

## COMMENTARY

### SCIENTIFIC THINKING ABOUT SCIENTIFIC THINKING

*David Klahr and Sharon M. Carver*

Readers come to the Commentaries of these *Monographs* with three questions in mind. First, is the *Monograph* worth reading? Second, does it raise specific points that deserve further attention, emphasis, or criticism? Finally, are there broad issues raised by the *Monograph* that are sufficiently important to warrant further discussion, almost independent of the content of the *Monograph* itself? With respect to the present *Monograph*, the short answers to these questions are “yes,” “yes,” and “yes.” The longer answers follow.

Kuhn, Garcia-Mila, Zohar, and Andersen address the question of how people generate evidence about multivariable causal systems and then form hypotheses about the relevant variables on the basis of that evidence. They investigated this issue in two broad domains (physical and social), using two groups of subjects (preadolescents and adults), and they focused on how, over the course of 10 weekly experimental sessions, subjects acquired not only domain-specific knowledge (e.g., the factors that make for fast cars or effective television commercials) but also domain-general strategies for making valid inferences from data. The design enabled Kuhn et al. to compare performance within and across domains and subject populations, and the analysis revealed important similarities and differences in the use of valid and invalid strategies that we will describe below.

Kuhn et al.’s ambitious and unprecedented undertaking embraces a densely interwoven tapestry of fundamental methodological issues and central topics within the area of cognitive development. The methodological issues include transfer of training, microgenetic analysis, and the relative merits of quantitative and qualitative analysis of children’s behavior. The

topical areas include scientific reasoning, strategy acquisition and choice, and metacognition.

In order to deal with this array of interconnected issues and topics, we have organized this Commentary into four sections. In the first, we provide a broad context for research on scientific discovery in which to situate the Kuhn et al. project, and we emphasize its considerable strengths. In the second section, we view the Kuhn et al. work from another perspective: as a transfer of training study. In the third, we raise both methodological and theoretical questions about the work. Finally, in the fourth section, we offer suggestions for addressing some of the questions stimulated by this remarkable investigation.

#### *Investigating the Scientific Discovery Process*

The Kuhn et al. work is about many things, but it is, to our way of thinking, primarily an investigation of the scientific discovery process. The general paradigm used by psychologists who are interested in scientific reasoning is to present people with situations crafted to isolate one or more essential aspects of “real-world” science and to observe their problem-solving processes carefully. There are, of course, other ways to study scientific thinking, including historical analyses, retrospective reports, and “in vivo” studies of ongoing scientific work (Dunbar, 1994). However, the laboratory approach exemplified by the Kuhn et al. work has several important merits:

1. It allows the researcher great latitude in selecting the subject population under investigation.
2. It enables the researcher to exert substantial control over subjects' prior knowledge, through the type of selection mentioned above and through various levels of background training in the domain under investigation.
3. It facilitates the observation of the dynamic course of scientific discovery in great detail and the corresponding use of a variety of assessment methodologies.
4. It allows control over the “state of nature,” that is, the thing to be discovered by the subjects. Such studies have presented subjects with a variety of things to be discovered, including (a) an arbitrary rule that the experimenter has in mind (Gorman, 1992; Wason, 1960), (b) a computer microworld that embodies some realistic causal factors and some arbitrary ones (such as the race cars microworld developed by Schauble, 1990, and used by Kuhn et al.), (c) the causal factors in a real physical domain, such as the boat task used by Kuhn et al. (adapted from a task created by Schauble, Klopfer, & Raghavan, 1991) or the investigation of sinking rates of objects dropped in water (Penner &

## STRATEGIES OF KNOWLEDGE ACQUISITION

Klahr, in press), (d) the physics of a complex artificial universe (Mynatt, Doherty, & Tweney, 1977), and (e) a computer microworld designed to capture the essential features of a historical discovery (e.g., Dunbar's, 1993, microworld in which subjects attempted to [re]discover the mechanisms of genetic inhibition).

5. Perhaps the most valuable characteristic of laboratory studies of scientific reasoning is that they included a well-documented record of the unsuccessful, as well as the successful, discoverers. Because there is a scant historical or biographical record of the myriad failures of discovery, historical approaches to the psychology of scientific discovery can catalog only sufficient causes for discovery. They cannot tell us anything about necessary causes. Laboratory studies allow us to look at both successful and unsuccessful subjects and enable us to determine what distinguishes them.

The challenge posed by investigating the psychology of scientific discovery in "real time" is to find a way to evoke the cognitive processes inherent in scientific discovery while maintaining the experimental rigor that supports sound inferences about human cognition. Despite the difficulty of this task, the Kuhn et al. project is unusually successful in using a set of domains having all the above characteristics. In the next section, we place their work in the context of other attempts to study various aspects of scientific thinking.

### *Laboratory Investigations of the Cognitive Psychology of Science*

Laboratory investigations of scientific reasoning can be classified along two dimensions: one representing the degree of domain specificity or domain generality and the other representing the type of processes involved. Table 1 depicts this characterization of the field. The two rows correspond to the difference between domain-general knowledge and domain-specific knowledge, and the three columns correspond to the major components of the overall discovery process: searching a space of hypotheses, searching a

TABLE 1  
TYPES OF FOCI IN PSYCHOLOGICAL STUDIES OF SCIENTIFIC REASONING PROCESSES

	Hypothesis Space Search	Experiment Space Search	Evidence Evaluation
Domain-specific knowledge and strong methods . . . . .	A	B	C
Domain-general knowledge and weak methods . . . . .	D	E	F

SOURCE.—Klahr (1994).

space of experiments, and evaluating evidence. Psychologists' attempts to disentangle the relative influence of general versus specific knowledge have produced two distinct literatures: one on domain-specific knowledge and "strong methods" and the other on domain-general reasoning processes and "weak methods." This distinction corresponds to the two rows in Table 1.

The three columns in Table 1 reflect a view of scientific discovery as a type of problem-solving process involving search in a problem space (Newell & Simon, 1972). In the case of scientific discovery, there are two primary spaces to be searched: a space of hypotheses and a space of experiments. These spaces are sufficiently different that they require different representations, different operators for moving about in the space, and different criteria for what constitutes progress in the space. Without getting into detail here (see Klahr & Dunbar, 1988), we can convey the importance of the distinction between searching the hypothesis space and searching the experiment space by noting that, in most of the natural sciences, the difference between experimental work and theoretical work is so great as to have individuals who claim to be experts in one but not the other aspect of their discipline.

It is clear that the problems to be solved in each space are different, even though they have obvious and necessary mutual influences. Thus, in our characterization of research on scientific discovery, we emphasize three major interdependent processes: hypothesis space search, experiment space search, and evidence evaluation. In searching the hypothesis space, the initial state consists of some knowledge about a domain, and the goal state is a hypothesis that can account for some or all of that knowledge. When one or more hypotheses are active, it is not immediately obvious what constitutes a "good" experiment. In constructing experiments, subjects are faced with a problem-solving task paralleling their search for hypotheses. That is, they must search in the experiment space for an informative experiment.

The third process—evidence evaluation—involves a comparison of the predictions derived from the current hypothesis with the results obtained from experimentation. In the studies reported in this *Monograph*, the considerable emphasis on strategies for valid inferences deals mainly with this phase of the process.

During the course of scientific discovery, the various cells in Table 1 are traversed repeatedly. However, it is very difficult to study thinking processes that involve all of them simultaneously. Consequently, the early research in the field started with investigations designed to constrain the topic of interest to just one or two cells. As the field has matured, more complex contexts involving multiple cells have been used. We can best illustrate this with a few examples of investigations that involve various cells from Table 1.

*Cell A.*—Investigations falling into this cell are exemplified by McClos-

## STRATEGIES OF KNOWLEDGE ACQUISITION

key's (1983) well-known investigation of people's naive theories of motion. In this kind of study, subjects are asked about their knowledge about a specific domain, but they do not run experiments, and they do not evaluate evidence.

*Cell B.*—In some investigations (e.g., Tschirgi, 1980), subjects are asked to decide which of a set of prespecified experiments will demonstrate the correctness of a prespecified hypothesis. There is no search for hypotheses, and the experiment space search is limited to choosing among alternative experiments.

*Cells D, E, and F.*—Bruner, Goodnow, and Austin (1956) created their classic concept-learning task in order to better understand people's appreciation of the logic of experimentation and their strategies for discovering regularities. Their subjects had to generate hypotheses, choose among "experiments" (i.e., select different cards that displayed specific combinations of attributes), and evaluate the evidence provided by the yes/no feedback that they received. Because the task is abstract and arbitrary, none of the domain-specific cells are involved. Another venerable task that spans cells D, E, and F is Wason's (1960) 2-4-6 task.

*Cell E.*—Studies of people's ability to design factorial experiments (e.g., Case, 1974; Siegler & Liebert, 1975) focus almost entirely on effective search of the experiment space. Domain knowledge is minimized, as are hypothesis space search and evidence evaluation.

*Cells C and F.*—Studies in this category focus on people's ability to decide which of several hypotheses is supported by evidence. Typically, such studies present tables of covariation data and ask subjects to decide which of several hypotheses is supported or refuted by the data. In some cases, the factors are abstract and arbitrary (e.g., Shaklee & Paszek, 1985)—in which case we classify the studies in Cell F—and in others they refer to real-world factors (e.g., plant growth in the context of different amounts of sunlight and water; Bullock & Ziegler, in press). In such cases, subjects have to coordinate their prior domain knowledge with the covariation data in the tables (e.g., Ruffman, Perner, Olson, & Doherty, 1993).

### *Integrative Investigations of Scientific Reasoning*

Research focusing on either domain-specific or domain-general knowledge has yielded much useful information about scientific discovery. However, such efforts are, perforce, unable to assess the interaction between the two types of knowledge. Similarly, the isolation of hypothesis search, experimentation strategies, and evidence evaluation begs some fundamental questions. How are the three main processes integrated? How do they mutually influence one another?

Although many investigations focus on one or two of the cells depicted in Table 1, few studies attempt to traverse the entire matrix. Such investigations are necessary to really understand scientific reasoning because in “real science” both domain-specific knowledge and domain-general heuristics guide scientists in designing experiments and evaluating their outcomes. More informative are tasks requiring coordinated search in *both* the experiment space and the hypothesis space as well as the evaluation of evidence produced by subject-generated experiments.

Kuhn and her colleagues have pioneered this kind of research (cf. Kuhn, 1989; Kuhn, Amsel, & O’Loughlin, 1988; Kuhn, Schauble, & Garcia-Mila, 1992; Schauble, 1990; Schauble, Glaser, Raghavan, & Reiner, 1991), and the present *Monograph* represents yet another valuable extension of the approach. Although others have created similar “discovery contexts” in which to investigate the development of scientific reasoning processes (Dunbar, 1993; Klahr & Dunbar, 1988; Klahr, Fay, & Dunbar, 1993), only Kuhn and her colleagues have combined this integrated approach to scientific discovery with a microgenetic approach. Moreover, with respect to domain-specific reasoning, the Kuhn et al. work represents the first such study that simultaneously utilizes two distinct types of domains and examines the mutual influence of reasoning in one domain on reasoning in the other. One can depict the Kuhn et al. study as a series of layers of 2 × 3 tables, as shown in Figure 1. Each layer represents a single session in which all the cells are traversed, and the series of layers represent the time course of densely connected repeated assessments of how subjects traverse these spaces.

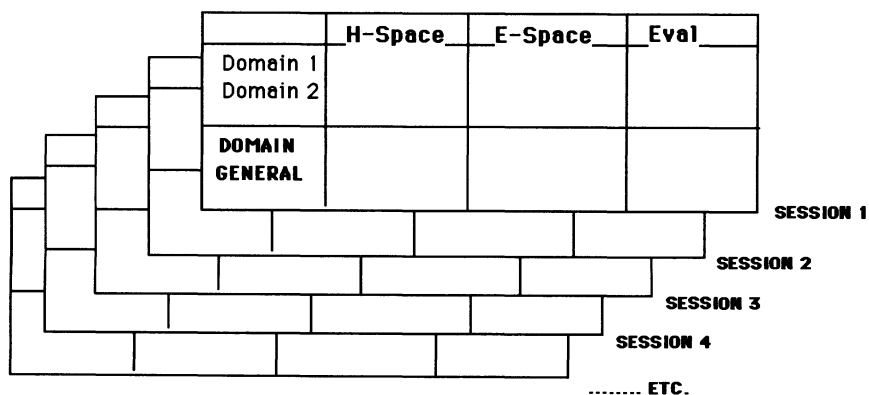


FIG. 1.—Microgenetic study of scientific reasoning

*Transfer of Training*

Another unique and valuable feature of this work is the way in which it focuses on the temporal course of *both* of the knowledge types depicted in Figure 1: knowledge about the specific domain (e.g., boats, school achievement) and domain-general knowledge about scientific reasoning (e.g., valid inclusion). It is these domain-general processes that Kuhn et al. emphasize in their title: *Strategies of Knowledge Acquisition*. Moreover, Kuhn et al.'s ingenious design allows them to assess the extent to which the knowledge acquisition strategies acquired in one domain transfer to another domain. In other words, Kuhn et al. have a direct measure of how domain general such skills really are.

Kuhn et al. comment on the mixed picture provided by earlier investigations of transfer, and they ask, "Why did our subjects show transfer of newly developing cognitive strategies when transfer so often fails to occur in both children and adults?" (p. 100). We find this question particularly intriguing because we have previously described what we believe to be sufficient conditions for transfer: "If the domain is properly analyzed, if instruction is based on the formal analysis, and if what is learned in the base domain and what is transferred to more remote domains are also grounded in the formal analysis, then a powerful idea . . . can be taught and can have an impact on general problem-solving capacities" (Klahr & Carver, 1988, pp. 364–364).

Given the design and results of the present study, it appears that our conditions are, at best, a statement of sufficiency rather than necessity, for the procedure used by Kuhn et al. seems to honor none of our conditions for transfer. In fact, not only is transfer surprising, but so is learning, because the experimenter provides the subjects with no feedback about the efficacy of the inference strategies currently being used to evaluate experimental outcomes. Although the experiments that subjects run give them a substantial amount of feedback about the domain under investigations, there is *no* direct feedback about the next level of knowledge acquisition—the inferencing strategies—yet this is what subjects learned, and this is what transferred from one context to the next.

On the other hand, a careful look at the Kuhn et al. procedure reveals that the kind of carefully elaborated goal structure that we include in our sufficiency list is inherent in the sequence of questions and probes that precede and follow each of the subject's experiments. Although subjects do not get feedback in the traditional sense, they do receive a kind of Socratic dialogue as they are walked through the goal structure that underlies valid inferencing and the coordination of theory and evidence. It is quite likely that subjects have never been presented with such a highly structured sequence of probes about how and why they examined specific pieces of evi-

dence or drew specific conclusions from that evidence. For this reason, the Kuhn et al. study should be viewed not only as a study of transfer but also as a study of transfer of *training* (and a successful one at that!).

### *Theoretical and Methodological Questions*

In addition to the substantial strengths listed thus far, the *Monograph* raises important points that warrant further attention, and we address several of them in this section. We start with the observation that only a limited subset of the many forms of knowledge acquisition were actually studied here: selection of instances and the formation of valid inferences from patterns of covariation and noncovariation. But how much of people's knowledge is acquired through the coordination of theory and evidence? Although attempts to answer this question precisely may founder on the problem of quantification of "amount" of knowledge, it seems to us that very little of one's overall knowledge base comes from experimentation. Instead, most of it comes from generalizations over particular instances, or from reading, or from direct instruction from parents and teachers. It seems that little of what we know about what factors contribute to fast cars or fast boats comes from running experiments (confounded or not) in those domains. Moreover, it is even less likely that we acquire knowledge via experimentation in the social domain than it is in the physical domain. Thus, we claim that the work reported here is not so much about "knowledge acquisition" as it is about the narrower—but still important—context of scientific reasoning.

### *Strategies and Metacognition*

Our second question is also related to another key term in the title of this *Monograph*—*strategies*. Over the past decade or so, the term *strategy* has undergone a transformation from its original use in game theory—that is, a deliberate, rational, intentional scheme to achieve an end—to a more amorphous notion describing any set of organized processes or rules, intentional or not, explicit or not. In its present usage, exemplified here and also by Siegler's focus on "strategy choice" (Siegler & Shipley, 1995), one could easily replace *strategy* with the generic term *process*. Indeed, we question the extent to which it is productive to label these and other knowledge acquisition processes as strategies. Would it make sense to call associative learning, learning from instruction, or learning from reading *strategies*?

This transformation of *strategy* from a well-defined to a generic notion results in even more ambiguity when one discusses *metastrategic* issues, as do Kuhn et al. in their characterization of subjects using strategies "selectively



## STRATEGIES OF KNOWLEDGE ACQUISITION

and variably.” This implies some higher-level agent that controls the selection and variation process and justifies the various levels of metacognition that Kuhn et al. invoke. The problem with this conception is that those processes of strategy selection and strategy variation are not explained at all. This lack of specification is not an uncommon result of invoking metacognitive processes. As Siegler & Shipley (1995) note:

Such metacognitive models are useful for conveying hypotheses about relations among different types of knowledge and for pointing to one way in which intelligent strategy choices can be generated. However, they also have a number of weaknesses. . . . As statements of theory, they generally have been vague regarding the mechanisms that produce the phenomena of interest. Do people make explicit judgments about their intellectual capacities, available strategies, and task demands every time they face a task they could perform in multiple ways? If not, how do they decide when to do so? Do they consider every strategy they could use on the task, or only some of them? If only some, how do they decide which ones? How do people know what their cognitive capacity will be on a novel task or what strategies they could apply to it? The apparent simplicity of metacognitive models masks a world of complexity. (p. 41)

### *Microgenetic Method*

Kuhn and her colleagues represent one of the primary influences on the current reemergence of microgenetic methods, and the current *Mono-graph* is yet another valuable example of the approach. But there are some important differences between the way that Kuhn et al. use the method and the way it is characterized by some of its other advocates (Siegler & Crowley, 1991). Kuhn et al. argue that a primary goal of the microgenetic method is “to accelerate the change process by providing a subject with frequent opportunities over a period of weeks or months to engage the particular cognitive strategies that are the object of investigation” (p. 8). The idea is to run a sort of cognitive “summer camp” that includes extensive exercise of the cognitive skill to be acquired. Siegler and Crowley, on the other hand, do not view the method itself as the necessary cause for the acceleration. Instead, they propose conducting a preliminary analysis of the natural developmental course in a domain and then ensuring that “observations span the entire period from the beginning of the change to the time at which it reaches a relatively stable state” (Siegler & Crowley, 1991, p. 606). For them, the repeated exposures are not so much a way of stimulating or prodding the change process as they are a procedure for generating a high “sampling rate” so that the change process can be observed in detail.

Another important distinction—albeit not a disagreement—is Siegler

and Crowley's focus on cognitive skills that just about everyone acquires in the natural course of development (such as quantity conservation or the min method for single-digit addition). This focus leads Siegler and Crowley to associate with the microgenetic method the determination of a critical window of opportunity during which to observe the changes of interest. In contrast, Kuhn et al. have focused on a skill that few people master without formal training (valid inferences from empirical data). Thus, rather than seek a critical period in which to locate their observations, they made the strategic bet that both adolescents and adults would start at similar levels of knowledge. As their results show, this was indeed the case. Not only did both groups start at similar levels, but both also demonstrated significant changes in their knowledge level over the course of the microgenetic observations.

*On the Logic of Confirmation and Disconfirmation*

The conceptual core of this *Monograph* is Kuhn et al.'s analysis of "inductive causal and noncausal inference." Given the fundamental importance of the strategies that support such inferences, it is surprising that Kuhn et al. make no contact with the extensive literature on "confirmation bias" or with Klayman and Ha's (1987) elegant analysis of the role of confirmation and disconfirmation strategies in rule discovery tasks. Although Klayman and Ha focus on the classic 2-4-6 rule discovery task, in which subjects have to discover a rule that is being used to classify triples of integers (Wason, 1960), their analysis has implications for the work reported here.

Kuhn et al. note that a single instance in which a feature and an outcome co-occur can lead, at best, to what they call a "co-occurrence false inclusion inference" (p. 19). However, if subjects construe their task as a rule discovery task, then they may establish goals to create "experiments" that are (or are not) "instances" of, for example, the "fast rule." To the extent that subjects adopt this rule discovery stance, the literature on confirmation bias is highly relevant to the current *Monograph*. Nearly all previous investigations of the Wason task (e.g., Gorman, 1986, 1989; Wason, 1960) concluded that subjects approach rule discovery tasks with a strong "confirmation bias": a desire to select instances that confirm (" + Htests") rather disconfirm (" - Htests") the current hypothesis. In other words, "people tend to test hypotheses by looking at instances where the target property is hypothesized to be present" (Klayman & Ha, 1987, p. 225).

However, as Klayman and Ha note, there is no logical basis for interpreting + Htests as attempts to confirm or - Htests as attempts to disconfirm. Depending on the relation between the hypothesized rule and the true rule, both + Htests and - Htests can provide either conclusive falsifica-

## STRATEGIES OF KNOWLEDGE ACQUISITION

tion or ambiguous verification of the current hypothesis. Conclusive falsification occurs when a +Htest receives “no” feedback (e.g., when a subject who believes that big motors make for fast cars creates a car with a big motor that runs slowly) or when a –Htest receives “yes” feedback (e.g., when a subject who believes that big motors make for fast cars creates a car with a small motor that runs fast). Ambiguous verification occurs when a +Htest receives “yes” feedback (the big motor does produce a fast car) or when a –Htest receives “no” feedback (the small motor does produce a slow car). The conclusiveness or ambiguity of these outcomes derives from the standard falsificationist arguments (Popper, 1959). Thus, +Htesting “does not necessarily contradict the goal of seeking falsification” (Klayman & Ha, 1987, p. 225). In the general scheme of things, most factors are noncausal rather than causal. Consequently, from an efficiency point of view, one should focus on the plausibly causal factors and run +Htests on them while deferring consideration of the potentially infinite number of noncausal factors until it becomes necessary to explore further.

Further complicating the labeling of subjects’ strategies as *valid* or *invalid* is the fact that, for those cases in which subjects have very strong beliefs about the irrelevance of certain factors, it is inappropriate to fault them for running what an omniscient observer would characterize as confounded experiments. As experimenters, we routinely fail to control all possible confounds. For example, in the current *Monograph*, we can be fairly sure that the proportion of parents and married subjects was greater in the adult sample than in the adolescent sample or that the number of years since the last formal schooling was greater for the adults than for the adolescents. Thus, all comparisons between adults and children are confounded by these factors, yet it would be foolish to call any conclusions about adult-child differences here *false inclusions* or *false exclusions* because these covariates were not controlled.

### *The Equivalence of Experimentation in Social and Physical Domains*

One of the most interesting features of the Kuhn et al. study is the way in which it contrasts performance not only between different problems within a domain (i.e., boats vs. cars) but also between the social and the physical domains. In order to accomplish this, Kuhn et al. faced a formidable challenge in the creation of materials and procedures that would keep all the important aspects of “running experiments” equivalent in both domains. Kuhn et al.’s inventive solution to this problem was to create a set of records that have to be examined by subjects, thereby making it feasible to run “experiments” in the social domain. However, the solution is not entirely satisfactory because many potentially important differences between the two domains—other than the domain as such—remain.

One difference is that, in the physical domain, the causal mechanisms that determine the outcome operate only *after* subjects have selected a set of features, whereas, in the social domain, the features and their effects have been determined *prior* to the selection of the card that shows what those effects were. Although the two domains have identical causal structures, they differ with respect to this potentially important aspect. Recent emphasis on the importance of future-oriented thinking (Haith, Pennington, & Benson, 1994) suggests that this is not a trivial difference, yet all social-physical comparisons are compromised by this confound. The confound could have been eliminated if, in the physical domain, the experimental results were all precomputed and stored in the same kind of card catalog as was used in the social domain.

Furthermore, we would expect that, for most subjects, the process of making inferences (valid or not) on the basis of selected sets of instances is a rare activity in the social domain, although it may be a typical "science class" type of activity. In fact, subjects are likely to have strong personal opinions about the social domains, with a fair amount of associated affect, which is unlikely in the physical domain.

Another potentially important difference between the physical and the social domains is that, although in the physical domain categorical outcomes were used, it was clear to the subject that they were categorizations of things that had an underlying continuum on a ratio scale. In the social domain, categorical labels do not have clear correspondence to a scale. In other words, one distance can be twice as far as another in the race car domain, but, in the social domain, such quantification is impossible.

Finally, in at least one of the physical domains but none of the social domains, subjects understand that the outcomes that they see are not "rigged" by the experimenter. Instead, the laws of physics determine the outcomes. In the social domain, it is never clear the extent to which subjects really believe that the outcomes on the cards realistically portray the social world from which their initial beliefs derive.

All these differences between the social and the physical domains make the Kuhn et al. comparisons between them vulnerable to the very criticism of confounded experimentation and false inclusion that is the subject of investigation in this *Monograph*. (We are aware that all investigations of scientific reasoning are prey to this kind of reflexive criticism, but it seems so serious in this instance as to be unavoidable.)

#### *Levels of Analysis in Microgenetic Research*

Kuhn et al. note that "qualitative case study analysis deserves more respect and use than it receives from most developmental psychologists." They continue, "Qualitative and quantitative analysis can be used in comple-

## STRATEGIES OF KNOWLEDGE ACQUISITION

mentary ways that enhance understanding beyond what would be achieved by either method alone" (p. 000). Although we fully concur with the general claim, we believe that Kuhn et al.'s use of both types of analysis could have been more informative. On the one hand, their quantitative analyses are presented at very high levels of aggregation, with no quantitative information about how individual subjects acquired domain knowledge or how they utilized valid and invalid inclusion and exclusion strategies. For example, all the tables comparing subject's initial and final strategies fail to indicate the extent to which *individual* subjects changed or retained their initial domain theories. On the other hand, Kuhn et al.'s "qualitative analyses" fail to go beyond an informal narrative description of the experience of individual subjects. Some important information is not provided: what in Arnie's background might account for his outstanding performance? When and why did individual subjects use their notebooks in evaluating their experimental outcomes? In addition to providing more of this kind of information, one can perform a more systematic analysis of verbal protocols, seeking characteristic patterns of domain knowledge and inferencing strategies and quantifying aspects of individual protocols by segmenting the protocols into episodes that are, in turn, subject to quantitative analyses (cf. Ericsson & Simon, 1984).

### *Suggestions for Future Studies*

Kuhn et al. characterize their work as "bridge building" (p. 118). As inhabitants of a region with more than 500 bridges (White & von Bernewitz, 1928), we fully appreciate their value. We find, however, that, like many of those in and around Pittsburgh, some of the bridges suggested in this *Monograph* are in need of some skeptical examination. Kuhn et al. claim that their "most basic and possibly single most important result" is that "exercise of reasoning strategies can be a sufficient condition to effect their change" and that, "if reasoning is practiced, consistency in application of sound strategies is likely to improve" (p. 119). Although this view is consistent with much of the rhetoric surrounding radical constructivist approaches to learning, we find two problems with it. First, in its starkest form, the claim is simply not supported by the results reported in this *Monograph*. Second, it leaves the process of knowledge acquisition and strategy change unexplained.

The Kuhn et al. claims quoted suggest that the "exercise" that leads to learning is entirely learner directed. However, as we noted earlier, a crucial feature of their methodology was the use of a systematic set of probes that, in effect, indicated to subjects an underlying goal structure for searching the experiment space and the hypothesis space and for making valid infer-

ences when evaluating evidence. We believe that, absent this highly structured and often-repeated set of questions, the undirected “exercise of reasoning strategies” would lead to little learning. In fact, although Kuhn et al. allude to the “transfer on trial” literature, careful examination of that literature shows that the few successful cases of transfer are precisely those in which either the underlying goal structure of the domain or systematic feedback about the outcome of the strategy use, or both, is provided to subjects. Moreover, there are existing accounts of learning that explain why and how these conditions promote learning and transfer (Anderson, 1993; Anderson, Reder, & Simon, 1995; Singley & Anderson, 1989). Given the importance that Kuhn et al. attribute to efficacy of strategic “exercise,” it would be informative to replicate this type of study with and without the explicit probes that we claim are providing so much of the guidance for strategy change.

Finally, we suggest that, in future studies of this type, both the “qualitative” and the “quantitative” analyses be designed to dig deeper and probe further. To use another local metaphor, it strikes us that the data collected for this *Monograph* represent a potentially rich lode, but one that has only been strip-mined rather than deep mined.

## References

- Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.
- Anderson, J. R., Reder, L. M., & Simon, H. A. (1995). *Applications and misapplications of cognitive psychology to mathematics education*. Unpublished manuscript, Carnegie Mellon University.
- Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. New York: Science Editions.
- Bullock, M., & Ziegler, A. (in press). Scientific reasoning: Developmental and individual differences. In F. Weinert & W. Schneider (Eds.), *Individual development from 3 to 12: Findings from the Munich Longitudinal Study*. New York: Cambridge University Press.
- Case, R. (1974). Structures and strictures: Some functional limitations on the course of cognitive growth. *Cognitive Psychology*, *6*, 544–573.
- Dunbar, K. (1993). Concept discovery in a scientific domain. *Cognitive Science*, *17*, 397–434.
- Dunbar, K. (1994). How scientists really reason: Scientific reasoning in real-world laboratories. In R. J. Sternberg & J. Davidson (Eds.), *Mechanisms of insight*. Cambridge, MA: MIT Press.
- Ericsson, A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.
- Gorman, M. E. (1986). How the possibility of error affects falsification on a task that models scientific problem solving. *British Journal of Psychology*, *77*, 85–96.
- Gorman, M. E. (1989). Error, falsification and scientific inference: An experimental investigation. *Quarterly Journal of Experimental Psychology*, *41A*(2), 385–412.
- Gorman, M. E. (1992). *Simulating science: Heuristics, mental models, and technoscientific thinking*. Bloomington: Indiana University Press.
- Haith, M., Pennington, B., & Benson, J. (Eds.). (1994). *The development of future-oriented processes*. Chicago: University of Chicago Press.

## STRATEGIES OF KNOWLEDGE ACQUISITION

- Klahr, D. (1994). Searching for cognition in cognitive models of science. *Psychology* 5(94), scientific-cognition.12.klahr.
- Klahr, D., & Carver, S. M. (1988). Cognitive objectives in a LOGO debugging curriculum: Instruction, learning, and transfer. *Cognitive Psychology*, 20, 362–404.
- Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, 12(1), 1–55.
- Klahr, D., Fay, A. L., & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology*, 25(1), 111–146.
- Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation and information in hypothesis testing. *Psychological Review*, 94, 211–228.
- Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, 96, 674–689.
- Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific reasoning skills*. Orlando, FL: Academic.
- Kuhn, D., Schauble, L., & Garcia-Mila, M. (1992). Cross-domain development of scientific reasoning. *Cognition and Instruction*, 9, 285–327.
- McCloskey, M. (1983). Naive theories of motion. In D. Gentner & A. L. Stevens (Eds.), *Mental models*. Hillsdale, NJ: Erlbaum.
- Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology*, 29, 85–95.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Penner, D., & Klahr, D. (1995). *The interaction of domain-specific knowledge and domain-general discovery strategies: A study with sinking objects*. Working paper.
- Popper, K. R. (1959). *The logic of scientific discovery*. London: Hutchinson.
- Ruffman, T., Perner, J., Olson, D. R., & Doherty, M. (1993). Reflecting on scientific thinking: Children's understanding of the hypothesis-evidence relation. *Child Development*, 64, 1617–1636.
- Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology*, 49, 31–57.
- Schauble, L., Glaser, R., Raghavan, K., & Reiner, M. (1991). Causal models and experimentation strategies in scientific reasoning. *Journal of the Learning Sciences*, 1, 201–238.
- Schauble, L., Klopfer, L. E., & Raghavan, K. (1991). Students' transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching*, 28, 859–882.
- Shaklee, H., & Paszek, D. (1985). Covariation judgment: Systematic rule use in middle childhood. *Child Development*, 56, 1229–1240.
- Siegler, R. S., & Crowley, K. (1991). The microgenetic method: A direct means for studying cognitive development. *American Psychologist*, 46, 606–620.
- Siegler, R. S., & Liebert, R. M. (1975). Acquisition of formal scientific reasoning by 10- and 13-year-olds: Designing a factorial experiment. *Developmental Psychology*, 10, 401–402.
- Siegler, R. S., & Shipley, C. (1995). Variation, selection, and cognitive change. In G. Halford & T. Simon (Eds.), *Developing cognitive competence: New approaches to process modeling*. Hillsdale, NJ: Erlbaum.
- Singley, K., & Anderson, J. R. (1989). *The transfer of cognitive skills*. Cambridge, MA: Harvard University Press.
- Tschirgi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development*, 51, 1–10.
- Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, 12, 129–140.
- White, J., & von Bernewitz, M. W. (1928). *The bridges of Pittsburgh*. Pittsburgh: Carmer.