Psychological Review
1989, Vol. 96, No. 4, 674–689

0033-295X/89/$00.75

Children and Adults as Intuitive Scientists

Deanna Kuhn
Teachers College Columbia University

The metaphor of children and lay adults as intuitive scientists has gained wide acceptance. Although useful in one sense, pertaining to scientific understanding, in another, pertaining to the process of scientific thinking, the metaphor may be fundamentally misleading. Research is reviewed indicating that processes of scientific thinking differ significantly in children, lay adults, and scientists. Hence, it is the instruments of scientific thinking, not just the products, that undergo "strong restructuring" (Carey, 1986). A framework for conceptualizing development of scientific thinking processes is proposed, centering on progressive differentiation and coordination of theory and evidence. This development is metacognitive, as well as strategic. It requires thinking about theories, rather than merely with them, and thinking about evidence, rather than merely being influenced by it, and, hence, reflects the attainment of control over the interaction of theories and evidence in one's own thinking.

The metaphor of the lay adult—or the child—as an intuitive scientist has gained wide acceptance in the last decade. As the scientist explores the environment, constructs models as a basis for understanding it, and revises those models as new evidence is generated, so do lay people endeavor to make sense of their environments by processing data and constructing mental models based on these data. The highly influential volume by Nisbett and Ross (1980) and seminal research by Tversky and Kahneman that is highlighted in the Nisbett and Ross book have done much to promote the metaphor. More recently, it has been promoted as well by the literature on scientific understanding and conceptual change, discussed later, and to an extent by the general theory of induction proposed by Holland, Holyoak, Nisbett, and Thagard (1986), a theory whose principles are intended to apply to all forms of induction from the very simplest forms of concept formation to the thinking of scientists. As the research summarized by Nisbett and Ross documents, the particular inference rules that the intuitive scientist uses to interpret evidence and make inductive inferences are likely to be faulty. The intuitive scientist, for example, makes inferences based on insufficient sample size and overlooks base rates. Nevertheless, the process and goal of this intuitive scientific activity are thought to be analogous in the layperson and the professional scientist.

In this article, I have a particular concern with the thinking of children and the ways in which it may or may not resemble that of scientists. I therefore focus on the metaphor of child as scientist, although I consider as well the thinking of lay adults and its resemblance to that of scientists. I explore two quite different senses in which this metaphor might be taken and conclude that, although it may be useful and productive in one sense, the child-as-scientist metaphor in another sense may be fundamentally misleading.

Part of the research described in this article was supported by the Spencer Foundation.

Correspondence concerning this article should be addressed to Deanna Kuhn, Department of Psychology, Teachers College Columbia University, New York, New York 10027.

What is Scientific Thinking?

The usefulness of our conceiving of children or lay adults as intuitive scientists is rendered problematic from the start by lack of clarity regarding what it means to think scientifically. At one end of a broad spectrum is the view that scientific thinking is a mode of exploring and coming to know the world that is within the competence of very young children. At the other end of this spectrum is the view that scientific thinking makes cognitive demands that even professional scientists may be unable to fulfill (Faust, 1984). The present article offers some reconciliation of these contrasting characterizations by examining senses in which both views may be correct.

The view taken in this article is that the heart of scientific thinking is the coordination of theories and evidence. A central premise underlying science is that scientific theories stand in relation to actual or potential bodies of evidence against which they can be evaluated. Reciprocally, scientific "facts" stand in relation to one or more actual or potential theories that offer a vehicle for their organization and interpretation. No strong claims need be made here regarding the range of thinking processes that professional scientists actually use as they think about scientific problems, on the basis of the limited evidence available on this subject (Mahoney & Kimper, 1976). Nevertheless, some fundamental competencies of the scientist are clearly assumed, competencies that center on the coordination of theories and evidence. Although nonconscious, associative processes may play a role in a scientist's generation of ideas, it is by means of the former competencies that scientists reconcile their ideas with evidence and justify them to the community. The scientist (a) is able to consciously articulate a theory that he or she accepts, (b) knows what evidence does and could support it and what evidence does or would contradict it, and (c) is able to justify why the coordination of available theories and evidence has led him or her to accept that theory and reject others purporting to account for the same phenomena. Although they do not encompass all aspects of scientific thinking, these skills in coordinating theories and evidence arguably are the most central, essential, and general skills that define scientific thinking. Moreover, a high level of mastery of such skills is assumed not only in formal scientific inquiry but in a wide range of other highly skilled endeavors, such as medical diagnosis (Elstein, Shulman, & Sprafka, 1978) and legal practice (Kassin & Wrightsman, 1985).

The major concern of the present article is to examine the extent to which such skills are exhibited by children and lay adults in thinking outside of formal scientific contexts. Do children and lay adults in these respects think like scientists?

Scientific Thinking as Scientific Understanding

The most recent influential stream of research to promote the child-as-scientist metaphor has been the work on scientific understanding and conceptual change, and it clearly illustrates use of the metaphor in the first of the two senses that I consider. It portrays the child as a constructor of scientific theories and, as such, is not contradictory to the view of scientific thinking presented in the preceding section, although I claim below that it focuses only on one part of this process. Research by both cognitive psychologists and science educators has shown that children and adults hold a variety of naive, intuitive conceptions—usually misconceptions—about how the world works (Carey, 1985a, 1985b, 1986; Champagne & Klopfer, 1984; DiSessa, 1983; Gentner & Gentner, 1983; Larkin, 1983; McCloskey, 1983; Vosniadou & Brewer, 1987; West & Pines, 1985). These conceptions, although they are wrong, have been shown to be powerful and remarkably resistant to instruction, so that science educators must conceive of their task as making contact with these incorrect conceptions and working to modify them, rather than superimposing new, correct concepts. These mental models, as they are referred to, are more than the scripts (of event sequences) that Nelson (1985) has investigated, for they contain at least some element of explanation of how key elements in the model relate to one another and how phenomena operate. Multiple, inconsistent models of the same phenomenon are likely to co-exist within an individual and be activated in different contexts (Williams, Hollan, & Stevens, 1983).

Following from this mental model view of scientific thinking is the proposition that the development of scientific understanding consists of a progression of partially correct theories within individual conceptual domains, a progression that some theorists such as Glaser (1984) have suggested is the heart of cognitive development. A number of researchers have identified progressions of this sort for particular scientific concepts (Carey, 1985b; Kaiser, McCloskey, & Proffitt, 1986; Krupa, Selman, & Jacquette, 1985; Strauss & Stavy, 1982). Kaiser et al. (1986), for example, have shown that although preschool children most often correctly predict a straight path for a ball exiting a curved tube, school-age children often predict a curved path, which Kaiser et al. attribute to a *persistence of motion* theory. By adulthood, this theory typically is revised to acknowledge that curvilinear motion requires a continuing external force.

Significant in promoting the child-as-scientist metaphor in this first sense has been the explicit parallel drawn by Carey (1986), Gruber (1973), and others between the progression of scientific understanding in the child and the progression of scientific understanding in the history of science. In particular, the child's development of understanding has been likened to T. Kuhn's (1962) account of the history of science as a succession of paradigms not directly commensurate with one another, each prevailing (despite discrepant evidence) until a replacement is available. T. Kuhn's characterization of paradigms and paradigm shifts provides the foundation for Carey's (1986) distinction between strong and weak restructuring in conceptual change. The weaker sense of restructuring entails new relations among concepts and the addition of new, previously absent concepts. Restructuring in the strong sense, which is equivalent to Kuhn's paradigm shift and is the more important kind of restructuring to occur in conceptual change, entails not only these kinds of change but changes in the core concepts of the theory and their interrelation, so that the core concepts in the new theory may not be directly translatable into those of the old theory.

One sense in which the child-as-scientist metaphor can be taken, then, is as a parallel with respect to scientific understanding: Both child and scientist gain understanding of the world through construction and revision of a succession of models, or paradigms, that replace one another. But what of the *process* in terms of which these models, or theories, are revised (or, indeed, are constructed in the first place)? If the revision of theories is the heart of cognitive development, as Glaser (1984) has suggested, we ought to know something about how it occurs. And might not this process sense be another one in which the child-as-scientist metaphor should be examined? Are the processes in terms of which the child, the lay adult, and the scientist go about exploring the world, generating and interpreting the data that will inform their mental models, comparable?

Conceptual change researchers have had little to say about how successive understandings are transformed from one form to another (or how they were constructed initially)—that is, about the process by which theories are constructed and revised as a way of learning about the world. In justifying her approach, Carey (1985b) legitimately has claimed that the task of identifying and describing such sequences of understandings precedes any investigation of mechanism, and she has focused her work on very content-rich, complex domains in which characterization of the conceptual content of successive theories requires detailed, precise description and tends to overshadow any questions of mechanism. Carey (1985b) goes on to argue, however, that many of the developments that researchers have identified—developments in the child's way of knowing the world—were mistakenly identified and can in fact be explained in terms of the succession of conceptual changes within particular domains that her work and that of the researchers cited earlier have examined. Although she does not deny the possibility that such changes in the child's *cognitive machinery,* as she has termed it, occur, Carey (1985a, 1985b) has claimed that there exists no firm evidence for such changes, changes that cannot more readily be accounted for in terms of conceptual change within particular content domains. A similar view has been expressed by Keil (1984).

Evidence for Strong Restructuring in the Processes of Scientific Thinking

Recently, however, this picture has begun to change, as several investigators, including myself and my coworkers (Dunbar & Klahr, 1988; Karmiloff-Smith, 1984, 1986, 1988; Klahr & Dunbar, 1988; D. Kuhn, Amsel, & O'Loughlin, 1988; Schauble (in press); Schauble & Kuhn, 1989; Shute, Glaser, & Raghavan, 1989; Voss, Blais, Means, Greene, & Ahwesh, 1986), have investigated the processes of scientific thinking and obtained results to suggest that these processes are significantly different in the child, the lay adult, and the scientist. On the basis of such results, it is in a second sense—the sense of process of scientific thinking rather than understanding of scientific phenomena—that I shall argue that the child-as-scientist metaphor is misleading.

If the processes of scientific thinking displayed by children, adults, and scientists differ, a developmental framework for conceptualizing these differences is likely to be useful. Most desirable would be an approach that takes the developing knowledge structures studied by Carey (1985b) and others into account, without foregoing the search for strategy change as well (Chi & Ceci, 1987; Glaser, 1984). Such an approach is reflected in the conceptual framework and research described in this article. In recent work (Kuhn, 1989; Kuhn et al., 1988; Schauble (in press); Schauble & Kuhn, 1989), my coworkers and I have taken the subject's own theories as a starting point, in this respect aligning our work with that of the conceptual change researchers, such as Carey (1985b). Rather than focusing on the theories themselves, however, we have examined the process by which theories are revised in the course of encounters with new evidence, and our results indicate that this process itself undergoes change.

Adults in these studies included undergraduate students, nonacademic young adults attending a business training institute, older adults of mixed education level, and a small number of experts, either professional scientists or advanced Phd candidates in philosophy (the latter were regarded as experts in reasoning). Children were third, sixth, and ninth graders. In the initial series of studies reported by Kuhn et al. (1988), subjects' theories were assessed and they were then asked to generate and evaluate various forms of evidence, sometimes consonant with and sometimes conflicting with their own theories.

It was anticipated that theoretical beliefs would affect the evaluation and generation of evidence. The organizing influence of theoretical concepts on forms of cognition ranging from simple categorization to complex scientific thought has by now been widely acknowledged (Alloy & Tabachnik, 1984; Fischhoff & Beyth-Marom, 1983; Holland et al., 1986; Murphy & Medin, 1985; Neisser, 1987). Yet, my central claim will be that this influence operates in distinctly different ways in the child and the scientist.

Differentiating Theory and Evidence

The research focus that my coworkers and I adopted was less on documenting the relevance of theoretical belief than on examining in a microgenetic way how a subject attempted to reconcile his or her theory with accumulating discrepant evidence. However, the cases in which evidence and theory were congruent turned out to be equally revealing. Peter, a sixth grader, provides an example. In one of the studies, subjects evaluated graphically presented evidence depicting covariation or noncovariation of various foods that children at a hypothetical boarding school ate and their susceptibility to colds. The evidence was presented one instance at a time, cumulatively, until eight instances were on display together. When asked to evaluate the first instance of what would be covariation evidence, Peter first made a theory-based response with respect to the cake variable (chocolate or carrot), which he believed was causally implicated in catching colds:

> (Does the kind of cake make a difference?) Yes. Carrot cake is made with carrots, and chocolate cake is made with a lot of sugar. This [carrot cake] is made with some sugar too, but it's made with less sugar.

The interviewer then posed the evidence-focus probe, designed to direct the subject's attention to the presented evidence: "Do the *findings of the scientists* [italics added for verbal emphasis] show that the kind of cake makes a difference, doesn't make a difference, or can't you tell what the scientists' findings show?" In response, Peter merely elaborated his theory:

> Less sugar means your blood pressure doesn't go up. It makes a difference.

After the second instance of evidence was presented, Peter first reiterated the theory but then, in response to the evidence-focus probe, did finally refer to the evidence:

> (Do the *findings of the scientists* show it makes a difference . . . ?) Yes. Because these [children] are like "ugghh" with tissues [the children held to their noses], and children at table [instance] one have no tissues. [Children at table one ate carrot cake and those at table two ate chocolate cake.]

One might expect that having recognized and interpreted the fact that the evidence reflected covariation, Peter would continue to refer to the presented evidence, at least in his responses to the evidence-focus probe (Do the *findings of the scientists* show . . . ?). Following presentation of the third instance, however, he first reiterated the theory and then, in response to the evidence-focus probe, simply repeated the theory again:

> (Do the *findings of the scientists* show it makes a difference . . . ?) Yes, because it [chocolate cake] has a lot of sugar and a lot of bad stuff in it.

Over the next few instances, Peter again noted the covariation several times. Yet after the addition of the seventh instance, with the set of seven reflecting perfect covariation between variable and outcome, he again substituted a reiteration of his theory for evaluation of the evidence:

> (Do the *findings of the scientists* show it makes a difference . . . ?) Yes. Because the sugar; [it's] not [in] carrot cake.

Peter's theory of the relation between kind of cake and colds was completely compatible with the covariation evidence he was asked to evaluate. Yet the sequence of his responses is curious. Especially because they are compatible, perhaps, he appears not to clearly distinguish theory and evidence, responding to a request to evaluate the evidence with a reiteration of his theory even *after* he has attended to and interpreted the evidence. Both theory and evidence point to the same conclusion, and one thus seems to be the same as the other for purposes of justifying the conclusion. Put differently, a subject like Peter appears not to differentiate the different sources of support for his beliefs.

This vacillation between theory and evidence as the basis for justifying judgments was common among sixth graders, declined by adulthood, and never occurred among experts (Kuhn et al., 1988). Before accepting this response pattern as attributable to limited differentiation between theory and evidence, several control studies were performed to eliminate alternative explanations. In one, subjects were explicitly instructed to disregard their own beliefs and to consider only the evidence before them. This condition produced only a slight (nonsignificant) reduction of theory-based responses among sixth graders and no change among ninth graders (Kuhn et al., 1988). Thus, failure to convey adequately to subjects what they were being asked to do cannot be regarded as an explanation for their performance.

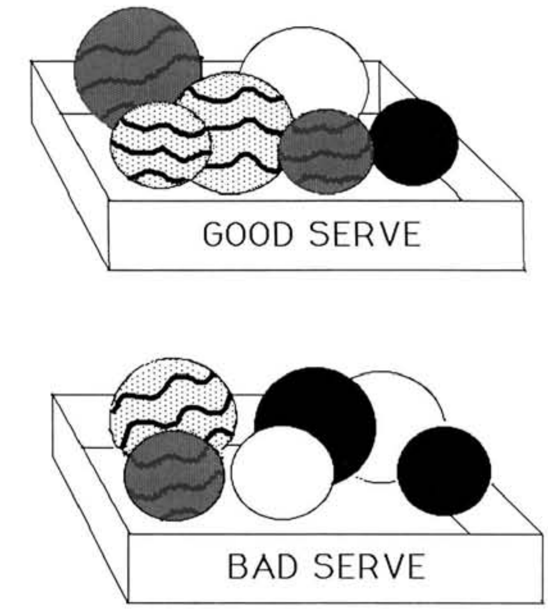

Figure 1. 4/2 evidence in the balls problem. (From *The Development of Scientific Thinking Skills* [p. 141] by D. Kuhn, E. Amsel, & M. O'Loughlin, 1988, Orlando, FL: Academic Press. Copyright 1988 by Academic Press. Reprinted by permission.)

Adjusting Evidence to Fit Theories

Consider now what happens when theory and evidence are at odds with one another. Doesn't this discrepancy force the subject into a clearer distinction between the two? The answer, in a word, is *no.* A ninth grader, Laura, for example, theorized that kind of relish was causal and kind of candy bar was not. When presented evidence depicting a pattern of noncovariation between both variables and outcome, she interpreted the evidence with respect to candy bar as follows:

> (Does the kind of candy bar make a difference?) With the Mars Bar you get a cold off and on, because here's one they got colds and over here they didn't. . . so it really doesn't matter.

The identical evidence with respect to relish, however, she interpreted like this:

> Yes [it makes a difference]. Mostly likely all the time you get a cold with the mustard. Like there you did [instance 2] and there you did [instance 7].

Laura ignored, of course, the equal number of cases in which mustard co-occurred with no colds.

Another in this series of studies had to do with the effects of features of a set of sports balls on the quality of a player's serve. Evidence was portrayed by the actual balls placed in baskets labeled *Good serve* and *Bad serve* (see the example in Figure 1), and the subject was asked to relate the evidence to two different theories—that of a Mr. (or Ms.) S, for example, who believed size makes a difference, and that of a Mr. (or Ms.) C, who believed color makes a difference. Two (of four possible) dimensions selected for questioning were (a) one the subject had earlier identified as one that he or she believed made a difference and (b) another that he or she believed made no difference.

Subjects' evaluations of minimal, insufficient evidence were especially interesting. For example, the following is the response of a third grader, Allen, who theorized that size was causal (with large balls yielding good serves and small balls yielding bad serves) and color was noncausal; he was shown a single large, light-colored ball in the *Good* basket (and no balls in the *Bad* basket).

> (Do these results help more to show that one person is right . . . ?) Mr. Size would win. (Why?) Because this ball is big. And it came out good. (Do these results prove that Mr. S is right?) Yes. (What do these results have to say about Mr. C's view?) He loses. (Why?) Because the color doesn't really matter. (And this ball coming out good, what does that say about Mr. C's view?) That says that the color doesn't matter.

Similar theoretical bias is evident in the response of a noncollege young adult, Matt, to the more extended evidence portrayed in Figure 1. Matt theorized that texture was causal (with rough balls yielding good serves and smooth balls yielding bad serves) and ridges were noncausal. As seen in Figure 1, the evidence reflects a very slight (4/2), and identical, association of both texture and ridges to outcome (although, in fact, the covariation of the two variables with one another precludes any causal inference).

> (Do these results help more to show that one person is right . . . ?) Texture, because you have more balls that have smooth texture that came out with bad serves than you do balls that have rough texture and bad serves. (Do these results prove that Mr. T is right?) Yes, because the balls with smooth texture, large and small, have bad serves. (What do these results have to say about Mr. R's view?) It's not showing nothing about ridges. (Why not?) Because you have balls that have ridges that have bad serves and balls that have ridges that have good serves.

Matt thus applied two different inference strategies to identical evidence (one focused on covariation as the basis for a causal inference and one focused on noncovariation as the basis for a noncausal inference), in a way that served his theoretical beliefs.

As these examples illustrate, subjects commonly exhibited strategies that served to bring theory and evidence into alignment with one another. A major, and anticipated, strategy was biased evaluation of the evidence to reduce its inconsistency with theory (although it was a strategy that was hard to maintain as discrepant evidence mounted). Subjects either failed to acknowledge discrepant evidence or attended to it in a selective, distorting manner. Identical evidence was interpreted one way in relation to a favored theory and another way in relation to a theory that was not favored, which suggests that the evidence is not sufficiently differentiated from the theory itself; it does not retain its own identity—its constancy of meaning—across a range of theories to which it might be related.

Now, with the Bayesian statistical model in mind, one might immediately object that it is perfectly rational, or correct, to treat identical evidence differently in the context of different theoretical beliefs. An assumption implicit in such a concept, however, is that an individual *chooses* to adjust the evaluation of evidence to take into account prior belief. If the individual wished to forego this adjustment and interpret the evidence independent of prior beliefs, he or she could do so. In sharp contrast, the theoretical beliefs of many of our subjects appeared to color their evaluation of evidence in ways outside their conscious control.

Several other kinds of evidence from these studies are consistent with this picture. A similar phenomenon occurred when subjects were asked to generate evidence that would show a theory to be correct or incorrect (a result not predicted by Bayesian theory). Subjects were often unable to generate evidence that did not accord with their own theories. For example, Freda, a noncollege adult, had no difficulty arranging the balls in a pattern of covariation with outcome for the texture variable, which she believed causal. When asked to generate the same evidence to show that the color variable makes a difference, she placed both light- and dark-colored balls in the *Good* basket and offered this explanation:

> It would make a difference because one is light and one is dark. (How does this show that color makes a difference?) Because one is lighter and one is darker and both came out good.

Clearly, Freda's theoretical belief compromised her ability to reason about the meaning of evidence. In offering explanations of the evidence they generated, even when it did not conflict with their theories, subjects showed the same confusion between theory and evidence as Peter's performance illustrated in the case of evidence evaluation. For example, when asked to arrange the evidence to show that her theory that texture makes a difference was correct, a ninth grader placed eight smooth balls in the *Good* basket and eight rough balls in the *Bad* basket. When asked to explain how this arrangement showed that texture made a difference, she responded:

> The rough texture will make the ball heavier so it won't go so far when hit.

Adjusting Theories to Fit Evidence

Further support for our interpretation comes from the fact that "adjustment" of evidence to fit a theory was only one of the strategies subjects displayed to maintain alignment between theory and evidence. The other was the adjustment of theory, to reduce its inconsistency with evidence. In the case of evidence generation, this took the form of modifying one's own theoretical belief to match that of the generated evidence (even though there was no requirement to do so). For example, a ninth grader, Melissa, initially declined to generate evidence that demonstrated correctness of the opposing theory (that color was effective):

> I would say that either of them [dark- or light-colored balls] would come out good.

The interviewer then repeated the task instruction, and Melissa correctly placed eight light-colored balls in the *Bad* basket and eight dark-colored balls in the *Good* basket.

> (Can you explain how this proves that color makes a difference?) These [dark in *Good* basket] are more visible in the air. You could see them better.

Subjects commonly exhibited this need to explain why the evidence they had just generated was plausible, or sensible, even though they had produced the evidence in response to an explicit instruction to generate evidence supporting an opposing theory and not as evidence that they believed to be true.

In the case of evidence evaluation, what was most noteworthy about adjustment of theory to fit the evidence was the likelihood that it occurred without the subject's awareness. For example, suppose a subject held the theory that type of cola (diet or regular) was unrelated to colds, and covariation evidence suggesting such an association was presented. After several responses in which this noncausal theory was expressed without acknowledgment of the evidence, a subject might again offer a theory-based response, but this time voice a new theory espousing a causal connection between kind of cola and outcome. Only with this new theory in place would the subject then acknowledge and interpret the covariation evidence. A similar pattern occurred in the case of an initial causal theory and noncovariation evidence. Such subjects appeared unwilling to acknowledge the implications of evidence unless they had a compatible theory in place that provided an explanation of this evidence.

Why, it might be asked, should someone need to discard their own very plausible theories that kind of cola or relish have nothing to do with getting colds and formulate new, often implausible theories about the relation of these variables to colds, before they are willing to acknowledge evidence showing covariation between these variables and colds? 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? The answer may be that doing so would leave theory and evidence not in alignment with one another and therefore needing to be recognized as distinct entities.

One other form of evidence for this interpretation comes from reconstruction data. Subjects' representations of both the evidence and their own theories were probed by asking them to recall each, following evidence evaluation. Subjects often recalled their own original theories inaccurately, representing them as consistent with the evidence that had been presented. Or, in physically reconstructing the evidence from available materials, they represented it as more consistent with their theories than it in fact had been. Both of these tendencies reflect additional mechanisms for maintaining theory and evidence in alignment with one another.

Patterns of Performance in Reconciling Theory and Evidence

Figure 2 summarizes patterns of response to theory-discrepant covariation evidence for the 35 sixth graders, 35 ninth graders, 20 average adults, and 5 philosophy graduates who participated in the study about foods and colds. (Data for 5 sixth graders, 3 ninth graders, and 3 adults whose theoretical beliefs vacillated are omitted from Figure 2.) Because number of subjects varied across groups, percentages (relative to the immediately preceding division) are included beneath the simple frequencies. As reflected in the initial column of Figure 2, roughly one third of the subjects generated a new theory to fit the evidence, most often before acknowledging the evidence. The other two thirds maintained their theories but, except for one ninth grader, one adult, and the philosophers, distorted the evidence to better fit the theory. The final column shows how these subjects resolved the discrepancy as the discrepant covariation evidence accumulated: A few acknowledged the mismatch between their own theories and the portrayed evidence, and about a third of the youngest subjects achieved no resolution, but most subjects, after initially attempting to distort the evidence, finally set aside their theories (i.e., made no further reference to them, giving only evidence-based responses) and acknowledged the covariation reflected in the evidence. These same resolution patterns occurred among subjects who constructed new theories but, as shown in Figure 2, earlier theories often resurfaced and sometimes no resolution was achieved.

Response patterns for discrepant noncovariation evidence (not shown) were similar; the major difference was that subjects were less likely to construct new theories (only 21% did so) and were more likely to reduce the discrepancy by maintaining their theories and distorting the evidence. Thus, subjects of all ages are more likely to resist the implications of noncovariation evidence showing their causal theories to be wrong than they are to resist the implications of covariation evidence showing their noncausal theories to be wrong.

Figure 2 does not provide information regarding the extent of bias displayed by subjects who distorted evidence. Some indication of this is provided by Table 1, for the 100 subjects (20 in each group) who participated in the sports balls study. Table 1 summarizes responses to 5/1 evidence—evidence in which 10 of the 12 balls displayed conform to a covariation pattern. Each subject made four different evaluations of 5/1 evidence, two regarding a variable the subject believed causal (subject's theory) and two regarding a variable the subject believed noncausal (other's theory). Reported in Table 1 are mean frequencies (of a possible 2.0) of evidence-based responses, inclusion (causal inference) responses, and exclusion (noncausal inference) responses, as a function of subject's belief. As reflected in the figures in the final row (total sample), subjects were more likely to acknowledge covariation evidence (make an evidence-based response) if they held a causal theory than they were if they held a noncausal theory, were more likely to evaluate covariation evidence as indicating causality (inclusion) if they held a causal theory, and were slightly (nonsignificantly) more likely to evaluate covariation evidence as indicating noncausality (exclusion) if they held a noncausal theory. Statistical analyses of the data in Table 1 showed, for evidence-based responses, a significant effect of theory (subject's vs. other's) and significant interaction of theory and age group (with effects of theory minimal among ninth graders and college adults). Effects of theory and age–theory interaction were likewise significant for inclusion responses; no effects reached significance for exclusion responses. Comparable results were obtained for the other forms of evidence included in the study. An important exception, however, are more complex forms of asymmetric evidence in which frequencies of neither the two outcomes nor the two variable levels are fixed in a symmetrical 6/6 ratio (making the problems equivalent to the correlation problems studied by Inhelder & Piaget, 1958, and more recently by Shaklee & Paszek, 1985, and others). In this case, college adults also succumb to the theory bias they resisted for simpler forms of evidence.

Development of Scientific Thinking as Progress in the Coordination of Theory and Evidence

Taken together, the findings that have been described suggest limitations in the differentiation, and hence coordination, of theory and evidence. When theory and evidence are compatible, there is a melding of the two into a single representation of "the way things are." The pieces of evidence are regarded not as independent of the theory and bearing on it, but more as *instances* of the theory that serve to illustrate it. The theory, in turn, serves to explain the evidence, to make sense of it. In other words, there is no concept of evidence as standing apart from the theory and bearing on it. In responding to a request to evaluate the evidence, articulating the theory is thus as good as making reference to the evidence. When theory and evidence are discrepant, subjects use a variety of devices to bring them into alignment: either adjusting the theory—typically prior to acknowledging the evidence—or "adjusting" the evidence by ignoring it or by attending to it in a selective, distorting manner.

The complete fusion of theory and evidence, of course, represents the extreme, reflected primarily in our younger subjects. Younger subjects were less likely than older ones to distinguish firmly between theory and evidence, and also, given that they did experience a conflict between the two, were less likely to be able to resolve it. As notable as this developmental change, however, is the presence in adults of all of the characteristics that have been described. The performance of college subjects was superior to that of noncollege adults, but even they showed the characteristics that have been described as soon as more complex problems were introduced. In contrast, none of these characteristics appeared in the performance of expert subjects.

What are the skills in coordinating theories and evidence that such subjects lack? The ability to evaluate the bearing of evidence on a theory at a minimum requires first that the evidence be encoded and represented separately from a representation of the theory. If new evidence is merely assimilated to a theory, as an instance of it, the possibility of constructing relations between the two as separate entities is lost. Second, the subject must represent the theory itself as an object of cognition, that is, think about the theory rather than with it (Moshman, 1979); otherwise, evidence cannot be evaluated in relation to it. Third, and paradoxically, coordination of theory and evidence requires temporary bracketing, that is, disregarding or setting aside one's acceptance of the theory, in order to assess what the evidence by itself would mean for the theory, were it the only basis for making a judgment.

The ability to reflect on one's own thought is a uniquely human capacity and one that figures importantly in many theories of cognition and its development (Piaget, 1950; Sternberg, 1985; Vygotsky, 1962). Although it has been used in various and sometimes poorly specified ways, the term *metacognition* can be used as a label for this capacity. Sternberg (1985) does not address development of the capacity but sees metacognitive control of one's cognitive actions as the most important component of intelligence. For Piaget (1950), ability to reflect on one's cognitive actions (operate on one's mental operations) marks

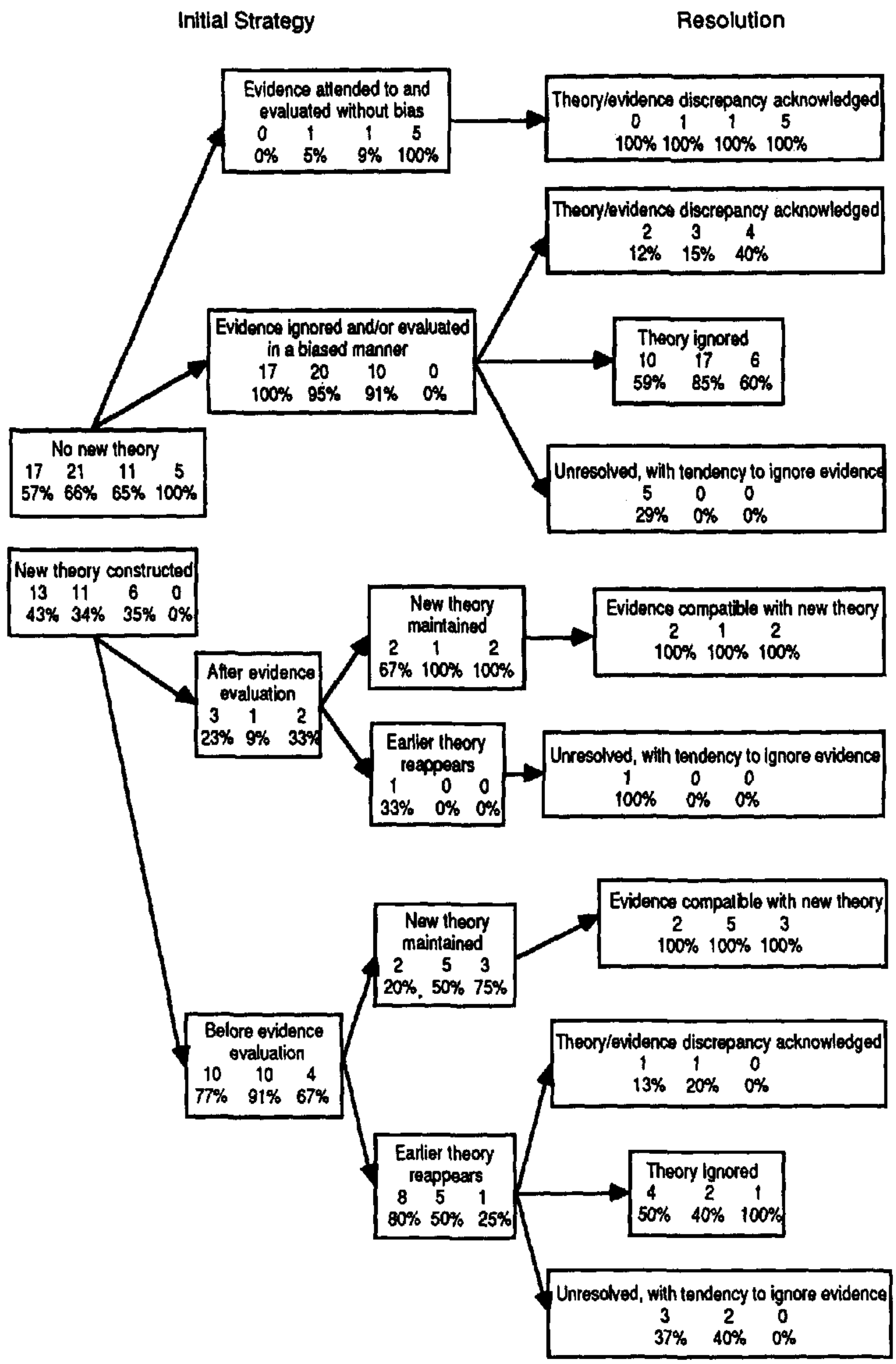

Figure 2. Patterns of performance in reconciling theories with discrepant covariation evidence. (Data for sixth graders appear first, ninth graders second, adults third, and philosophers fourth. From *The Development of Scientific Thinking Skills* [p. 87] by D. Kuhn, E. Amsel, & M. O'Loughlin, 1988, Orlando, FL: Academic Press. Copyright 1988 by Academic Press. Reprinted by permission.)

the attainment of formal operations, the second of only two new levels of cognition he believed to be acquired during development. (The other is the attainment of symbolic, or representational, operations during the second year of life.) Vygotsky (1962), like Piaget, regarded reflective, or metacognitive, capacity as a major developmental acquisition that does not emerge before middle to late childhood. In the work on coordination of theories and evidence, I have followed Vygotsky (1962) in regarding *reflective awareness* and *deliberate control* as the dual aspects of metacognition. Someone able to reflect on their own mental acts is also likely to be able to access and apply them in a manner under their voluntary control. Clearly, from an early

Table 1
Responses to 5/1 Evidence in the Balls Problem

Group	Evidence-based responses		Inclusion responses		Exclusion responses	
	S	O	S	O	S	O
Third graders	1.15	0.70	0.95	0.45	0.20	0.25
Sixth graders	1.95	1.70	1.35	0.85	0.60	0.80
Ninth graders	2.00	1.95	1.15	1.05	0.80	0.85
Noncollege adults	1.80	1.45	1.25	0.85	0.50	0.55
College adults	2.00	2.00	1.80	1.90	0.15	0.10
Total sample	1.78	1.56	1.30	1.02	0.45	0.51

Note. S = subject's theory; O = other's theory. From *The Development of Scientific Thinking Skills* (p. 149) by D. Kuhn, E. Amsel, & M. O'Loughlin, 1988, Orlando, FL: Academic Press.

age, children modify their primitive theories in the face of evidence. Yet, only through the development of the skills that have been described here do people attain control over the interaction of theory and evidence in their own thinking.

If one accepts the view that metacognitive awareness and control develop only gradually and, therefore, frequently exist in a partial state of attainment, it helps to explain a contradiction implicit in the data described earlier. In order to make the effort to bring theory and evidence into alignment with one another, a subject must at some level have recognized them as discrepant. Unlike subjects who are fully aware of and able to verbally acknowledge the discrepancy (and, hence, needn't seek to remove it), such subjects both know and do not know that there is a discrepancy between their theories and the evidence before them. At one level the discrepancy is processed, which it could not be if subjects had no ability to cognize their own theories. But this metacognitive capacity is not great enough to firmly maintain the differentiation between what derives from one's own thought and what derives from external sources, and hence, the characteristics that have been described appear.

The further clarification should be added that in the proposed framework, it is actually two sets of skills that are seen as codeveloping and reinforcing one another. One is the set of skills considered earlier that pertain to the differentiation and coordination of theory and evidence. The other is the set of skills involved in understanding the meaning of evidence once it is sufficiently differentiated from theory. Our research suggests that both are important and, furthermore, are closely connected. To the extent that an individual has acquired explicit, consistent criteria for interpreting evidence, these criteria are less likely to be compromised by the biasing effects of theory. Conversely, to the extent an individual is able to dissociate evidence from the context of his or her own theoretical beliefs and regard it as an independent entity in its own right, a concern for consistent and explicit criteria for interpreting evidence will be enhanced.

As an example, consistent throughout the results was the "overinterpretation" of covariation evidence, that is, the implication of any covariate as causal, with theoretical belief playing a major motivating role. A contributing factor, however, was inadequate understanding of the concept of covariation. Covariation of two variables means that the two vary together. If no variation occurs in one, the concept of covariation is meaningless. Yet, many subjects were quite willing to make causal inferences based on such lack of variation. Furthermore, when evidence regarding both variable levels was available, many subjects believed that a causal interpretation could be made separately for each. A subject might claim, for example, that rough-textured balls "make a difference" to outcome (as they were always associated with positive outcomes), whereas smooth-textured balls (which were evenly distributed across outcomes) do not. If asked about the relation of texture "overall" to outcome, the subject was likely to claim that it "sometimes makes a difference." Taken to its extreme, this "sometimes" conception means that each individual co-occurrence or non-co-occurrence of variable level and outcome can be interpreted in isolation.

Such errors are particularly likely to be theoretically motivated: "Here is some evidence I can point to as supporting my theory, and therefore the theory is right." Until the individual is willing to relinquish the versatile but false power (to serve as evidence for just about any theory one chooses) afforded by this inference strategy, it is very difficult to master the crucial valid inference strategy of exclusion—inferring that a variable has no effect on outcome based on detection of a pattern of noncovariation between variable and outcome over a set of instances. Some subjects explicitly recognized the power of the false inclusion strategy without recognizing its invalidity. Sixth-grader Randy, for example, concluded:

> Mr. C [for color] is half right because if he wanted light [to come out good], three light came out good and if he wanted dark three dark came out good.

Or, as another sixth grader put it:

> No matter which size Miss S likes, she's right!

Another example has to do with generation of evidence. These results are consistent with the many studies in the literature on formal operations (Inhelder & Piaget, 1958; Keating, 1980): Children, and even many adolescents and adults, fail to construct conclusive proofs of the effect of a variable because in demonstrating the covariation of one variable with outcome, other variables are left uncontrolled. This well-documented failure to control variables has been regarded as reflecting failure to attend to these variables and therefore to recognize the possibility that left uncontrolled they may exert their own effects on outcome. The explanations that subjects in this research offered of the evidence they generated, however, suggest another interpretation, one that implicates subjects' theories about effects of the variables. Rather than overlooking "uncontrolled" variables, our subjects appeared to be attempting to construct a body of evidence that would reflect the operation of these variables as well. For example, in responding to the request to demonstrate that size makes a difference, a subject might arrange several large balls in the *Good* basket and several small balls in the *Bad* basket. The large balls, however, would all have ridges and the small balls would have no ridges. In explaining the arrangement, the subject made it clear that ridges were deliberately covaried with outcome, because they too were expected to have an effect. In order to in fact demonstrate the effect of a variable, a subject must set aside his or her beliefs regarding the remaining variables and focus only on the single variable whose effect is to be demonstrated. This ability to set aside, or bracket, one's own theoretical beliefs is of course the same bracketing ability referred to earlier as central to successful coordination of theory and evidence.

Related Research

The framework that has been proposed here may be useful as well in conceptualizing the results of other research on processes of scientific thinking in children and adults. Although the studies of adults by Shute et al. (1989) and Voss et al. (1986) cited earlier revealed similar weaknesses, Klahr and Dunbar's (1988; Dunbar & Klahr, 1988) studies included both children and adults, and their work is hence of particular relevance given the developmental framework adopted here. Klahr and Dunbar observed college students and third- through sixth-grade children in self-directed exploratory activity as they engaged the goal of discovering the function of a particular key (the REPEAT key) on a computerized toy. They observed major differences in process between the children and adults. Children commonly accepted a hypothesis as true based on minimal evidence and overlooked conflicting evidence. Although children generated as much data as adults, their experiments were less well designed and they were less able to make use of their data in discovering the correct function. Instead, children appeared content with local interpretation, that is, making an inference consistent with the last result generated (ignoring earlier discrepant evidence).

Following Simon and Lea (1974), Klahr and Dunbar have conceptualized the scientific reasoning process as a search in two problem spaces, the space of hypotheses and the space of instances, or experiments. Their model offers a means of conceptualizing both adults' and children's activities in scientific exploration. Yet, what would add significantly to the value of their work with children is a means of conceptualizing the *difference* between children's and adults' behavior. The framework offered in the present article provides one such means. Klahr and Dunbar (1988) characterized their subjects as engaged in an effort to coordinate the two problem spaces of hypotheses and experiments, which their adult subjects were largely successful in doing. For the third- through sixth-grade children in their research, in contrast, the two problem spaces of hypotheses and experiments may have existed as a single, undifferentiated whole, limiting the children's ability to coordinate them and, hence, design informative experiments that would have led to successful problem solution. Instead, the lack of firm differentiation between the problem spaces of hypotheses and experiments allowed one or the other to often dominate, in an unstable, uncontrolled way.

Karmiloff-Smith's (1984, 1986, 1988) research on children's problem solving is also consistent with the framework that has been proposed here. In one of a series of studies she described, children of ages 5–12 years examined cubes made of various materials such as metal, wood, plastic, and sponge, and flat surfaces made of these same materials. A succession of cube and surface type combinations were presented, and the child was asked to explain the resulting phenomenon (e.g., why the sponge surface became slightly indented when the iron cube was placed on it). Like children in Kuhn et al.'s (1988) and Klahr and Dunbar's (1988) research, the youngest children in Karmiloff-Smith's work were satisfied with local explanation of isolated instances; that is, each cube–surface pair was treated as an isolated problem, with no concern for consistency between one explanation and the next. Older children, in contrast, looked for some overall principle as a basis for their individual explanations. As a result, however, they became so wedded to the theories they constructed for this purpose that the data themselves were compromised, that is, children reported phenomena that they had not observed in order to maintain their theories. Finally, the oldest children were more successful in coordinating the succession of instances they observed with the construction of a theory that was consistent with them.

Karmiloff-Smith's (1984, 1986, 1988) work is especially interesting because of the wide range of problem-solving contexts in which she noted a similar progression. In another study, for example, children of ages 4–9 were asked to tell a story to accompany a sequence of pictures. The youngest children were successful in the local task of describing the contents of each picture, but the successive descriptions were not integrated into a coherent narrative. Slightly older children were much more successful at the global task of producing a narrative, but they did so at the cost of fidelity to the details reflected in the individual pictures. The oldest children were more successful in producing a coherent narrative that incorporated the data represented in the pictures.

The latter study takes us far from the arena of scientific reasoning, but Karmiloff-Smith's (1984, 1986, 1988) studies as a whole are relevant here in several respects. First, her work illustrates that a wide range of developmental phenomena can be conceptualized as involving progressive coordination of theory and data. Second, the progression she described highlights the trade-off between data-bound and theory-bound procedures. The transition from her first, data-driven level to second, theory-driven level can appear to reflect a regression, she noted, in that following this transition, the child may represent the data that are observed less accurately. A trade-off is involved, however, as this transition also involves the emergence of a coherent theory that will guide the assimilation of new data. Third, Karmiloff-Smith emphasized that the process of achieving coordination between data-driven and theory-driven procedures is not something that is mastered once and for all in some domain-free manner. Instead, as her data on different tasks illustrate, it can be achieved at a relatively early age in some simple tasks such as the narrative task just described, whereas in other tasks even adults may have achieved it incompletely.

Scientific Thinking and Everyday Thinking

Thinking in scientific contexts represents only a minute portion of human thinking. In other research (Kuhn, 1989) I have undertaken to identify the broader range of contexts in which the skills that have been described appear. In what ways might the weaknesses in reasoning that have been described manifest themselves in everyday thought? I have explored this question by asking subjects of four age groups—teens, 20s, 40s, and 60s—to describe their own causal theories and to relate evidence to them. The topics—the cause of children failing in school, the cause of prisoners returning to crime after they have been released, and the cause of unemployment—were chosen as ones that people are likely to have occasion to think and talk about and ones about which people are able and willing to make causal inferences without a large base of technical knowledge. Nevertheless they involve phenomena the true causal structure of which is complex and uncertain. The interview was presented to subjects as eliciting their views on urban social problems. After eliciting a subject's causal theory, the subject was asked questions such as: How do you know that this is so—that this is the cause of prisoners returning to a life of crime? If you were trying to convince someone that your view is right, what evidence would you give to try to show this? Suppose that someone disagreed with your view; what might they say to show you were wrong? What would they say is the cause? What could you say to show they were wrong? In other words, subjects were asked to generate multiple, contrasting theories and coordinate evidence with them.

The weaknesses observed in the responses of many subjects parallel those described earlier. Subjects commonly related a scenario, or script, of how the phenomenon might occur. The request for supporting evidence typically produced merely an elaboration of the script, rather than evidence for its correctness. The attempt to elicit alternative theories or counterevidence was often unsuccessful. In a word, theory and evidence were fused into a script of "how it happens." For example, a subject in his 40s, responding to the question, "What causes prisoners to return to a life of crime after they're released from prison?" said:

> I think some of our laws today are really not strict enough. Maybe some of the rehabilitation supposedly that they are supposed to have did not do the job. And let's face it, today some of them even commit murder and they are out on the street the next day. Plus, they've got the overcrowding in the prisons and I don't think . . . well, in some places they are letting them out before time. They just don't have the room to put them anymore.

The causal script expressed by this subject is summarized in Figure 3. He was then asked, "How do you know that this is so, that this is what causes prisoners to return to crime?" This request for evidence produced merely an elaboration of the script:

> I think they feel it is a little bit lenient and I think today they get the feeling they can get away with more. (Can you explain exactly how this shows that this is the cause?) Our prison system today is more or less I think a little bit lenient, plus the judges and a lot of them, I don't think they give them the full amount in sentencing. I really think they are lenient with them, and they, you know, they get out and they get right back into the same rut again and they hold up a store or kill somebody or rap somebody then it's right back to day one again.

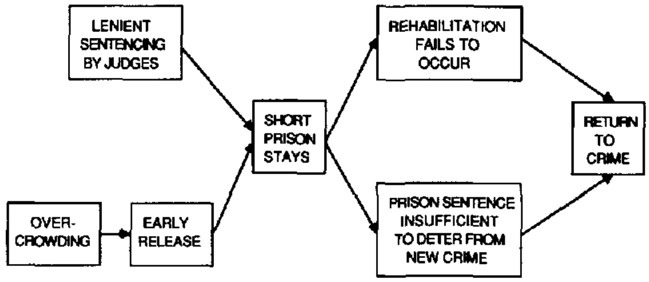

Figure 3. Illustration of causal script for crime topic.

After describing several examples that fit this script, the subject finally suggested, in response to the interviewer's probe, that a survey of prisons could be taken to show that they were overcrowded.

> (How would this prove that what you are claiming is right?) I think they feel they can do it again and it will be the same ball game all over again. In other words, you know, you live for a while and you are back out in the street and you do the same thing.

This subject exhibits a nondifferentiation of theory and evidence similar to that exhibited by the younger subjects quoted earlier. The two are fused into a script of "how it happens." Examples of the script are confused with evidence that would bear on its correctness. An adolescent subject put it explicitly in discussing the school failure topic:

> (How could you prove that what you are claiming is right?) I could give examples of people I've heard about that it happened to. I could ask them questions about what they've seen in their own classes. (How would this prove it?) Because if I could give examples they couldn't disprove my examples, since they really happened.

In contrast to this proof by examples, comments of some other subjects reflected a subscription to proof by simple assertion:

> If you want to convince someone, you show them what it is and you tell them that's why. You would tell them.

Or, in another subject's words:

> What evidence would I give? I would just talk about it.

Closely related to the fusion of theory and evidence into a script with accompanying instances is the inability to generate alternative theories or counterevidence for a theory. The first subject quoted was asked, "Suppose that someone disagreed with your view; what might *they* say to show that you were wrong?" Instead of generating a counterargument or counterevidence, he attempts to generate an alternative theory:

> Maybe they would feel the judges aren't strict enough.

But even here he is unsuccessful, as this proposition is part of his own original script (Figure 3).

A majority of subjects were successful in generating alternative theories when explicitly asked, although most did not do so spontaneously. Many, however, were not (suggesting that deficits were of both the competence and performance variety). A number of unsuccessful subjects acknowledged their inability to conceive of alternatives. For example:

> I don't know what they would say. I'd really have to get someone else's point of view. Cause I imagine my thoughts run in this direction and that's about it.

Other subjects were not even open to the possibility of alternatives:

> I don't think they could say anything. I don't think I'm wrong.

> They would agree with me. The majority of people think the way I do.

The key elements in subjects like the ones quoted developing skills in differentiating and coordinating theory and evidence are (a) recognition of the possibility of alternative theories, and (b) recognition of the possibility of evidence that doesn't fit a theory. The first achievement is likely to facilitate the second, as the presence of multiple, contrasting theories makes it difficult to assimilate the same evidence to both of them. In the case of each of these achievements, awareness that things could be otherwise is the key element. A script becomes a theory when its possible falsehood and the existence of alternative theories are recognized. Instances become evidence when the possibility of their lack of concordance with a theory is recognized.

Shown in Table 2 are percentages of subjects who were able to generate genuine evidence and percentages of subjects who were able to generate alternative theories for the school failure topic. Evidence categorized as genuine was by no means conclusive, but it was, at a minimum, evidence that was differentiated from the causal proposition itself and bore on its correctness (in contrast to the pseudoevidence reflected in the scriptlike responses illustrated earlier). As reflected in Table 2, generation of genuine evidence was the more difficult of the two skills. No age group or sex differences were significant, but there were consistent effects of education, a factor to which I turn in the next section. Topic differences also appeared, with subjects showing the best reasoning on the topic for which they were most likely to have personal knowledge (school failure) and reasoning least well on the topic for which they were least likely to have personal knowledge (return to crime). These differences, however, were small: Overall percentages for genuine evidence ranged from 39% (crime) to 48% (school), and for alternative theories from 58% (crime) to 69% (school). (Percentages for the unemployment topic were intermediate.) Also noteworthy, and supportive of the suggested theoretical connection between these skills, are the significant relations that appeared between the two skills shown in Table 2—despite the overall difference in difficulty level—for each of the three topics. Subjects who displayed one skill were more likely (than subjects who did not) to also display the other.

The extension of the present framework from scientific to everyday reasoning makes it relevant to psychologists studying belief formation and modification, as well as to those studying reasoning and inference. Particularly significant in this respect is the suggestion that developmental analysis may yield insight into adult functioning and, related to this idea, the concept of coordination of theory and evidence as a skill that may show progressive degrees of mastery. Belief bias phenomena, which have been conceptualized here as reflecting limitations in differentiation and coordination of theory and evidence, have been noted widely in the social psychology literature (Baron, 1985). As well as suggesting the importance of the often-ignored factor of conscious control, the present developmental framework offers a means of conceptualizing these phenomena and the developmental progression that effects their decline.

Table 2
Percentages of Subjects Generating Genuine Evidence and Alternative Theories for the School Failure Topic

	Age group				
Measure	Teens	20s	40s	60s	Total
Genuine evidence					
Noncollege	10	35	45	25	29
College	65	80	65	55	66
Alternative theories					
Noncollege	75	55	45	55	58
College	75	95	80	75	81

Note. $N = 160$.

Mechanisms of Development

How does the development of skill in coordinating theory and evidence take place? In this section, I first note some implications of the research already described that relate to this question and then turn to other research that addresses it more directly.

The performance of experts in the tasks that have been described provides a standard against which other subject groups can be evaluated. Comparisons of performance across these subject groups reflect both age and education differences. Children's performance improves between third grade and ninth grade. There is only slight improvement between ninth grade and adulthood, however, and at these ages performance is strongly influenced by education level. In the everyday thinking research described in the preceding section, I examined two education levels, those subjects who had some college education (or who were college-bound in the case of the adolescents) and those who did not. On all of the measures examined (e.g., forms of evidence offered, ability to generate alternative theories, ability to generate counterarguments), college subjects demonstrated a striking superiority to noncollege subjects. Caution must of course be exercised in interpreting such differences entirely in causal terms; education level is to some degree an effect as well as a cause of reasoning ability. These education differences within adult groups, nonetheless, accord with those found by Kuhn et al. (1988) and by Voss et al. (1986). (The other studies that have been discussed involving adults—by Klahr and Dunbar, 1988, and Shute et al., 1989—have been confined to college students.) Research by Nisbett and his colleagues (Lehman, Lempert, & Nisbett, 1988; Nisbett, Fong, Lehman, & Cheng, 1987) also indicates relations between college education and reasoning skill, although no association was found by Perkins (1985). (The Nisbett work further indicates an association with professional specialization, a finding also reported by Amsel, Langer, & Loutzenhiser, in press.)

Education differences are particularly interesting in the kinds of skills examined in the present article because such skills ordinarily are not taught in any direct, explicit way as part of the school curriculum. Some more general forms of experience associated with schooling are thus implicated. One way in which school experience may more indirectly foster the skills that have been examined is in affording practice in bracketing one's own experience, or beliefs, in order to infer a conclusion that follows from information given. This bracketing ability, which was suggested earlier to be central to successful coordination of theory and evidence, has been portrayed by Scribner (1977) and others as the discourse mode of those who follow academic paths—the "assume that" stance that is central to hypothetico-deductive thinking and common in problems posed to students in academic settings. Thus, an academic environment may well support the kinds of reasoning examined in this article. It does not follow, however, that the value or utility of such reasoning is confined to academic contexts, as the research described in the preceding section illustrates.

Nor does it follow that the skills in question cannot be developed outside of academic settings. In research that is more directly addressed to the question of mechanism, my coworkers and I have adopted a microgenetic approach (Kuhn & Phelps, 1982; Schauble, in press; Schauble & Kuhn, 1989). Subjects are engaged in self-directed exploration in weekly sessions over a period of several months, enabling us to observe how strategies change over time with exercise. Although no instruction or feedback is provided, we have found the method successful in eliciting progressive change in most (although not all) subjects. This result is not only significant in itself, but has allowed us to observe something about the nature of the change process.

In the Kuhn and Phelps (1982) study, fourth and fifth graders were engaged in scientific investigation to discover which elements in a mixture of liquids were responsible for a chemical reaction (either a color change or formation of a precipitate). This neutral content was chosen so as to minimize content involvement and focus examination on strategies and strategy change. In more recent work (Schauble, in press; Schauble & Kuhn, 1989), with fifth and sixth graders, the picture has been complicated by the choice of a knowledge-rich domain. As a result, both knowledge change and strategy change are likely to take place during the period of involvement with the materials. This is the situation most likely to exist in natural settings. Strategy changes occur in the context of efforts to acquire knowledge. Ideally, the exercise of knowledge-acquisition strategies both enhances knowledge and improves the strategies themselves.

The problem domain in this research consisted of a microcomputer racetrack on which cars that have different features (e.g., color, engine size, presence or absence of a tail fin) travel at different speeds in test runs that the subject can construct. The subject's task is to determine what makes a difference in how fast the cars travel. The design also afforded us the opportunity to observe how subjects' theories about the effects of the cars' features (assessed at the outset) influence the discovery process, as subjects' understanding of the racetrack microworld increases over the period of weeks.

Some of the preadolescent subjects in this research simultaneously engaged in separate weekly sessions that provided them exercise in generating and evaluating evidence in the sports balls domain. This parallel activity made it possible to assess the extent to which the same evidence generation and evaluation skills exercised in the more structured sports balls task format are identifiable in self-directed scientific activity (i.e., when subjects are free to generate and evaluate evidence toward the goal of acquiring knowledge within a particular microworld). Most important, however, the dual-task design made it possible to address the critical question of transfer, or skill generality, by examining the extent to which strategy change observed in one task domain tends to co-occur with strategy change in the other. Microgenetic designs involving multiple contexts, which provide a dynamic picture of change across two or more contexts, arguably offer the best means of investigating issues of transfer and skill generality (Schauble & Kuhn, 1989). Such issues are integral to the question of mechanism.

Results of the race-car research show that subjects acquire knowledge despite their reliance on strategies that are far from optimal. Consonant with the earlier study of self-directed scientific investigation (Kuhn & Phelps, 1982) and the work with children by Dunbar and Klahr (1988), and in sharp contrast to the performance of several professional scientists to whom the task was presented, subjects had difficulty designing informative experiments and readily made invalid inferences based on the experiments they did conduct. Only 31% of all experiments and 38% of all inferences were coded as valid (Schauble, in press). Yet, almost all of the subjects gradually discovered simple effects (but not more complex curvilinear and interactive effects) despite their generally poor procedures (although subjects with better procedures discovered them faster).

Although the extent to which subjects' theoretical beliefs distorted their interpretation of evidence declined over time, this progression could not accurately be characterized as one in which subjects gradually set aside their own theories in order to attend to and interpret the evidence. Instead, there was a close interlocking of theory and evidence. Rather than setting aside theories, which were often wrong, and simply interpreting the evidence before them, subjects typically replaced their original theories with new ones, most often before acknowledging the corresponding pattern in the evidence (a pattern also found in our earlier studies and reflected in Figure 2). What subjects seemed unwilling to do was to interpret evidence of the effectiveness or ineffectiveness of a feature until they had a compatible theory in place that made sense of that evidence.

Although significant improvement occurred over time, a majority of the scientific exploration activity of these preadolescent subjects fell into two categories, both of which reflect failures in the coordination of theory and evidence. These categories resemble those described by Karmiloff-Smith (1984, 1988) and by Klahr and Dunbar (1988). They can be referred to as *theory-bound* and *data-bound* (similar to Klahr and Dunbar's, 1988, characterization of their adult subjects as theorists or experimenters), but here I describe unsuccessful, rather than successful, procedures. Subjects engaged in theory-bound investigation were so bound to their original theories that they had difficulty either in attending to evidence at all or in generating evidence that would provide useful information, and when they did interpret evidence, they typically distorted it to fit their theories. Subjects engaged in data-bound investigation, in contrast, were overly bound to the evidence. They likewise had difficulty in generating informative experiments, but each piece of evidence they did generate, they felt obliged to explain. Like Dunbar and Klahr's (1988) child subjects, however, they confined themselves to local interpretation of isolated results, rather than searching for a broader pattern of results over a number of instances, as more successful subjects did. Stated differently, their evidence generation was insufficiently guided by a theoretical representation that would have enabled them to make better sense of their results.

Progress with respect to procedures, then, requires not only mastery of appropriate experimentation and inference strategies, but also improved coordination of theories and evidence, attainments that I have argued are interrelated and have a strong metacognitive component. The race-cars study provided a test of the hypothesis that subjects' procedural weaknesses were metacognitive as well as strategic, as each subject was provided a notebook for the course of the study and was asked to use it to keep records to aid in the discovery of how the cars' features affected their speed. Although these fifth and sixth graders were well past the age at which researchers have inferred presence of a *covariation strategy* (Shultz & Mendelson, 1975; Siegler, 1975), not one subject recorded covariation data (i.e., feature combinations and corresponding speed outcomes), data that were essential to the causal inferences to be made. Some subjects recorded only the cars' features, without outcomes, and others recorded outcomes without noting the cars' features, and some recorded neither. Clearly, these subjects did not know what they needed to know to master the problem.

How, then, does strategy change take place? The changes observed in this microgenetic research rarely consist of a simple transition in which an inferior strategy is replaced by a superior one (although a small minority of subjects show this pattern). Instead, during a long period of variable performance, a subject typically uses more and less advanced experimentation and inference strategies in conjunction with one another. My hypothesis is that during this period of variable usage, subjects are not only gaining practice in the use of the more advanced strategies, but they are also gaining in the metacognitive understanding of them—that is, why they are the most correct or efficient strategies to apply, what their range of application is, and so forth. Mastery of new strategies, however, is only part of the task at hand. The other is the relinquishment of less adequate strategies. Discarded and replaced by a better strategy on one occasion, incorrect strategies repeatedly turn up again when the context becomes more complicated, or even when it does not. Thus, even more important than the attainment of new strategies (which are typically already in the subject's repertory, although they may appear infrequently) may be the abandonment of old, less adequate strategies—a reversal of the way one typically thinks about development.

The results of this research thus indicate a number of respects in which the picture of what it is that is developing is complex, rendering unlikely any simple conception of mechanism. First, there is rarely a discrete point at which acquisition can be said to occur. Instead, more and less advanced strategies co-exist in an individual's repertory, with the more advanced only gradually overpowering the less advanced. Second, both mastery of new strategies and discarding of old ones involve metacognitive as well as strategic understanding. Third, as was also noted in the discussion of Karmiloff-Smith's (1984, 1986, 1988) research and the Kuhn et al. (1988) studies, this development takes place not once but many times over, as the skill mastered or the error avoided in one context remains to be conquered in others.

The last fact points to the need to study transfer across contexts as integral to the question of mechanism. To what extent is it a domain-general versus a domain-specific entity that is developing? Clearly, little progress can be made in understanding mechanism without addressing this question. Important in this respect is a distinction noted by Baron (1985), one that is seldom made in the extensive discussion on the generality of thinking skills (Chipman, Segal, & Glaser, 1985). Clearly, the skills considered in this article are definable in a general sense, that is, without reference to specific content, in contrast, for example, to the scientific thinking studied by Carey (1985b) and other domain-specific researchers, which is characterizable only within the context of specific subject matter. It is a separate question, however, whether the skills examined here are general in an empirical sense.

The empirical results do indicate a significant degree of generality when the same skill is examined across different content or contexts of usage (Kuhn, 1989; Schauble & Kuhn, 1989). In the microgenetic study across the two problem domains of the sports balls and race cars (Schauble & Kuhn, 1989), subjects who exhibited improvement in a skill in one domain (vs. those who did not) were more likely to exhibit improvement in the other. Some skills showed more generality than did others, and the data by no means support a completely domain-general model of acquisition. Yet, the data also stand as evidence against the other extreme—that there exist no cognitive strategies that extend beyond a specific context or knowledge domain. Instead, these results suggest some transfer of skill development across contexts.

The likelihood that metacognition is implicated in such transfer should be emphasized. Heightened metacognitive awareness of a strategy makes it more available and, hence, increases the probability of the user's recognizing the applicability of the strategy in other contexts. Note that this account of transfer differs from traditional ones. Accounts of transfer within traditional learning paradigms have focused on stimulus similarity as the critical feature governing the transfer of learned behaviors. To the extent that a new stimulus situation is similar to the one in which the behavior was originally learned, the behavior is likely to be elicited by the new situation. Instead, focus on the behavior itself, and particularly metacognitive awareness and control of it, as the critical features that determine transfer may be more productive.

In sum, then, metacognition plays a central role in our account of developmental change in the cognitive skills involved in the successful coordination of theory and evidence. Exercise of these skills enhances metacognitive awareness and control of them, which in turn promote their generalization. Increased generality may, in turn, promote metacognitive awareness in a relationship of reciprocal support. Furthermore, for the skills of concern here (and in contrast to some other areas of cognitive development), it is not the discovery of new strategies that requires explanation so much as the relinquishment of old ones. The new strategies are for the most part very simple and already in the subject's repertory. The major issue of interest is how superior strategies gradually increase in strength and displace inferior ones, as the subject's understanding of their superiority (and the inadequacies of old strategies) grows.

The fact that our microgenetic research has shown exercise to be a sufficient factor to induce change allows certain inferences but not others. It does not imply that exercise is necessarily the optimal means for inducing change, an issue to which I return shortly. It does imply that formal training, modeling, or other kinds of specific experience are not necessary conditions for the development of the elementary skills in generating and evaluating evidence and coordinating it with theories that have been the subject of this article.

Exercise of these skills may be a sufficient mechanism to induce change in part because it provides the practice that has been found important in the improvement of a wide range of cognitive skills. With such exercise, or practice, of the kinds of skills of concern here, I have suggested, comes enhanced metacognitive awareness and hence control of these skills. However, another more specific reason that exercise may be sufficient to induce change is that, in the case of these skills, it is naturally self-corrective. In other words, the contradictions that need to be overcome are for the most part inherent in the evidence the subject is contemplating. For every false inclusion inference, there exists an opposing inference implicit in the evidence (Might it not be the balls' texture rather than their size that is responsible for the outcome?). Evidence that is ignored or distorted does not disappear but rather remains to be confronted again. What may need to be contributed from an external source, however, is the possibility of alternative theories. As noted, subjects often cannot conceive of the possibility of an alternative theory to their own. Once alternatives are established and the subject is engaged in relating evidence to each of them, the contradictions inherent in applying different standards for different theories may eventually come to be recognized.

My reliance on a microgenetic method focused on exercise of skills as a means of investigating the question of mechanism is aimed at enhancing understanding of the nature of the subject's (as opposed to an external agent's) activity in the development in question. It does not imply that dydactic methods of teaching reasoning are inappropriate. Although concerned primarily with statistical inference principles that are considerably less basic than the skills considered in the present article, Nisbett and his colleagues (Cheng, Holyoak, Nisbett & Oliver, 1986; Lehman et al., 1988; Nisbett et al., 1987) have reported considerable success in teaching certain reasoning principles (but not others) by means of explicit instruction. Certain of the skills considered here undoubtedly could be promoted by explicit instruction, for example, teaching rules regarding the forms of evidence required for valid inferences of inclusion or exclusion. Nisbett et al.'s (1987) research suggests that a combination of abstract rule training (which would serve to heighten metacognitive awareness of the rule) and practice with specific instances is most successful. In such cases, metacognitive and strategic development most likely support one another. Although I believe that less directive research methods of the sort I have relied on here will continue to yield important insights into mechanisms of change, research involving more dydactic methods is of such potential practical significance that it should not be neglected. Although the present research indicates that exercise can be a sufficient mechanism to induce change, the fact that many adults reason at no more advanced a level than sixth graders indicates that sufficient exercise may often not be available. Whether increased exercise or more dydactic forms of intervention would yield greater long-term success in enhancing such subjects' reasoning skills remains to be determined.

Conclusions

One sense in which the child-as-scientist metaphor can be interpreted is with respect to scientific understanding. Both child and scientist gain understanding of the world through construction and revision of mental models. Recent research evidence described in this article, however, suggests that the process in terms of which mental models, or theories, are coordinated with new evidence is significantly different in the child, the lay adult, and the scientist. In this sense, then, the metaphor of child as scientist may be fundamentally misleading. In some very basic respects, children (and many adults) do not behave like scientists.

The differences between child, lay adult, and scientist in the process of scientific thinking can be usefully conceived in a developmental framework. In this article I have described research to support a proposed framework, one in which the lower end of a developmental continuum is conceived of as reflecting the nondifferentiation of theory and evidence, precluding the construction of relations between the two. In Klahr and Dunbar's (1988) terms, the problem spaces of hypotheses and evidence exist as a single, undifferentiated whole. When theory and evidence are compatible, the two are melded into a single representation of "the way things are." When they are discrepant, subjects exhibit strategies for maintaining their alignment—either adjusting the theory, typically without acknowledging having done so, or "adjusting" the evidence, by ignoring it or attending to it in a selective, distorting manner.

At the other end of this developmental continuum is the full differentiation and coordination of theories and evidence and the elevation of the theory–evidence interaction to the level of conscious control. It is in these crucial respects that the professional scientist and child as scientist differ.

In scientific exploration activities, lack of differentiation and coordination of theory and evidence is likely to lead to uncontrolled domination of one over the other. Exploration may be so theory-bound that the subject has difficulty "seeing" the evidence, or so data-bound that the subject is confined to local interpretation of isolated results, without benefit of a theoretical representation that would allow the subject to make sense of the data.

In everyday reasoning, lack of differentiation and coordination of theory and evidence are reflected in the failure to conceive alternatives. Theories take the form of scripts of "how it happens," and instances of the script are confused with evidence for its correctness. Only when its possible falsehood and existence of alternative theories are recognized does a script become a theory. Only when the possibility of lack of concordance with a theory is recognized do instances become evidence.

On this basis, I conclude that the development of processes of scientific thinking entails strong restructuring of the concepts of theory and evidence, in the sense that Carey (1985b, 1986) has used this term and contrasted it to the more common case of weak restructuring. In the case of strong restructuring, the two terms do not have the same meanings or relation to one another at two different points in development. My claim, however, is that it is the *instruments* of scientific thinking, not just the products, that undergo strong restructuring.

The development I have described requires thinking about theories, rather than merely with them, and thinking about evidence, rather than merely being influenced by it. This development is thus metacognitive, as well as strategic. From a very early age, children modify their primitive theories in the face of evidence, but only through the development that has been the topic of this article does one attain control over the interaction of theory and evidence in one's own thinking. It is a development that occurs not once but many times over, as theories and evidence repeatedly come into contact with one another. It is also, however, a development that is incompletely realized in most people.

To the extent that it centers on reflection on one's own thought, the framework that has been proposed for conceptualizing the development of scientific thinking builds directly on the pioneering work of Inhelder and Piaget (1958). Too often, both theorists and practitioners embrace new concepts without appreciating their history. A number of science educators in recent years have renounced Piaget's construct of formal operations as no longer useful to them, claiming that *metareasoning*, defined as the ability to reason about one's own reasoning, now appears to offer greater promise. While appreciating the connection of the framework that has been proposed here to Piaget's, I believe that Inhelder and Piaget (1958) regarded the formal operational reasoning skills they studied in too "formal" a way, that is, as operating in a uniform manner irrespective of the subject's own beliefs and understanding regarding the content being reasoned about. Moreover, unlike Inhelder and Piaget (1958), I do not see the skills examined here as deriving from an underlying logical competence, and rather regard them as emerging more in the form of the pragmatic, goal-related schemes described by Cheng and Holyoak (1985). Such pragmatic schemes both provide an inductive apparatus through which new information is interpreted and, I claim, themselves undergo development.

Shafer and Tversky (1985) have made the important distinction between evidence and information, with evidence, unlike information, intricately tied to a process of reasoning. As I have tried to show, evidence can exist only insofar as the beholder has attained mastery of the reasoning skills examined in this article. The ability to coordinate evidence with theories tends to be taken for granted in scientific thinking. The very elementary skills in relating evidence to theories that are examined in this article are simply assumed as givens, even when the scientists being considered are children. The research that has been described here indicates that such assumptions are mistaken and makes clear the need to study these skills developmentally. It was Galileo who said, "In questions of science, the authority of a thousand is not worth the humble reasoning of a single individual." How important, then, that we pursue the effort to understand that reasoning and how it develops.

References

Alloy, L., & Tabachnik, N. (1984). Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information. *Psychological Review, 91,* 112–149.

Amsel, E., Langer, R., & Loutzenhiser, L. (in press). Causal inquiry and the acquisition of expertise in law and psychology. In R. Sternberg & P. Frensch (Eds.), *Complex problem solving: Principles and mechanisms.* Hillsdale, NJ: Erlbaum.

Baron, J. (1985). *Rationality and intelligence.* New York: Cambridge University Press.

Carey, S. (1985a). Are children fundamentally different kinds of thinkers and learners than adults? In S. Chipman, J. Segal, & R. Glaser (Eds.), *Thinking and learning skills* (Vol. 2, pp. 485–517). Hillsdale, NJ: Erlbaum.

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

Carey, S. (1986). Cognitive science and science education. *American Psychologist, 41,* 1123–1130.

Champagne, A., & Klopfer, L. (1984). Research in science education: The cognitive psychology perspective. In D. Holdzkom & P. Ludz (Eds.), *Research within reach: Science education* (pp. 171–189). Washington, DC: National Science Teachers Association.

Cheng, P., & Holyoak, K. (1985). Pragmatic reasoning schemas. *Cognitive Psychology, 17,* 391–416.

Cheng, P., Holyoak, K., Nisbett, R., & Oliver, L. (1986). Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology, 18,* 293–328.

Chi, M., & Ceci, S. (1987). Content knowledge: Its representation and restructuring in memory development. In H. Reese (Ed.), *Advances in child development and behavior* (Vol. 20, pp. 91–141). Orlando, FL: Academic Press.

Chipman, S., Segal, J., & Glaser, R. (Eds.). (1985). *Thinking and learning skills* (Vol. 2). Hillsdale, NJ: Erlbaum.

DiSessa, A. (1983). Phenomenology and the evolution of intuition. In D. Gentner & A. Stevens (Eds.), *Mental models* (pp. 15–33). Hillsdale, NJ: Erlbaum.

Dunbar, K., & Klahr, D. (1988). Developmental differences in scientific discovery strategies. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon (Proceedings of the 21st Carnegie-Mellon Symposium on Cognition,* pp. 109–144). Hillsdale, NJ: Erlbaum.

Elstein, A., Shulman, L., & Sprafka, S. (1978). *Medical problem solving: An analysis of clinical reasoning.* Cambridge, MA: Harvard University Press.

Faust, D. (1984). *The limits of scientific reasoning.* Minneapolis: University of Minnesota Press.

Fischhoff, B., & Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review, 90,* 239–260.

Gentner, D., & Gentner, D. (1983). Flowing waters or teeming crowds: Mental models of electricity. In D. Gentner & A. Stevens (Eds.), *Mental models* (pp. 101–129). Hillsdale, NJ: Erlbaum.

Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist, 39,* 93–104.

Gruber, H. (1973). Courage and cognitive growth in children and scientists. In M. Schwebel & J. Raph (Eds.), *Piaget in the classroom* (pp. 73–105). New York: Basic Books.

Holland, J., Holyoak, K., Nisbett, R., & Thagard, P. (1986). *Induction: Processes of inference, learning, and discovery.* Cambridge, MA: MIT 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: Curvilinear motion in the absence of external forces. *Developmental Psychology, 22,* 67–71.

Karmiloff-Smith, A. (1984). Children's problem solving. In M. Lamb, A. Brown, & B. Rogoff (Eds.), *Advances in developmental psychology* (Vol. 3, pp. 39–90). Hillsdale, NJ: Erlbaum.

Karmiloff-Smith, A. (1986). From meta-processes to conscious access: Evidence from children's metalinguistic and repair data. *Cognition, 23,* 95–147.

Karmiloff-Smith, A. (1988). The child is a theoretician, not an inductivist. *Mind and Language, 3,* 1-13.

Kassin, S., & Wrightsman, L. (Eds.). (1985). *The psychology of evidence and trial procedure.* Beverly Hills, CA: Sage.

Keating, D. (1980). Thinking processes in adolescence. In J. Adelson (Ed.), *Handbook of adolescent psychology* (pp. 211-246). New York: Wiley.

Keil, F. (1984). Mechanisms in cognitive development and the structure of knowledge. In R. Sternberg (Ed.), *Mechanisms of cognitive development* (pp. 81-99). New York: Freeman.

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

Krupa, M., Selman, R., & Jacquette, D. (1985). The development of science explanations in children and adolescents: A structural approach. In S. Chipman, J. Segal, & R. Glaser (Eds.), *Thinking and learning skills* (Vol. 2, pp. 427-455). Hillsdale, NJ: Erlbaum.

Kuhn, D. (1989). *The skills of argument.* Manuscript in preparation.

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

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, T. (1962). *The structure of scientific revolutions.* Chicago: University of Chicago Press.

Larkin, J. (1983). The role of problem representation in physics. In D. Gentner & A. Stevens (Eds.), *Mental models* (pp. 53-73). Hillsdale, NJ: Erlbaum.

Lehman, D., Lempert, R., & Nisbett, R. (1988). The effects of graduate training on reasoning. *American Psychologist, 43,* 431-442.

Mahoney, M., & Kimper, T. (1976). From ethics to logic: A survey of scientists. In M. Mahoney (Ed.), *Scientist as subject: The psychological imperative* (pp. 187-194). Cambridge, MA: Ballinger.

McCloskey, M. (1983). Naive theories of motion. In D. Gentner & A. Stevens (Eds.), *Mental models* (pp. 299-324). Hillsdale, NJ: Erlbaum.

Moshman, D. (1979). To really get ahead, get a metatheory. In W. Damon (Series ed.), D. Kuhn (Ed.), *New directions in child development: Vol. 5. Intellectual development beyond childhood* (pp. 59-68). San Francisco: Jossey-Bass.

Murphy, G., & Medin, D. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289-316.

Neisser, U. (Ed.). (1987). *Concepts and conceptual development: Ecological and intellectual factors in categorization.* New York: Cambridge University Press.

Nelson, K. (1985). *Making sense: The acquisition of shared meaning.* Orlando, FL: Academic Press.

Nisbett, R., Fong, G., Lehman, D., & Cheng, P. (1987). Teaching reasoning. *Science, 30,* 625-631.

Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice-Hall.

Perkins, D. (1985). Postprimary education has little impact on informal reasoning. *Journal of Educational Psychology, 77,* 562-571.

Piaget, J. (1950). *The psychology of intelligence.* London: Routledge & Kegan Paul.

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

Schauble, L., & Kuhn, D. (1989). *Applying scientific thinking skills in self-directed exploration of a microworld.* Manuscript submitted for publication. (Available from L. Schauble, LRDC, University of Pittsburgh, Pittsburgh, PA 15260)

Scribner, S. (1977). Modes of thinking and ways of speaking. In P. Johnson-Laird & P. Wason (Eds.), *Thinking* (pp. 483-500). Cambridge, England: Cambridge University Press.

Shafer, G., & Tversky, A. (1985). Languages and designs for probability judgment. *Cognitive Science, 9,* 309-339.

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

Shultz, T., & Mendelson, R. (1975). The use of covariation as a principle of causal analysis. *Child Development, 46,* 394-399.

Shute, V., Glaser, R., & Raghavan, K. (1989). Discovery and inference in an exploratory laboratory. In P. Ackerman, R. Sternberg, & R. Glaser (Eds.), *Learning and individual differences* (pp. 279-326). San Francisco: Freeman.

Siegler, R. (1975). Defining the locus of developmental differences in children's causal reasoning. *Journal of Experimental Child Psychology, 20,* 512-525.

Simon, H., & Lea, G. (1974). Problem solving and rule induction: A unified view. In L. Gregg (Ed.), Knowledge and cognition (pp. 105-127). Hillsdale, NJ: Erlbaum.

Sternberg, R. (1985). *Beyond IQ: A triarchic theory of human intelligence.* Cambridge, England: Cambridge University Press.

Strauss, S., & Stavy, R. (1982). U-shaped behavioral growth: Implications for theories of development. In W. Hartup (Ed.), *Review of child development research* (Vol. 6, pp. 547-599). Chicago: University of Chicago Press.

Vosniadou, S., & Brewer, W. (1987). Theories of knowledge restructuring in development. *Review of Educational Research, 57,* 51-67.

Voss, J., Blais, J., Means, M., Greene, T., & Ahwesh, E. (1986). Informal reasoning and subject matter knowledge in the solving of economics problems by naive and novice individuals. *Cognition and Instruction, 3,* 269-302.

Vygotsky, L. (1962). *Thought and language.* Cambridge, MA: MIT Press.

West, L., & Pines, A. (Eds.). (1985). *Cognitive structure and conceptual change.* Orlando, FL: Academic Press.

Williams, M., Hollan, J., & Stevens, A. (1983). Human reasoning about a simple physical system. In D. Gentner & A. Stevens (Eds.), *Mental models* (pp. 131-154). Hillsdale, NJ: Erlbaum.

Received September 27, 1988
Revision received February 15, 1989
Accepted March 3, 1989 ■