COGNITIVE PSYCHOLOGY 8, 481–520 (1976)

Three Aspects of Cognitive Development

ROBERT S. SIEGLER

Carnegie-Mellon University

An attempt was made to characterize and explain developmental differences in children's thinking, specifically in their understanding of balance scale problems. Such differences were sought in three domains: existing knowledge about the problems, ability to acquire new information about them, and process-level differences underlying developmental changes in the first two areas. In Experiment 1, four models of rules that might govern children's performance on balance scale problems were proposed. The rules proved to accurately describe individual performance and also to accurately predict developmental trends on different types of balance scale problems. Experiment 2 examined responsiveness to experience: it was found that older and younger children, equated for initial performance on balance scale problems, derived different benefits from identical experience. Experiment 3 examined a potential cause of this discrepancy, that younger children might be less able than older ones to benefit from experience because their encoding of stimuli was less adequate. Independent assessment procedures revealed that the predicted differences in older and younger children's encoding were present; it was also found that these differences were not artifactual and that reducing them also reduced the previously observed differences in responsiveness to experience. It was concluded, therefore, that the encoding hypothesis explained a large part of the developmental difference in ability to acquire new information.

The purpose of this article is to characterize and explain developmental differences in thinking. The focus is upon three aspects of development: specific knowledge governing task performance, responsiveness to experience, and basic processes that underlie differences in the other two areas. The goal is to make both conceptual and experimental distinctions among the three domains and to map out the interrelationships among them.

An example may clarify the conceptual basis for the trichotomy. Consider the familiar conservation of liquid quantity problem. Nonconservers

This study was supported in part by Public Health Service Grant MH-07722 from the National Institute of Mental Health. Thanks are due to Miss Eleanor Tucker, Headmistress of the Winchester Thurston School; Mr. James Burke, Superintendent of the Fox Chapel School District; and Mrs. Louise Brennan, Head of Elementary Schools in the Pittsburgh Public School District. The extensive and invaluable suggestions of David Klahr and the insightful comments of Herbert Simon on earlier versions of the manuscript are also acknowledged with thanks. Special thanks to Elliott Simon, who prepared the materials and ran subjects in Experiment 1, and to Jean Lea and Renee Silberner who performed similar duties in Experiments 2 and 3. Requests for reprints should be sent to Robert S. Siegler, Psychology Department, Carnegie-Mellon University, Pittsburgh, PA. 15213.

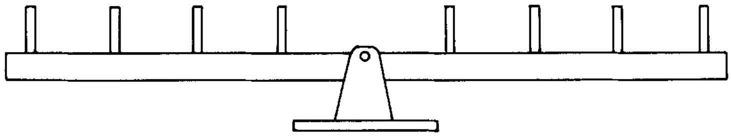


FIG. 1. The balance scale used in Experiments 1 and 2.

as well as conservers seem to have rules for solving such tasks—they may believe that the taller liquid column invariably is the one with more water, or that the container of greater circumference always has more, or that the height cue ordinarily points to the correct answer except in the case where the height of the liquid in the two beakers is equal, in which case the circumference must also be considered. These are examples of specific knowledge that governs task performance. However, two children who at present rely on the identical rule may be differentially "ready" to become conservers. A brief explanation of the logic of conservation might move one child, while repeated explanations, examples, and threats might not influence the other. This corresponds to the construct of responsiveness to experience. Finally, children's current conservation knowledge and their responsiveness to experience with conservation problems presumably are not accidental; they are rooted in more basic differences in such areas as short-term memory, ability to comprehend instructions, ability to control attention, and so on. This is the third domain of inquiry.

In the present study, this three-part framework is applied to characterizing and explaining developmental changes in children's understanding of balance scale problems (Fig. 1). In Experiment 1, four models of task-relevant knowledge that children might use to perform balance scale problems are proposed. The primary goal of the Experiment is to test the fit of these rule models to the performance of 5- to 17-year-old children. In Experiment 2, different-aged children's responsiveness to experience is examined. Older and younger children whose initial performance on the balance scale task is governed by identical rules are presented identical experience; the question is whether their final performance will be comparable. Finally, Experiment 3 focuses on whether differential encoding might underlie developmental change in responsiveness to experience with balance scale problems.

The balance scale task presented a number of advantages for this type of analysis. It is an interesting task mathematically, being related to the concept of proportionality. It occupies an important place within Piagetian theory, and this has led to a moderate-sized body of empirical work on the problems. It is applicable over a very wide age-range; children as young as 5-years often know that balances such as teeter-totters tend to fall toward the side with more weight, while even 16-year-olds often do not know the formal rules determining the balance's behavior (Jackson, 1965; Lee, 1971; Lovell, 1961). Finally, the balance scale task would seem to share an interesting characteristic with many other scientific induction problems—the rule for generating correct solutions, once known, is trivially easy to execute, but inducing the rule in the first place is quite difficult.

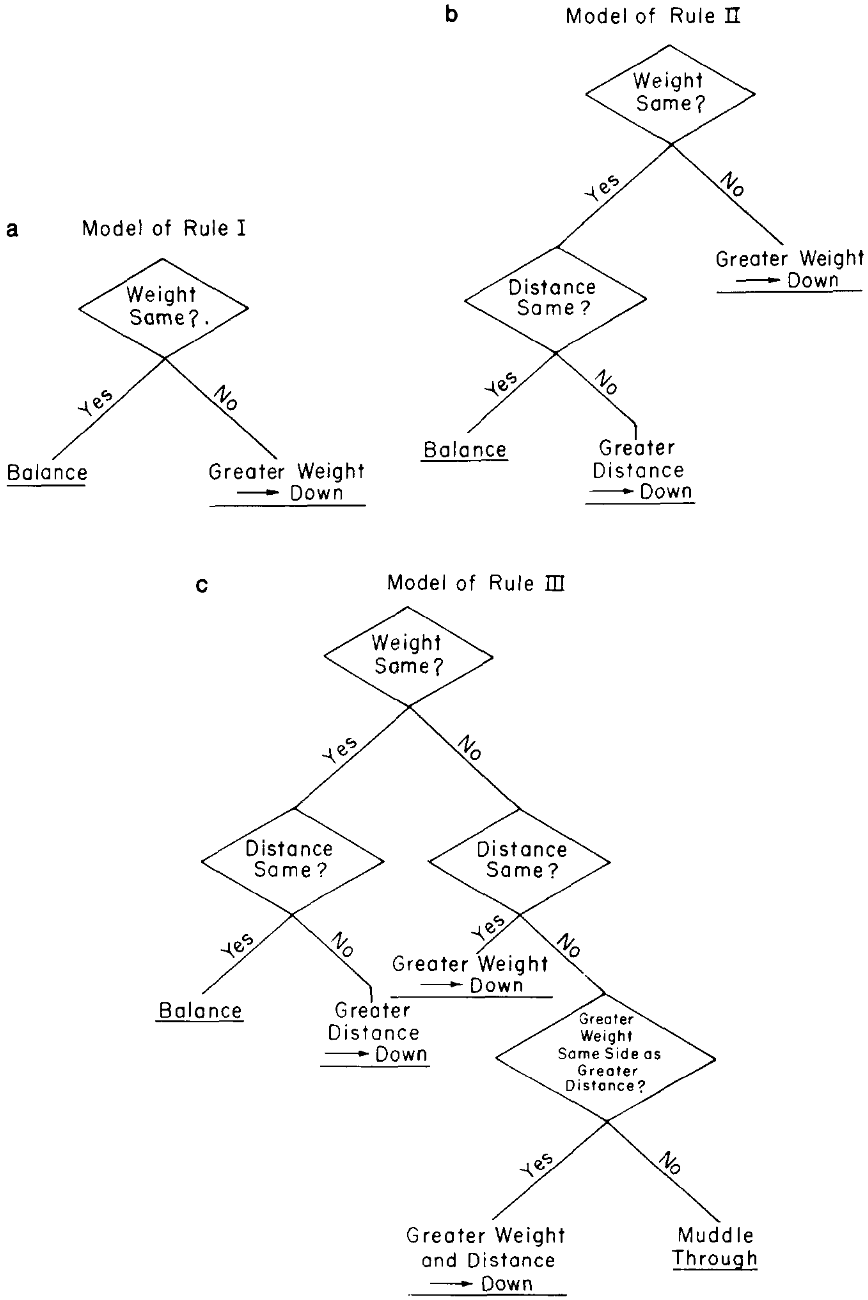
The balance scale apparatus that was used is shown in Fig. 1. On each side of the fulcrum were four pegs on which metal weights could be placed; the arm of the balance could tip left or right or remain level depending on how the weights were arranged. However, blocks of wood (not shown in Fig. 1) were placed underneath each side of the balance, thus preventing it from tipping regardless of the weights' configurations. The children's task was to predict which (if either) side would go down if the blocks were removed.

EXPERIMENT 1

Models of Children's Specific Knowledge about Balance Scale Problems

The main purpose of Experiment 1 was to determine whether children's knowledge about balance scale problems could be characterized accurately and unambiguously. Specifically, the experiment was a test of the utility of the four rule system characterizations shown in Figure 2 (a–d). The model of mature performance (Rule IV) was suggested by a rational task analysis of balance scale problems (cf. Resnick, 1976); the models of less sophisticated performance (Rules I–III) were derived from Inhelder and Piaget's (1958) and Lee's (1971) empirical findings, and from pilot work with the present problems. In the most advanced system, Rule IV, both the amount of weight and the distance of the weights from the fulcrum are always considered, and if the cues suggest different outcomes, the sum of cross products rule is invoked. For example, if, as in the fifth problem in Table 1, there were three units of weight on the third peg to the left of the fulcrum, and if there were two units of weight on the first peg to the right and three units of weight on the second peg to the right of the fulcrum, the distance cue would point to the left side's going down and the weight cue would suggest the reverse. Therefore, the product of distance and weight would be taken on each peg, the results summed for each side, and the two sums compared—$(3 \times 3) = 9$; $(1 \times 2) + (2 \times 3) = 8$; $9 > 8$, therefore left side down.

Rule IV directly suggested a number of less differentiated approaches to the problem. Children following Rule I consider only a single dimension; Inhelder and Piaget's (1958) work indicates that it would generally be weight, though from the viewpoint of the complexity of the rules involved it could as easily be distance. Rule II represents an advance over



FIGURES 2a–2d. Decision tree model of rule for performing balance scale task.

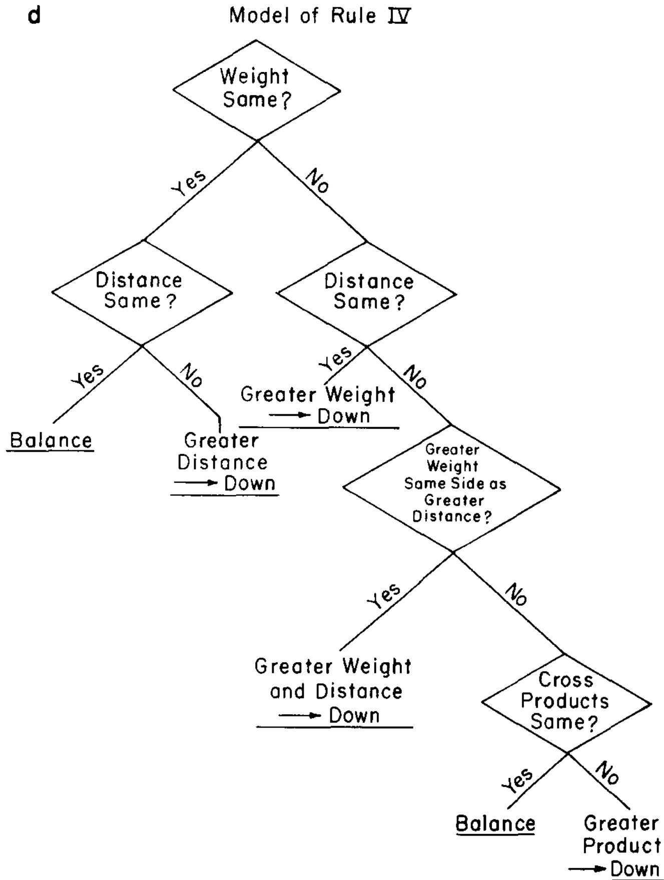


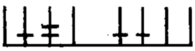
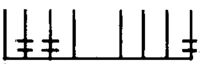
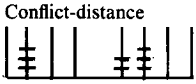
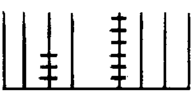
FIGURE 2D

Rule I in that distance from the fulcrum as well as amount of weight is considered whenever the weight on the two sides is equal, though not when the weights are unequal. Children using Rule III always consider both weight and distance, but when the cues are discrepant they do not have a rule for resolving the conflict. They therefore "muddle through" or guess. Within this system, use of Rule II should never precede use of Rule I nor should use of Rule IV precede use of Rule III in any child's development. This is for logical rather than psychological reasons, the relationship among rules conforming to Flavell's (1972) inclusion model; all of the questions posed in Rule I are included in Rule II, all of the questions in Rule III are included in Rule IV, etc.

These rule characterizations are related to Inhelder and Piaget's (1958)

TABLE 1

PREDICTIONS FOR PERCENTAGE OF CORRECT ANSWERS AND ERROR PATTERNS ON POSTTEST FOR CHILDREN USING DIFFERENT RULES

	Rules				
Problem type	I	II	III	IV	Predicted developmental trend
Balance 	100	100	100	100	No change—all children at high level
Weight 	100	100	100	100	No change—all children at high level
Distance 	0 (Should say "balance")	100	100	100	Dramatic improvement with age
Conflict-weight 	100	100	33 (Chance responding)	100	Decline with age Possible upturn in oldest group
Conflict-distance 	0 (Should say "right down")	0 (Should say "right down")	33 (Chance responding)	100	Improvement with age
Conflict-balance 	0 (Should say "right down")	0 (Should say "right down")	33 (Chance responding)	100	Improvement with age

analysis of the balance scale task, but differ in several regards. In Inhelder and Piaget's Stage I, children do not follow any consistent rule; in the present Rule I, they consistently rely on the amount of weight. There is no indication in any stage of Inhelder and Piaget's system of an approach comparable to Rule II in which children consider distance from the fulcrum only if the amounts of weight are equal. Finally, while Inhelder and Piaget's highest stage (III) emphasized recognition of proportionality in creating balances (e.g., one weight placed three units to the left of the fulcrum balances three weights placed one unit to the right of the fulcrum), this realization would not necessarily lead to understanding of the current Rule IV; children would also need to know the composition rule of summing the products of weight and distance on each side of the fulcrum.

The present rule analysis suggested six different types of problems for assessing a child's knowledge (Table 1). Three are solvable without any arithmetic computation: *balance* problems, with equal amounts of weight equidistant from the fulcrum; *weight* problems, with unequal amounts of weight equidistant from the fulcrum; and *distance* problems, with equal amounts of weight different distances from the fulcrum. The three additional types of problems had more weight on one side but the weight on the other side was farther from the fulcrum—for example, three weights on the second peg to the right of the fulcrum versus six weights on the first peg to the left of the fulcrum; thus they required computation. These "conflict" problems were distinguished by their outcomes: on *conflict-weight* items, the side with the greater amount of weight would go down, on *conflict-distance* problems, the side with the weight farther from the center would tip, and on *conflict-balance* tasks, the two influences would cancel out, leaving the arm of the scale level. Thus, the example in this paragraph was a conflict-balance problem.

Children whose performance conformed to different rules would display dramatically different patterns of successful and unsuccessful predictions on the six types of problems (Table 1). Those using Rule I would consistently make correct predictions on balance, weight, and conflict-weight problems (or on balance, distance, and conflict-distance problems) and would never be correct on the other three types of tasks. Children conforming to Rule II would behave similarly on five of the six problem-types, but would correctly solve distance problems. Those following Rule III would consistently make accurate predictions on weight, balance, and distance problems and would perform at a roughly chance level on all conflict tasks. Those using Rule IV would solve all problems of all types.

The analysis makes specific predictions about error patterns as well as correct and incorrect answers. All of the errors of children adhering to Rule I should conform to the weight cue (or all to the distance cue). There should be no differences in number of errors between children following Rule II and those following Rule III, but the pattern of errors for children using Rule III should be more complex.

To the degree that older children more often use Rule III and younger ones Rules I and II, there should be a developmental decrement in the number of accurate predictions on conflict-weight problems; younger children using Rules I and II should perform virtually perfectly, while older children using Rule III should perform at roughly a chance level. By a similar logic, performance on conflict-balance and conflict-distance problems should proceed from below chance for the youngest children to approximately chance for older ones adhering to Rule III, to above chance if many of the oldest children follow Rule IV. Additionally, to the extent that the correlation between age and rule-system is present, there should be a particularly dramatic increase in performance on distance problems; children using Rule I should get virtually none right, while older children using Rules II, III, and IV should get virtually none wrong. On the other hand, there should be little or no developmental trend on balance and weight problems, since they are solvable by any of the four rule systems. Overall, balance and weight problems, solvable by any rule, should be most often predicted correctly; distance and conflict-weight problems, solvable by three of the four rules, should be next most often correctly responded to; and conflict-distance and conflict-balance items, solvable consistently only by Rule IV, should elicit the fewest correct predictions. By omission, the model also implies that despite substantial differences in the number of weights involved and their distribution over different pegs, there should be no substantial differences in performance on the tasks fitting under any given problem-type (e.g., among the six conflict-balance tasks).

In addition to testing the rule models, a second purpose of Experiment 1 was to examine the impact of different types of experience on children's understanding of balance scale problems. Several previous studies have demonstrated that even 9- and 10-year-olds can master formal operations problems if provided directive instruction (Case, 1974; Kuhn & Angelev, 1975; Siegler & Atlas, 1976; Siegler, Liebert, & Liebert, 1973; Siegler & Liebert, 1975). Relatively little is known, however, about how children go about inducing formal operations relationships for themselves, nor about how they draw conclusions from observing the activities of others. In order to learn about the effects of these types of experiences, and about developmental changes in the effects, 5- and 6-year-olds, 9- and 10-year-olds, 13- and 14-year-olds, and 16- and 17-year-olds were exposed to one of three experiential conditions: *a priori, experimentation,* or *observation*. Children in the a priori condition were simply presented the posttest problems, the aim being to assess their existing knowledge. Children in the experimentation group were told that there were rules by which they could know which way the balance would tip and that they should "experiment" by placing the metal weights on the pegs in as many different ways as they needed to learn how the balance worked. Those in the observation group were provided similar instructions except that the experimenter would decide how to put the weights on the pegs and the children would watch and try to learn the rules.

Previous investigations indicated that full understanding of balance scale problems grows slowly, remaining below 50% through age 17-years (Jackson, 1965; Lee, 1971; Lovell, 1961). Even lower levels of proficiency were expected in the a priori groups of the present study; this was because the task, unlike those previously used, required that children know the composition rule of summing the products on each side of the balance and comparing the sums, in addition to the usual requirement that they know individual ratio equivalences. By contrast, the success of previous instructional procedures, along with Inhelder and Piaget's (1958) emphasis on the development of logical problem solving strategies during the formal operations period, suggested that adolescents would benefit substantially from the experimentation and observation sequences and that 9- and 10-year-olds might also benefit. Finally, because the experimentation condition demanded that children generate informative problems and grasp the implications of the results while the observation condition demanded only the latter data analysis skill, it was predicted that observation would be more helpful than experimentation, especially to children aged 10-years and under.

Method

Participants. Subjects in Experiment 1 were 120 female students at a predominantly upper-middle-class private school in Pittsburgh. Fifteen kindergarteners and 15 first graders (mean CA = 73.48 months; range = 62–85 months), 15 fourth and 15 fifth graders (mean CA = 119.75 months; range = 109–129 months), 15 eighth and 15 ninth graders (mean CA = 169.27 months; range = 156–186 months), and 15 eleventh and 15 twelfth graders (mean CA = 207.17 months; range = 196–219 months) made up the experimental sample. Within each of these age groups, 10 children, five in each grade level, were assigned to the experimentation, 10 to the observation, and 10 to the a priori condition. A male and a female research assistant, each of them 22-years-old, served as experimenters.

Materials. Materials included a wooden balance scale, 10 different colored metal weights, and two wood blocks. The balance scale's arm was 32 in. long, with four pegs on each side of the fulcrum. The first peg on each side was 3 in. from the fulcrum and each subsequent peg was 3 in. from the peg before it. The arm could swing freely from the point of attachment to the fulcrum, 4 in. above the fulcrum's base. Each metal weight weighed 1.4 ounces, measured 1 in. in diameter, and had a hole in its middle so that it would fit on the pegs; as many as six weights could be placed on any one peg. Weights of the same color were never stacked adjacently on a given peg. The two blocks of wood, each 4.5 in. high, could be placed under the arm of the balance scale to prevent it from moving regardless of the configuration of the metal weights on the pegs.

Posttest. Children's knowledge was assessed through a 30 item posttest. On each problem the experimenter started with an empty balance, the arms of which were supported by the two wooden blocks. Then the metal weights were placed on the pegs on the two sides of the balance scale, and the child was asked to predict which side would go down or whether the scale would balance if the two wooden blocks, underneath the arms of the balance, were not there. Among the 30 items were four balance, four weight, four distance, six conflict-weight, six conflict-distance, and six conflict-balance tasks of the types shown in Table 1. On balance problems, equal numbers of weights were arranged identically on the two sides of the fulcrum (e.g, two weights on the first and one on the second peg to the left of the fulcrum versus two weights on the first and one on the second peg to the right of the fulcrum). Weight problems had differing numbers of weights placed identically on the balance (e.g., two weights on the first peg and one on the second peg on the left versus one weight on the first and one weight on the second peg on the right). Distance tasks had equal numbers of weights different distances from the fulcrum (e.g., three weights on third peg on left versus three weights on second peg on right). On conflict-weight, conflict-distance, and conflict-balance problems, one side would include a greater number of weights while the weights on the opposite side would be farther from the fulcrum. On conflict-weight problems, the balance would tip toward the side with the greater number of weights (e.g., two weights on third peg and two weights on second peg, versus two weights on fourth peg); on conflict-distance problems the balance would tip toward the side with the weights farther from the fulcrum (e.g., three weights on third peg versus two on first and three on second peg); and on conflict-balance items the scale would remain level (e.g., three weights on second peg versus six on first peg).

The problems were assigned to the 30 positions in the testing sequence through use of a random number table. The median number of weights used on each problem was six, with a range of between two and 10. There were no substantial differences in the number of weights used in the six types of problems. On one-half of the problems of each type there were weights on the fourth peg, the farthest one from the fulcrum, as well as on others closer to it—on the other one-half the weights were distributed only over the first three pegs.

Procedure. Children were brought individually to a conference room adjacent to the school library and were asked to sit next to the experimenter at a table with the balance scale in front of them. The experimenter's initial instructions were the same in all treatment conditions. "Today we are going to play with this balance scale. The balance scale has these pieces of wood that are all the same distance from each other (pointing to the pegs) and these pieces of metal that all weigh the same." At this point children were encouraged to hold the weights to see that they weighed the same amount and to observe the equal distance between adjacent pegs. *Experimentation group* members were then told:

> There are rules that you can learn that will tell you whether it will tip this way or this way or stay level, even if you have never seen the weights on the wooden pegs in the particular way before. So put the weights on the wooden pegs in all the ways you can think of that might help you learn how the balance scale works.

Before they started to experiment, children were shown an example of the type of problem they would be asked about later (three weights on the second peg versus two on the third) and told to learn what would happen on these kinds of problems. During the experimentation period three rather than four pegs were present on the balance scale; in this way, children in the experimentation group might learn directly some or all of the three peg problems on the posttest but it would be impossible for them to have previously encountered any of the 15 posttest problems in which weights were on the fourth peg.

Observation group members also were told that their job was to figure out the rules for how the scale worked. On each of the 36 trials, the experimenter arranged weights on the pegs, removed the wooden blocks holding up the balance scale arm, and allowed 10 sec for the child to consider the outcome. Three of the 36 problems in the observation sequence—one conflict-weight, one conflict-distance, and one conflict-balance—were also included in the posttest as another way of determining whether children were learning particular responses or general rules. Problems in the sequence were ordered by increasing levels of hypothesized complexity to encourage realization that simple rules did not always work and to facilitate discovery of the Rule IV approach. The initial group of five items illustrated simple balance, weight, and distance relationships. The emphasis then shifted for the next 20 items to demonstrating how balanced arrangements could be created when weight and distance cues conflicted. For example, on item six there was one weight on the first peg to the left of the fulcrum, and one weight on the third peg to the right. Weights were added one per trial to the first peg on the left until on item nine there were four disks on that peg, thus creating an imbalance in the opposite direction from the one initially in effect. On item 10, the last weight that had been added was removed, thus restoring the balance present on item 8. After four such sequences (item 25), the balanced arrangements that had been created were redisplayed on four successive items to focus attention on what they had in common. Finally, seven high-difficulty conflict problems were presented, involving weights on several pegs on each side of the fulcrum and thus demanding knowledge of the composition rule (e.g., two weights on fourth peg and two weights on third peg

TABLE 2

MEAN NUMBER[a] OF CORRECT PREDICTIONS IN EXPERIMENT I ACCORDING TO AGE AND TREATMENT

	Treatment condition			
Age	A Priori	Experimentation	Observation	Means
5–6	14.0	13.8	13.8	13.9
9–10	18.5	17.8	18.8	18.4
13–14	17.9	18.5	19.8	18.7
16–17	19.3	20.9	20.5	20.2
Means	17.4	17.8	18.2	17.8

[a] Of a possible 30.

to left of fulcrum, versus four weights on first peg and two weights on second peg to right of fulcrum).

Children in all groups were presented the same posttest instructions. They were told:

> Let's see what you know about the balance scale. I'll put the weights on the pegs in different ways and you tell me whether this side would go down or this side would go down or whether they would stay like they are now if I took the wood

TABLE 3

DEVELOPMENTAL TRENDS OBSERVED AND PREDICTED ON DIFFERENT PROBLEM-TYPES IN EXPERIMENT I[a]

	Age				Predicted developmental trend[b]
Problem type	5–6	9–10	13–14	16–17	
Balance	94	99	99	100	No change—all children at high level
Weight	88	98	98	98	No change—all children at high level
Distance	9	78	81	95	Dramatic improvement with age
Conflict-weight	86	74	53	51	Decline with age—possible upturn for oldest
Conflict-distance	11	32	48	50	Improvement with age
Conflict-balance	7	17	26	40	Improvement with age

[a] Percentage of problems predicted correctly.
[b] From Table 1.

> blocks away. The balance scale won't actually move, but you tell me how the scale would go if the pieces of wood were not there.

Following the posttest, children were asked for explanations of their responses. Specifically, they were queried, "How could you tell which way the balance would tip?" and "What do you think made a difference in the way it tipped?" Participation in the a priori and observation groups took approximately 15 and 30 min, respectively. Depending on the number of "experiments" they performed, children spent between 20 and 30 min participating in the experimentation condition.

Results

Predictions Data

Number of correct predictions. A 4 (Age: 5- and 6-year-olds, 9- and 10-year-olds, 13- and 14-year-olds, or 16- and 17-year-olds) × 3 (Treatments: experimentation, observation, or a priori) × 2 (Fourth peg status: occupied or vacant) revealed a single significant main effect for age [$F(3,108) = 32.38$, $p < .001$]. The performance of 9- and 10-, 13- and 14-, and 16- and 17-year-olds consistently exceeded that of 5- and 6-year-olds; however, the performance of 9- to 17-year-olds did not differ (Table 2).[1] No significant main effects for treatments or fourth peg status were present nor did these variables interact reliably with each other or age (all $Fs < 2.00$).

Analysis by problem-type. In contrast to the simple pattern for total number of correct answers, substantial complexity emerged when the type of problem was included as a variable. A 4 (Age) × 3 (Treatment) × 6 (Problem-type: weight, balance, distance, conflict-weight, conflict-balance, or conflict-distance) analysis of variance again revealed a significant main effect for age [$F(3,108) = 26.71$, $p < .001$], and in addition, a main effect for problem-type [$F = 188.67$, $p < .001$] and interactions between age and problem-type [$F(15,540) = 16.92$, $p < .001$], and between treatment and problem-type [$F(10,540) = 5.33$, $p < .001$].

The main effect for problem-type resulted from weight and balance problems being more often solved than distance and conflict-weight tasks, which in turn were more often solved than conflict-distance and conflict-balance items (Table 3). The significant age by problem-type interaction can be understood in terms of different developmental patterns for the six types of problems. For balance and weight problems there was no developmental trend; performance of all children was virtually perfect. The greatest developmental change occurred on distance problems; 5- and 6-year-olds predicted correctly on less than 10% of such items, 9- and 10-year-olds and 13- and 14-year-olds on more than 75%, and 16- and 17-year-olds on 95%. On conflict-weight problems, performance

[1] Unless otherwise indicated, all differences cited in analyses of main effects and interactions differ at or beyond the .05 significance level by Duncan Multiple Range Tests.

actually declined with increasing age; both 5- and 6-year-olds and 9- and 10-year-olds made a greater number of correct predictions than 13- and 14- or 16- and 17-year-olds. The other two problem-types—conflict-distance and conflict-balance—showed the more typical positive correlation between age and percentage correct.

The treatment by problem-type interaction can be understood through a similar analysis. Children in the observation condition made a greater number of correct predictions on conflict-distance and conflict-balance problems than did peers in the experimentation and a priori conditions. However, on conflict-weight problems, the pattern was reversed; children in the latter two groups were more often correct than those in the observation condition. No differences among treatment groups were present on balance, weight, and distance tasks.

There was also substantial consistency on items within each problem-type. Only on conflict-weight problems did accurate prediction decrease with age and within this category such decrements occurred on all six problems. The magnitude of the improvement over age on the four distance problems was unmatched by that on any of the 26 other items. On all eight of the balance and weight items, but on no other tasks, was performance at a very high level for all age groups.

As mentioned above, three problems in the observation sequence—one conflict-weight, one conflict-distance, and one conflict-balance—also appeared on the posttest; this allowed an examination of whether children were learning specific responses or general rules. If they were learning specific responses, the feedback they received during the observation sequence would presumably help them on the three problems; if they were learning general rules, posttest performance on the previously presented problems would be expected to be no better than performance on other posttest problems of the same type. This latter possibility proved to be true. In no case was the number of accurate predictions significantly different (for the conflict-weight problems, $\chi^2(1) = .82, p > .10$; for the conflict-distance problems, $\chi^2(1) = .02, p > .10$; for the conflict-balance problems, $\chi^2(1) = .15, p > .10$).

Fit of rule models. The models of rules underlying performance on the balance scale task predicted how children using Rules I, II, and IV would answer each of the 30 problems; for Rule III, determinate predictions were made on 12 problems and essentially chance performance was anticipated on the remainder. The following standards were arbitrarily chosen as evidence that a child was using a particular rule: for Rule I, that at least 26 of the 30 responses conform to the "weight" cue, including at least three predictions of "balance" on the four distance problems; for Rule II, that 26 of the 30 responses correspond to the "if there are an unequal number of weights consider only weight—if the number of weights is equal also consider distance" formula, including three of the four distance problems; and for Rule IV, 26 of 30 correct predictions.[2] The multinomial probability that any one of these three criteria would be met by a random responder was less than 5×10^{-9}.[2] Use of Rule III entailed 10 of 12 correct responses on the nonconflict problems, including three of four on the distance problems, and more than four departures in 18 trials from complete reliance on the weight (distance) cue as indicating the correct answers on the conflict problems.[3] The probability that a child would by chance match the Rule III criterion was less than 5×10^{-4}.

As shown in Table 4, 107 of the 120 children could be classified as using one of the four rules: 29 used Rule I, 22 used Rule II, 48 used Rule III, and eight used Rule IV.[4] Although children could have as many as four deviations from the 30 predicted responses and still be said to use Rules I, II, or IV, most children had far fewer; the average number of deviations among those using Rule I was .52; among those using Rule II, 1.86; and among those using Rule IV, .50. There was a clear relationship between children's age and the rule they used. All 23 of the 5- and 6-year-olds who used any rule used Rule I. The distribution of rules used by 9- to 17-year-olds was relatively constant but there was some change toward 16- and 17-year-olds using Rules III and IV more often than 9- and 10-year-olds [$\chi^2(1) = 4.37$, $p < .05$]. The type of treatment condition also tended to affect the rules that were used [$\chi^2(6) = 11.31$,

[2] The formula used to compute the probability of a random responder falling into Rule I, II, or IV was:

$$\binom{30}{26}\left(.33^{26}\right)\left(.67^{4}\right) + \binom{30}{27}\left(.33^{27}\right)\left(.67^{3}\right) + \cdots \binom{30}{30}\left(.33^{30}\right)\left(.67^{0}\right).$$

The formula used to compute the probability of Rule III was:

$$\binom{12}{10}\left(.33^{10}\right)\left(.67^{2}\right) + \binom{12}{11}\left(.33^{11}\right)\left(.67^{1}\right) + \binom{12}{12}\left(.33^{12}\right)\left(.67^{0}\right).$$

It should be noted that each of these formulas overestimated the probability of a random responder falling within a rule, as not all criteria are included.

[3] The impact of using this arbitrary criterion was not large. If a criterion of three deviations in the 18 conflict problems had been used, six children would have been classified differently—two 9- and 10-year-olds, one 14-year-old, and three 16- and 17-year-olds would have been classified as using Rule III rather than Rule II. If the criterion had been five deviations in 18 conflicts problems, then seven children—two 9- and 10-year-olds, three 13- and 14-year-olds, and two 16- and 17-year-olds—would have been said to use Rule II rather than Rule III. Changing the criteria for Rules I and IV to either three of 30 or five of 30 deviations would not have changed the classification of any child.

[4] Of the 13 children who did not fit into any rule by the predictions criterion, seven were in the observation condition (five kindergarten and first graders, one fourth grader, and one eighth grader); four were in the control condition (one first grader, two fourth and fifth graders, and one eighth grader); and two were in the experimentation group (one kindergartener and one fifth grader). Among the 13, seven were unclassifiable because they responded inconsistently to distance items; five were unclassifiable because they consistently responded "same" to distance items, but did not consistently use either weight or distance cues on other items; and one child failed to solve the balance items which were solvable by any of the four rule systems.

TABLE 4

NUMBER OF CHILDREN IN EXPERIMENT 1 USING EACH RULE—PREDICTIONS DATA

Age	Treatment condition	Rule I	Rule II	Rule III	Rule IV	Unclassifiable
5–6	Experimentation	9	0	0	0	1
	Observation	5	0	0	0	5
	A Priori	9	0	0	0	1
9–10	Experimentation	2	3	4	0	1
	Observation	0	2	6	1	1
	A Priori	1	4	2	1	2
13–14	Experimentation	1	2	7	0	0
	Observation	1	1	6	1	1
	A Priori	1	4	4	0	1
16–17	Experimentation	0	2	6	2	0
	Observation	0	0	8	2	0
	A Priori	0	4	5	1	0
	Total	29	22	48	8	13

$p < .10$]. Children in the observation condition were relatively more likely to conform to Rules III and IV while those in the experimentation and a priori conditions more often used Rules I and II [$\chi^2(1) = 7.95$, $p < .01$]. These changes in the rules that were used substantially accounted for the age by problem-type and treatment by problem-type interactions in the predictions data described above.

Explanations Data

The criteria used to score children's explanations of how they knew the way the balance scale would tip were derived from the rule system hypothesized to underlie different levels of performance. A child was classified as using Rule I if her explanation referred only to the number of weights, Rule II if she referred to distance from the fulcrum specifically as a means of solving problems on which there were equal amounts of weight and in no other context, Rule III if she referred to both amount of weight and distance from the fulcrum but did not allude to the general sum-of-cross-products rule, and Rule IV if she stated the proper mathematical formula. Altogether, 117 of the 120 children's explanations fit one of the four criteria (Table 5).[5] As shown in Table 6, the two measures of rules—one derived from children's predictions, the other from their

[5] All three of the children who did not fit within a rule by the explanations criteria were first graders, one in the experimentation and two in the observation group. One child was vague as to what should be done if the weights were equal, one said she tried to remember particular combinations from the observation sequence, and one child's statements could not be readily interpreted.

TABLE 5

NUMBER OF CHILDREN IN EXPERIMENT 1 USING EACH RULE—EXPLANATIONS DATA

Age	Treatment condition	Rule				Unclassifiable
		I	II	III	IV	
5–6	Experimentation	9	0	0	0	1
	Observation	7	0	1	0	2
	A Priori	10	0	0	0	0
9–10	Experimentation	3	1	6	0	0
	Observation	0	0	9	1	0
	A Priori	1	4	4	1	0
13–14	Experimentation	1	1	8	0	0
	Observation	1	0	8	1	0
	A Priori	1	2	7	0	0
16–17	Experimentation	0	1	7	2	0
	Observation	0	0	8	2	0
	A Priori	0	0	9	1	0
	Total	33	9	67	8	3

explanations—were closely related. Of particular note, all of the 23 children judged to be using Rule I by the predictions data were judged as using Rule I by the explanations criterion and all eight of the children classified as using Rule IV on the predictions measure—and only those eight—were classified as using Rule IV on the explanations measure as well.[6]

Discussion

The results of this experiment provided considerable support for the descriptive accuracy of the proposed rule models. Of the 120 children, 107 could be classified unambiguously as following one of the four rules. Each of the rules provided an accurate description of at least eight children's performance. The alternative measures of children's knowledge, explanations and predictions, were highly correlated, and each was

[6] The only substantial discrepancy between rule placements as assessed by the predictions and by the explanations data was in the rather large percentage of children classified at Rule II by the predictions criteria and at Rule III by the explanations criteria. This discrepancy seemed largely due to the nature of the explanations criteria; a child who merely mentioned distance and weight as important factors would by default be placed at Rule III; to be placed at Rule II, he would have needed to add that he only considered distance when the weights on the two sides were equal. Many children who predicted according to Rule II may have thought this elaboration of their strategy unnecessary or may have failed to reflect adequately on the subtleties of their approaches.

related to children's age. Even the youngest children's behavior appeared to be rule governed; a large majority of 5- and 6-year-olds' predictions fit the Rule I model, performance on particular items within a given problem-type was very similar for all age groups, and there were no differences between performance on items that had been or could have been seen previously and items that could not have been directly experienced. Finally, the rule-system analysis helped to clarify the consistent but complex pattern of results—stability on balance, weight, and distance problems; improvements on conflict-distance and conflict-balance problems; and decrements on conflict-weight problems—that occurred among 9- to 17-year-olds with increasing age and with exposure to the observation condition.

Thus, Experiment 1 illustrated some of the advantages of the rule-model level of analysis compared either to an entirely quantitative conception of development or to a broad stage level approach. A purely quantitative measure of the total number of correct predictions revealed no effects for treatment conditions, no developmental change between ages 9- and 17-years, and no differences in the effects of the treatments on different aged children; it missed all of the richness of the pattern exposed by analysis in terms of problem-types, an analysis directly suggested by the rule-system model. The advantages over the more traditional general stage approach were in the precision with which the developmental progression was portrayed and in the number and concreteness of the predictions. The model also made evident the logical rather than psychological nature of the rule orderings.

A major unanticipated finding in Experiment 1 was that the experimentation and observation treatments did not produce a greater movement toward Rule IV. This was quite surprising since in previous investigations even 9- and 10-year-olds benefited greatly from training on formal operations level tasks. The primary difference between the present experiential manipulations and earlier training efforts was in the explicitness with which the specific operating rules for solving the task were presented. Previous investigations provided direct instruction and feedback in the output rules needed to solve the particular or related problems (e.g., Case, 1974; Kuhn & Angelev, 1975; Siegler, Liebert, & Liebert, 1973; Siegler & Liebert, 1975). In both experimentation and observation conditions of the present experiment, however, children needed to generate the steps in the solution rules for themselves. One reason that such nondirective experience was provided was the relative lack of complexity of Rule IV; it seemed that it would be trivially easy to teach it to any child who knew how to multiply and add. To verify this belief, a brief, directive, balance scale training procedure was given to a small number of fourth graders. Three girls, all 10-year-olds, were told the sum of cross products rule and were given six trials on which to apply it and receive feedback.

TABLE 6

NUMBER OF CHILDREN IN EXPERIMENT 1 USING EACH RULE—PREDICTIONS AND EXPLANATIONS CRITERIA

		Rule by predictions criterion			
		I	II	III	IV
Rule by explanations criterion	I	23	1	0	0
	II	0	7	1	0
	III	0	13	46	0
	IV	0	0	0	8

Then they were given the standard posttest. Both in their predictions and in their explanations, all three performed at the Rule IV level, a feat achieved by less than 20% of the 16- and 17-year-olds in the larger study. This finding suggests a need to carefully distinguish between the processes of learning and discovery; on this and on other scientific induction tasks, it appears that 9- and 10-year-olds may be able to learn easily relationships that children (adults?) twice their age cannot discover even with difficulty (cf. Lovell, 1961; Siegler & Liebert, 1975).

EXPERIMENT 2

Responsiveness to Experience

A second unanticipated finding in Experiment 1 was the rather small effect of the observation and experimentation procedures. This left open the question of how children using a particular rule might come to induce more advanced approaches. General cognitive developmental theory suggested at least one prediction; that inducing new knowledge would be easier, the closer the new material was to the learner's current level. In the present context, children using Rule I would be expected to benefit more from experience with distance problems, hypothesized to be the next developmental acquisition, than from experience with conflict problems, further beyond their current level. This was the point of departure for Experiment 2.

A second, more general, question about development was also posed in Experiment 2. Stated hypothetically, "If two children of different ages but with identical task specific initial knowledge were presented the same learning experiences, would they emerge with the same final knowledge about the task?" In other words, is assessing a child's knowledge about a problem sufficient to predict his subsequent responsiveness to experience with it, or must additional age-related factors be taken into account? If developmental changes are primarily due to younger children's knowing less than older ones about particular problems, no differences in responsiveness would be expected. On the other hand, if there are differences in the way that children of different ages go about learning that are independent of their knowledge about particular tasks, differences in what is learned might well be anticipated.

In Experiment 2, 5- and 8-year-olds, equated for not possessing knowledge beyond Rule I, were provided either experience with distance problems, experience with conflict problems, or one of two control procedures. Distance problem experience focused on the type of problems solvable by Rule II but not by Rule I; it thus was geared one "step" above the learners' initial level. Conflict problem experience, emphasizing problems not understood even qualitatively until Rule III, was intended to be two or more "steps" advanced. According to Piagetian theory, the fit between the child's existing knowledge and the new information presented is a critical determinant of when, how much, and what kind of learning will occur (cf. Piaget, 1971). Support for this view has been found by Turiel (1966) and Blatt (1971) in the area of moral development, and by Kuhn (1972) on a class inclusion task. Therefore, it was predicted that a greater number of children would benefit from experience with distance problems than from experience with conflict problems.

Deriving the prediction for possible differential responsivity to experience of older and younger children was less straightforward. There was nothing in the rule models that would suggest such differences. On the other hand, older children have proven more adept than younger ones at mastering numerous novel psychological problems on which specific knowledge about the task must have been equally lacking for all subjects (e.g., Siegler, 1975; Siegler & Liebert, 1974). In those experiments, there appeared to be developmental differences in children's ability to extract information from new experiences independent of previous knowledge about the particular task. Therefore, no prediction was made concerning possible differential responsivity to experience of older and younger children.

Method

Experiment 2 included three segments: pretest, experience, and posttest.

Pretest. The pretest consisted of eight items: two weight, two distance, two conflict-weight, and two conflict-distance. The task and apparatus were the same as those used in the posttest in Experiment 1. At no time during the pretest was any type of feedback provided.

Experience. All experiential conditions except the bias control (see below) included 16 trials on which children were presented various types of balance scale problems. Children were asked to predict what would happen and why they thought so; then the wood blocks supporting the scale were removed so that the prediction was confirmed or disconfirmed. After a 10 sec interval the weights were removed and placed on the scale in a different arrangement. The ordering of problems was random for all experiential conditions; within each group, one-half of the children received the problems in one order and the other one-half in the reverse order.

Conflict problem experience involved presentation of six conflict-weight, six conflict-distance, one distance, two balance, and one weight problem. *Distance problem experience* included 12 distance, two balance, and two weight problems. Thus, each experiential condition included 12 problems of the type being emphasized; the additional four problems of other types were intended to prevent children from acquiring strategies too narrowly suited to the demands of the majority of items.

Within the control condition there were two subgroups: *the exposure control* and *the bias control*. The exposure condition was designed to control for the possibility that *any* experience with the balance scale could improve performance; children in this condition were presented a sequence composed of 14 weight and two balance items that would familiarize them with the balance scale's workings but would not directly engender knowledge of Rules II or III. However, this control procedure might itself bias children toward a greater reliance on Rule I than if they had been left untutored. Therefore, a bias control was included in which children simply received the pretest and posttest. Within each age groups's control condition, one-half of the children were assigned to the exposure control and one-half to the bias control.

Posttest. The posttest included 24 items, four each of balance, weight, distance, conflict-weight, conflict-distance, and conflict-balance types. The procedure was identical to that used on the pretest, with no feedback being given. Tasks were ordered randomly within the 24 item set, and one-half of the children received the problems in the opposite order from the other one-half. Unlike some of the tasks used in Experiment 1, all Experiment 2 pretest and posttest tasks involved weights on only one peg on each side of the fulcrum.

The pretest took approximately 10 min, the experience 25 min, and the posttest 15 min. Eight-year-olds were given the three parts in succession: 5-year-olds were given the pretest one day and the experience and posttest in a second session within the next 48 hr.

Participants. Pretests were given to 109 children, 56 kindergarteners (5-year-olds) and 53 third graders (8-year-olds), attending a middle-class school in suburban Pittsburgh. To equate the initial knowledge of participants, children whose responses were consistent with the distance cue on both distance tasks or on more than two of the six distance and conflict items were excluded from further participation in the experiment. This eliminated 21 children—one male and one female 5-year-olds, eight male and 11 female 8-year-olds. An additional 28 children—17 female and seven male 5-year-olds, and four female 8-year-olds—were excluded randomly from experiential and posttest phases in order to equate age and sex characteristics of the age and treatment groups. These children differed in no systematic way from their peers who did participate. Finally, the remaining 60 children, 30 5-year-olds and 30 8-year-olds, were randomly assigned within age and sex to the three treatment groups. All groups had equal numbers of males and females except for the 8-year-old control group that included four boys and six girls. The mean CA of kindergarteners was 70 months (range = 66–75 months), while the mean CA of the third graders was 106 months (range = 101–117 months). The experimenter, a 22-year-old female research assistant, served for all children.

Results

Data were collected on children's predictions for the eight pretest, 16 experience, and 24 posttest items. Children in the bias and exposure control groups performed quite similarly on all measures; for example, members of the exposure control solved 50% of posttest problems, members of the bias control 51%. Their percentage of solutions was similar on each type of problem. Boys and girls also behaved comparably;

TABLE 7

PERCENTAGE OF CORRECT PREDICTIONS—EXPERIMENT 2[a]

Age and treatment condition		Problem-type						Totals for condition
		Balance	Weight	Distance	Conflict-weight	Conflict-distance	Conflict-balance	
Control	5-years	75	98	8	98	0	0	
	8-years	84	98	40	98	8	0	51
Distance problems	5-years	88	98	55	95	12	2	
	8-years	98	92	92	92	20	2	62
Conflict problems	5-years[b]	72	86	14	92	8	0	
	8-years	88	90	78	58	42	8	53
Totals for problem-types		84	94	48	88	15	2	

[a] See text, footnotes 7 and 8.
[b] See text, footnote 6.

boys responded correctly to 56% of all posttest problems, girls to 55%. Therefore, in subsequent analyses, data were not distinguished by the type of control group that children were in nor by their sex.

Pretest data

Accuracy of predictions. A 2 (Age: 5-years or 8-years) × 3 (Experiential condition: conflict problems, distance problems, or control) × 4 (Problem-type: weight, distance, conflict-weight, or conflict-distance) analysis of variance on the number of correct pretest predictions revealed a single significant main effect for problem-type [$F(3,162) = 797.95$, $p < .001$]. Children solved a far greater number of weight and conflict-weight problems (96% of each type) than distance problems (14%), and more distance problems than conflict-distance problems (0%). In all cases, the percentages were similar for 5- and 8-year-olds—97 vs. 95% on weight problems; 10 vs. 18% on distance problems; 98 vs. 93% on conflict-weight problems; and 0 vs. 0% on conflict-distance problems.

Rule analysis. Forty-six of the 60 children, 77%, could be classified as using a rule. As noted previously, the selection process specifically excluded from further participation children who used Rule II or III on the pretest. Thus, all 77% of the children whose performance fit any classification followed Rule I. A chi-square test revealed no significant differences among age and treatment groups in the number of children using the rule ($\chi^2(5) = 5.92, p > .10$).

Experience

The experiences of the different groups were not directly comparable as they involved different types of problems. However, analysis of performance within each group for the first and second half of trials is revealing. Children in the exposure control who received only weight and balance items performed virtually perfectly. They solved 97% of problems in the first half of their training and 100% of the items in the second one-half. Performance of children who were presented conflict problems was also similar for the first and second halves of training, but at a much lower level. They predicted correctly on 50% of items in the first half and 52% of items in the second half. However, there was improvement within the distance problem group from 75 to 85% correct [$t(19) = 2.33$, $p < .05$). All trends were comparable for 5- and 8-year-olds and for the two orders of item presentation within each procedure.

Posttest

Number of correct predictions. A 2 (Age) × 3 (Treatment) × 6 (Problem-type) analysis of variance on the number of correct posttest predictions revealed significant main effects for age [$F(1,54) = 29.39$, $p < .001$], for experiential condition [$F(2,54) = 12.00$, $p < .001$], and for problem-type [$F(5,270) = 169.22, p < .001$].[7,8] Significant interactions between age and problem-type [$F(5,270) = 8.85$, $p < .001$] and between experiential condition and problem-type [$F(10,270) = 6.84$, $p < .001$] were also present (Table 7). As in Experiment 1, these main effects and interactions could be explained directly in terms of the children using different rules; therefore in the interest of brevity, no finer grain analysis will be reported (these analyses are available from the author upon request).

Rule analysis. A standard of 20 or more predicted responses among the 24 Experiment 2 posttest items was adopted as the standard for judgment that a child was adhering to Rule I, II, or IV. The criteria for Rule III were 10 predicted responses among the 12 for which specific

[7] A decision was made to exclude one 5-year-old in the conflict training condition from the analyses of posttest predictions in Experiment 2. This child used Rule I, but relied on the distance rather than on the weight cue. Thus, he solved, four balance, four distance, three conflict-distance, and no weight, conflict-weight, or conflict-balance items; his performance conformed to the distance cue on 23 of the 24 problems. Including him in the reporting of percentages of correct posttest predictions would have distorted the group pattern of results to no good purpose. This decision to exclude children conforming to the distance cue was later applied to the one kindergartener in Experiment 3e and one third grader in Experiment 3d who conformed to the distance cue; no children had conformed to it in Experiment 1.

[8] The 16 problems presented in the conflict problem sequence of Experiments 2 and 3 did not include any conflict-balance problems. Thus, whenever weight and balance cues conflicted, the scale always tipped one way or the other. Children who adopted Rule III apparently noted this, for on the posttest they very rarely predicted that the scale would balance when weight and distance cues conflicted (less than 2% of the time). This did not seem to change the basic nature of the muddling through process in Rule III, but did effectively change the domain of muddling from three to two choices.

TABLE 8

NUMBER OF CHILDREN USING DIFFERENT RULES—EXPERIMENT 2

Age and treatment		Rule I	Rule II	Rule III	Unclassifiable
5-year-olds	Control	8	0	0	2
	Distance problems	3	4	1	2
	Conflict problems	5	0	0	5
8-year-olds	Control	5	3	0	2
	Distance problems	0	8	1	1
	Conflict problems	0	2	5	3
	Total	21	17	7	15

predictions were made and at least three exceptions to the simple weight (distance) cue on the other 12 items. By formulas analogous to those used in Experiment 1 the probability of a random responder being judged as using Rule I, II, or IV was less than 5×10^{-7}; the probability of random placement in Rule III was less than 5×10^{-4}.

As shown in Table 8, 45 of the 60 children were judged to use one of the rules: 21 using Rule I, 17 using Rule II, and 7 using Rule III. A Chi-square test indicated that significant differences were present in the type of rule placements achieved by children in the six age by experience groups [$\chi^2(10) = 45.54$, $p < .001$]. More specific analyses revealed that 5-year-olds more often used Rule I and 8-year-olds more often Rules II or III [$\chi^2(1) = 12.91$, $p < .001$], and that children exposed to the control procedure more often used Rule I while those exposed to conflict or to distance problems more often used Rule II or III [$\chi^2(1) = 13.20$, $p < .001$].

An interactive relationship between experiential procedure and age was also apparent. Fisher Exact tests indicated that among 5-year-olds, experience with distance problems led to more adoptions of Rules II and III than did experience with conflict problems or the control conditions ($p < .01$). As can be seen in Table 8, the effect was almost exclusively to promote attainment of Rule II; no condition led to many children's attaining Rule III. Among the 8-year-olds, however, both distance and conflict problem experience led to more adoptions of Rules II and III than did the control procedures ($p < .001$) and conflict problem experience led to greater use of Rule III than did the distance problems and control conditions ($p < .01$).

This pattern of rule usage provided a framework for interpreting the age by problem-type and experience by problem-type interactions in the posttest predictions data cited above. Suppose for purposes of simplicity that all children used the modal rule for their age and experiential group (i.e., control group members used Rule I; distance problem group members used Rule II; 5-year-olds exposed to conflict problems used Rule I, and 8-year-olds exposed to conflict problems used Rule III). The following distinct pattern would be implied: Older children would excel over younger ones on distance and conflict-distance problems, but would do less well on conflict-weight tasks. Children given distance problem experience would predict more accurately on distance problems than children exposed to conflict problems, who in turn would predict more accurately on distance problems than children in the control group. On conflict-distance problems, children in the conflict problems condition would be more likely to respond correctly than children in the distance problems or control groups, who would not differ. On conflict-weight tasks, the reverse pattern would be found; children in the control and distance problems groups would not differ, and would excel over children who received conflict problems. As shown in Table 7, these predictions were in close accord with the data; taken together, they account for essentially the entire basis of the interactions.

Discussion

The most striking finding of Experiment 2 was that 5- and 8-year-olds, equated for using identical initial rules on the balance scale task, derived radically different lessons from experience with conflict problems. Older children benefited greatly from such experience, tending to adopt Rule III; younger children benefited not at all, tending to remain with Rule I. The obvious question was "Why?" Traditionally, such failures to learn are attributed to a lack of "readiness"; this attribution, however, only labels the phenomenon, it does not explain it. In hopes of generating more genuinely explanatory hypotheses, detailed protocol analyses were undertaken for three young children. Each of them was given pretest, experience, and posttest sessions that were similar to the conflict problem groups' except for the addition of impromptu questions and other attempts by the investigator to elicit the child's underlying logic. Videotapes were made to allow detailed and repeated scrutiny of all that transpired.

The theme that appeared consistently was that young children seemed to misencode problems, sometimes ignoring distance relationships altogether and other times attending to absolute rather than relative distance from the fulcrum. In other words, rather than viewing the problems in terms of continuous dimensions of weight and distance, 5-year-olds either formulated absolute hypotheses about the role of distance such as "when the weights are on the third peg they always go down" or ignored distance altogether.

These observations, together with previous work, suggested the *encoding hypothesis*: *Five-year-olds are less able to acquire new* *information than 8-year-olds because their encoding of stimuli is less adequate*. The vagueness of the phrase "less adequate" is deliberate; it is due to the variety of phenomena that the hypothesis subsumes. For example, Gelman's (1969) conservation training procedure emphasized discrimination between length and number dimensions; her view was that young children's encoding on these two dimensions was indistinct. Siegler (1975), considering children's formation of causal inferences, suggested that distracting stimuli prevented 5-year-olds from attending to a temporally delayed but regular relationship; here, the problem was viewed as a failure to encode on the relevant dimension at all. Bruner (1966), based his "perceptual screening" training procedure on the hypothesis that young children's failure to solve liquid quantity conservation problems was due to their overly rigid encoding along the single dimension of the liquid column's height. Finally, Hagen (1972), working with memorial problems, found that young children were poorer than older ones at recalling material they expected to be asked about, but better at recalling incidental material; he concluded that young children were less likely than older peers to limit their encoding to the relevant dimensions. Thus, 5-year-olds are seen to be less able than older children to encode separately along different dimensions, to encode certain dimensions at all, to avoid overly rigid encoding along a single dimension, and to encode rigidly enough along the relevant dimensions. In all cases, the 5-year-olds' encoding is seen as less adequate to meet the demands of the tasks.

Despite this convergence of supporting evidence, however, the encoding hypothesis' status as an explanation of developmental change is weak; it occupies more the place of a restatement than an explanation of data. Experiment 3 represented an attempt to determine whether the encoding hypothesis could also meet more rigorous standards in explaining the differential "readiness" to benefit from experience that was observed in Experiment 2.

EXPERIMENT 3

The Encoding Hypothesis

A first step toward the goal of more rigorous explanation was independent assessment of the proposed explanatory variable, to determine if performance on it actually followed the hypothesized pattern. The reconstruction paradigm introduced by Chase and Simon (1973) suggested a means by which children's encoding could be assessed independent of their predictive performance. Chase and Simon's experiment involved briefly presenting chess masters and nonmasters with either organized or disorganized arrangements of chess pieces, and then asking them to reproduce the exact configuration of pieces that they had observed. In Experiment 3 of the present study, 5- and 8-year-olds were briefly presented balance scale configurations of disks on pegs. Then the scale was hidden from sight, and a second identical scale was presented; the task was to reproduce on the second balance scale the arrangement of disks on pegs that had been observed on the first apparatus. It should be noted that this paradigm allowed fully independent assessment of encoding on the weight and distance dimensions. Consider an example in which the initial arrangement had three weights on the third peg to the left of the fulcrum and three weights on the second peg to the right. A child's performance could vary in four ways: both weight and distance could be accurately reproduced (the initial configuration); weight but not distance could be reproduced (three weights on the third peg to the left vs. three weights on the third peg to the right); distance but not weight could be reproduced (two on the third peg and three on the second peg); or neither dimension might be (four on the third peg vs. four on the third peg).

The encoding hypothesis suggested the following pattern of results. Older children, presumed to encode on both weight and distance dimensions, would be relatively accurate on each; no differences would be expected between their correct reproductions of weight and distance dimensions. Younger children, however, presumed to encode on weight but not on distance, would be relatively proficient only on the former dimension; a substantial difference would exist between their ability to reproduce weight and distance arrangements. The pattern would be expected to hold not only for overall differences between the age groups, but also when only those children of each age using approaches below the Rule II level were considered.

Method

Essentially the same procedure was followed in all phases of Experiment 3. This basic procedure will be described first, followed by descriptions of the variations pursued in Experiments 3a–3e.

Participants. Overall, 70 children, 40 kindergarteners (5-year-olds) and 30 third graders (8-year-olds), participated in Experiment 3: 10 kindergarteners and 10 third graders in Experiment 3a, another 10 kindergarteners in Experiment 3b, 10 more of each age in Experiment 3c, and an additional 10 of each age (the same 10) in Experiments 3d and 3e. The mean age for kindergarteners was 65 months (range = 56–70 months), while that for third graders was 100 months (range = 93–105 months). All children were from two elementary schools in Pittsburgh, each school including students of widely varying backgrounds. Within the experimental sample, roughly two-thirds of the students were white and one-third black. This was a similar proportion as in their schools overall. A 21-year-old white, female research assistant served as experimenter.

Materials. The two balance scales used in Experiment 3 were slightly different from the one used previously. They included seven rather than four pegs on each side of the fulcrum with the pegs 2 rather than 3 in. apart from each other. The new scales were also somewhat smaller than the old one, measuring 30 in. wide and 6.5 in. high. In addition, each scale's arm was held steady by a lever rather than by two pieces of wood. In basic function, though, the new equipment was identical to the old; weights could be placed on pegs in a variety of configurations, and children could be asked which side would go down or whether the scale would balance if the lever were released.

Another new material used in Experiment 3 was a white, Styrofoam board, 36 in. wide by 10 in. high, that was used in the encoding test.

Encoding test. The encoding test included 16 problems. On each problem, there were three, four, or five weights on a peg on one side of the fulcrum and three, four, or five weights on a peg on the other. The weights on each side were always placed on the third, fourth, or fifth peg from the fulcrum. Numerical values on the distance and weight dimensions were equated, so that if one problem included four weights on the third peg to the left of the fulcrum and three weights on the fifth peg to the right, then another, corresponding item would have three weights on the fourth peg to the fulcrum's left and five weights on the third peg to the right. Problems were randomly assigned to the 16 positions within the test sequence.

Predictions test. The predictions test was identical to the posttest used in Experiment 2.

Procedure. Children were brought individually by the experimenter to a vacant room within their school. All were presented the encoding test first and the predictions test second. As an introduction to the encoding task, children were told:

> The idea of the first game is for you to look how the weights are set on the pegs on my balance scale and then to make the same problem by putting the weights on the pegs on yours. First I'll put the weights on the pegs on my scale. You should watch closely to see how the weights are set on the pegs. Then I'll put the Styrofoam board back up so you can't see my scale. You will then need to put the weights on the pegs on your scale in the same way that you saw them on my scale. Don't worry about the different colors of the weights. Just put the weights on the pegs so it's just like the problem you saw on my scale.

After the first encoding trial, children were again told, "Remember, you should watch closely to see how the weights are on the pegs on my scale so that you can put the weights on the pegs on your scale in the same way." Children were allowed 10 sec to observe the initial balance scale arrangement on each trial, and then were immediately allowed to reproduce the arrangement on the other scale. There was no time limit for reproduction, though children usually finished quickly.

Following the last encoding problem, children were presented the predictions tasks with the following instructions:

> Now let's play another game. I'll put the weights on the pegs on this scale in different ways and you'll tell me whether this side would go down or this side would go down or they would both stay level like they are now if I released the lever. I'll have this lever set like this so that the balance scale won't actually move, but you tell me how the scale would go if I opened the lever.

Following these instructions, children were presented the 24 predictions trials, told they had done a good job, awarded a prize, and returned to their classrooms. The encoding and predictions tasks were given in a single session that lasted approximately 25 min.

EXPERIMENT 3a

In Experiment 3a, the encoding and predictions tasks were presented as described above to 20 haphazardly selected children, 10 5-year-olds and 10 8-year-olds.

Results and Discussion

Encoding data. Encoding performance was scored separately on the weight and distance dimensions. The criterion for a correct encoding was

TABLE 9

PERCENTAGE OF CORRECT ENCODINGS—EXPERIMENT 3

	5-year-olds		8-year-olds	
Experiment	Weight encodings	Distance encodings	Weight encodings	Distance encodings
3a	51	16	73	56
3b	54	9		
3c	54	19	64	73
3d	52	51	72	76

perfect performance on both the left- and right-hand values of the dimension being considered. (Other means of scoring that allowed partial credit for partially correct responses had no effect on the basic pattern of results. However, the error data are considered in terms of confusion matrices at the end of Experiment 3, and in that context are revealing.)

As shown in Table 9, the pattern of results was entirely consistent with the encoding hypothesis. There were substantial differences in the 5-year-olds' ability to correctly reproduce the weight and distance dimensions; they were correct more often on the weight dimension (51%) than on the distance dimension (16%) [$t(9) = 3.83, p < .01$]. Eight-year-olds, however, did not differ in their ability to encode weight (73% correct) and distance (56% correct) [$t(9) = 1.58, p > .10$].

These differing patterns of encoding were the product of individual 8-year-olds encoding on both dimensions and of individual 5-year-olds encoding on only one, rather than of some 8-year-olds encoding only on weight, others only on distance, and all 5-year-olds only on weight. To illustrate this point, a comparison was made between the number of older and younger children with comparable numbers of correct encodings on the two dimensions (comparable was here and throughout Experiment 3 arbitrarily defined as an absolute difference of four or less in the number of correct encodings on distance and weight dimensions). Eight of the 10 8-year-olds met this "comparability" criterion, versus only three of 10 5-year-olds (Fisher Exact probability < .05). Each of the seven other 5-year-olds produced at least seven more correct encodings on the weight than on the distance dimension.[9]

No differences were evident in the older and younger children's predictions performance (Table 10). Of the younger children, nine used Rule

[9] The absolute numbers were of no importance: for example, if the cutoff was a difference of three, six older and three younger children would meet the criterion; if the cutoff was five, eight older and three younger children would meet the criterion, etc.

TABLE 10

PERCENTAGE OF CORRECT PREDICTIONS—EXPERIMENT 3

		Problem type					
Experiment	Age	Balance	Weight	Distance	Conflict-weight	Conflict-distance	Conflict-balance
3a	5-year-olds	95	100	8	100	2	0
	8-year-olds	98	100	5	100	0	0
3b	5-year-olds	85	85	18	92	8	2
3c	5-year-olds	72	90	18	72	12	15
	8-year-olds	100	98	30	90	20	0
3d[a]	5-year-olds	72	92	22	86	17	6
	8-year-olds	100	100	22	100	0	0
3e[a]	5-year-olds	92	89	72	89	33	0
	8-year-olds	100	100	94	67	50	0

[a] See text, footnote 7.

I and one was not classifiable (Table 11). Of the older children, eight used Rule I, one used Rule II, and one did not fit into any category. (In the interest of brevity, the age by problem-type analysis of variance will not be reported in Experiment 3, except in the final section, Experiment 3e, where it constitutes the datum of primary interest. In all cases the analyses of variance provided a similar picture to that of the rule system analyses, as is suggested by the Table 10 percentages.)

These results indicated that there were developmental differences on the proposed explanatory variable (encoding) and that these differences were not dependent on differing predictive knowledge. Thus, the encoding hypothesis remained a viable explanation for the previously observed finding of different-aged children with identical predictive knowledge being differentially able to acquire new information about the balance scale task.

EXPERIMENT 3b

One possible interpretation was that younger children failed to encode the distance dimension simply because they were too slow in focusing their attention or in counting. In this view, the 5-year-olds might only have had time to encode one dimension, and therefore chose to encode the dimension they viewed as more important, weight (cf. Chi & Klahr, 1975).

The obvious test of this alternative was to allow 5-year-olds a greater amount of time. In Experiment 3b, 10 of them were provided 15 rather than 10 sec. Otherwise the procedure was identical to the one used in Experiment 3a.

TABLE 11

NUMBER OF CHILDREN USING DIFFERENT RULES—EXPERIMENT 3

Experiment	Age	Rules used			Unclassifiable
		I	II	III	
3a	5-years	9	0	0	1
	8-years	8	1	0	1
3b	5-years	7	0	0	3
3c	5-years	7	0	0	3
	8-years	6	2	1	1
3d	5-years	6	0	0	4
	8-years	6	1	0	3
3e	5-years	1	3	4	2
	8-years	0	3	7	0

Results and Discussion

As shown in Table 9, the encoding pattern was very similar to that obtained in Experiment 3a. The 5-year-olds correctly reproduced 54% of weight relationships but only 9% of distance relationships. The difference was significant [$t(9) = 9.73$, $p < .001$]. Only two children met the criterion for comparable encoding on the two dimensions. On the predictions test, seven followed Rule I and three did not conform to any of the proposed systems (Table 11). Thus, the "lack of time" explanation could be discarded.

EXPERIMENT 3c

Another possible reason for the developmental differences in patterns of encoding was that 5-year-olds might not have understood the instructions. That is, they might have thought that the request to "make the same problem" meant that they should reproduce only what was important (in their view) in the problem. Therefore, the instructions were changed to the following.

> The idea of the first game is for you to look how the weights are set on the pegs on my balance scale and then to make the same problem by putting the weights on the pegs on yours. You want it to be the same problem in two ways. You want the same number of weights on each side of your scale as I had on my scale, and you want the weights on each side of your scale to be the same distance from the center as they were on my scale. . . .

Later in the instructions, children were again told that they should "watch closely to see how the weights are set on the pegs—how many there are on each side and how far from the center the weights on each side are." Finally, at the end of the instructions, children were asked to indicate the two ways their arrangements should be like the experimenter's. This was to ensure that they understood what they had been told. The few children who did not understand were again presented the instructions and asked the identical question until they could answer appropriately. In all other ways, the procedure was the same as that used in Experiment 3b.

It should be noted that this proposed explanation for the developmental difference is not entirely artifactual. Not only were the instructions being clarified, but children were being told *what* to encode. This seemed a logically insuperable dilemma, given the nature of the interpretation being tested (that children might not have understood from the previous instructions what they were to attend to).

Results and Discussion

The results for the encoding task differed hardly at all from those of Experiments 3a and 3b. As shown in Table 9, 5-year-olds correctly encoded the weight dimension on 54% of problems, but the distance dimension on only 19% of problems [$t(9) = 3.42$, $p < .01$]. Eight-year-olds encoded correctly on weight 64% of the time and on distance 73% of the time ($t < 1$). Eight older children met the comparability of encoding criterion, versus only two younger ones ($p < .025$). On the predictions measure, children again tended to use Rule I (Table 11).

The consistency of the encoding data with the previous pattern indicated that misinterpretation of instructions was not the problem. Telling children what to encode left intact the differences between 5-year-olds' encoding of the weight and distance dimensions. The absolute percentage of correct encoding on weight and distance dimensions also was unchanged for both 5- and 8-year-olds. This suggested that the difficulty might reside in the younger children not knowing *how* to encode the distance dimension, or perhaps in their not knowing how to encode any two quantitative dimensions simultaneously. The possibility was tested in Experiment 3d.

EXPERIMENT 3d

In Experiment 3d, children were provided direct training in both what and how to encode. Their instructions were identical to those provided in Experiment 3c until they had been told that they wanted "the same number of weights on each side of your scale as I had on my scale and . . . the weights on each side of your scale to be the same distance from the center as they were on my scale." Then they were told:

> You do it like this. First you count the number of weights on this side—one, two, three, four (this action was carried out on a balance scale that was on the

TABLE 12a

PERCENTAGE OF NEAR AND FAR ERRORS—ENCODING OF WEIGHT

Distribution of weights on two sides of fulcrum[a]	Experiment 3(a–c)				Experiment 3d			
	5-year-olds		8-year-olds		5-year-olds		8-year-olds	
	Near errors	Far errors	Near errors	Far errors	Near errors	Far errors	Near errors	Far errors
3–3	9	3	13	1	13	10	15	3
3–4	28	13	25	0	40	0	30	0
4–3	47	5	15	5	55	15	15	0
4–4	30	3	28	0	55	5	10	0
4–5	60	18	43	15	50	5	55	0
5–3	53	27	40	0	30	10	50	0
5–4	53	17	25	5	70	10	10	10
5–5	58	15	55	5	40	10	20	15
Averages	37	10	28	4	39	8	24	3

[a] Percentages in table are based on total number of problems—the percentage of correct reproductions on each configuration = 1 − (percentage near errors) − (percentage far errors).

table between experimenter and child). Then you count the number of pegs the weights are from the center—first, second, third. So you say to yourself "four weights on the third peg." Then you would do the same for the other side—one, two, three, four, five weights on the first, second, third peg. So it would be five weights on the third peg. Then you would say "four weights on the third peg and five weights on the third peg." Then you would put the right number of weights on the right pegs on each side. Let's practice one.

The experimenter modeled this procedure, then child and experimenter together completed a problem, then the child did seven practice tasks on which the experimenter provided feedback when the counting of either weights or pegs was omitted or incorrect. Following this, children were given the usual encoding and predictions tasks. The direct implication of the encoding hypothesis was that instruction in how to encode should reduce or eliminate differences in the 5-year-olds' ability to encode weight and distance dimensions. Eight-year-olds, presumably already knowing how to encode each dimension, were not expected to be affected by the training. Nor was the procedure expected to influence either age groups' predictions performance. Older children in the previous parts of Experiment 3 had been observed to encode accurately on each dimension yet to possess Rule I level knowledge. This was the situation expected for both older and younger children after the encoding instruction.

Results and Discussion

Encoding instruction changed the pattern of 5-year-olds' encoding substantially and as predicted. As shown in Table 9, they now correctly encoded 52% of problems on the weight dimension and 51% on the distance dimension. The corresponding figures for 8-year-olds were 72 and 76%

TABLE 12b

PERCENTAGE OF NEAR AND FAR ERRORS—ENCODING OF DISTANCE

Distribution of pegs on two sides of fulcrum[a]	Experiments 3(a–c)				Experiment 3d			
	5-year-olds		8-year-olds		5-year-olds		8-year-olds	
	Near errors	Far errors	Near errors	Far errors	Near errors	Far errors	Near errors	Far errors
3–3	42	38	20	7	33	0	13	0
3–4	37	47	50	5	50	0	40	0
4–3	23	50	30	5	60	0	30	0
4–4	52	43	47	2	53	0	27	0
4–5	45	34	28	3	15	10	10	0
5–3	41	54	38	8	45	10	30	5
5–4	37	60	45	5	50	10	20	10
5–5	47	34	17	7	40	27	13	0
Averages	43	37	32	5	41	8	21	1

[a] Percentages in table are based on total number of problems—thus, the percentage of correct reproductions on each configuration = 1 − (percentage near errors) − (percentage far errors).

($ts < 1$). Thus, while the absolute levels of performance differed between the two age groups, the pattern of encoding on weight and distance was very similar. In addition, all 10 older children and all 10 younger ones met the comparability criterion. Also as predicted, both groups remained predominantly at the Rule I level on predictions performance (Table 11).

A contrast of children's encoding errors in Experiments 3(a–c) with their errors in Experiment 3d provided additional rather striking corroboration for the conclusion that age-related differences in encoding were substantially reduced by the encoding instruction. Table 12 (a and b) shows the number of "far" and "near" errors on weight and distance dimensions that children made given different stimulus configurations. Near errors were defined as errors in which one side was reproduced correctly and the other side was off by one weight or peg, or as errors in which an initial equality was preserved and both sides were mistaken by one unit. Thus, if the initial configuration had four weights on each side, a child could make a near error by placing three weights on each side, five weights on each side, four weights on one side and five on the other, or four weights on one side and three on the other. All other errors were considered far errors.

As shown in Table 12, the distribution of near and far encoding errors in Experiment 3 (a–c) differed substantially for the 5- and 8-year-olds, and in a way reminiscent of their patterns of correct and incorrect encodings. The 8-year-olds' pattern of errors on both weight and distance dimensions and the 5-year-olds' pattern of errors on the weight dimension were quite similar—an overwhelming majority of errors were near errors. As previously, however, 5-year-olds performed differently on the distance dimension; in this case, they produced an almost even division between near and far errors. In fact, 5-year-olds made a greater percentage of far errors on the distance dimension on each of the eight stimulus configurations than they made on any of the configurations on the weight dimension or than 8-year-olds made on any configuration on either dimension (Tables 12a and b). This pattern changed in Experiment 3d, with the provision of encoding training. Now, the errors made by children of both ages on both dimensions were overwhelmingly near errors (Table 12).

EXPERIMENT 3e

The same children who participated in Experiment 3d were brought back to the experimental room either 2 or 3 days afterwards. They were told that they would be playing with the balance scale again, and reminded that both the number of weights on the pegs and the distance of the pegs from the center were important. Following this, the Experiment 2 conflict problem procedure was presented. If the encoding hypothesis was correct, both 5- and 8-year-olds would now be expected to benefit from experience with conflict problems that only had benefited the 8-year-olds previously.

Results and Discussion

As shown in Table 9, conflict problems experience now aided both age groups. Virtually all children of both ages progressed as far as Rule II, and seven older and four younger children progressed to Rule III. A 2 (Age) $\times$ 6 (Problem-type) analysis of variance on the number of correct posttest predictions revealed a single significant main effect for problem-type [$F(5,80) = 45.85$, $p < .001$]. Balance, weight, and distance problems were more often solved than conflict-weight problems, conflict-weight problems were more often solved than conflict-distance problems, and conflict-distance problems were more often solved than conflict-balance problems. There was no significant main effect for age, and no interaction between age and problem-type. A Chi-square test indicated that there were also no differences in the distribution of rules adopted by the two age groups [$\chi^2(2) = 1.28$, $p > .05$], though as shown in Table 11, 8-year-olds were somewhat more likely than 5-year-olds to adopt Rule III.

The Table 10 data indicate that children of both ages predicted correctly on a large majority of distance problems though the older children were even more proficient than the younger ones. There were also differences in conflict-weight and conflict-distance problems, with 5-year-olds doing better on conflict-weight problems and less well on conflict-distance problems. These reflected the greater number of older children who attained Rule III; the performance of the seven older children and four younger ones who used this rule was closely comparable on all problem types.

Thus, the qualitative differences in reaction to conflict problem experience were eliminated by prior training in encoding. Although somewhat fewer younger than older children adopted Rule III, their encoding performance also had not reached as high a level. At the very least, reducing the differences in encoding substantially reduced the differential responsiveness to experience previously exhibited by 5- and 8-year-olds.

It should be noted that differences did remain between younger and older children's number of correct encodings. Absolute memorial capacity did not appear to be the source of these differences; in Experiment 3d, when the accuracy of 5-year-olds' encoding of the distance dimension increased greatly from the level of Experiments 3a–c, there was no corresponding decrement in the number of correct weight encodings. However, young children are known to differ from older ones in a wide variety of other ways that might have influenced performance on the encoding task: susceptibility to fatigue, ability to focus attention, accuracy of counting, and knowledge and use of memorial strategies, to name a few. Such causes might account for the directional difference toward 8-year-olds' deriving greater benefits than 5-year-olds from experience with distance problems in Experiment 2 and with conflict problems in Experiment 3e.

In general, though, the encoding hypothesis seemed to explain a large part of 5- and 8-year-olds' differential reactions to experience. The finding was consistent with previous interpretations of empirical work (Bruner, 1966; Gelman, 1969; Hagen, 1972; Siegler, 1975) and also with previous theoretical accounts (Bruner, 1964; Pascual-Leone & Smith, 1969; Piaget, 1971; White, 1967). However, the present study suggested a modification of the theoretical emphasis. Previous accounts indicate that changing children's attentional or encoding strategies will directly change their knowledge. For example, Bruner (1964) hypothesized that children would immediately conserve if only they encoded conservation problems at a symbolic rather than at an iconic level. By contrast, the present study, in which encoding and predictive knowledge were assessed independently, suggests that the relationship between encoding and problem solving often may be more dynamic. Experiment 3d demonstrated that changing children's encoding had no direct effect on their predictive knowledge; however, Experiment 3e showed that the changes in encoding allowed children to acquire a new knowledge that they would not otherwise have been able to acquire. It seems likely, then, that improved encoding can either produce direct changes in knowledge or can produce the conditions necessary for such changes. Which type of phenomenon occurs in a specific situation may provide an index of how far children are from acquiring the relevant concept. If changes in encoding produce new knowledge directly, then children might be presumed to be reasonably close to inducing the knowledge on their own; if intervening experience is also necessary then the acquisition may be more removed from the child's level. It is interesting in this regard that the 4-year-olds in Bruner's perceptual screening experiment did not derive the conservation knowledge directly, although 5- and 6-year-olds did.

At a very general level, the encoding hypothesis suggests a three step view of development. First, knowledge is at a particular point, and encoding of relevant stimuli is well adapted to the constraints of that knowledge. Second, the range of dimensions that are encoded expands, but knowledge remains unchanged. Third, knowledge grows and becomes consistent with the new encoding. As might be apparent, this formulation is closely related to the Piagetian construct of equilibration, with the present term encoding playing a similar role to the Piagetian term assimilation. However, there is one crucial difference between the two formulations; it is possible to independently measure encoding, whereas no means have been devised to measure assimilation. With measurement comes the possibility of prediction; thus, if encoding and knowledge can be independently measured it should be possible to identify individuals who are in a state of disequilibrium on a particular concept and thereby to predict which individuals are and which individuals are not ready to benefit from experience. This approach would avoid the circularity of the equilibration and readiness concepts as they are typically used.

The explanatory method used in Experiment 3 would also seem to have utility beyond the particular hypothesis it was used to test. Its steps include proposing an explanatory factor, independently assessing whether developmental differences of the appropriate form are present on it, demonstrating that these differences are not artifactual, demonstrating that the differences can be lessened by theoretically relevant manipulations, and finally showing that reducing the developmental differences on the explanatory factor also reduces the originally to-be-explained difference. One strength of this method is its high degree of testability; the explanatory hypothesis can be falsified at any of a number of steps. Philosophers of science have frequently emphasized the importance of methodologies that provide numerous opportunities for the disconfirmation of hypotheses (cf. Platt, 1964; Popper, 1968). Another advantage is that the method makes explicit a number of relationships that are implied by typical interpretations of developmental change but that are rarely examined directly.

With specific reference to the investigation of cognitive development, the explanatory method promises to serve a weeding-out function. The need for such a sharp-edged tool is aptly illustrated by Beilin's (1971) conclusion concerning the lessons of conservation and class inclusion training studies, "What emerges from the data is the striking fact that a wide variety of methods—in fact, practically all types of experimental methods—lead to successful improvement in performance, even if not in every experiment" (p. 113). It may be that the large number of successful training approaches reflects an equally large number of alternative paths to acquiring cognitive capacities, but another possibility is that the numerous successful training studies reflect the common presence of one or a few basic features. Brainerd and Allen (1971) suggested a synthesis of this form when they noted that a variety of successful conservation training techniques with different labels (e.g., verbal rule, logical multiplication, and learning set training) all referred to the reversibility of conservation transformations; Brainerd and Allen therefore hypothesized that emphasis on reversibility might be the active factor in the success of all of the procedures. Such similarities among training approaches might be more readily discovered if near the outset investigators attempted to independently demonstrate developmental differences in their proposed explanatory variables, and then showed that training explicitly geared to reducing these differences also reduced differences on the original tasks of interest.

This approach has a built-in bias in favor of certain types of explanatory factors. It is reductionistic in the sense of trying to reduce developmental differences in conceptual functioning and reasoning to more process-level differences in discrimination, understanding of instructions, encoding, and memory. There is also a bias favoring experimentally manipulable explanatory factors, and factors that can be independently assessed. Finally, implicit in the approach is the recognition that developmental differences exist at multiple levels and that the level chosen in any particular case will depend on the data to be explained. The present study illustrates this point. When only the predictions data were in need of explanation, the decision rule models sufficed. When differential response to experience among children using identical rules was being considered, however, another level of analysis became necessary. Even more detailed data would probably need to be considered to account for reaction time on the balance scale tasks; consideration of both finer and coarser grained explanatory factors might be necessary to account for children's degree of generalization among problem isomorphs. Cognitive development clearly occurs on multiple levels; enumerating these levels and outlining the interrelationships among them are among the most profound tasks facing students of development.

CONCLUSIONS

The goals of this study were to characterize and explain developmental differences in children's understanding of balance scale problems. Investigation centered on three aspects of development: children's existing knowledge, their ability to benefit from relevant experience, and their encoding of stimuli. The different focuses yielded a consistent and plausible account of developmental change.

The initial aim was to characterize children's knowledge of how balance scales operated. It was hypothesized that this knowledge could be precisely represented in terms of decision rules, differing only in their consideration of weight and distance factors. Six types of problems were devised to test the rule models; a child adhering to any one of the rules would produce a distinctive pattern of predictions, unlikely to be the result of a random response process or of adherence to a different rule. Summing over the three experiments, the proposed rules accurately described the performance of more than 80% of children. The descriptive accuracy was virtually as high for 5-year-olds as for 17-year-olds. The rules characterized children's knowledge following change-inducing experience as well as their pre-existing knowledge. Performance on problems within each problem-type was quite consistent, while performance on problems of different types was quite distinct. Although the rules were proposed as models of individual performance, they suggested predictions concerning the relative difficulty of problem-types as well as developmental trends in performance on them. These predictions were consistently supported.

Experiment 2 concentrated on another aspect of children's understanding of balance scale problems, their ability to induce more mature rules from experiences of different types. It was already known that children of different ages differed in the rules they used; the further question was whether children of different ages, equated for using the same rule initially, would differ in their ability to induce new rules. The results of Experiment 2 demonstrated that there were developmental differences, beyond the rules on which predictions were based, that influenced children's ability to acquire new information. Five- and eight-year-olds, initially using the same predictive rules, derived radically different lessons from experience with conflict problems. Eight-year-olds in the condition made greater progress toward Rule III than members of any other experimental group; 5-year-olds in it made no progress whatsoever. Protocol analyses of children given conflict problems, together with the results of previous research, suggested the encoding hypothesis as an interpretation of the diverging patterns—5-year-olds were hypothesized to be less able to induce new information than 8-year-olds because their encoding of stimuli was less adequate.

This interpretation was tested in Experiment 3. An independent assessment procedure revealed that 5- and 8-year-olds did encode balance scale problems differently; older children encoded both weight and distance dimensions while younger ones encoded only the amount of weight. The finding was not attributable to the younger children's being slower or to their misunderstanding the instructions. However, tuition in how to encode produced the same pattern of encoding in the 5- as in the 8-year-olds, and once the 5-year-olds encoded both dimensions they too benefited from experience with conflict problems. Thus, the encoding hypothesis seemed to explain a good part of the developmental change in responsiveness to experience.

REFERENCES

Beilin, H. The training and acquisition of logical operations. In M. F. Rosskopf, L. P. Steffe, & S. Taback (Eds.), *Piagetian cognitive developmental research and mathematics education*. National Council of Teachers of Mathematics, 1971. Pp. 78–118.

Blatt, M. The effects of classroom discussion upon children's level of moral judgment. In L. Kohlberg & E. Turiel (Eds.), *Recent research in moral development*. New York: Holt, Rinehart & Winston, 1971.

Brainerd, C. J., & Allen, T. W. Experimental inductions of the conservation of "first order" quantitative invariants. *Psychological Bulletin*, 1971, **75**, 128–144.

Bruner, J. S. The course of cognitive growth. *American Psychologist*, 1964, **18**, 1–15.

Bruner, J. S. On the conservation of liquids. In J. S. Bruner, R. R. Olver, and P. M. Greenfield (Eds.), *Studies in cognitive growth*. New York: Wiley, 1966. Pp. 183–207.

Case, R. Structures and strictures: Some functional limitations on the course of cognitive growth. *Cognitive Psychology*, 1974, **6**, 544–574.

Chase, W. G., & Simon, H. A. The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press, 1973. Pp. 215–281.

Chi, M. T. H., & Klahr, D. Span and rate of apprehension in children and adults. *Journal of Experimental Child Psychology*, 1975, **19**, 434–439.

Flavell, J. H. An analysis of cognitive developmental sequences. *Genetic Psychology Monographs*, 1972, **86**, 279–350.

Gelman, R. Conservation acquisition: A problem of learning to attend to relevant attributes. *Journal of Experimental Child Psychology*, 1969, **7**, 167–187.

Hagen, J. W. Strategies for remembering. In S. Farnham-Diggory (Ed.), *Information processing in children*. New York: Academic Press, 1972. Pp. 66–78.

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

Jackson, S. The growth of logical thinking in normal and subnormal children. *British Journal of Educational Psychology*, 1965, **35**, 255–258.

Kuhn, D. Mechanisms of change in the development of cognitive structures. *Child Development*, 1972, **43**, 833–844.

Kuhn, D., & Angelev, J. *An experimental study of the development of formal operational thought*. Paper presented at Society for Research in Child Development Conference, Denver, Colorado, 1975.

Lee, L. C. The concomitant development of cognitive and moral modes of thought. A test of selected deductions from Piaget's theory. *Genetic Psychology Monographs*, 1971, **85**, 93–146.

Lovell, K. A follow-up study of Inhelder and Piaget's 'The growth of logical thinking.' *British Journal of Psychology*, 1961, **52**, 143–153.

Pascual-Leone, J., & Smith, J. The encoding and decoding of symbols by children. *Journal of Experimental Child Psychology*, 1969, **7**, 328–355.

Piaget, J. *The child's conception of time*. New York: Ballantine Books, 1971.

Platt, J. R. Strong inference. *Science*, 1964, **146**, 347–353.

Popper, K. R. *The logic of scientific discovery*. New York: Harper & Row, 1968.

Resnick, L. B. Task analysis of instructional designs: Some cases from mathematics. In D.

Klahr (Ed.), *Cognition and instruction*. Hillsdale, NJ: Lawrence Erlbaum, 1976.

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

Siegler, R. S., & Atlas, M. Acquisition of formal scientific reasoning by 10- and 13-year-olds: Detecting interactive patterns in data. *Journal of Educational Psychology*, 1976, **68**, 360–370.

Siegler, R. S., & Liebert, R. M. Effects of contiguity, regularity and age on children's causal inferences. *Developmental Psychology*, 1974, **10**, 574–579.

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

Siegler, R. S., Liebert, D. E., & Liebert, R. M. Inhelder and Piaget's pendulum problem: Teaching preadolescents to act as scientists. *Developmental Psychology*, 1973, **9**, 97–101.

Turiel, E. An experimental test of the sequentiality of developmental stages in the child's moral judgment. *Journal of Personality and Social Psychology*, 1966, **3**, 611–618.

White, S. H. Evidence for a hierarchical arrangement of learning principles. In L. P. Lipsitt and J. Spiker (Eds.), *Advances in child development and behavior*. Vol. 2, 1965.

(Accepted April 20, 1976)