The Interaction of Domain-Specific Knowledge and Domain-General Discovery Strategies: A Study with Sinking Objects

David E. Penner and David Klahr
Carnegie Mellon University

PENNER, DAVID E., and KLAHR, DAVID. *The Interaction of Domain-Specific Knowledge and Domain-General Discovery Strategies: A Study with Sinking Objects*. CHILD DEVELOPMENT, 1996, 67, 2709–2727. Recent work on scientific reasoning has largely focused on either domain-specific content knowledge or domain-general reasoning knowledge. This study investigated the interaction between the 2 types of knowledge in a real-world domain in which strict control of variables was not possible. We used a context, sinking objects, in which 10-, 12-, and, 14-year-old children's strong a priori beliefs could be revealed by participant-designed experiments. The results showed that most children initially believed weight alone determined an object's sinking rate. Older, but not younger, participants typically viewed experimentation as a means of exploring the effects of attributes other than weight. However, experimentation did help all children to understand the effects of object shape and material on sinking rates. The results suggest a number of questions for further research, including how children come to understand experimentation as a matter of evaluation rather than demonstration, and the role of unexpected experimental results in driving conceptual understanding.

The development of scientific reasoning skill encompasses two types of knowledge: (*a*) domain-specific knowledge about the natural world, and (*b*) domain-general procedures for generating, assessing, and integrating that knowledge. The former includes substantive knowledge about particular domains (e.g., physics, biology, chemistry) and the latter includes a complex set of cognitive skills used to support scientific discovery, including the search for hypotheses via induction, abduction, or analogy; the design, execution, and interpretation of experiments; and the revision of hypotheses.

Most developmental studies of scientific reasoning have focused on one or the other of these two components. The domain-specific focus is exemplified by Chi and Koeske's (1983) investigation of novice-expert differences in children's knowledge about dinosaurs, and Carey's (1985) study of the development of biological concepts. In contrast, other researchers have used knowledge-lean tasks to investigate children's ability to use domain-general reasoning skills, such as the ability to design factorial experiments (e.g., Case, 1974; Siegler & Liebert, 1975) or the understanding of determinacy and indeterminacy (Fay & Klahr, 1996; Piéraut-Le Bonniec, 1980).

This strategy of decomposition into domain-general investigations and domain-specific investigations seems both reasonable and tractable. However, when people reason about real-world contexts, their prior knowledge is likely to impose strong theoretical biases. These biases may influence the initial choice of hypotheses, and the strength with which they are held (Klayman & Ha, 1987). Additionally, prior domain knowledge may influence the experimental strategies utilized to gather new evidence, as well as which features of the evidence are attended to (Evans, 1989).

Pazzani and his colleagues (Pazzani, 1991; Pazzani, Dyer, & Flowers, 1986) argued that without prior knowledge, or, more

This research was supported in part by grants to the second author from the National Institute of Child Health and Human Development (R01-HD25211) and the A. W. Mellon Foundation. We thank Shari Ellis, Anne Fay, Leona Schauble, and four anonymous reviewers for comments on an earlier draft. Thanks also to the students, parents, teachers, and administrators at The Carnegie Mellon Children's School and The Ellis School, Pittsburgh, PA, for their enthusiastic cooperation and participation.

Correspondence concerning this article should be addressed to David Penner, who is now at the Wisconsin Center for Education Research, UW-Madison, 1025 West Johnson St., Madison, WI 53706.

[*Child Development*, 1996, 67, 2709–2727. 0009-3920/96/6706-0024$01.00]

specifically, prior causal theories, people are reduced to using covariation evidence when faced with novel problems. Similarly, Schauble, Glaser, Raghavan, and Reiner (1991) have pointed out that the discovery process is marked by developing links between knowledge and the processes of experimentation; prior beliefs and theories guide initial exploration by framing the problem and highlighting potentially important variables to investigate. Experimental results then provide information for modifying one's domain knowledge. Central to both of these arguments is the coordination of theory and evidence.

Kuhn and her colleagues (Kuhn, 1989; Kuhn, Amsel, & O'Loughlin, 1988; Kuhn & Phelps, 1982) have argued that young children and many adults find it difficult to make a distinction between theory and evidence and between testing hypotheses and producing results. Further, Carey, Evans, Honda, Jay, and Unger (1989) have shown that adolescents believe that science involves the passive acquisition of knowledge that reveals the true nature of the world. In contrast, Sodian, Zaitchik, and Carey (1991) showed that children as young as 6 years of age are able to determine which of two possible experiments will produce a conclusive test of two competing hypotheses in a simple story context. They argued that participants' performance in Kuhn et al.'s study may reflect their understanding of the task, not their ability to coordinate theory and evidence. Similarly, Ruffman, Perner, Olson, and Doherty (1993) suggested that young children may overly depend on other sources of information, such as prior beliefs, when evaluating experimental evidence. Thus, the use of familiar stimuli may have biased some of Kuhn et al.'s participants to view some of the hypotheses they were evaluating as either highly plausible or implausible, and consequently not worthy of an unconfounded experimental test.

Recently, researchers have begun to investigate scientific reasoning in contexts where this interaction between domain knowledge and general discovery processes can be observed (e.g., Klahr & Dunbar, 1988; Schauble, 1990). However, even these studies have used relatively limited and somewhat arbitrary laboratory simulations of real-world events. Therefore, participants could not assume that their real-world knowledge would be represented in the regularities they were attempting to discover. For example, although children have beliefs about what constitutes a fast car, only some of the causal variables in Schauble's (1990; Kuhn, Schauble, & Garcia-Mila, 1992) race car micro-world were consistent with these beliefs.

The Micro-Domain of Sinking Objects

Our goal in the present study was to further explore developmental differences in the influence of domain-specific knowledge on domain-general experimentation strategies. We used a task in which participants experimented with real objects whose behavior is influenced only by a few natural forces (i.e., a micro-domain of the larger physical world) rather than by experimenter-determined "physical laws" (see Mynatt, Doherty, & Tweney, 1978). Our participants were given the task of discovering which factors determine the rate at which objects sink in water. They could do this by selecting pairs of objects from a predetermined set, dropping them into water-filled cylinders, and observing their relative sink times.

Children (and adults) have strong beliefs about the determinants of sinking rates, derived from their everyday experiences with objects sinking in bathtubs, sinks, and swimming pools, as well as with objects falling through air. However, the underlying physics of objects sinking in liquids is quite complex (Daily & Harleman, 1966). To illustrate, consider what happens as a ball sinks through water. As the ball sinks, its acceleration is determined by the difference between the downward force of gravity and the countervailing forces of friction and buoyancy. The frictional force is determined by several factors—such as object size, cross-sectional area, and surface texture—and increases as the velocity of the ball increases. The buoyant force on the ball depends on the difference between the density of the water and the density of the ball. When the downward gravitational force equals the combined resisting forces of friction and buoyancy, the ball continues to sink at a constant, "terminal" velocity. Because the coefficient of friction associated with an object in a given liquid is not known with any precision, it is extremely difficult to calculate the terminal velocity and, consequently, to predict the sinking time of most objects.

Many investigators (Piaget, 1930/1972; Piaget & Inhelder, 1942/1974; Smith, Carey, & Wiser, 1985; Smith, Snir, & Grosslight, 1992) have argued that density concepts emerge from an initial undifferentiated weight-density concept. This conflation of weight and density is reflected in children's use of the label "heavy" to refer variously to "heavy for them, heavy for objects of its type, or heavy for its size" (Smith et al., 1992, p. 224). Smith et al. (1985) concluded that children come to differentiate weight and density between the ages of 8 and 10.

Recently, however, Kohn (1993) argued that children demonstrate a differentiated density concept around age 5. Kohn had 3- through 5-year-olds and adults predict whether or not objects would sink or float. Four- and 5-year-olds and the adults accurately predicted the outcome for very low- and high-density objects. In contrast, the 3-year-olds responded inconsistently throughout the study. Kohn concluded that although 4- and 5-year-olds do not have a formal understanding of density, they do demonstrate an intuitive understanding that, in questions of buoyancy, what matters is object density.

The present study differs from the work cited above in two respects: (*a*) Although density is *the* factor determining if something will float or sink, it is only one of many factors that determine how quickly something will sink; (*b*) moreover, although previous research has focused on *when* concepts, such as density, emerge, it has not addressed the question of *how* such emerging knowledge affects, and is affected by, children's experiments in a domain. For example, if children hold strong biases about the irrelevance of a dimension, or if they do not recognize it as a dimension at all, then it might appear to an all-knowing observer to be varied unsystematically (e.g., many studies confound day of the week with other experimental variables). Thus, what might be interpreted as faulty domain-general understanding of the logic of factorial design might actually derive from biased domain-specific knowledge (see Klahr & Carver, 1995, for a detailed framework for examining the integration of domain-general and domain-specific knowledge during scientific reasoning tasks).

It would seem to be relatively straightforward for participants to test hypotheses in this micro-domain; they need only drop pairs of objects in water and compare the relative sinking rates. The results could then be used to revise their domain knowledge. However, because some attributes are correlated (i.e., object weight depends both on object size and density), it can be difficult, or impossible, to isolate all of the variables. Consequently, determining causal effects depends on coordinating and evaluating evidence over multiple experiments.

In summary, there are several reasons why this micro-domain is well suited for investigating the interaction of domain-general and domain-specific knowledge. First, as we have already noted, participants are likely to have strong prior beliefs about the causal structure of the domain; second, the influence of this prior knowledge on participants' hypothesis testing and experimentation skills can be assessed with appropriate procedural manipulations; third, the implementation of an experiment is very straightforward (participants simply drop pairs of objects and see which one sinks the fastest); fourth, the phenomenon under investigation is caused by natural laws. That is, in contrast to many studies of children's domain knowledge (e.g., Carey, 1985; Chi & Koeske, 1983; Schauble, 1990) or experimental strategies (e.g., Kuhn et al., 1988; Siegler & Liebert, 1975; Tschirgi, 1980), the experimental outcomes in the present study are conveyed by the materials themselves, rather than by a pictorial, verbal, or computer micro-world depiction.

Pilot Study

Refinement of our research questions was facilitated through a pilot study investigating 4- through 20-year-olds' beliefs about sinking objects. Our goal was to obtain converging evidence about initial knowledge structures and the evolution of such during experimentation by presenting participants with a variety of tasks and repeatedly eliciting their beliefs.

Participants were first asked to explain why objects sink and which attributes are important for sinking quickly in water. They were then shown a set of six common objects (e.g., steel washers, rubber mouse, etc.) that were used throughout the remainder of the study. The 15 possible pairs of objects were presented one at a time, and participants were asked to predict which object of the pair would sink faster in water, and to explain the basis of their prediction. During the final phase of the study, participants were encouraged to explore the domain by dropping objects—singly or in pairs—into tall containers of water. For each experiment, participants explained what dropping the object(s) would tell them about sinking fast in water; following each experiment, they were asked what the outcome told them about sinking fast in water. At the conclusion of the experimental phase, participants were asked to summarize what factors they believed were important determinants of how fast objects sink in water.

The results of the pilot study indicated that the majority of participants of all ages initially believed object weight to be the most important attribute in predicting sinking rate: heavier objects sink fastest. This belief was reflected both in participants' responses to the initial probes, and their justifications during the pairwise task.

The experimental phase was included in order to see how participants' beliefs changed when given opportunity to explore the domain. The results show that the youngest children spent most of their effort attempting to demonstrate that heavy objects sink fastest. The majority of these children concluded that weight alone was the most important attribute for sinking quickly. In contrast, following experimentation, older participants concluded that multiple attributes (e.g., weight, shape, and size) interacted to influence the sinking rate.

The pilot work suggested the following research questions:

1. Do children consider attributes other than weight when exploring this microdomain?

2. Do children interpret the task as one of empirically assessing the impact of various attributes or one of demonstrating the correctness of their beliefs?

3. How does experimentation affect belief revision?

In summary, the work reported below was designed to address questions about the interaction of domain-specific knowledge and domain-general scientific discovery skills in a real-world micro-domain where unambiguous results are not always possible. We explored the manner in which children negotiated the complex relations among object attributes, hypotheses, and possible experiments when these relations were not explicitly pointed out. We used a context in which children had to (*a*) decide which object attributes were of interest, (*b*) generate hypotheses to explore, (*c*) design experiments to test their hypotheses, (*d*) interpret experimental outcomes with respect to the current hypotheses, and (*e*) realize that a single test cannot conclusively answer the question of interest. We focused on an age range, 10- to 14-year-olds, for which the pilot study suggested changes in knowledge structure and experimental strategies occurred.

Method

Participants

Thirty children in three age groups participated: 9- to 10-year-olds (M = 9-10, range 9-7 to 10-3), 11- to 12-year-olds (M = 11-11, range 11-1 to 12-7), and 13- to 14-year-olds (M = 13-11, range 13-1 to 14-8). For the sake of simplicity the three groups will be referred to below as 10-, 12-, and 14-year-olds. All participants were volunteers from a private girls' school in western Pennsylvania.

Materials

Materials consisted of a set of objects and a pair of identical Plexiglas cylinders 91 cm high × 12 cm in diameter. During the experimentation phase, participants could observe sinking rates by dropping objects in the cylinders, which were filled with water to a height of 85 cm.

Obviously, there is no limit to the types of objects that could be dropped in the cylinders. Consequently, we designed an object set that would allow children to utilize a control-of-attributes strategy if they so desired. The object set consisted of eight objects designed to vary along three dimensions: shape (cube or sphere), size (large or small), and material (stainless steel or white Teflon). Table 1 lists the weight, volume, and mean sink times of the eight objects.[1] Stainless steel and Teflon were chosen be-

---

[1] Times are approximate. They are based on the means of 10 "drops" of each object conducted outside of the experimental context. Table 1 also shows the standard deviations in sink times for these 10 trials. In most cases, the differences in sink times between pairs of objects can be reliably distinguished by an attentive observer. The two exception pairs are the large and small steel spheres and the large and small steel cubes, whose mean sink times differ by less than a tenth of a second. However, these same comparisons with Teflon objects (large and small Teflon spheres and large and small Teflon cubes) can be easily discriminated. Thus, comparisons of the steel spheres and cubes may yield inconsistent evidence about the effect of size, although comparisons of the Teflon spheres and cubes produce unambiguous and reliable evidence that the larger object sinks faster.

TABLE 1

WEIGHT, VOLUME, AND SINK TIME OF EACH OBJECT

MATERIAL AND OTHER ATTRIBUTES	SPHERE Large	SPHERE Small	CUBE Large	CUBE Small
Steel:				
Weight (g)	65.40	18.90	64.20	17.60
Volume (cc)	8.18	2.35	7.94	2.20
Sink time (sec)	.58	.62	.83	.91
(SD)	.02	.06	.04	.04
Teflon:				
Weight (g)	18.80	5.60	18.50	5.10
Volume (cc)	8.18	2.35	7.94	2.20
Sink time (sec)	1.23	1.38	1.71	2.04
(SD)	.05	.07	.07	.04

NOTE.—Sink times increase reading from left to right and top to bottom.

cause they differ substantially in density (approximately 2.3 g/cc for Teflon and 8.0 g/cc for stainless steel) and in overall "look": color, surface smoothness, reflectivity, texture, etc. The pilot study showed that participants tend to believe that an object's weight is the most important determinant of its sink time. Therefore, because our procedure did not allow participants to accurately weigh the objects, a list showing the weight of each object was available for their reference throughout the course of the study.

In order to determine the effect of the design attributes on sink times, we ran a step-wise multiple regression of physical attributes against mean sink times. Material alone accounted for 75% of the variance in sink times, reflecting the relative impact of density. Including shape in the regression equation accounted for an additional 17% of the variance. Adding the third design attribute—size—did not increase the adjusted R^2 (because of the small increment relative to the additional degree of freedom), but it did allow accurate prediction of the absolute sink times for each object, yielding the following regression equation: sink time = .45 + .855 (Teflon) + .42 (cube) + .15 (small).

How well would participants do if they followed a strategy that focused solely on weight? Regression analysis suggests that weight is a poor predictor, accounting for only 32% of the variance in sink time for this set of objects. Nevertheless, weight alone correctly predicts the relative sink times in 23 of the 28 possible pairwise comparisons, and in 9 of the 12 same-material pairs. For example, the large steel sphere is both the heaviest and the fastest of all eight objects, the large steel cube is heavier and faster than the five slowest items, and so on (see Table 1). When violations of this simple "heavier is faster" rule do occur, they are quite striking, as when a small sphere is compared with the large cube of the same material: for both steel and Teflon the object with three times the weight of the other has at least a 25% greater sink time. However, if participants randomly sampled only a few of the 28 possible pairs of objects, they might find no violations of the weight strategy.

Design and Procedure

The study was run as a four-phase structured interview that took approximately 25 min to complete. Participants were interviewed individually by the first author. Interviews were videotaped for subsequent analysis.

Phase 1: Initial questions and probes.—Participants were asked to explain why things (in general) sink, whether or not falling in air was the same as sinking in water, and which object attributes are important for sinking fast in water.

Phase 2: Sinking predictions.—Participants were introduced to the object set and familiarized with the list of object weights. Participants were next presented with all 28 possible pairs of objects (in one of two arbitrary orders). For each pair, they were asked to predict which object would sink faster in water, and to justify their answer. Participants then ordered the eight objects according to what they believed to be the slowest to fastest sinking.

Phase 3: Experimentation.—Participants were given an opportunity to test their hypotheses. For each experiment, participants chose one or two objects to drop. Before dropping the object(s), participants were asked (*a*) what dropping the object(s) would tell them about sinking fast in water (i.e., they were asked to generate a hypothesis), and (*b*) to predict which object would sink faster. The experimenter then dropped the object(s) while the participant observed. Participants were asked to describe what happened, and what this outcome "told them about what was important for sinking fast in water."

Participants were required to conduct a minimum of four experiments. If, following their fourth experiment, children indicated they wished to terminate this phase of the study, they were asked, "Are you sure you know what is important for sinking quickly in water? Do you want to drop anything else?" If children expressed uncertainty about the contributing factors to sinking quickly, they were encouraged to conduct more experiments.

Phase 4: Consolidation and summary.—Following the experimentation phase, participants summarized their beliefs about the relevant causal factors. If multiple attributes were mentioned, participants were asked whether or not the attributes were of equal importance. Participants were also asked to describe what they thought the fastest possible (i.e., "ideal") sinking object would be like.

Results

Before providing a quantitative analysis of participants' performance, we provide a qualitative analysis in the form of the detailed behavior of two participants, a 10-year-old (BA) and a 14-year-old (KS). BA and KS were chosen to provide a flavor of the typical range of behavior across the groups. Their experiments, as well as the resulting outcomes and inferences, are shown in Table 2.

BA (10-Year-Old

In the initial probe phase, BA stated that weight was important for sinking quickly; a heavier object "pushes down, so it can't float." Her belief in the importance of weight was also reflected in the pairwise task. In this task BA referred to "heavier weight" in justifying 72% of her selections.

Although participants were asked to generate an explicit hypothesis before each experiment (e.g., does size make a difference?), they often replied with simple predictions (e.g., the large steel cube is going to sink fastest because it is heavier). Predictions may reflect participants' implicit hypotheses, or may simply indicate attempts to demonstrate the correctness of their current beliefs. BA's protocol exemplifies this confusion of hypothesis testing and prediction.

BA's belief in the primacy of weight motivated her first experiment, in which she chose to drop the small Teflon cube and small Teflon sphere (see Table 2). She replied to the hypothesis probe with a prediction: the sphere would sink faster because it was heavier and spherical. BA's reasoning about small differences in weight was typical of many participants: small weight differences were initially believed to greatly affect an object's sinking rate.

After seeing the sphere sink faster, BA stated that it sank faster "because of its shape, and it's a little heavier." In response to the conclusion probe (i.e., what is important for sinking quickly?), BA concluded that being spherical and heavier made it easier to "go through the water."

BA's second experiment involved the small steel and large Teflon cubes. As in her first experiment, BA chose not to state an explicit hypothesis. Rather, she predicted that the Teflon cube would sink faster, since it was both heavier and larger. After observing the steel cube sink faster, BA stated, "I think it would go faster if it was smaller, because it doesn't have as much to go through."

For her third experiment, BA chose both large steel objects. Once again she answered the hypothesis probe with a prediction: the sphere will sink faster because it is heavier. Following this experiment, she stated that the sphere sank faster because its shape allowed it to "slide through the water easier . . . if it's round it will go down faster."

Although BA's final experiment replicated her first experiment (i.e., small Teflon sphere and cube), there is no evidence in her protocol that she was aware of this fact. As in her previous experiments, BA responded to the hypothesis probe with a prediction: the sphere will sink faster, since "the other circles have been going . . . I just think it will cut through the water faster. Also, it's a little bit heavier."

At the conclusion of her experimental phase, BA responded to the consolidation

TABLE 2

SEQUENCE OF EXPERIMENTS AND INFERENCES FOR TWO SUBJECTS

	BA (10-Year-Old)				KS (14-Year-Old)			
TRIAL	Weight (g)	Object[a]	Fastest?	Inference(s)[b]	Weight (g)	Object	Fastest?	Inference(s)
1	5.1	▫		Sphere (√)	64.5	●	X	Sphere (√)
	5.6	○	X	Heavy (√)	64.2	▩		
2	17.6	•	X	Small (√)	5.1	▫		Sphere (√)
	18.5	□			5.6	○	X	Heavy (√)
3	65.4	●	X	Sphere (√)	18.9	•	X	Heavy material (√)
	64.2	▩			18.5	□		Sphere (√)
4	5.6	○	X	Sphere (√)	65.4	●	Tie	Small = heavy (√)
	5.1	▫			18.9	•		

[a] ▫ = small teflon cube; ● = large steel sphere, etc.
[b] √ = inference supported by outcome.

probe by stating that the object's shape was important: spheres were "smoother" and did not have "corners" like cubes; that "if it's heavier, I think it will go down faster"; and "the smaller it is, the faster it will go down." When asked if being spherical, heavy, and small were equally important, BA stated that being heavy and small were the most important for sinking quickly. Interestingly, she attempted to illustrate the trade-off between weight and size by referring to the large and small steel spheres, a pair of objects that BA did not drop. BA predicted that the two objects would sink at the same rate because "this [small sphere] is smaller so it can cut through faster, but this [large sphere] is heavier."

How reasonable are BA's conclusions? In two of her three unique experiments, the heavier object did sink faster; however, in her second experiment, the lighter object sank faster. Thus, although the majority of the evidence supports BA's belief that heavier objects sink faster, there is evidence that weight alone cannot determine an object's sinking rate. As for BA's conclusion regarding shape, two of her three unique experiments varied on shape. In both cases the sphere sank faster than the cube, supporting her conclusion that spheres sink faster than cubes. However, BA's conclusion about the effect of size is based on a single experimental outcome.

BA's choice of experiments, interpretation of experimental results, and final conclusions reflect the effects of both her experimental strategies and her prior beliefs. First, rather than stating hypotheses, and then designing experiments to test them, BA selected object pairs that would *demonstrate* that heavy, spherical objects sink faster than other objects. That is, BA adopted what Schauble, Klopfer, and Raghavan (1991) call an engineering, rather than a scientific, stance.

Second, the impact of BA's prior beliefs is reflected in the way in which she interpreted the outcomes of her experiments: large differences in sinking rate were at times attributed to small differences in weight. This led BA to conclude that any difference in weight has a major impact on sinking rate. Finally, we see that BA's adoption of the engineering stance led her to conclude that smaller objects sink faster, even though she had only a single outcome to support this belief.

In summary, BA, as did most of our participants, understood that her task was not simply to find the fastest sinking object. Further, she realized that to evaluate attributes, she needed to look at pairs of, rather than single, objects. BA's protocol also illustrates the difficulty of integrating knowledge across experiments. For example, in this study, weight is correlated with size, shape, and material. Consequently, understanding the effect of weight requires coordinating outcomes from multiple experiments. Although BA's second experiment provided evidence that the heaviest object does not always sink the fastest, she failed to note this fact at the conclusion of the experiment.

KS (14-Year-Old)

In response to the initial probe, KS stated that objects should be heavy, large, and have a small cross-section in order to sink quickly. However, KS justified 71% of her pairwise predictions solely on the basis of which object was heavier; a further 10% of her justifications involved both heavier weight and spherical shape; none of her justifications referred to size.

KS chose the large steel sphere and the large steel cube for her first experiment (see Table 2). In response to the hypothesis probe, she stated that she was interested in seeing how object shape affected the sinking rate. When asked to predict the outcome, KS stated that the sphere would sink faster because of its shape. After watching the sphere sink faster, KS inferred that the sphere sank faster because it was "circular, and it weighs a little more." Note that although KS did not explicitly indicate that she was investigating the effects of weight, she concluded that *all* differences between objects were important. When asked to explain why being "circular" is important for sinking quickly, KS stated that spherical objects sink faster because the water "just sort of goes off the sides, and doesn't get trapped underneath."

For her second experiment, KS again stated that she was interested in testing the effect of object shape. For this experiment, KS selected the small Teflon cube and the small Teflon sphere, predicting that the sphere would sink faster because it "weighs more, and has a circular shape." After watching the sphere sink faster, KS stated that being spherical and heavier were important for sinking quickly, once again asserting that being spherical allows the water "somewhere to go, and the square doesn't."

KS chose the small steel sphere and large Teflon cube for her third experiment, because "they weigh about the same amount, but they're totally different shape; the steel ball is probably more dense than the other one." Note that in contrast to her first two experiments, KS appears to believe in this case that a small weight difference will have little effect on the relative sinking rates. Moreover, she did not test a hypothesis for her third experiment; rather, she predicted that the sphere would sink faster because of its shape and material. After noticing that the sphere did sink faster, KS stated that a "dense material" and spherical shape were important for sinking quickly.

For her final experiment KS selected the large and small steel spheres, indicating that she wanted to see "whether size or weight matters more." That is, KS attempted to test for an interaction between attributes. However, in response to the prediction probe, KS predicted that the large sphere would sink faster (as does our regression model) because it weighed more than the small sphere. This suggests that she believed weight to be more important than size.

In contrast to all other possible pairs of objects in this set, the two steel spheres sink at virtually the same rate.[2] When asked to explain this outcome, KS stated, "I have no idea. Well, the small one may go faster because it's smaller, and the large one because it weighs more. It might equal out." That is, KS speculated that sinking rate, at least in this case, is the result of an interaction between attributes.

Following the experimental phase, KS summarized her current understanding of the micro-domain: spherical shape, small size, and dense material equally affect an object's sinking rate.

As with BA, KS's experimental outcomes provided differing degrees of support for her final conclusions. For example, only on one experiment did KS receive evidence that object material might make a difference; however, on three experiments she received evidence that spheres sink quicker than cubes. Although KS concluded that shape, size, and material were equally important, her final experiment revealed that the relation between size and sinking rate is not that simple. Thus, although KS conducted some potentially informative experiments, especially her final one, she had some difficulty making sense of the results.

KS's choice of experiments, interpretation of experimental results, and final conclusions reflect the effects of both her prior beliefs and experimental strategies. Although BA primarily attempted to demonstrate the correctness of her beliefs, KS's explicit hypotheses reflect her understanding of the task as one of evaluating the effects of the different attributes. However, her predictions show that she is still somewhat influenced by her a priori belief that being heavy is most important for sinking quickly: in two of three experiments, small weight differences were seen as having an important effect on differences in sinking rates.

KS's changing understanding of the role of object weight is reflected in her final two experiments. In her third experiment, KS explicitly stated that the small difference in weight was unimportant; in her final experiment, she attempted to directly contrast weight and size. Moreover, although she initially believed object weight to be important, KS did not include this attribute in her final summary.

Overall, KS's protocol reflects behaviors characteristic of many of the older participants: she understood that her task was to explore the space of possible experiments. However, like many participants, she was also influenced by her a priori beliefs, particularly the effect of weight, in interpreting some of her experimental outcomes.

One final note. KS's behavior during her final experiment is interesting more for what she did *not* do, rather than what she did do. Given the confusing and surprising outcome of her fourth experiment, KS might have applied a heuristic to "explore surprising results" that has been identified in other studies of scientific reasoning (Klahr, Fay, & Dunbar; 1993; Klayman & Ha, 1987; Kulkarni & Simon, 1988). But she did not; rather, she concluded with an ad hoc explanation of how different attributes of the two objects produced equivalent sinking rates. We will return to the issue of children's reluctance to explore surprising results in the discussion section.

---

[2] As discussed above, sinking involves two phases: acceleration, followed by constant velocity. For the given depth of water used in this study, the two steel spheres essentially remain within the acceleration phase.

TABLE 3

ATTRIBUTES DEEMED CAUSAL PRIOR TO AND AFTER EXPERIMENTATION BY PARTICIPANT AND EXPERIMENTAL ORIENTATION

		INITIAL CAUSAL ATTRIBUTES					FINAL CAUSAL ATTRIBUTES			
AGE	PARTICIPANT	W	SH	S	M	HYPOTHESIS ORIENTATION?	W	SH	S	M
10 …	s1	+				No	+		+	
	s2	+				No	+	+		
	s3	+				Yes	+	+		+
	s4	+				No	+			+
	s5	+				No	+	+		
	s6	+				No		+		
	s7	+				No	+		+	
	s8	+				Yes		+		+
	s9	+				Yes	+	+	+	
	s10	+				Yes	+	+		
12 …	s11	+				No	+	+		+
	s12	+				No				+
	s13	+				Yes	+	+		+
	s14	+				Yes		+		+
	s15	+				No	+	+		
	s16	+				Yes	+	+		
	s17	+	+			Yes	+	+	+	
	s18	+	+			Yes	+	+		
	s19	+	+			Yes	+		+	
	s20		+	+		Yes	+	+		+
14 …	s21					Yes	+	+		
	s22	+				No	+	+		
	s23	+				Yes	+	+		
	s24	+				Yes	+	+	+	
	s25	+	+			Yes	+	+		+
	s26	+	+			Yes	+	+		
	s27	+		+		Yes	+	+		
	s28	+	+	+		Yes		+	+	+
	s29			+		Yes	+	+		+
	s30				+	No		+		+

NOTE —Attributes are W: weight, SH: shape, S: size, M: material Initial causal attributes represent Phase 1; final causal attributes represent Phase 4

Having provided a flavor for the participants' behavior, we now present results for all of the participants Specifically, we will report on participants' initial beliefs, experimental intentions, inferences, and final knowledge states.

Initial Beliefs

It is important that we assess not only the beliefs children hold for this microdomain prior to experimentation, but also the consistency with which they apply their beliefs across tasks (Carey, 1985; diSessa, 1988). Phases 1 and 2 of the interview addressed these issues.

Participants' initial beliefs about the attribute(s) that determine sinking rates are listed in Table 3. In response to the initial probe, all of the 10-year-olds, six of the 12-year-olds, and three of the 14-year-olds claimed that weight determined the sinking rate: heavier objects sink faster than light objects. Three of the 12-year-olds and two of the 14-year-olds believed that objects should be both heavy and spherical. The remaining 12-year-old believed that objects need to be spherical and small in order to sink quickly. Four of the remaining five 14-year-olds believed that (*a*) being heavy and large; (*b*) being large, heavy, and spherical; (*c*) having a small cross-sectional area (e.g., "as skinny as possible"); or (*d*) being made out of a heavy material was important for sinking quickly. The remaining 14-year-old refused to answer the probe.

These results are in accord with the findings from our pilot study: most participants believe that weight is the major determinant of sinking rate. However, diSessa (1988) has argued that people's beliefs about physical laws are often fragmented and applied inconsistently. In order to assess the overall stability of the weight bias suggested by the initial probe, we examined the extent to which children's responses to the pairwise comparisons revealed the same bias. The 10-, 12-, and 14-year-olds invoked heavier weight to justify 78% (SD = 20%), 78% (SD = 21%), and 79% (SD = 16%) of their pairwise selections, respectively.

Virtually all of the pairwise comparisons involved objects for which the heavier object does sink faster. However, for five of the problems, the lighter object sinks as fast or faster than the heavier object. Participants' behavior on this subset of problems may be especially revealing, since the problems violate the "heavier is faster" belief. However, analyzing the justifications revealed a pattern similar to that found for the larger problem set: 88% (SD = 22%), 86% (SD = 21%), and 86% (SD = 19%) of the 10-, 12-, and 14-year-olds, respectively, referred to heavier weight in justifying their object selection. Moreover, seven of the 10-year-olds and six each of the 12- and 14-year-olds used the heavier weight justification on all five of these problems. Thus, performance on the pairwise task provides further evidence of the pervasiveness of the "heavier-sinks-faster" belief.

The consistency with which participants applied their beliefs prior to experimentation was assessed by computing the proportion of their pairwise comparisons that were consistent with the implicit pairwise comparisons derived from the rank-ordering task in phase 2. For example, SD, a 12-year-old, gave inconsistent responses to the two types of assessment: On the ranking task she chose the large steel sphere as the fastest sinking object; in contrast, on the pairwise task she predicted that the small Teflon sphere would sink faster than the large steel sphere. For each participant, each pairwise comparison was scored as being consistent or inconsistent with that participant's rank ordering of the objects. There was no main effect of age with respect to consistency, $F(2, 27) = 1.61, p > .05$). Mean consistency was 68% (SD = 17%), 81% (SD = 10%), and 73% (SD = 18%) for the 10-, 12-, and 14-year-olds, respectively. That is, all three groups had some difficulty consistently applying their a priori beliefs to a concrete task.

The initial probe, pairwise comparisons, and ranking task were all designed to identify children's initial beliefs, and the consistency with which they reasoned on the basis of these beliefs. Results of the three tasks reflect children's predominant belief that an object's weight determines its sinking rate. Given this finding, the question becomes, how will children explore the micro-domain of sinking objects, and how will experimental outcomes affect their beliefs?

Experimentation

As in previous developmental studies of scientific reasoning (e.g., Schauble, 1990), children in the present study quickly concluded that they understood the micro-domain. Few participants conducted more than the required four experiments, even though they (*a*) were free to do so, and (*b*) had seen all 28 unique pairs during the pairwise comparison phase. On average, the 10-, 12-, and 14-year-old participants conducted 4.5 (SD = .5), 4.3 (SD = .5), and 4.4 (SD = .5) experiments, respectively. Only five of the 10-year-olds, three of the 12-year-olds, and four of the 14-year-olds conducted more than four experiments. Thus, at all ages participants extracted much less information from the experimental phase than they could have.[3]

Previous research has shown that people often view experimentation as a means of demonstrating the correctness of their beliefs (Klayman & Ha, 1987, 1989; Schauble, Klopfer, & Raghavan, 1991). Given children's initial weight bias, we were interested in seeing which attributes children chose to explore; moreover, would they attempt to confirm their beliefs, or investigate the effects of other attributes?

Experimental rationale.—Each experiment was coded with respect to the participant's explicitly stated intent. *hypothesis-oriented* experiments included statements about investigating the effect of an attribute, such as, "I wonder if shape has something to do with it." In contrast, *prediction-oriented* experiments were accompanied only by

---

[3] Although subjects were free to run single-object experiments, it is clear that they realized that, absent a timing device, such experiments provided little useful information: Only three of 131 experiments involved single objects.

statements about the predicted outcome, such as, "the bigger one will drop faster." The protocols were scored by a second person; there was 90% agreement between assessments. Differences were settled through discussion.

If children begin experimentation with an engineering stance (Schauble, Klopfer, & Raghavan, 1991), this should be reflected in a prediction-oriented, rather than hypothesis-oriented, approach to their first experiment. Three, seven, and nine of the 10-, 12-, and 14-year-olds, respectively, stated a hypothesis for their first experiment, $\chi 2(2, N = 30) = 8.04, p < .05$. Post-hoc analysis of the hc cell contributions showed that fewer 10-year-olds, but more 14-year-olds than would be expected by chance, stated an explicit hypothesis for their first experiment, $p < .05$ for both comparisons. This result suggests that around the age of 12 children begin to understand that their task is to test attributes rather than predict results.

We classified children as having a hypothesis orientation if they generated an explicit hypothesis on three or more of their experiments; otherwise they were classified as having a prediction orientation. Four, seven, and eight of the 10-, 12-, and 14-year-olds, respectively, were classified as having a hypothesis orientation. Moreover, there was an interaction between age and orientation with respect to the number of initial causal attributes (see Table 3). All of the 10-year-olds, regardless of orientation, initially believed a single attribute to be important. In contrast, in each of the two older groups, approximately 50% of the hypothesis-oriented children, but none of the prediction-oriented children, initially believed multiple attributes to be causal. This suggests that the older children's prior knowledge may have influenced their decision to conduct experiments that allowed them to test the relative effects of different attributes.

Together, the results provide some evidence that though all children held similar pre-experimental beliefs, older children viewed experimentation as a means to explore the effects of different attributes. In contrast, the youngest children appear to approach experimentation as an opportunity to demonstrate the correctness of their domain beliefs.

Although children might hold a hypothesis orientation, deciding which attributes to explore is independent of experimental intent. That is, children might still focus solely on testing their a priori beliefs, regardless of experimental orientation. Given the widespread belief in the primacy of object weight, we wished to see whether children investigated other attributes, and if so, when. Two participants at each age conducted only weight-based trials. There was no main effect of age with respect to the first non-weight-based trial for the remaining children, $F(2, 21) = 3.27, p > .05$. The mean trial on which the first non-weight-based experiment was conducted was 2.9 (SD = 1.4), 2.3 (SD = 1.0), and 1.5 (SD = .8) for the 10-, 12-, and 14-year-olds, respectively. This trend suggests that the older children were somewhat less fixated on demonstrating the effect of weight than were the younger children.

Control of attributes.—One focus of many studies of scientific reasoning has been on children's developing understanding of a control-of-attributes strategy, and the inferential power such a strategy affords (e.g., Kuhn, 1989; Sodian et al., 1991; Tschirgi, 1980). A complete control of attributes was not possible in the current study. However, by focusing on the three design attributes, shape, size, and material, participants could attempt to minimize the number of uncontrolled attributes (e.g., weight) in a given experiment. We were interested in seeing if there was an increase in a control-of-variables strategy with increasing age.

Only 12 of the 28 distinct object pairs exemplify an unconfounded contrast of a single design attribute. Therefore, if participants chose pairings at random, then, on average, 43% of their experiments would be "unconfounded." There was no main effect of age with respect to the proportion of unconfounded experiments conducted, $F(2, 27) = 2.88, p > .05$. However, there was a trend for older children to choose such unconfounded pairs more often than did the younger children: 58% (SD = 28%), 78% (SD = 25%), and 82% (SD = 16%) of the 10-, 12-, and 14-year-olds' experiments, respectively. The 12- and 14-year-olds conducted significantly more single-attribute experiments than would be expected by chance, $t(9) = 4.30, p < .001$, and $t(9) = 7.46, p < .001$, for the two groups, respectively.

Although we previously argued that 10-year-olds have a differentiated understanding of weight and density, other work (e.g., Piaget & Inhelder, 1974) argues that this understanding continues to develop into early adolescence. It is possible that the 10-year-olds' difficulty generating unconfounded experiments may be the result of this continuing conceptual development. However, isolation of shape is independent of the density-weight issue. Thus, a difference in the proportion of unconfounded shape-based experiments would provide some support for a separation of domain-general and domain-specific knowledge. There was a main effect of age with respect to the proportion of such experiments, $F(2, 27) = 4.34$, $p < .05$. Only 57% (SD = 32%) of the 10-year-olds' shape-based experiments were unconfounded, compared with 81% (SD = 27%) and 90% (SD = 16%) for the 12- and 14-year-olds, respectively. Fisher PLSD tests showed that the two older groups differed significantly from the 10-year-olds, $p < .05$ for both comparisons.

A hypothesis orientation suggests the use of a control-of-attributes strategy. However, participants' orientation may have little impact on how they test their hypotheses. To explore this issue, we tested the relation between participants' experimental orientation and their generation of single-attribute experiments. There was a trend for participants with a hypothesis orientation to utilize a control-of-attributes strategy more often than those with a prediction orientation, 76% (SD = 24%) and 64% (SD = 28%) for the two groups, respectively, $F(1, 28) = 1.43$, $p > .05$. This suggests that an understanding of a control-of-variables strategy develops somewhat independently of an understanding of science as the testing of hypotheses.

Inferences

Effective discovery depends both on experimental design and on inferences based on experimental outcomes. Considerable research (e.g., Kuhn, 1989; Kuhn et al., 1988; Kuhn, Garcia-Mila, Zohar, & Andersen, 1995; Siegler & Liebert, 1975; Sodian et al., 1991; Tschirgi, 1980) has focused on people's construction of valid inferences in experimental situations where perfect control of attributes is possible. However, in the current study, attributes covary. For example, weight and size covary, making it impossible to design a single experiment to test for the effect of size on sinking rate. Consequently, determining the effect of a given attribute may require examining the results of multiple experiments. Thus, rather than speaking of valid or invalid inferences, it is more appropriate to consider the degree to which inferences are supported by a body of evidence.

Overall, participants' initial beliefs were reflected in their causal inferences: 57% (SD = 25%), 39% (SD = 18%), and 37% (SD = 16%) of the 10-, 12-, and 14-year-olds' inferences, respectively, were based on weight; collapsing weight and shape together accounted for 82% (SD = 13%), 84% (SD = 20%), and 83% (SD = 18%) of the inferences across the three groups. Although all participants generated weight and shape inferences, only four of the youngest children and one child in each of the other groups made size-based inferences, although two or three participants at each level made a material-based inference. Although we intentionally avoided using the term "density" during the study, four of the oldest children referred to density during experimentation.

Table 4 reflects the degree of confirming, disconfirming, or mixed (i.e., participant received both confirming and disconfirming support) experimental support participants received for their final conclusions. The degree to which participants' final conclusions are supported by experimental evidence is a reflection of their understanding of the need to attend to all the evidence they generate. As the table shows, conclusions about material were based solely on confirmatory experimental outcomes. Although the majority of participants' decisions about shape were supported by multiple confirmatory outcomes, one 10- and one 12-year-old had a single disconfirming outcome in addition to two or more confirmatory outcomes. Additionally, one 14-year-old concluded that a cubic shape was important, although she had no evidence supporting this conclusion.

Although relatively few participants concluded that size was important, 50% of the participants who did received only confirmatory support for this conclusion. Of the remaining children, one 10-year-old had two confirmatory outcomes and one disconfirmatory outcome; one 14-year-old received one confirmatory and one disconfirmatory outcome; a final 14-year-old received two disconfirmatory outcomes.

The majority of participants concluded that heavier objects sink faster than light objects. Across groups, between 20% and 50% of participants received only confirmatory support for this conclusion; all but two of the remaining participants received a single disconfirmatory outcome in addition to three

TABLE 4

EXPERIMENTAL SUPPORT FOR FINAL ATTRIBUTES BY NUMBER OF PARTICIPANTS PER AGE GROUP

ATTRIBUTE AND TYPE OF OUTCOME	NUMBER OF PARTICIPANTS		
	10-Year-Olds	12-Year-Olds	14-Year-Olds
Weight:			
All confirmatory	4	5	2
Mixed	4	3	6
All disconfirmatory			
Shape:			
All confirmatory	8	8	9
Mixed	1	1	
All disconfirmatory			1
Size:			
All confirmatory	1	2	
Mixed	1		1
All disconfirmatory			1
Material:			
All confirmatory	2	5	4
Mixed			
All disconfirmatory			

or more supporting outcomes. However, two 10-year-olds decided that heavy objects sink faster even though the majority of their evidence did not support this conclusion.

As the two sample protocols illustrate, participants varied in their beliefs about whether or not small differences in weight are important. Although weight is a continuous attribute, the objects used in this study can be grouped into three categories: heavy (approximately 65 grams), medium (approximately 18 grams), and light (approximately 5 grams). This categorization would allow participants to ignore small weight differences (e.g., approximately 1 to 2 grams) between objects. Alternatively, participants might consider *any* weight difference to be relevant.

To explore participants' sensitivity to small weight differences, we focused on those experiments in which objects differed by less than 2 grams. Between 60% and 80% of the experiments at each age level involved such pairs of objects. We analyzed these near-weight experiments in terms of whether or not participants explicitly stated that any difference in weight contributed to differences in sinking rates. The analysis revealed that the 10-year-olds claimed that small weight differences mattered on 72% (SD = 42%) of their near-weight experiments, compared with 47% (SD = 48) and 20% (SD = 16%) for the 12- and 14-year-olds, respectively, $F(2, 26) = 3.62$, $p < .05$. Fisher PLSD tests revealed that the 14-year-olds disregarded near-weight differences significantly more often than did the 10-year-olds, $p < .05$. This suggests that the oldest, but not the youngest, children usually considered the magnitude of the weight difference when interpreting experimental outcomes.

Surprising Results

Our finding that children prefer to demonstrate their knowledge of a domain rather than rigorously test their hypotheses is consistent with much of the literature on the psychology of scientific reasoning. However, even a propensity to demonstrate knowledge can lead to an unexpected experimental outcome. A number of researchers (Klahr et al., 1993; Klayman & Ha, 1987; Kulkarni & Simon, 1988) have argued that surprising results play an important role during scientific discovery. Since surprising results may suggest a critical misunderstanding of a domain, they may lead scientists to establish a new goal: the exploration of alternative causal mechanisms. These explorations may induce scientists to generate new hypotheses, gather new data, or modify their current beliefs.

In the current study, dropping the large and small steel spheres produced a surprising outcome for participants holding a "heavier objects sink faster" belief: The sink times for the two objects are indistinguishable even though the large sphere weighs more than three times as much as the small sphere. Given the purported importance of surprising results, we were interested in seeing how children responded to this experimental outcome.

Eight participants (one 10-year-old, one 12-year-old, and six 14-year-olds) experimented with the two steel spheres. Although all of these participants acknowledged that the outcome was unexpected, only one participant, a 12-year-old, appeared to utilize the "exploit surprising results" heuristic described by Klahr et al. (1993). SD followed up the unexpected result by generating an analogous experiment with the large and small Teflon cubes. She explained that she wanted to see if size did not matter, as long as shape and material were held constant. In contrast to the steel spheres, the two Teflon cubes do sink at different rates. Thus, following this experiment SD had contradictory outcomes regarding the effect of size on sinking rate. However, SD concluded that size has no effect on the sinking rate, and subsequently dropped this line of experimentation. Her final conclusions made no reference to object size.

Given that all eight participants were surprised by the fact that the two steel spheres sank at the same rate, why did only one person explore this outcome? Exploring a surprising result requires two independent steps. First, a person must recognize the inadequacy of her current knowledge to explain the result. Once a person recognizes the inadequacy of her current knowledge, she must then set an explicit goal to explore the surprising outcome (Klahr et al., 1993). However, Kuhn (1989) has noted that many people have difficulty separating theory and evidence. In response to unexpected results they often simply add the new information to their knowledge base, producing an amalgamation of ad hoc, and sometimes contradictory, beliefs that diSessa (1988) refers to as knowledge in pieces. Thus, if an individual is able to fit an unexpected outcome into her current beliefs, there is no need to explore the outcome—it may be surprising, but it is explainable. All eight of the participants were surprised by the fact that the two steel spheres hit the bottom at the same time. However, only one participant, SD, believed that the outcome did not fit within the scope of her current beliefs. Consequently, she explicitly set out to explore the effect of size once material and shape were held constant.

In contrast to SD, the remaining seven participants simply explained the unexpected result by referring to isolated pieces of their current beliefs. Three participants concluded that the experiment showed that being spherical was important for sinking quickly; two participants explained that the weight of the heavy object was offset by the size of the small sphere, "the smaller one, it was easier for gravity to pull it down, and the big one because it was heavier." Of the two remaining participants, one believed the experiment showed that weight and shape were equally important, although the remaining participant claimed that the outcome showed weight did not affect the sinking rate. Note that the outcome of the experiment does not support conclusions about shape or weight; further, all seven participants had conducted experiments that contradicted these conclusions.

Knowledge Revision

In our pilot study, even physics majors failed to come to a complete understanding of this micro-domain; therefore, we did not expect children to come to one either. However, we were interested in seeing how experimentation affected their beliefs. More specifically, we wanted to explore developmental differences in the extent to which participants use experimentation as an opportunity either to evaluate the effects of different attributes or to demonstrate their a priori beliefs.

One way to assess belief revision is to examine the number and type of attributes each participant thought important before and after experimentation (see Table 3). In response to the initial probe, participants listed a mean of 1.3 attributes, predominantly weight, as important for sinking. Following experimentation, they listed 2.3 attributes, most often weight and shape, as causal, $F(1, 27) = 59.60$, $p < .001$. This suggests that most participants modified their conceptualization of the micro-domain. However, it does not reflect the relative importance children assigned to the different attributes following experimentation. Three 10-, five 12-, and one 14-year-old concluded that weight was the most important attribute. In contrast, two 10-, two 12-, and seven 14-year-olds believed that each of their stated attributes was equally important.

Participants' description of their "ideal" objects (i.e., combinations of attribute values that will produce the fastest sinking object possible) provides a measure of the consistency with which they applied their new knowledge. We measured consistency by comparing participants' postexperimental list of attributes, and the attributes they stated for their ideal objects. For example, AK concluded that a "heavy material and spherical shape" were important for sinking quickly, describing her ideal object as "a ball, made from very heavy material." Thus there is a 100% consistency between AK's summary statement and her description of her ideal object.

There was no main effect of age with respect to consistency, $F(2, 27) = .84$, $p > .05$. Mean consistency was .48 (SD = .39), .60 (SD = .31), .68 (SD = .34) for the 10-, 12-, and 14-year-olds, respectively. That is, although children learned that an object's sinking rate is determined by more than one attribute, their descriptions of their ideal objects often included attributes not found in their summary statements. Most commonly, participants described their ideal object with respect to a certain type of material, often steel, although they had not previously mentioned material as an important attribute. It is possible that asking the children to describe their ideal object may have prompted their use of an "object" schema that includes a "material" slot. In contrast, during the experimentation phase, children were testing individual attributes, not objects. Consequently, the difference in the two tasks may be responsible for the low consistency. Alternatively, the relatively low degree of consistency across groups may reflect that although children increased their knowledge during the study, this knowledge is to some degree still fragmented (e.g., diSessa, 1988).

Discussion

The goal of this study was to investigate developmental differences in the interaction between children's prior domain beliefs and their experimental strategies. Although there are limitations to any cross-domain generalizations possible from the study of a single micro-domain, this initial foray enabled us to explore the extent to which children were able to negotiate the complex relations among object attributes, hypotheses, and possible experiments when these relations were not explicitly pointed out. This research may provide a framework with which to study other domains. We used a micro-domain in which children held strong and consistent (although incorrect) prior beliefs, and in which they had to decide which attributes were of interest, generate hypotheses to test, design experiments to test their hypotheses, and interpret experimental outcomes.

The study enabled us to address a number of questions. First, how do children construe the task they are presented with? Carey et al. (1989) have argued that children neither view science as the construction of explanations of real-world phenomena nor understand that knowledge grows through a process of cumulative testing and revision. Similarly, Schauble, Klopfer, and Raghavan (1991) proposed that people often interpret scientific reasoning tasks as attempting to produce a desirable result, such as finding the fastest sinking object.

Children in our study varied in their understanding of the task. Unlike many of the children in Schauble's (1990) race car study, none of our participants interpreted the task as simply finding the fastest sinking object. That is, they all recognized that the task demanded evaluating pairs of objects. However, the 10-year-olds' tendency to run experiments without explicit hypotheses challenges Sodian et al.'s (1991) claim that 6-year-olds understand experimentation as means of acquiring information about a domain.

It is possible that our pairwise task may have modeled choosing pairs of objects for the children. However, Schauble's methodology also included a procedure in which children were shown pairs of stimuli. There is a significant difference in the two procedures: our study required children to choose objects from a fixed set; Schauble's participants had to construct test objects. Although it is hard to resolve this issue at this time, it is reasonable to believe that young children are sensitive to such procedural variations.

Second, how does entrenched knowledge influence the experimental strategies used to explore the micro-domain? Prior beliefs had a strong effect on participants' experimental intent; the majority of participants across groups focused largely on demonstrating the primacy of object weight. There is some indication that a bias to demonstrate the correctness of one's beliefs does decrease with age: older children were more likely to begin testing nonfavored attributes earlier in the experimental phase than were younger children. Moreover, the older children were more consistent in their approach to experimentation as a matter of testing hypotheses. Regardless of initial beliefs and experimental strategies, all participants concluded the experimental phase believing that weight alone was insufficient for determining an object's sinking rate.

However, our data are not conclusive about the nature of participants' knowledge change. It is possible that even the youngest participants initially believed that multiple attributes mattered, but chose to focus solely on weight. Consequently, experimentation may not have led participants to revise their beliefs so much as affecting the relative weighting they gave to the different attributes.

Research has shown that even preschool children can recognize the logical requirements necessary for testing a hypothesis (Kuhn et al., 1988; Siegler & Liebert, 1975; Sodian et al., 1991; Tschirgi, 1980). However, these studies have largely presented participants with a few discrete categorical attributes. In such cases, people have little difficulty utilizing a factorial design. However, as Klayman and Ha (1987) have shown, most real-world situations, such as the current study, involve correlated attributes. In such cases, it is not always possible to utilize a strict control of attributes strategy. Rather, inferences must be based on a body of experimental results. As the study shows, children's conclusions, with some exceptions, were supported by results from multiple experiments.

Finally, this study raises the issue of how people deal with surprising results. A number of researchers have stressed the utility of such results in highlighting potentially useful regions of the hypothesis space to explore (Klahr et al., 1993; Klayman & Ha, 1987; Kulkarni & Simon, 1988). However, as other research has shown, an individual's theoretical commitment is the best predictor of how anomalous data will be treated (Chinn & Brewer, 1993; Penner & Klahr, in press). Consequently, capitalizing on surprising results requires satisfying three conditions. First, people must notice that something unexpected has occurred. Second, they must realize that their current understanding is insufficient to explain the outcome. Third, they must set a new goal to investigate possible mechanisms for the surprising result.

Each of the eight participants who dropped the two steel spheres expressed surprise that the heavier sphere did not sink faster. None of them ignored or misrepresented the unexpected result; they accurately reported the outcome, and expressed surprise that the experiment did not turn out as expected. However, only one participant appeared to realize the need to explore why this result occurred. Consequently, she set a goal to explore whether or not all pairs of objects that varied only on material sank at the same rate. The other seven participants simply applied their current knowledge in a piecemeal fashion to explain the anomalous result.

In summary, this study attempted to capture two aspects of real-world science. First, this study incorporated a difficulty common to real-world science, that of trying to isolate attributes to test. Although studies have shown that very young children can understand the logic of hypothesis testing (e.g., Sodian et al., 1991), the work reported here illustrates how fragile this understanding can be, and how easily it can be overwhelmed in "messy" tasks where it is not always possible to isolate attributes.

Second, we used a task for which people had strong prior, albeit incorrect, beliefs. In this manner we were able to investigate how prior beliefs affect experimental behavior, and how knowledge changes during the course of experimentation. We found a growing ability with age to hold prior beliefs in abeyance in order to determine the effects of all attributes. That is, younger children construe experimentation as a request to demonstrate the correctness of their initial beliefs (i.e., heavier objects sink faster); older children attempt, albeit with mixed results, to investigate the effects of attributes other than those implicated in their initial beliefs.

This study raises a number of questions for further investigation. What is the process by which children come to understand experimentation as a matter of evaluation rather than demonstration? How do children come to consider some attributes, such as weight, as being continuous and not just discrete (e.g., "heavy" or "light"), and what effect does this have on their exploration of complex phenomena? Finally, further attention needs to focus on how a priori domain-specific knowledge and understanding of scientific reasoning affect the interpretation of, and response to, surprising experimental outcomes.

References

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

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

Case, R (1974) Structures and strictures: Some functional limitations on the course of cognitive growth *Cognitive Psychology*, 6, 544–573

Chi, M T H , & Koeske, R. D (1983) Network representations of a child's dinosaur knowledge. *Developmental Psychology*, 19, 29–39

Chinn, C A., & Brewer, W F (1993) Factors that influence how people respond to anomalous data *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society* (pp 318–323) Boulder, CO: Erlbaum

Daily, J W., & Harleman, D R (1966) *Fluid dynamics* Reading, MA: Addison-Wesley

diSessa, A (1988) Knowledge in pieces In G Forman & P Pufall (Eds), *Constructivism in the computer age* (pp 49–70) Hillsdale, NJ: Erlbaum

Evans, J St B T (1989) *Bias in human reasoning: Causes and consequences* London: Erlbaum

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

Klahr, D , & Carver, S M (1995) Commentary: Scientific thinking about scientific thinking *Monographs of the Society for Research in Child Development*, 60(4, Serial No 245)

Klahr, D., & Dunbar, K (1988) Dual space search during scientific reasoning *Cognitive Science*, 12, 1–55

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

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

Klayman, J , & Ha, Y (1989) Hypothesis testing in rule discovery: Strategy, structure, and content *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 596–604

Kohn, A (1993) Preschoolers' reasoning about density: Will it float? *Child Development*, 64, 1637–1650

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

Kuhn, D , Amsel, E D , & O'Loughlin, M (1988) *The development of scientific reasoning skills* New York: Academic Press

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

Kuhn, D , & Phelps, E (1982) The development of problem-solving strategies In H Reese (Ed), *Advances in child development and behavior* (Vol 17, pp 1–44) New York: Academic Press

Kuhn, D , Schauble, L , & Garcia-Mila, M (1992) Cross-domain development of scientific reasoning? *Cognition and Instruction*, 9, 285–327

Kulkarni, D , & Simon, H (1988) The processes of scientific discovery: The strategy of experimentation *Cognitive Science*, 12, 139–175

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

Pazzani, M (1991) Influence of prior knowledge on concept acquisition: Experimental and computational results *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 416–432

Pazzani, M , Dyer, M , & Flowers, M (1986) The role of prior causal theories in generalization In *Proceedings of the National Conference on Artificial Intelligence* (pp 545–550) Philadelphia: Morgan Kaufmann

Penner, D E , & Klahr, D (in press) When to trust the data: Further investigations of system error in a scientific reasoning task *Memory and Cognition.*

Piaget, J (1972) *The child's conception of physical causality* Totowa, NJ: Littlefield, Adams (Original work published 1930)

Piaget, J , & Inhelder, B (1974) *The child's construction of quantities* London: Routledge & Kegan, Paul (Original work published 1942)

Piéraut-Le Bonniec, G (1980) *The development of modal reasoning: Genesis of necessity and possibility notions* New York: Academic Press

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

Schauble, L (1990) Belief revision in children: The role of prior knowledge and strategies for generating evidence *Journal of Experimental Child Psychology*, 49, 3–57

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 , & Raghavan, K (1991) Students' transition from an engineering model to a science model of experimentation

Journal of Research in Science Teaching, 28, 859–882.

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

Smith, C., Carey, S., & Wiser, M (1985) On differentiation: A case study of the development of the concepts of size, weight, and density *Cognition*, 21, 177–237

Smith, C., Snir, J., & Grosslight, L (1992) Using models to facilitate conceptual change: The case of weight-density differentiation. *Cognition and Instruction*, 9, 221–283

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

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