledge (e.g., Kuhn et al., 1988; Shaklee & Paszek, 1985). Recent approaches to the study of inductive causal inference situate participants in a simulated discovery context, in which they discover laws or generalities in the multivariable causal system through active experimentation. The development of both strategies and conceptual knowledge can be monitored. These two aspects of cognition bootstrap one another, such that experimentation and inference strategies are selected based on prior conceptual knowledge of the domain. These strategies, in turn, foster a deeper understanding of the system via more sophisticated causal models, which (iteratively) foster more sophisticated strategy usage. Participants discover and use both invalid strategies (e.g., HOTAT and CA isolation-of-variables, false inclusion and exclusion inferences, inferences based on single instances or one level of a variable, and justifications that appeal to prior theories) and valid strategies (e.g., inclusion and exclusion inferences based on VOTAT and multiple observations and justifications that appeal to experimental evidence).

Klahr and Dunbar (1988) acknowledged that the separation between knowledge and strategy in scientific reasoning was and is highly artificial and posited a framework to integrate these knowledge types. Psychological research in general has been influenced by the idea that it is difficult to study context-free cognition. For example, one of the most robust findings in the reasoning literature is the facilitative effects of semantic content (e.g., Almor & Sloman, 1996; Cheng & Holyoak, 1985, 1989; Cosmides, 1989; Cummins, 1995, 1996; Cummins, Lubart, Alksnis, & Rist, 1991; Evans & Over, 1996; Girotto, 1990; Girotto & Politzer, 1990; Hawkins, Pea, Glick, & Scribner, 1984; Klaczynski, Gelfand, & Reese, 1989; Lawson, 1983; Markovits & Vachon, 1990; Oaksford & Chater, 1994, 1996; O'Brien, Costa, & Overton, 1986; Overton, Ward, Noveck, Black, & O'Brien, 1987; VanLehn, 1989; Ward & Overton, 1990; Wason & Johnson-Laird, 1972). That is, the substitution of concrete rather than abstract content affects performance on tasks involving categorical syllogisms or conditional reasoning. Given the repeated demonstrations of the effect of knowledge on reasoning, it appears that an approach designed to minimize prior knowledge, or one that instructs

participants to disregard prior knowledge, will do little to inform us of the nature or development of reasoning in general, and inductive scientific reasoning in particular, especially given that a primary goal of scientific investigation is to extend and elaborate our *knowledge* of the world.

My third goal was to review the integrative approaches to scientific reasoning in order to demonstrate to researchers in science education that developmentalists have been studying “situated cognition.” Strauss (1998) asserted that “Traditional cognitive developmental psychologists often choose to investigate the development of concepts located at a level of mental organization that is so general and deep, and so tied to phylogenetic selection that it is virtually unchangeable” (p. 360). The studies reviewed here illustrate that this assertion is overstated. In fact, Siegler (1998) has written about the “main themes” that pervade the contemporary study of cognitive development. One of the themes that has emerged is that children are far more competent than first suspected, and likewise, adults are less so. This characterization describes cognitive development in general and scientific reasoning in particular.

. . . The complex and multifaceted nature of the skills involved in solving these problems, and the variability in performance, even among the adults, suggest that the developmental trajectory of the strategies and processes associated with scientific reasoning is likely to be a very long one, perhaps even lifelong. Previous research has established the existence of both early precursors and competencies . . . and errors and biases that persist regardless of maturation, training, and expertise. (Schauble, 1996, p.118)

Contrary to Strauss’s impression, a significant number of developmental psychologists have focused their efforts on the study of cognitive development and its relation to academic skills (Siegler, 1998). For over 2 decades, the academic skills of primary interest have been the traditional “R’s”: reading, writing, and arithmetic (e.g., Gibson & Levin, 1975). These are considered the basic *skills* that children learn in school. In contrast, science typically is taught as though it were a *content domain* like Social Studies or History. It is my assertion that the research reviewed here suggests that it is possible to teach both of the key features of science, that is, the content of the scientific disciplines (e.g., biology, physics) and the experimentation and evidence evaluation skills. Gibson and Levin (1975) asked “Could we do research that would, in the end, help children learn to read?” (p. xi). I suggest that researchers in scientific reasoning have been progressing toward research that could help children to become better science students and scientifically literate adults (whether or not they become professional scientists). There has already been movement in this direction. For example, Chen and Klahr (1999) have examined the effects of direct and indirect instruction of the *control-of-variables* (CVS) strategy. Direct instruction about CVS improved children’s ability to design informative experiments which in turn facilitated conceptual change in a number of domains. Raghavan and her colleagues

(e.g., Raghavan & Glaser, 1995; Raghavan, Sartoris, & Glaser, 1998a,b) have used simulated discovery environments similar to those discussed here in the model-based analysis and reasoning in science (MARS) curriculum. In the MARS curriculum, experimentation skills (e.g., prediction, manipulation of variables) and evidence evaluation are used to foster an understanding of fundamental scientific concepts such as force and mass. Perhaps it is time to reconsider the core academic skills and focus on at least *four* “R’s”: reading, writing, arithmetic, and scientific *reasoning* and, moreover, to situate these skills within meaningful content domains.

REFERENCES

- Ahn, W., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition*, **54**, 299–352.
- Aikenhead, G. (1989). Scientific literacy and the twenty-first century. In C. K. Leong & B. S. Randhawa (Eds.), *Understanding literacy and cognition: Theory, research and application* (pp. 245–254). New York: Plenum.
- Almor, A., & Sloman, S. A. (1996). Is deontic reasoning special? *Psychological Review*, **103**, 374–380.
- American Association for the Advancement of Science (1990). *Science for all Americans: Project 2061*. New York: Oxford Univ. Press.
- Amsel, E., & Brock, S. (1996). The development of evidence evaluation skills. *Cognitive Development*, **11**, 523–550.
- Anderson, J. R. (1993). Problem solving and learning. *American Psychologist*, **48**, 35–44.
- Baillargeon, R., Kotovsky, L., & Needham, A. (1995). The acquisition of physical knowledge in infancy. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 79–116). Oxford: Clarendon Press.
- Baron, J. (1994). *Thinking and deciding*. New York: Cambridge Univ. Press.
- Bassoff, E. (1985). Neglecting the negative: Shortcomings in reasoning. *Journal of Counseling & Development*, **63**, 368–371.
- Bauer, H. H. (1992). *Scientific literacy and the myth of the scientific method*. Urbana, IL: Univ. of Illinois Press.
- Bisanz, J., & LeFevre, J. (1990). Strategic and nonstrategic processing in the development of mathematical cognition. In D. F. Bjorkland (Ed.), *Children's strategies: Contemporary views of cognitive development* (pp. 213–244). Hillsdale, NJ: Erlbaum.
- Brewer, W. F., & Samarapungavan, A. (1991). Children's theories vs. scientific theories: Differences in reasoning or differences in knowledge. In R. R. Hoffman & D. S. Palermo (Eds.), *Cognition and the symbolic processes* (pp. 209–232). Hillsdale, NJ: Erlbaum.
- Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. New York: Wiley.
- 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 to see 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.
- Case, R. (1974). Structures and strictures: Some functional limitations on the course of cognitive growth. *Cognitive Psychology*, **6**, 544–573.
- Chen, Z., & Klahr, D. (1999). All other things being equal: Children's acquisition of the control of variables strategy. *Child Development*.

- Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, **17**, 391–416.
- Cheng, P. W., & Holyoak, K. J. (1989). On the natural selection of reasoning theories. *Cognition*, **33**, 285–313.
- Chi, M. T. H., & Koeske, R. D. (1983). Network representations of children's dinosaur knowledge. *Developmental Psychology*, **19**, 29–39.
- Clement, J. (1983). A conceptual model discussed by Galileo and used intuitively by physics students. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 325–339). Hillsdale, NJ: Erlbaum.
- Copi, I. M. (1986). *Introduction to logic* (7th ed.). New York: Macmillan.
- Corrigan, R., & Denton, P. (1996). Causal understanding as a developmental primitive. *Developmental Review*, **16**, 162–202.
- Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, **31**, 187–276.
- Cummins, D. D. (1995). Naive theories and causal deduction. *Memory & Cognition*, **23**, 646–658.
- Cummins, D. D. (1996). Evidence of deontic reasoning in 3- and 4-year-old children. *Memory & Cognition*, **24**, 823–829.
- Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & Cognition*, **19**, 274–282.
- DeBoer, G. E. (1991). *A history of ideas in science education: Implications for practice*. New York: Teachers College Press.
- diSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and Instruction*, **10**, 105–225.
- 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. (1998). Problem solving. In W. Bechtel & G. Graham (Eds.), *A companion to cognitive science*. Malden, MA: Blackwell.
- Dunbar, K., & Klahr, D. (1989). Developmental differences in scientific discovery strategies. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 109–143). Hillsdale, NJ: Erlbaum.
- Eisenhart, M., Finkel, E., & Marion, S. F. (1996). Creating the conditions for scientific literacy: A re-examination. *American Educational Research Journal*, **33**, 261–295.
- Evans, J. St. B. T. (1993). The cognitive psychology of reasoning: An introduction. *Quarterly Journal of Experimental Psychology*, **46A**, 561–567.
- Evans, J. St. B. T., & Over, D. E. (1996). Rationality in the selection task: Epistemic utility versus uncertainty reduction. *Psychological Review*, **103**, 356–363.
- Farris, H. H., & Revlin, R. (1989). Sensible reasoning in two tasks: Rule discovery and hypothesis evaluation. *Memory & Cognition*, **17**, 221–232.
- Feist, G. J., & Gorman, M. E. (1998). The psychology of science: Review and integration of a nascent discipline. *Review of General Psychology*, **2**, 3–47.
- Flavell, J. H. (1963). *The developmental psychology of Jean Piaget*. Princeton, NJ: Von Nostrand.

- Gelman, S. A. (1996). Concepts and theories. In R. Gelman, & T. Kit-Fong Au (Eds.), *Perceptual and cognitive development: Handbook of perception and cognition* (2nd ed., pp. 117–150). San Diego, CA: Academic Press.
- Gentner, D., & Stevens, A. L. (Eds.) (1983). *Mental models*. Hillsdale, NJ: Erlbaum.
- Gholson, B., Shadish, W. R., Neimeyer, R. A., & Houts, A. C. (Eds.) (1989). *Psychology of science: Contributions to metascience*. Cambridge, MA: Cambridge Univ. Press.
- Gibson, E. J., & Levin H. (1975). *The psychology of reading*. Cambridge, MA: MIT Press.
- Giere, R. N. (1979). *Understanding scientific reasoning*. New York: Holt, Rinehart & Winston.
- Giroto, V. (1990). Biases in children's conditional reasoning. In J. Caverni, J. Fabre, & M. Gonzalez (Eds.), *Cognitive biases* (pp. 155–167). Amsterdam: North-Holland.
- Giroto, V., & Politzer, G. (1990). Conversational and world knowledge constraints on deductive reasoning. In J. Caverni, J. Fabre, & M. Gonzalez (Eds.), *Cognitive biases* (pp. 87–107). Amsterdam: North-Holland.
- Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist*, **39**, 93–104.
- Gorman, M. E. (1989). Error, falsification and scientific inference: An experimental investigation. *Quarterly Journal of Experimental Psychology*, **41A**, 385–412.
- Haberlandt, K. (1994). *Cognitive psychology*. Boston, MA: Allyn & Bacon.
- Hatano, G., & Inagaki, K. (1994). Young children's naive theory of biology. *Cognition*, **56**, 171–188.
- Hawkins, J., Pea, R. D., Glick, J., & Scribner, S., (1984). "Merds that laugh don't like mushrooms": Evidence for deductive reasoning by preschoolers. *Developmental Psychology*, **20**, 584–594.
- Hirschfeld, L. A., & Gelman, S. A. (Eds.) (1994). *Mapping the mind: Domain specificity in cognition and culture*. New York: Cambridge Univ. 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.
- Hume, D. (1758/1988). *An enquiry concerning human understanding*. Buffalo, NY: Prometheus Books.
- Hunt, E. (1994). Problem solving. In R. J. Sternberg (Ed.), *Thinking and problem solving* (pp. 215–232). San Diego, CA: Academic Press.
- Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.
- Kaiser, M., McCloskey, M., & Proffitt, D. (1986). Development of intuitive theories of motion. *Developmental Psychology*, **22**, 67–71.
- Kareev, Y., & Halberstadt, N. (1993). Evaluating negative tests and refutations in a rule discovery task. *Quarterly Journal of Experimental Psychology*, **46A**, 715–727.
- Karmiloff-Smith, A., & Inhelder, B. (1974). If you want to get ahead, get a theory. *Cognition*, **3**, 195–212.
- Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.
- Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist*, **28**, 107–128.
- Keys, C. W. (1994). The development of scientific reasoning skills in conjunction with collaborative writing assignments: An interpretive study of six ninth-grade students. *Journal of Research in Science Teaching*, **31**, 1003–1022.

- Klaczynski, P. A., Gelfand, H., & Reese, H. W. (1989). Transfer of conditional reasoning: Effects of explanations and initial problem types. *Memory & Cognition*, **17**, 208–220.
- Klahr, D. (1992). Information-processing approaches to cognitive development. In M. H. Bornstein, & M. E. Lamb (Eds.), *Developmental psychology: An advanced textbook (3rd ed.)*, pp. 273–336. Hillsdale, NJ: Erlbaum.
- Klahr, D. (1994). Searching for the cognition in cognitive models of science. *Psychology*, **5**(94), scientific-cognition.12.klahr.
- Klahr, D. (2000). *Exploring science: The cognition and development of discovery processes*. Cambridge, MA: MIT Press.
- Klahr, D., & Carver, S. M. (1995). Scientific thinking about scientific thinking. *Monographs of the Society for Research in Child Development*, Serial No. 245, **60**(40), 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., & Simon, H. A. (1999). Studies of scientific discovery: Complementary approaches and convergent findings. *Psychological Bulletin*.
- Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis-testing. *Psychological Review*, **94**, 211–228.
- Klayman, J., & Ha, Y. (1989). Hypothesis testing in rule discovery: Strategy, structure, and content. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 596–604.
- Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*. Cambridge, MA: MIT Press.
- Koslowski, B., & Maqueda, M. (1993). What is confirmation bias and when do people actually have it? *Merrill-Palmer Quarterly*, **39**, 104–130.
- Koslowski, B., & Okagaki, L. (1986). Non-Humean indices of causation in problem-solving situations: Causal mechanisms, analogous effects, and the status of rival alternative accounts. *Child Development*, **57**, 1100–1108.
- Koslowski, B., Okagaki, L., Lorenz, C., & Umbach, D. (1989). When covariation is not enough: The role of causal mechanism, sampling method, and sample size in causal reasoning. *Child Development*, **60**, 1316–1327.
- Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, **96**, 674–689.
- Kuhn, D. (1991). *The skills of argument*. New York: Cambridge Univ. Press.
- Kuhn, D. (1993a). Science as argument: Implications for teaching and learning scientific thinking. *Science Education*, **77**, 319–337.
- Kuhn, D. (1993b). Connecting scientific and informal reasoning. *Merrill-Palmer Quarterly*, **39**, 74–103.
- Kuhn, D. (1995). Migenetic study of change: What has it told us? *Psychological Science*, **6**, 133–139.
- Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills*. Orlando, FL: Academic Press.
- Kuhn, D., Garcia-Mila, M., Zohar, A., & Andersen, C. (1995). Strategies of knowledge acquisition. *Monographs of the Society for Research in Child Development*, Serial No. 245, **60**(40), 1–128.

- Kuhn, D., & Phelps, E. (1982). The development of problem-solving strategies. In H. Reese (Ed.), *Advances in child development and behavior* (Vol. 17, pp. 1–44). New York: Academic Press.
- Kuhn, D., Schauble, L., & Garcia-Mila, M. (1992). Cross-domain development of scientific reasoning. *Cognition & Instruction*, **9**, 285–327.
- Langley, P., Simon, H. A., Bradshaw, G. L., & Zytkow, J. M. (1987). *Scientific discovery: Computational explorations of the creative processes*. Cambridge, MA: MIT Press.
- Lawson, A. E. (1983). The effects of causality, response alternatives, and context continuity on hypothesis testing reasoning. *Journal of Research in Science Teaching*, **20**, 297–310.
- Levin, I., Siegler, R. S., & Druyan, S. (1990). Misconceptions about motion: Development and training effects. *Child Development*, **61**, 1544–1557.
- Mahoney, M. J., & DeMonbreun, B. G. (1977). Psychology of the scientist: An analysis of problem-solving bias. *Cognitive Therapy and Research*, **1**, 229–238.
- Markovits, H., & Vachon, R. (1990). Conditional reasoning, representation, and level of abstraction. *Developmental Psychology*, **26**, 942–951.
- McCloskey, M. (1983). Naive theories of motion. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 299–324). Hillsdale, NJ: Erlbaum.
- Miller, J. D. (1983). Scientific literacy: A conceptual and empirical review. *Daedalus*, **112**, 29–48.
- Miller, J. L., & Bartsch, K. (1997). The development of biological explanation: Are children vitalists? *Developmental Psychology*, **33**, 156–164.
- Murphy, G., & Medin, D. (1985). The role of theories in conceptual coherence. *Psychological Review*, **92**, 289–316.
- Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1978). Consequences of information and disconfirmation in a simulated research environment. *Quarterly Journal of Experimental Psychology*, **30**, 395–406.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.
- Norris, S. P. (1995). Learning to live with scientific expertise: Towards a theory of intellectual communalism for guiding science teaching. *Science Education*, **79**, 201–217.
- Norris, S. P. (1997). Intellectual independence for nonscientists and other content-transcendent goals of science education. *Science Education*, **81**, 239–258.
- Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, **101**, 608–631.
- Oaksford, M., & Chater, N. (1996). Rational explanation of the selection task. *Psychological Review*, **103**, 381–391.
- O'Brien, D. P., Costa, G., & Overton, W. F. (1986). Evaluations of causal and conditional hypotheses. *Quarterly Journal of Experimental Psychology*, **38A**, 493–512.
- Okada, T., & Simon, H. A. (1997). Collaborative discovery in a scientific domain. *Cognitive Science*, **21**, 109–146.
- Overton, W. F., Ward, S. L., Noveck, I. A., Black, J., & O'Brien, D. P. (1987). Form and content in the development of deductive reasoning. *Developmental Psychology*, **23**, 22–30.
- Pauen, S. (1996). Children's reasoning about the interaction of forces. *Child Development*, **67**, 2728–2742.
- 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.
- Perfetti, C. A. (1992). The representation problem in reading acquisition. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 145–174). Hillsdale, NJ: Erlbaum.
- Pfundt, H., & Duit, R. (1988). *Bibliography: Students' alternative frameworks and science education (2nd ed.)*. Kiel: Inst. for Sci. Ed.
- Piaget, J. (1970). *Genetic epistemology*. New York: Columbia Univ. Press.
- Popper, K. R. (1959). *The logic of scientific discovery*. London: Hutchinson.
- Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Journal of Research in Science Teaching*, **66**, 211–227.
- Raghavan, K., & Glaser, R. (1995). Model-based analysis and reasoning in science: The MARS curriculum. *Science Education*, **79**, 37–61.
- Raghavan, K., Sartoris, M. L., & Glaser, R. (1998a). Impact of the MARS Curriculum: The mass unit. *Science Education*, **83**, 53–91.
- Raghavan, K., Sartoris, M. L., & Glaser, R. (1998b). Why does it go up? The impact of the MARS curriculum as revealed through changes in student explanations of a helium balloon. *Journal of Research in Science Teaching*, **35**, 547–567.
- 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.
- Ryan, A. G., & Aikenhead, G. S. (1992). Students' preconceptions about the epistemology of science. *Science Education*, **76**, 559–580.
- Samarapungavan, A., & Wiers, R. W. (1997). Children's thoughts on the origin of species: A study of explanatory coherence. *Cognitive Science*, **21**, 147–177.
- 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 scientific reasoning in knowledge-rich contexts. *Developmental Psychology*, **32**, 102–119.
- Schauble, L., & Glaser, R. (1990). Scientific thinking in children and adults. *Contributions to Human Development*, **21**, 9–27.
- Schauble, L., Glaser, R., Raghavan, K., & Reiner, M. (1991a). Causal models and experimentation strategies in scientific reasoning. *Journal of the Learning Sciences*, **1**, 201–238.
- Schauble, L., Klopfer, L. E., & Raghavan, K. (1991b). Students' transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching*, **28**, 859–882.
- Schauble, L., Glaser, R., Raghavan, K., & Reiner, M. (1992). The integration of knowledge and experimentation strategies in understanding a physical system. *Applied Cognitive Psychology*, **6**, 321–343.
- Schauble, L., Glaser, R., Duschl, R. A., & Schulze, S. (1995). Students' understanding of the objectives and procedures of experimentation in the science classroom. *Journal of the Learning Sciences*, **4**, 131–166.
- Schustack, M. W., & Sternberg, R. J. (1981). Evaluation of evidence in causal inference. *Journal of Experimental Psychology: General*, **110**, 101–120.

- Science Council of Canada. (1984). *Science for every student* (Report No. 36). Ottawa, ON: Science Council of Canada.
- Shadish, W. R., Houts, A. C., Gholson, B., & Neimeyer, R. A. (1989) The psychology of science: An introduction. In B. Gholson, W. R. Shadish, R. A. Neimeyer, & A. C. Houts (Eds.), *Psychology of science: Contributions to metascience* (pp. 1–16). Cambridge, MA: Cambridge Univ. Press.
- Shaklee, H., & Mims, M. (1981). Development of rule use in judgments of covariation between events. *Child Development*, **52**, 317–325.
- Shaklee, H., & Paszek, D. (1985). Covariation judgment: Systematic rule use in middle childhood. *Child Development*, **56**, 1229–1240.
- Shaklee, H., Holt, P., Elek, S., & Hall, L. (1988). Covariation judgment: Improving rule use among children, adolescents, and adults. *Child Development*, **59**, 755–768.
- Shamos, M. H. (1995). *The myth of scientific literacy*. New Brunswick, NJ: Rutgers Univ. Press.
- Shultz, T. R., Fisher, G. W., Pratt, C. C., & Rulf, S. (1986). Selection of causal rules. *Child Development*, **57**, 143–152.
- Shultz, T. R., & Mendelson, R. (1975). The use of covariation as a principle of causal analysis. *Child Development*, **46**, 394–399.
- Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology*, **8**, 481–520.
- Siegler, R. S. (1978). The origins of scientific reasoning. In R. S. Siegler (Ed.), *Children's thinking: What develops?* (pp. 109–149). Hillsdale, NJ: Erlbaum.
- Siegler, R. S. (1981). Developmental sequences within and between concepts. *Monographs for the Society for Research in Child Development*, **46** (whole No. 189).
- Siegler, R. S. (1995). Children's thinking: How does change occur? In F. E. Weinert & W. Schneider (Eds.), *Memory performance and competencies: Issues in growth and development* (pp. 405–430). Mahwah, NJ: Erlbaum.
- Siegler, R. S. (1998). *Children's thinking* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.
- Siegler, R. S., & Crowley, K. (1991). The microgenetic method: A direct means for studying cognitive development. *American Psychologist*, **46**, 606–620.
- 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.
- Siegler, R. S., & Shipley, C. (1995). Variation, selection, and cognitive change. In T. J. Simon & G. S. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 31–76). Hillsdale, NJ: Erlbaum.
- Siegler, R. S., & Shrager, J. (1984). Strategy choices in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *Origins of cognitive skill* (pp. 229–293). Hillsdale, NJ: Erlbaum.
- Siegler, R. S., & Thompson, D. R. (1998). "Hey, would you like a nice cold cup of lemonade on this hot day?": Children's understanding of economic causation. *Developmental Psychology*, **34**, 146–160.
- Simon, H. A. (1973). Does scientific discovery have a logic? *Philosophy of Science*, **40**, 471–480.
- Simon, H. A. (1986). Understanding the processes of science: The psychology of scientific discovery. In T. Gamelius (Ed.), *Progress in science and its social conditions* (pp. 159–170). Oxford: Pergamon Press.
- Simon, H. A. (1989). The scientist as problem solver. In D. Klahr & K. Kotovsky (Eds.),

- Complex information processing: The impact of Herbert A. Simon* (pp. 375–398). Hillsdale, NJ: Erlbaum.
- 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: Erlbaum.
- Slowiaczek, L. M., Klayman, J., Sherman, S. J., & Skov, R. B. (1992). Information selection and use in hypothesis testing: What is a good question, and what is a good answer? *Memory & Cognition*, **20**, 392–405.
- Sodian, B., Zaitchik, D., & Carey, S. (1991). Young children's differentiation of hypothetical beliefs from evidence. *Child Development*, **62**, 753–766.
- Sophian, C., & Huber, A. (1984). Early developments in children's causal judgments. *Child Development*, **55**, 512–526.
- Solomon, G. E. A., Johnson, S. C., Zaitchik, D., & Carey, S. (1996). Like father, like son: Young children's understanding of how and why offspring resemble their parents. *Child Development*, **67**, 151–171.
- Spelke, E. S., Phillips, A., & Woodward, A. L. (1995). Infants' knowledge of object motion and human action. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 44–78). Oxford: Clarendon Press.
- Sperber, D., Premack, D., & Premack, A. J. (Eds.) (1995). *Causal cognition: A multidisciplinary debate*. Oxford: Clarendon Press.
- Strauss, S. (1998). Cognitive development and science education: Toward a middle level model. In W. Damon (Series Ed.), I. E. Sigel, & K. A. Renninger (Vol. Eds.), *Handbook of child psychology: Volume 4, Child psychology in practice* (5th ed., pp. 357–399). New York: Wiley.
- Thagard, P. (1989). Explanatory coherence. *Behavioral & Brain Sciences*, **12**, 435–502.
- Tschirgi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development*, **51**, 1–10.
- Tukey, D. D. (1986). A philosophical and empirical analysis of subjects' modes of inquiry in Wason's 2-4-6 Task. *The Quarterly Journal of Experimental Psychology*, **38A**, 5–33.
- Tweney, R. D. (1991). Informal reasoning in science. In J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.), *Informal reasoning and education* (pp. 3–16). Hillsdale, NJ: Erlbaum.
- Tweney, R. D., Doherty, M. E., & Mynatt, C. R. (Eds.). (1981). *On scientific thinking*. New York: Columbia Univ. Press.
- VanLehn, K. (1989). Problem solving and cognitive skill acquisition. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 527–579). Cambridge, MA: MIT Press.
- Varnhagen, C. K. (1995). Children's spelling strategies. In V. W. Berninger (Ed.), *The varieties of orthographic knowledge II: Relationships to phonology, reading, and writing* (pp. 251–290). Dordrecht, Netherlands: Kluwer Academic.
- Varnhagen, C. K., McCallum, M., & Burstow, M. (1997). Is children's spelling naturally stage-like? *Reading and Writing: An Interdisciplinary Journal*, **9**, 451–481.
- Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, **24**, 535–585.
- Voss, J. F., Wiley, J., & Carretero, M. (1995). Acquiring intellectual skills. *Annual Review of Psychology*, **46**, 155–181.
- Ward, S. L., & Overton, W. F. (1990). Semantic familiarity, relevance, and the development of deductive reasoning. *Developmental Psychology*, **26**, 488–493.
- 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. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, **20**, 63–71.
- Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content*. London: B. T. Batsford.
- Wellman, H. M., & Gelman, S. A. (1992). Cognitive development: Foundational theories in core domains. *Annual Review of Psychology*, **43**, 337-375.
- Wharton, C. M., Cheng, P. W., & Wickens, T. D. (1993). Hypothesis-testing strategies: Why two goals are better than one. *Quarterly Journal of Experimental Psychology*, **46A**, 743–758.
- White, P. A. (1988). Causal processing: Origins and development. *Psychological Bulletin*, **104**, 36–52.
- Wolpert, L. (1993). *The unnatural nature of science*. London: Faber & Faber.

Received: July 30, 1998; revised: December 14, 1998, June 7, 1999