Klahr, D., & Siegler, R. S. (1978). The representation of children's knowledge. In H. Reese & L. Lipsitt (Eds.), *Advances in child development and behavior, Vol. 12* (pp. 61-116). New York: Academic Press.

THE REPRESENTATION OF CHILDREN'S KNOWLEDGE

David Klahr and Robert S. Siegler
CARNEGIE-MELLON UNIVERSITY

I. Introduction

In this chapter we will discuss two related issues. One issue concerns the ways that children from 5 to 17 years perform a scientific induction task. We will summarize a series of experiments designed to investigate questions about initial knowledge, instructional effectiveness, and individual differences in both initial performance and responsiveness to instruction. The second issue is methodological: its focus is not on *what* we can say about children's knowledge of a task, but rather on *how* we can say it. That is, the second issue we will address is the representation of children's knowledge.

The two issues are related simply because the researcher's decision about how to represent knowledge plays a central role in guiding both the kind of theory that gets formulated and the kind of experiment that gets run. We have found this to be the case in our own studies, and we believe that it might be worthwhile to direct attention to some properties of different representations and criteria for choosing among them.

Our discussion will move back and forth between general conceptual issues and some very specific examples of both empirical techniques and theoretical statements. We will start by describing the historical trend in instructional psychology that has made the representation of knowledge a central issue, and then we introduce criteria that we believe might be useful in choosing and evaluating different representations. In addition to a set of evaluative criteria, we will list five central questions for research in developmental and instructional psychology.

Next we will introduce a specific task that has interesting psychological and instructional properties: a variant of Piaget's balance scale problem. We will present a formal model—using a particular representation—for different levels of knowledge that children might have about how to do the task. The task will provide the concrete reference for the rest of our chapter.

Having described the formal properties of the task, and some predictions about the performance of different aged children on it, we will then describe our first experiment. Based upon the results of the experiment, we will evaluate the initial hypotheses, as well as examine the merits and limitations of the representation in which the initial models are stated.

The initial representation and the associated experiment enable us to make certain predictions about the effects of an instructional sequence. In the second experiment we will explore some instructional issues, and this in turn will reveal some limitations of the initial representation. In particular, we find that older and younger children who are initially classified by our models as having identical task-specific knowledge show a striking differential responsiveness to instruction. This presents a serious challenge to our initial representation of children's knowledge. It is clear that the initial formulation does not tell the whole story about differences in task-specific knowledge.

A revised representation of the knowledge required to perform at different levels is introduced. The representation is a production system and some of the general properties of production systems are discussed. Then we will present an analysis of the problem-by-problem performance of two children during a training sequence, and formulate a more detailed production system model of the knowledge of one of them. The model is actually run as a computer simulation and its results are compared with the child's performance. This fine-grained analysis suggests that the initial encoding of the stimulus may be a crucial difference between older and younger children, and that this may account for the results of Experiment 2.

This encoding hypothesis states that differential responsiveness to training between 5- and 8-year-old children is due to differences in the way they encode the balance scale dimensions. The explanatory power of the hypothesis is illustrated in detail in Experiment 3.

Finally, we will summarize the preceding discussion in terms of how well it answers our initial set of questions. Then we will briefly discuss the several types and levels of knowledge that might be important in instructional investigation. Different levels of aggregation of both model and data are obviously appropriate for different scientific questions. The direct and explicit consideration of some ways to represent knowledge will provide useful guidelines for further empirical work.

II. From Behavioral to Cognitive Objectives

The goal of any instructional effort is the production of new knowledge in the learner. Over the last 15 years of instructional research there has been an increasing emphasis on stating such goals as clearly as possible. The trend was to move from an emphasis on simply describing educational means—the sequence of instructional activities—to a prior statement of the desired ends of instruction. The elaboration of behavioral objectives was perhaps the most extensive formalization of this trend (Mager, 1962). Behaviors were typically specified in great detail, although the underlying processes were not. However, even behavioral objectives have implicit in them an underlying cognitive theory: Behavior is simply an observable indicator of underlying cognitive processes.

One well-known normative model for instruction (Glaser, 1968) stresses the need to determine the learner's initial state as well as the desired end state. In the original formulations of this approach, both initial and final learner states were described primarily in terms of tasks and subtasks arranged in a Gagne-like hierarchy. There was little mention of how one might characterize the underlying psychological processes that acted upon them to produce the task behavior in question. As that approach has developed, however, it has focused increasingly upon such cognitive representations (see Resnick, 1976, for a summary of this trend in the area of elementary mathematics instruction).

Perhaps the strongest statement of the desirability and feasibility of describing the learner in terms of internal psychological representations is Greeno's view of "cognitive objectives." Greeno (1976) argued that cognitive psychology has now developed powerful and flexible methods for the representation of knowledge. Using an example from instruction in elementary fractions, Greeno showed how two different views of the conceptual content of the subject matter can be represented explicitly by two quite distinct cognitive structures, which in turn lead to differential predictions about problem difficulty, problem-solving strategies, and optimal procedures for instruction. Without such a representation, these predictions might never have been made.

This is essentially the same point stressed by Klahr and Wallace (1976) with respect to the need for explicit and precise models in cognitive development: "A theory of transition can be no better than the associated theory of what it is that is undergoing that transition" (p. 14). According to this view, the first step in the formulation of developmental theories is the creation of a precise model of the initial and final form of the cognitive process under investigation. Studies by Baylor and Gascon (1974), Young (1973), Klahr and Wallace (1976), and several others, summarized in Siegler (1978), provide developmental models of this type.

III. Some Criteria for Choosing a Representation

What considerations might guide us in the choice of a representation? What kinds of representations are available, and what are their relative merits? Knowledge representation has become an important topic in the emerging field of

"cognitive science" (Bobrow & Collins, 1975), and some initial attempts to address it can be found in Bobrow (1975), Becker (1975), Moore and Newell (1974), and Reddy and Newell (1974). These efforts constitute the first steps toward a full-fledged theory of representation, and they have already yielded a reasonable set of dimensions with which to characterize different representations.

Although such taxonomic systems allow us to classify representations, they do not make any statements about their relative merits. Regardless of the final location of a representation along the dimensions of importance, the ultimate evaluation of the quality of a representation depends upon the set of questions being addressed.

We believe that in the area of instruction and development the important questions are:

Question 1: What are the differences in knowledge that underlie different levels of task performance?

Question 2: What are the alternative strategies that might result in any given level of task performance?

Question 3: For a given level of performance, what is the optimal level of difficulty for an instructional sequence?

Question 4: What are the critical features of an instructional sequence that enable it to have any effect?

Question 5: When and why will two learners at the same initial performance level learn differently from the same instructional sequence?

Or, to summarize our concerns: What do children know about a task, how do they learn about it, and why do some know more and/or learn more than others?

Given this set of questions, there are four criteria that we believe to be most important in choosing a representation:

1. The representation must be sufficient to account for behavior. Thus, it must have a clear mapping onto the empirical base it is supposed to account for.
2. It should be amenable to multiple-level analyses. That is, it should be easy to aggregate and disaggregate the grain of explanation. For the design of well-controlled experiments or curriculum design, the representation will have to be stated in terms of averages across many subjects; it must be a modal form. For detailed study of individual strategies and component processes, it must be capable of disaggregation without drastic revision.
3. The representation should conform to the relevant properties of the human information-processing system as determined by laboratory studies of human processing capacities.
4. The representation should have "developmental tractability" (Klahr & Wallace, 1970). It should allow us to state both early and later forms of competence and provide an easy interpretation of each model as both a precursor and successor of other models in a developmental sequence (see Resnick, 1976, for a similar viewpoint).

IV. Balance Scale Task

The type of balance scale used throughout our investigation consisted of a two-arm balance, with several pegs located at equal intervals along each arm. Small circular disks, all of equal weight, were placed on the pegs in various configurations (as shown in Table I), while the balance was prevented from tipping. The subjects' basic task was to predict the direction in which the balance scale would move if it were allowed to. In order to answer some of the questions listed above, several variations on this basic theme were introduced. These included: asking children to explain their predictions; allowing the scale to move to its equilibrium position (thus providing feedback about the accuracy of the predictions); observing an experimenter-controlled series of configurations and their effects; constructing one's own configurations; and reconstructing initial configurations from memory. (A more complete report of these experiments is presented in Siegler, 1976.)

The basic physical concept that underlies the operation of the balance scale is torque: The scale will rotate in the direction of the greater of the two torques acting on its arms. The total torque on each arm is determined by summing the individual torques produced by the weights on the pegs, and the individual torques are in turn computed by multiplying each weight by its distance from the fulcrum. Since the pegs are at equal intervals from the fulcrum, and the weights are all equal, a simpler calculation is possible. It consists of computing the sum of the products of number of weights on a peg times the ordinal position of the peg from the fulcrum. This is done for each side, and the side with the greater sum of products is the side that will go down. (If they are equal, the scale will balance.)

The components of this knowledge are acquired over a remarkably long span of experience and education; even 5-year-olds often know that balances such as teeter-totters tend to fall toward the side with more weight, while many 16-year-olds do not know the appropriate arithmetic computations for determining the balance's behavior (Jackson, 1965; Lee, 1971; Lovell, 1961). It even seems likely that most college-educated adults could not easily state the physical principles that underlie the sum-of-products algorithm. Furthermore, for many configurations, there are shortcuts that eliminate the need to do any arithmetic computation (e.g., identical configurations on each arm will balance; if both weight and distance are greater on one side, that side will go down).

Note that the balance scale task shares a common property of many scientific problems: The universal rule for generating correct predictions is easy to describe, and once known, it is easily remembered and executed. However, the formulation of the rule—either by induction from empirical examples or by deduction from general physical principles—is quite difficult.

A. Representation of Children's Knowledge about the Balance Scale

Siegler (1976) suggested that the different levels of knowledge that children have about this task could be represented in the form of binary decision trees (see Fig. 1). The model of mature knowledge (Model IV, Fig. 1D) was suggested by a task analysis of balance scale problems; the models of less sophisticated knowledge (Models I–III, Figs. 1A–C) were derived from the empirical results of Inhelder and Piaget (1958) and Lee (1971), and from our own pilot studies. A child using Model I considers only the number of weights on each side: If they are the same, the child predicts balance, otherwise he predicts that the side with the greater weight will go down. For a Model II (Fig. 1B) child, a difference in weight still dominates, but if weight is equal, then a difference in distance is sought. If it exists, the greater distance determines which side will go down, otherwise the prediction is balance. A child using Model III (Fig. 1C) tests both

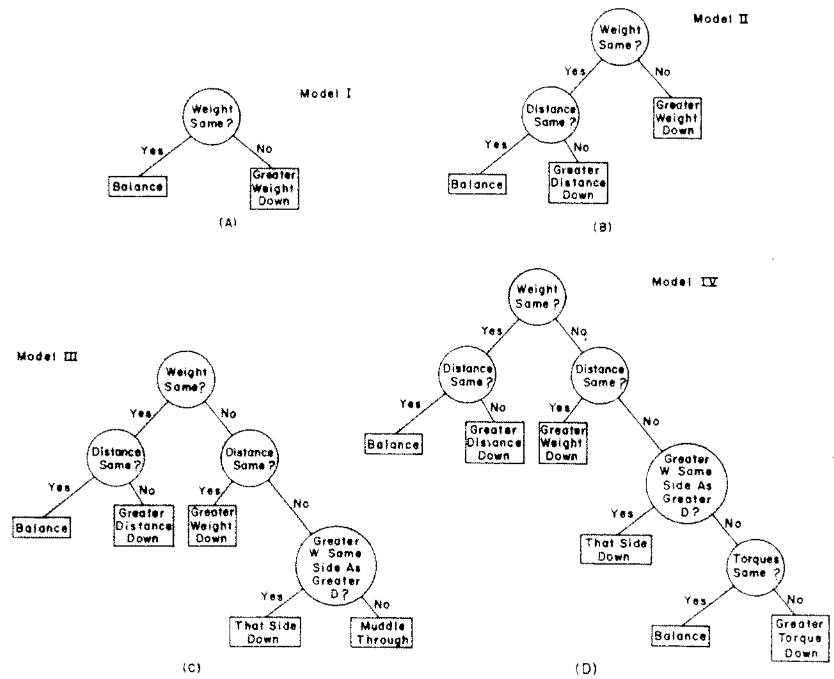


Fig. 1A–D. Decision tree representations for Models I–IV of balance scale predictions. A, Model I; B, Model II; C, Model III; D, Model IV. D = distance; W = weight.

weight and distance in all cases. If both are equal, the child predicts balance; if only one is equal, then the other one determines the outcome; if they are both unequal, but on the same side with respect to their inequality, then that side is predicted to go down. However, in a situation in which one side has the greater weight, while the other has the greater distance, a Model III child, although recognizing the conflict, does not have a consistent way to resolve it. This child simply "muddles through" by making a random prediction. Model IV represents "mature" knowledge of the task: Since it includes the sum-of-products calculation, children using it will always make the correct prediction. Note, however, that if they can base their prediction on simpler tests, they will do so.

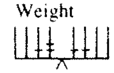
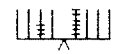
B. ASSESSING THE ACCURACY OF THE REPRESENTATIONS

It is possible to determine which, if any, of these four models accurately characterizes a child's knowledge about the balance scale task by examining his pattern of predictions for six types of problems (see Table I for an example of each type): (1) balance problems, with the same configuration of weights on pegs on each side of the balance; (2) weight problems, with unequal amounts of weight equidistant from the fulcrum; (3) distance problems, with equal amounts of weight different distances from the fulcrum; (4) conflict-weight problems, with more weight on one side and "more distance" (i.e., occupied pegs further from the fulcrum) on the other, and the configuration such that the side with more weight goes down; (5) conflict-distance problems, similar to conflict-weight, except that the side with more distance goes down; (6) conflict-balance problems, like other conflict problems, except that the scale remains balanced.

Children whose knowledge corresponded to different models would display dramatically different patterns of predictions on the six types of problems just listed. Those using Model I would consistently make correct predictions on balance, weight, and conflict-weight problems, and they would never be correct on the other three problem types. Children using Model II would behave similarly to those using Model I on five of the six problem types, but they would correctly solve distance problems. Those following Model 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 Model IV would solve all problems of all types.

To the extent that there is a correlation between age and the level of the model which best represents a child's knowledge, there should be clear developmental patterns for each problem type. The most interesting is the predicted decrement in performance on conflict-weight problems. Children using Models I or II will get these problems right even though they do not see them as conflict problems, whereas children using Model III will attend to the conflicting cues of weight and distance, but they will have to muddle through, and their resulting predictions will be at a chance level of performance. Another prediction, shown in Table I, is that performance on distance problems should improve dramatically with age. The youngest subjects, using Model I will err on every problem, while children using Models II, III, or IV will never err. By a similar logic each of the problem types yields a predicted developmental course, the results of which are shown in Table I. (See Siegler, 1976, for a complete analysis.)

TABLE I
Predictions for Percentage of Correct Answers and Error Patterns on Posttest for Children Using Different Models

	Models				
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

V. Experiment 1: Assessing Initial Knowledge

The purpose of Experiment 1 was to assess the validity of the foregoing analysis for a group of children spanning a wide age range.

A. METHOD

Subjects were 120 female students from a private school in Pittsburgh. Fifteen students from each of eight grade levels were grouped as shown at the top of Table II.

Materials included a wooden balance scale, 10 different colored metal weights, and two wood blocks. The balance scale's arm was 80 cm long, with four pegs on each side of the fulcrum. The first peg on each side was 7.6 cm from the fulcrum and each subsequent peg was 7.6 cm from the peg before it. The arm could swing freely from the point of attachment to the fulcrum, 10 cm above the fulcrum's base. Each metal weight weighed 40 gm, measured 2.5 cm 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. The two blocks of wood, each 11.4 cm 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.

Children's knowledge was assessed through a 30-item test. 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

TABLE II
Developmental Trends Observed and Predicted on Different Problem Types in Experiment I[a]

Number of each type	Grade / Age (years) / Mean age (months)	K-1st / 5-6 / 73	4th-5th / 9-10 / 120	8th-9th / 13-14 / 169	11th-12th / 16-17 / 207	Predicted developmental trend (from Table I)
	Problem type					
4	Balance	94	99	99	100	No change—All children at high level
4	Weight	88	98	98	98	No change—All children at high level
4	Distance	9	78	81	95	Dramatic improvement with age
6	Conflict-weight	86	74	53	51	Decline with age—Possible upturn for oldest
6	Conflict-distance	11	32	48	50	Improvement with age
6	Conflict-balance	7	17	26	40	Improvement with age
	Weighted mean %	46	61	62	67	

[a]Percentage of problems predicted correctly

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 I; they were presented in the same random order for each child.

Children were tested individually in a quiet room in their school. The experimenter's initial instructions were:

> 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 the children were encouraged to hold the weights to see that they weighed the same amount and to observe the equal distances between adjacent pegs.

Children's knowledge was then assessed by presenting them with the 30 problems described above. The problems were introduced with the following instructions:

> 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 they would both stay like they are now if I took the wood 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 this test, children were asked to explain their responses.[1] Children spent between 15 and 30 minutes on the entire task.

B. RESULTS

The percent of correct predictions for each problem type by each age group is shown in Table II. A 4 (age) by 6 (problem type) analysis of variance revealed that both main effects and their interaction were significant ($p < .001$). Note that the developmental patterns are very close to those predicted in Table I. In particular there is a dramatic improvement in distance problems and a decrement in conflict-weight problems. The conflict-weight problems never did show an upturn, although performance appears to have leveled off for the older age groups.

Not apparent in Table II is the substantial consistency that existed in performance on items within each problem type. Only on conflict-weight problems did accurate prediction decrease with age, and within this category such decrements

[1]See Siegler (1976) for the criteria used to classify explanations.

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 the developmental trend minimal.

With one class of exceptions, the four models make exact predictions about which of the three possible responses (left–down, right–down, balance) the subject will make on each one of the 30 problems. (The exception class contains the 18 conflict problems for Model III; here the prediction is a lack of consistency, i.e., essentially chance responding.) Thus, we can compare the response pattern of each child to the predicted patterns for each of the models, and classify the child according to which, if any, model she was using in making her predictions. Using very strict criteria that had vanishingly small probabilities of misclassifying a random responder, it was possible to classify 107 of the 120 children. The results are shown in Table III.

Children's explanations were also used to determine which model the child was using. The criteria for classifying according to explanations were derived from a literal interpretation of the models. Altogether, 117 of the 120 children's explanations fit one of the four models. As shown in Table IV, the two classifications—one derived from children's predictions, the other from their explanations—were highly correlated ($r = .89, p < .001$). All of the 23 children judged to be using Model I by the predictions data were judged as using Model I by the explanations criterion, and all eight of the children classified as using Model IV on the predictions measure—and only those eight—were classified as using Model IV on the explanations measure as well. On the other hand, many children were classified as using Model II by the predictions measure who were placed in Model III by the explanations measure.

One interpretation of this discrepancy between the explanations and predictions criteria is that there were some children who used Model III tests, but consistently resolved the conflict by relying on the weight cue. Further evidence that children knew more about the balance scale than is revealed by the predictions classification comes from an analysis of the content of their explanations.

TABLE III
Percentage of Children in Each Age Range Fitting Each Model

Age (years)		Model I	Model II	Model III	Model IV	Unclassified
5–6	(n = 30)	77	0	0	0	23
9–10	(n = 30)	10	30	40	6.7	13.3
13–14	(n = 30)	10	23.3	56.7	3.3	6.7
16–17	(n = 30)	0	20	63.3	16.7	0
Total	(n = 120)	24.2	18.3	40	6.7	10.8

TABLE IV
Number of Children in Experiment I Fitting Each Model—Predictions and Explanations Criteria

Model Used: Classified by Explanations Criterion	Model Used: Classified by Predictions Criterion I	II	III	IV
I	23	1	0	0
II	0	7	1	0
III	0	13	46	0
IV	0	0	0	8

Fully one-third of the children advancing Model III explanations cited the ratio properly of conflict-balance problems (e.g., one on the third peg equals three on the first peg), but not the composition rule necessary for a Model IV placement.

C. EVALUATION OF DECISION TREE REPRESENTATIONS

How well do the decision trees used in Fig. 1 represent children's knowledge on the balance scale task? The representation appears to fare well on the first, second, and fourth criteria listed earlier. With regard to the first criterion, the models are clearly sufficient to account for the predictions data: The problem type by model analysis provided an exhaustive and unambiguous mapping between behavior and theory. With regard to the second criterion, the data can be analyzed at the level of either individual subjects or group averages, and also can be considered at either the level of response patterns or sheer number of correct responses. Regarding the fourth criterion, the formal relationship between the models is one of strict inclusion: Model I tests are included in Model II, etc. This logical structure predicts an invariant developmental sequence (although we could not test this directly in a cross-sectional study).

In terms of the third criterion—integration of psychological parameters—the merits of the representation are less clear. Although a number of implications could be drawn, few explicit statements about psychological, as opposed to logical, properties of the knowledge required to do the task have yet been advanced.

Another orientation from which to evaluate the representation is to ask how well it answers the five questions about instruction and development that were listed above. Thus far, it has answered only Question 1: The difference between high and low performers is represented by differences among the four models. The models are silent on Question 2, which addresses the issue of alternative paths to the same performance. This inadequacy was most noticeable in the discussion of the idiosyncracies that are masked by the "muddle through" category on Model III. Since the models do not have any representation for their own induction, they are unable to say anything about Question 4, which asks about critical features of the instructional sequence. However, the models do suggest some straightforward ways to empirically investigate Question 3 (How difficult should an instructional sequence be?), and they imply that there should be no differences in responsiveness to instruction, thus providing an assertion that refutes the premise of Question 5 (When and why will differential learning occur?). In Experiment 2, these issues suggested by Questions 3 and 5 were addressed.

VI. Experiment 2: Training on the Balance Scale Task

In the second experiment, 5- and 8-year-olds were equated for performing at a level not beyond Model I. Then they were provided with experience on either distance or conflict problems, or with 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 a child's existing knowledge and the new information presented is a critical determinant of when, how much, and what kind of learning will occur (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) in class-inclusion training. Therefore, we predicted that our Model I children would benefit from distance problems, while they would learn little, if anything, from conflict problems.

As we already noted, there is nothing in the models that would predict differential responsivity to instruction of older and younger children. But both intuition and empirical evidence support the notion that older children are more adept than younger ones at mastering many novel problems on which task-specific knowledge is equally lacking (cf. Siegler, 1975; Siegler & Liebert, 1974, 1975). Thus, there were no clear grounds on which to base a prediction about age differences in response to the training sequences.

A. METHOD

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

1. Pretest

The pretest consisted of eight items: two weight, two distance, two conflict-weight, and two conflict-distance. The tasks and apparatus were similar to those used in Experiment 1; on each trial, the child was shown a configuration and

asked to predict which of the three possible outcomes would occur if the wood blocks were removed. There was no feedback during the pretest.

2. Experience

All experiential conditions except the bias control (see below) included 16 trials on which children were presented a randomized sequence of 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-second interval the weights were removed and placed on the scale in a different arrangement.

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, 2 balance, and 2 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 2 balance items that would familiarize them with the balance scale's workings but would not directly engender knowledge of Models II or III. However, this control procedure might itself bias children toward a greater reliance on Model 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 group's control condition, one-half of the children were assigned to the exposure control and one-half to the bias control.

3. Posttest

The posttest included a randomly ordered, no-feedback presentation of 24 items, four each of balance, weight, distance, conflict-weight, conflict-distance, and conflict-balance types. The pretest took approximately 10 minutes, the experience 25 minutes, and the posttest 15 minutes. 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 hours.

4. Participants

Sixty children, 30 5-year-olds and 30 8-year-olds, all with less than Model II knowledge, were randomly assigned within age and sex to the three treatment groups.[2] All groups had equal numbers of males and females except for the

[2]See Siegler (1976) for the details of this selection procedure.

8-year-old control group, which included four boys and six girls. The mean CA of kindergartners was 70 months (range = 66–75 months), while the mean CA of third graders was 106 months (range = 101–117 months). The experimenter, a 22-year-old female research assistant, served for all children.

B. RESULTS

Responses to the 24-item posttest were classified according to a scheme similar to the one used in Experiment 1. (There were no differences between the two control groups, so the data from both of them were combined.) As shown in Table V, 45 of the 60 children behaved according to the models: 21 using Model I, 17 using Model II, and 7 using Model III. A Chi-Square test indicated that significant differences were present in the type of rules used by children in the six age-by-experience groups ($\chi^2 = 45.54$, $df = 1$, $p < .001$). More specific analyses revealed that 5-year-olds more often used Model I and 8-year-olds more often Models II or III ($\chi^2 = 12.91$, $df = 1$, $p < .001$), and that children exposed to the control procedure more often used Model I, while those exposed to conflict or to distance problems more often used Model II or III ($\chi^2 = 13.20$, $df = 1$, $p < .001$).

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

TABLE V
Number of Children Using Different Models—Experiment 2

Age group	Model I	Model II	Model III	Unclassifiable
5-Year-olds				
Control	8	0	0	2
Distance training	3	4	1	2
Conflict training	5	0	0	5
Total	16	4	1	9
8-Year-olds				
Control	5	3	0	2
Distance training	0	8	1	1
Conflict training	0	2	5	3
Total	5	13	6	6
Grand total	21	17	7	15

Models II and III than did the control procedures ($p < .001$), and conflict problem experience led to greater use of Model III than did the distance problems and control conditions ($p < .01$).

In summary, then, Table V shows that both age groups can learn from training that is only one level beyond their current level (i.e., distance training). However, given training that is two levels beyond (i.e., conflict training), the 5-year-old children learned nothing, while the 8-year-olds benefited substantially. Thus, it is clear that older and younger children derived different lessons from the same experience, even when they had identical initial predictive knowledge about the task.

VII. Revised Representations for Balance Scale Knowledge

These empirical results raise questions that reveal some of the limitations of the decision tree representation used thus far to represent children's knowledge of balance scale tasks. Since the four models purport to represent all of what a child knows about the task, they predict that children classified according to one of the models should be identical on all task-related performance, including learning about the task. Thus, they predict that the differential responsiveness to experience with conflict problems that we observed between the two age groups should not have occurred. Of course, the models make this prediction by default, since they have no representation of the learning process as such. That is, they contain no representation of the way that positive and negative information obtained during the training sequence is treated, nor about the ways in which the models might undergo transformation from one level to the next.

Another limitation of the representation is that it allows no way to describe the many different means utilized by subjects to arrive at the same end. We have already alluded to this in our discussion of the Model III explanations data, and now it is time to address it directly.

We need a representation that can account for not only the logical form of the decision rules used to make predictions, but also the psychological properties of the rules. That is, we need a representation that enables us to clearly indicate the perceptual and mnemonic demands of actually using the decision rules. In this section we will introduce such a representation for children's knowledge about this task, and we will present examples of the kinds of questions the representation enables us to ask. Then, in the next section, we will describe an experiment that provides some answers to these questions.

A. PRODUCTION SYSTEM REPRESENTATION

In Fig. 2 we have restated the four models of Fig. 1 as production systems. [See Newell (1973) for an extensive introduction, and Klahr (1976b) for some

```
Model I
        P1:((Same W) --> (Say "balance"))
        P2:((Side X more W) --> (Say "X down"))

Model II
        P1:((Same W) --> (Say "balance"))
        P2:((Side X more W) --> (Say "X down"))
        P3:((Same W) (Side X more D) --> (Say "X down"))

Model III
        P1:((Same W) --> (Say "balance"))
        P2:((Side X more W) --> (Say "X down"))
        P3:((Same W) (Side X more D) --> (Say "X down"))
        P4:((Side X more W) (Side X less D) --> muddle through)
        P5:((Side X more W) (Side X more D) --> (Say "X down"))

Model IV
        P1:((Same W) --> (Say "balance"))
        P2:((Side X more W) --> (Say "X down"))
        P3:((Same W) (Side X more D) --> (Say "X down"))
        P4':((Side X more W) (Side X less D) --> (get Torques))
        P5:((Side X more W) (Side X more D) --> (Say "X down"))
        P6:((Same Torque) --> (Say "balance"))
        P7:((Side X more Torque) --> (say "X down"))
```

	Transitional requirements	
	Productions	Operators
I -> II	add P3	add distance encoding and comparison
II -> III	add P4, P5	
III -> IV	modify P4; add P6, P7	add torque computation and comparison

Fig. 2. Production system (P) representations for Models I–IV. D = distance; W = weight. Written in a special language called PSG. See text for further explanation.

examples from cognitive development.] A production system consists of a set of rules—called productions—written in the form of condition—action pairs; the conditions are symbolic expressions for elements of knowledge that might be present at some instant. A production system operates via a recognize-act cycle. During the recognition cycle, all the condition sides of all the productions are compared with the current contents of the immediate knowledge state. We will refer to this immediate knowledge as the contents of working memory (WM). It can be interpreted as primary or short-term memory (Waugh & Norman, 1965), M-space (Pascual-Leone, 1970), short-term plus intermediate-term memory (Bower, 1975; Hunt, 1971), or more generally as the currently activated portion of long-term memory, or simply as the current state of awareness of the system.

The productions whose conditions are matched by elements in WM are placed into the conflict set, a conflict resolution principle is applied, and one production fires. The act cycle executes the actions that are associated with the fired production. Then the next recognition cycle commences.

Thus, the conditions are tests on the momentary state of WM. A sequence of condition elements on the left side of a production is interpreted as a test for the simultaneous existence of the conjunction of the individual knowledge elements. If, for a given production, all the condition elements happen to be true at some instant, we say that the production is "satisfied." If only one production is satisfied, then it "fires": the actions associated with it, written to the right of the arrow (see Fig. 2) are taken. These actions can modify the knowledge state by adding, deleting, or changing existing elements in it, or they can correspond to interactions with the environment—either perceptual or motor.

If more than one production is satisfied at a given moment, then the system needs to invoke some conflict resolution principle. In the systems shown here all conflicts are assumed to be resolved such that special cases have priority over general cases. For example, suppose that the two productions in the conflict set are:

P1: (a b ——→ x)
P2:(b ——→ y)

P1 is a special case of P2, since P2 is satisfied whenever P1 is satisfied, but not vice versa. That is, P2 is satisfied when element b is in WM, but P1 is satisfied only when both b and a are present. The special case conflict resolution principle will choose P1. [Further discussion of conflict resolution in production systems can be found in McDermott and Forgy (in press), Newell (1973), Newell and McDermott (1975), and Rychener (1976).]

Consider, for example, Model II in Fig. 2. It is a production system consisting of three productions. The condition elements in this system are all tests for sameness or difference in weight or distance. The actions all refer to behavioral responses. None of the models in Fig. 2 contain a representation for any finer grain knowledge, such as the actual amount of weight or distance, or the means used to encode that information. Nor is there any explicit representation of how the system actually produces the final verbal output. It is simply assumed that the system has access to encoded representations of the relational information stated in the conditions. We will return below to further consideration of the way that this information becomes available to the system. Returning to Model II, notice that on any recognize cycle, only one production will fire. If the weights are unequal, then P2 will fire; if the weights are equal and the distances are not, then both P1 and P3 will be satisfied, but since P3 is a special case of P1, the conflict resolution principle will choose P3 to fire; finally, if both weights and distances are equal, then only P1 will be satisfied and it will fire. (The numbers attached to the productions [e.g., P1, P2, etc.] are not supposed to have any psychological meaning. They serve simply as labels for the reader; note that a production maintains its label across the four models.)

We can compare the four models to determine the task facing a transition model. At the level of productions the requisite modifications are straightforward: a transition from Model I to Model II requires the addition of P3; from Models II to III, the addition of P4 and P5; and from Models III to IV, the addition of P6 and P7 and the modification of P4 to P4′. (This modification changes the action side from random muddling through to "get torques".)

We can compare the four models at a finer level of analysis by looking at the implicit requirements for encoding and comparing the important qualities in the environment. Model I tests for sameness or difference in weight. Thus, it requires an encoding process that either directly encodes relative weight, or encodes an absolute amount of each and then inputs those representations into a comparison process. Whatever the form of the comparison process, it must be able to produce not only a same-or-different symbol, but if there is a difference, it must be able to keep track of which side is greater. Model II requires the additional capacity to make these decisions about distance as well as weight. This might constitute a completely separate encoding and comparison system for distance representations, or it might be the same system except for the interface with the environment.

Model III needs no additional operators at this level. Thus, it differs from Model II only in the way it utilizes information that is already accessible to Model II. Model IV requires a much more powerful set of quantitative operators than any of the preceding models. In order to determine relative torque, it must first determine the absolute torque on each side of the scale, and this in turn requires exact numerical representation of weight and distance. In addition, the torque computation would require access to the necessary arithmetic production systems to actually do the sum of products calculations.

Although we have compared the four models at two distinct levels—productions and operators—the levels are not really that easily separated. Missing from these models is a set of productions which would indicate the interdependence: productions that explicitly determine which encoding the system will make. That is, in these models, there are almost no productions of the form: (want to compare weights) ⟶ (attend to stimulus and notice weight). The sole exception to this occurs in P4′ in Model IV. When this model is confronted with a nonconflict problem, either P1, P2, P3, or P5 will fire on the first recognize cycle. However, if it is a conflict problem, then P4′ fires, and the system attempts to "get torques." The result of this unmodeled action, as described above, would be to produce a knowledge element that could satisfy either P6 or P7 on the next recognize cycle.

Each of the four production system models in Fig. 2 makes precisely the same prediction as its counterpart in the decision tree representation of Fig. 1. Thus, on the first of the evaluative criteria listed above—accounting for behavior—the production system model fares as well as the decision tree model. With respect to the second criterion—multiple-level analysis—and the fourth—developmental tractability—the production systems are somewhat more explicit than the decision trees about the requirements for both the encoding operations and the rules (i.e., productions) that utilize the symbolic elements produced by the operators. They also clarify the developmental differences between models in terms of these two kinds of entities.

The major advantage of the production system representation lies in its integration of general psychological principles—the third of our evaluative criteria. Production systems of the type used here incorporate a theory of the control structure and general representation that underlies a broad range of human problem-solving ability (Newell & Simon, 1972). As Newell (1973) put it: "The production system itself has become the carrier of the basic psychological assumptions—the system architecture of . . . [the production system] is taken to be the system architecture of the human information processing system" (p. 516). Thus, models written in this form can be viewed as variants within a general psychological theory, and to the extent that such a general theory is consistent with the empirical results from experimental psychology, then these models are also consistent with them.

With respect to the five questions listed earlier, the production systems have enabled us to be very explicit about Question 1 (differences that underlie performance), and in particular about the important role of encoding operators. They have indicated some potential sources of variation for each level of performance (Question 2), although since they are written as modal types, this is merely suggestive at this point. Similar comparisons of the relative efficacy of the two forms of representations for answering the other three questions yield the same result. Thus, while the new representation does not provide much of an advantage over the old for understanding the results of Experiment 2, it does provide some guidance about where to look for an explanation—in the encoding of the stimulus.

In order to model the conditions under which one or another aspect of the stimulus is attended to and encoded, we would need to augment the models in Fig. 2 with productions like P4′. These productions would transform the models from simple discrimination nets into active problem solvers, and they would enable us to make predictions about such things as eye movements and solution latencies for different classes of problems. However, before we can make such an extension, we must first determine the varieties of possible encoding schemes that subjects are actually using. As a first step in that direction, we undertook a detailed examination of the problem-by-problem protocols of a few children in a training sequence.

C. PROTOCOL ANALYSIS

Several children, ranging in age from 5 to 10 years, were unsystematically selected to be run individually in a conflict training sequence. They were given instructions about the balance scale and about the fact that there were rules underlying the balance scale's behavior that they could discover if they "watched carefully and thought about it." In addition, following their prediction on each trial, they were asked to state their reasons for the prediction. Then the blocks were removed, the children observed the scale's movement and if they were incorrect, they were again asked, "Why do you think that happened?"

These entire sessions were videotaped, and then all the verbal comments, as well as major physical activities, were transcribed into the form shown in Appendix A. At the beginning of each problem, there is an indication of the problem number, the configuration, and the elapsed time (in minutes and seconds) since the start of the session. Problem numbers T1, T3, etc., correspond to items from the training sequence, and problem numbers E7, E8, etc. (see lines 11300 and 14700) are from an exploratory session which followed the training sequence. In the exploratory session, the children were encouraged to build interesting problems or to explain to the experimenter what kinds of problems would achieve certain outcomes. The problem configuration is indicated by a numerical code that is a near-pictorial representation of the problem. In T1 (line 00400) the code 0001/2000 indicates one weight on the first peg (from the fulcrum) on the left side, and two weights on the first peg on the right side. In T3 (line 02200), the code 0100/1000 indicates a single weight on the third peg on the left, and a single weight on the first peg on the right.

Excerpts from the protocol of Lisa, a 5-year-old female, are shown in Appendix A. The protocol provides a rich data source from which to select "observations." However, in this discussion we will focus only on those aspects that indicate the kind of encoding of distance and weight that Lisa appears to use.

Lisa was first given the standard instructions and pretest described earlier. Her response pattern did not conform to any of the four models. However, if Model I were modified such that heavy things went up instead of down, then she was a perfect Model I subject. The first problem in the training sequence confirms this interpretation (lines 00400–01200). Lisa knows which side has more weight, but her prediction is based upon the assumption that more weight goes up. However, when confronted with the contrary evidence, she changes the "sign" of the correlation between weight and direction of tipping. This single-feedback trial

was sufficient: For the remainder of this half hour session, she never again errs in her understanding of the direction of the effect of weight differences. As we will see, the correct encoding of distance and its effect required a much longer series of trials.

The second training problem (not shown) was a balance problem, so T3 (0100/1000) was the first instance in which Lisa received feedback indicating that equality of weight is not a reliable predictor. Her own verbalization of the problem captures her puzzlement: "Well why are they both the same thing [same weight] and one's up and one's down?" (line 3300).

Another distance problem followed immediately (T4: 0020/0020), and Lisa's first response is to say balance, but she quickly corrects herself, having detected the distance difference. Her encoding of distance is correct in that it is based on the fulcrum, rather than the end points, as the zero reference point (lines 04400–04500). However, she incorrectly associates greater distance with the side that goes up rather than the opposite, in the same way that she initially had the sign wrong for weight effects. This is her first attempt to utilize distance information, and she gets negative feedback. At this point she might abandon distance as a useful cue, or she might—as she did with weight—simply change the sign of the relation. As we will see, she does neither.

T5 was a complex distance problem (0101/1100) (not shown), and T6 (0102/2010) a balance problem, neither one of which yielded a useful protocol. In T7 (0200/2000) we return to a distance problem. It is clear from the protocol that Lisa is still attempting to use distance (lines 09600–10000). She still encodes direction of distance from the fulcrum correctly, but she has not changed her erroneous assumption about the effect of this difference. Note also that she has not yet made any statement about absolute amount of distance; all her statements are about relative distance.

In order to focus on the issue of distance encoding, we skip over about 15 minutes of conflict training in which the problems were mainly complex conflict-weight and conflict-distance (i.e., two or more pegs occupied on each side) from which no clear pattern emerged. We pick up the protocol again in an excerpt from the exploratory phase in which Lisa was allowed to construct problems according to various experimenter requests or hints. In E7 (lines 11300–14400), she has been asked to construct some problems such that she will not be quite sure what the result will be. In general, Lisa does no such thing, and instead tends to construct problems about which she is very confident. Thus, her initial configuration is 0003/0004, a problem in which both weight and distance indicate that the right side will go down. Then the experimenter modifies it to a distance problem (0004/0004), and Lisa apparently forgets all about distance differences, reverting to a Model I prediction of "balance" (lines 12000–12300). With a little prompting from the experimenter (lines 13500–14000), she invokes a (post hoc) distance explanation (lines 14100–14200). Notice that the distance description is not just a relative judgment, but instead is stated in terms of two absolute (albeit approximate) quantities.

It appears that, even after almost 30 minutes of experience with the balance scale, Lisa knows that distance is an important factor, but she has not yet developed a reliable rule about the effect of distance differences. Then, over the next 2-minute period, she begins to demonstrate a stabilizing grasp of this concept. First she creates a balance problem and makes the correct prediction (lines 14900–15900). Then a new experimenter enters, and feigning ignorance, asks how the scale works. Lisa creates (0003/0003) and predicts correctly, and for the right reasons (lines 16800–17300). Then, at the experimenter's request, she correctly creates a balance problem (0003/3000). It is interesting that she does this in the "easiest" way, given the configuration from which she was starting, but it is also the case that this is the same balance configuration that was used in the preceding problem. Then she creates a distance problem such that the scale tips in a desired direction (19000–19400) and gives the correct explanation and, finally, she initiates yet another balance problem, one unlike any she has ever seen before (3000/0003).

Recall that this protocol analysis was undertaken after a discussion of the production system representation of knowledge about the balance scale (Fig. 2). In that representation, we tried to emphasize the differences between the encoding of information about the environment (the undefined operators) and the combination rules [cf. Gelman's (1972a, 1972b) operator–estimator distinction, and Klahr and Wallace's (1973) operator–rule dichotomies] for acting on that information (the productions). The protocols tell us something about the nature of the representations that are being used by the child, and hence something about the encoding operators that produce them. It is clear that Lisa extracts information from the training series that will enable her to improve both the encoding operators and the combination rules. With respect to weight, she has no difficulty in formulating an appropriate encoding based on counting the number of weights. Although there is an initial error with respect to the relation between weight differences and the direction of the scale, this is quickly corrected and remains stable for the rest of the session.

Distance encoding follows quite a different course. Initially it is ignored. Then differences in distance are noted, but their effect is quite unstable in the face of negative feedback, and as we saw, they are occasionally ignored well into the training sequence. However, it appears that by the very end of the exploratory trials, an appropriate encoding of distance, and a concomitantly appropriate rule for utilizing it (at least on distance problems), has been formulated.

Learning about the balance scale, then, would seem to require much more than is suggested by a comparison of adjacent models in Fig. 1. The production system representation of Fig. 2 has enabled us to make explicit the difference between encoding operators and decision rules, and it has guided our search for

instances of both of these kinds of learning in the protocol. The analysis suggested that there is a point in the development of knowledge about this task during which the dimensions may be encoded in idiosyncratically incorrect ways, and that the form of the encoding may depend upon trial-to-trial feedback. In the next section we will introduce a model that attempts to capture these phenomena for an individual subject.

D. REVISED PRODUCTION SYSTEM FOR A MODEL III CHILD

Thus far, the production system representation has been used only to suggest some of the complexities of learning about the task. In this section we will work toward the creation of a production system model of a single child's behavior during a training sequence. The representation will be more than suggestive, for it will be specific enough to run as a computer simulation. The simulation will serve two purposes. First, it will demonstrate the sufficiency of the model to account for the data it purports to explain. Second, the particular simulation language in which the model is stated is based upon, and incorporates in its structure, very specific assumptions about the nature of the human information-processing system. Thus, the model to be described here is a particular instance of a much broader theory of human problem solving.

Our subject, Jan, was a female second-grader, age 7 years, 11 months. Her performance on an 8-item pretest and a 16-item training series is shown in Table VI. In Table VI, each row corresponds to a problem. The columns indicate, respectively, problem number, problem configuration, problem type (Distance (D), Balance (B), Conflict-Weight (CW), Weight (W), etc.), Jan's response (Left (L)- or Right (R)-down, Balance (B)), feedback from the scale (if the subject's prediction was inconsistent with what the scale did, it is indicated by a −), predictions from three of the previously described models (Models IV, II, and I), and finally, two columns corresponding to the model to be described in this section. The first of these columns—IIIA—contains the model's prediction, and the second contains the value of a variable criterion that is used to make the prediction. For example, Problem 7 has three weights on the first peg on the left and two weights on the third peg on the right; it is a conflict-distance problem. Jan predicted that the left side would go down, but as Model IV (which is always correct) predicted, the right side went down so the subject got negative feedback. The other three models shown here (Models II, I, and IIIA) all make the same prediction as the subject: left–down. The number at the bottom of each of the four model columns shows the number of mismatches between Jan's predictions and the model's.

Jan's responses to the pretest make her a perfect Model II subject. Her responses during the training sequence provide a poor fit to Models I, II, and IV. Recall that the criterion for fitting Model III was that the responses be essentially

TABLE VI
Jan on Training Sequence, and Predictions from Four Models

Problem				Prediction					
Number	Configuration	Type	S2	Feedback	Model IV	Model II	Model I	Model IIIA	Criterion
Pretest									
1	100¦100	D	L			L			
2	010¦300	CW	R			R			
3	100¦200	CD	R			R			
4	010¦020	W	R			R			
5	020¦002	D	R			R			
6	200¦400	CD	R			R			
7	100¦200	CD	R			R			
8	030¦020	W	L			L			
Training series									
1	0200¦0200	D	L	+	L	L	B	L	W
2	0020¦0200	B	B	+	B	B	B	B	
3	0020¦3000	CD	R	−	L	R	R	R	
4	0003¦0100	CW	L	+	L	L	L	L	D
5	0200¦0400	CW	L	−	R	R	R	L	
6	0102¦2010	B	B	+	B	B	B	B	W
7	0003¦0020	CD	L	−	R	L	L	L	
8	0100¦0200	CW	L	−	R	R	R	L	D
9	0040¦1020	CW	L	+	L	L	L	L	W
11[a]	0001¦2000	W	R	+	R	R	R	R	
12	0013¦1020	CD	L	−	R	L	L	L	
13	0120¦2200	CD	L	+	L	R	R	L	D
14	0200¦1300	CW	L	−	R	R	R	L	
15	0002¦0010	CD	R	+	R	L	L	R	W
16	0023¦1110	CW	R	−	L	L	L	L	
					7[b]	6[b]	7[b]	1[b]	

[a]Problem 10 was omitted.
Abbreviations: B = balance; C = conflict; D = distance; W = weight; R = right; L = left.
[b]Number of mismatches between Jan and model.

random for conflict problems. Thus, although the "muddle through" prediction of Model III does not make an exact prediction on any trial, it predicts the absence of a consistent pattern over the set of conflict problems. And indeed, this is what we find in Table VI: On 5 of the 11 conflict problems Jan responds as if she were relying on the weight cue, and on the other six she conforms to the distance cue. Thus, we could simply classify Jan as a Model III subject and leave it at that.

Such an interpretation has several deficiencies. First, the classification scheme itself is unsatisfactory when compared with the others. Model III subjects get so classified as a residual category, by the absence of any pattern in their responses to conflict problems, while all other classification is based on the occurrence of things that were predicted to happen, rather than the absence of things that should not. In addition to this "taxonomic" weakness, Model III's "muddle through" prediction tells us nothing about the psychological processes that actually operate when subjects detect conflict but do not yet know how to deal with it correctly. We have already cited some of the idiosyncratic strategies that different subjects bring to bear on this situation. Finally, it is important to emphasize that Table VI represents responses during a training sequence, a situation in which the child was presumably attempting to integrate the feedback from the balance scale's actual behavior with her current hypothesis about how it worked. None of the four models described thus far have any mechanism to represent and utilize such information. Thus, the model to be described represents our first steps toward remedying these deficiencies.

Jan was run under the same conditions as Lisa, and an analysis of her trial-by-trial explanations provided the initial evidence for the model that we eventually formulated. The most striking feature of her comments was the way she appeared to represent distance and weight on conflict problems. Both of them were treated as dichotomous: More than two weights was treated as "big," otherwise weight was "little," and if the third or fourth peg were occupied, then distance was "big," otherwise it was "little." Rather than present another lengthy protocol analysis here, we will show just two examples of this dichotomous encoding of distance.

On Problem 12 (0013/1020), the child predicts left–down; upon seeing the result, she says:

> Oh, now I think I know why. . . . I think I know because. . . it's supposed to be a rule that they usually go down more if they're on that side [pointing to the extreme right of the balance scale]. So that one went down cause it's two there [pointing far right] and none there [pointing far left].

If we encode each arm of the balance scale into a near segment (Pegs 1 and 2) and a far segment (Pegs 3 and 4), then this protocol is easily interpreted. "They usually go down more if they're on that side" means that if the far segment is occupied ("big distance") then the scale will tip in that direction. "Two there and none there" means that the far segment on the right is occupied by two weights, whereas the far segment on the left is unoccupied.

The second example comes from Problem 14 (0200/1300), just before the child gets feedback. She says:

> This side's gonna go down [pointing left] . . . Even though this one has four [pointing right] and this one only has two [pointing left]. . . . Even though this one has [pointing right] twice as much as this [pointing left], that means that because this one's more [waves to far left] over, and that's [pointing right] all on that side.

In this case, we garner support for the dichotomous distance encoding from the comment that the weights on the right arm of the scale are "all on that side." "That side" of what? By our interpretation, they are on "that side" of the midpoint of the right arm, thus making distance "little," rather than "big" on the right.

In order to determine whether this interpretation of the protocols is valid, we need to construct a model that is consistent with Jan's actual predictions on each trial, as well as her explanations. Based upon many such comments and our interpretations of them, we constructed the model whose predictions are shown in Table VI. In order to provide a clear overview of the model we will describe it first in terms of a binary decision tree, plus a few ad hoc mechanisms. Then we will present a running production system for a more complete model based on the same underlying logic.

Figure 3 shows the binary decision tree representation for Model IIIA; Jan's performance on the training sequence is shown in Table VI. The numbers under the terminal nodes correspond to the problems from Table VI that are sorted to those nodes. The first three tests are the same as those in Model III (Fig. 1), and they account for balance, weight, and distance problems. If neither weight nor distance is "same," then the model begins to test for "big" values. If either

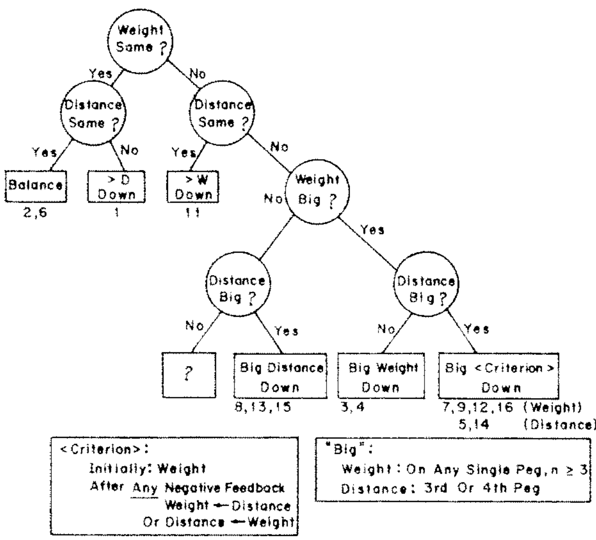


Fig. 3. Decision tree representation for Jan's prediction model. D = distance; W = weight. The numbers under the terminal nodes correspond to the problems from Table VI that are sorted to those nodes.

weight or distance—but not both—is big, then the side with the big value determines the prediction. If both are big, then Model IIIA favors whichever one is currently its criterion value. The criterion value starts as weight, but whenever negative feedback is received the criterion switches from one value to the other. The state of the criterion value is indicated in the last column in Table VI. Note that it changes after any negative feedback, not just on conflict trials with negative feedback. (The terminal node labeled "?" in Fig. 3 is never reached by the set of problems in Table VI. Such a problem would be a conflict problem with neither weight nor distance "big." We have no evidence upon which to base a prediction about what the subject would do with such a problem.)

E. PRODUCTION SYSTEM FOR MODEL IIIA

The production system for Jan is shown in Fig. 4. The representation contains the actual computer listing (with a few inessential details not shown) for the production system, which is written in a special language called PSG (Newell & McDermott, 1975). Appendix B contains a trace of this model running on a sequence of four problems from Table VI; one of them—Problem 7—is also shown in Fig. 5. Before we embark on a detailed description of the model, we will make a few comments about the properties of this rather complex representation of knowledge about the balance scale task.

This model represents, in addition to the child's knowledge about how the balance scale operates, her knowledge about the immediate experimental context in which she is functioning. The trial-by-trial cycle during the training phase comprises (1) observation of the static display, (2) prediction of the outcome, (3) observation of the outcome, (4) comparison of the outcome with the prediction, and (5) revision if necessary, of the criterion. The production systems shown previously (Fig. 2) represented knowledge sufficient to execute only the second of these five steps, while the present model (Fig. 4) explicitly represents all of this task-relevant knowledge in a homogeneous and integrated manner. This model utilizes, in one way or another, representations of knowledge about when and how to encode the environment, which side has *more* weight or distance, which side has a *big* weight or distance, what the current criterion value is, what the scale is expected to do, what the scale actually did, whether the prediction is yet to be made or has been made, and whether it is correct or incorrect.

F. PRODUCTION SYSTEM INTERPRETATION

Some general properties of production systems were described earlier. In this section we will add a few more details about how the model in Fig. 4 operates. Recall that the basic cycle for a production system is recognize–act. During a recognition cycle, all the productions compare their condition elements with an

```
<dimension.1>:(CLASS weight distance)    <dimension.2>:(CLASS weight distance)

<side.1>:(CLASS left right both)          <side.2>:(CLASS left right both)

<direction>:(CLASS up down level)


P1:((predict) (weight same) --> (made **)  (expect both level) say.b)

P2:((predict) (weight more <side.1>) --> (made **)  (expect <side.1> down) say.d)

P3:((predict) (weight same) (distance more <side.1>) --> (made **)  (expect <side.1> down) say.d)

P4:((predict) (weight more)(distance more) -->   find.big)

P5:((predict) (criterion <dimension.1>)(<dimension.1> big <side.1>)
          (<dimension.2> big <side.2>) --> (made **) (expect <side.1> down)say.d)

P6:((predict) (weight big <side.1>) --> (made **)  (expect <side.1> down) say.d)

P7:((predict) (distance big <side.1>) --> (made **)  (expect <side.1> down) say.d)

P8:((predict)(<dimension.1>) abs --> ATTEND)


E1:((expect) --> look)

E2:((expect <side.1> <direction>)(see <side.1> <direction>) --> (did **)(see ===> saw)(result correct))

E3:((expect <side.1> <direction>)(see <side.1> <direction>) abs (see) --> (did **)(see ===> saw)
                                                                      (result wrong))


SW1:((result wrong)(criterion distance) --> (old **)(distance ===> weight))

SW2:((result wrong)(criterion weight) --> (old **)(weight ===> distance))

SW3:((result correct)(criterion) --> (old **))


find.big:(OPR CALL) ;returns (weight|distance big left|right), one or two such.

look:(OPR CALL) ; looks for result of balance tipping;
                  returns (see left|right down)

attend:(OPR CALL) ; initial encoding of same or difference on distance & weight;
                    returns (weight|distance same|more left|right)
```

Fig. 4. Production system (P) for Jan. ABS = Absent. Written in a special language called PSG. See text for further explanation.

ordered list of elements in WM. The trace in Fig. 5 shows the state of WM after each cycle. For example, at the beginning of the second cycle in Fig. 5, we see that WM has four elements in it: (DISTANCE MORE RIGHT), (WEIGHT MORE LEFT), (PREDICT), and (CRITERION WEIGHT). An examination of the productions in Fig. 4 reveals that P1 is the only production whose condition elements are completely matched by working memory elements, so in this case, it fires. We can interpret


```
        (0003|0020)
Cycle 1
  WM: ((PREDICT) (CRITERION WEIGHT))

  Fire P8: ((PREDICT) (<DIMENSION.1>) ABS --> ATTEND)
        Output from ATTEND (input to WM) ::> (weight more left)(distance more right)

Cycle 2
  WM: ((DISTANCE MORE RIGHT) (WEIGHT MORE LEFT) (PREDICT) (CRITERION WEIGHT) )
     CONFLICT.SET: (P2 P4)
  Fire P4: ((PREDICT) (WEIGHT MORE) (DISTANCE MORE) --> FIND.BIG)
        Output from FIND.BIG (input to WM) ::> (distance big right)(weight big left)

Cycle 3
  WM: ((WEIGHT BIG LEFT) (DISTANCE BIG RIGHT) (PREDICT) (WEIGHT MORE LEFT)
        (DISTANCE MORE RIGHT) (CRITERION WEIGHT))
     CONFLICT.SET: (P2 P4 P5 P6 P7)
     CONFLICT.SET: (P4 P5) AFTER SPECIAL.CASE.ORDER
     CONFLICT.SET: (P5) AFTER WM.ORDER
  Fire P5: ((PREDICT) (CRITERION <DIMENSION.1>)(<DIMENSION.1> BIG <SIDE.1>)
          (<DIMENSION.2> BIG <SIDE.2.) --> (MADE **) (EXPECT <SIDE.1> DOWN) SAY.D)

********** LEFT down

Cycle 4
  WM: ((EXPECT LEFT DOWN) (MADE (PREDICT)) (CRITERION WEIGHT) (WEIGHT BIG LEFT)
        (DISTANCE BIG RIGHT) (WEIGHT MORE LEFT) (DISTANCE MORE RIGHT))
  Fire E1: ((EXPECT) --> LOOK)
        Output from LOOK (input to WM) ::> (see right down)

Cycle 5
  WM: ((SEE RIGHT DOWN) (EXPECT LEFT DOWN) (MADE (PREDICT)) (CRITERION WEIGHT)
       (WEIGHT BIG LEFT) (DISTANCE BIG RIGHT) (WEIGHT MORE LEFT) (DISTANCE MORE RIGHT))
     CONFLICT.SET: (E1,E3)
  Fire E3: ((EXPECT <SIDE.1> <DIRECTION>)
              (SEE <SIDE.1> <DIRECTION>) ABS (SEE) --> (DID **) (SEE ===> SAW)
(RESULT WRONG))


Cycle 6
  WM: ((RESULT WRONG) (DID (EXPECT LEFT DOWN)) (SAW RIGHT DOWN) (MADE (PREDICT))
        (CRITERION WEIGHT) (WEIGHT BIG LEFT) (DISTANCE BIG RIGHT) (WEIGHT MORE LEFT)
         (DISTANCE MORE RIGHT))
  Fire SW2: ((RESULT WRONG) (CRITERION WEIGHT) --> (OLD **) (WEIGHT ===> DISTANCE))

Cycle 7
  WM: ((OLD (RESULT WRONG)) (CRITERION DISTANCE) (DID (EXPECT LEFT DOWN))
       (SAW RIGHT DOWN) (MADE (PREDICT)) (WEIGHT BIG LEFT) (DISTANCE BIG RIGHT)
        (WEIGHT MORE LEFT) (DISTANCE MORE RIGHT))
```

Fig. 5. Trace of Jan's production system (P) running on a conflict-distance problem. ABS = Absent; WM = working memory. Written in a special language called PSG. See text for further explanation.

a production P:(ABC——→DE) as "If you know A and B and C (i.e., if they are currently in WM, in any order), then do actions D and E."

There are two conflict resolution principles. The first one to be applied, special case order, has already been described. If, after applying special case order, there are still two or more productions in the conflict set, then a second resolution principle, WM order, is applied. This principle chooses the productions with the frontmost element in WM. New information always enters the "front" of WM, pushing all else down a "notch." Furthermore, when a production fires, its evoking elements are moved to the front of WM (automatic rehearsal). Thus, the WM order conflict resolution principle says, in effect, "when in doubt, respond to the most recently important information."[3]

There are several different types of actions:

1. *WM additions*. These simply add new elements to the front of WM. For example, if E3 fired, (result wrong) would be added to the front of WM. Other sources of new information are the encoding operators (described below).

2. *WM modifications*. Elements in WM can be altered directly. The action (A ——→ B) changes symbol A to symbol B in the second element in WM. The action (X**) changes the first element in working memory from A to (X (A)), [e.g., (OLD**) would change (DOG) to (OLD (DOG))].

3. *Output*. These actions are surrogates for action on the external environment. The only ones used here are say.b (say "balance") and say.d (say "left [or right] down").

1. *Description of Model (Fig. 4)*

There are three major functional groups of productions.

a. Pn. These correspond to the major nodes in the decision tree representation. P1–P4 are essentially the same as P1–P4 in Fig. 2; P5, P6, and P7 correspond to the tests for "big" values in Fig. 3. Some of the productions use variables that can be matched by specific values in WM elements. These variables are defined in the first three lines of Fig. 4 in terms of the members of the class on values that the variable can take on. Thus, <dimension.1> and <dimension.2> can take on the values "weight" or "distance"; <side.1> and <side.2> can take on the values of "left," "right," or "both"; and <direction> can take on the value "up," "down," or "level."

b. En. These control the model's viewing of the balance scale after it tips, and compare what it expected to see with what it actually sees.

c. SWn. These change the criterion whenever the system determines (via the E productions) that it has made an incorrect prediction. There are three

[3]In Fig. 5 and Appendix B, special case order is usually adequate to resolve conflicts. All instances in which WM order is also used are explicitly indicated in the trace.

encoding operators. None are modeled, but their conditions of evocation are explicit, as is the form of the encoding they produce.

d. Attend. "Attend" does initial encoding of weight and distance. This operator can detect sameness or difference of weight or distance and can indicate the side on which weight or distance is greater. Thus, it is only an encoding of relative quantity. The model assumes that in the first instance this is all that is encoded.

e. Find.big. "Find.big" encodes big weight or big distance and side on which they occur (if they occur).

f. Look. "Look" encodes direction of tipping of scale.

2. *Dynamics of the Model*

The general procedure is as follows. First weight and distance differences, if any, are encoded. If there is no conflict, then a prediction is made, an expectation is formed, and the scale's actual behavior is observed. If it is inconsistent with the prediction, then the criterion is changed. If initial encoding reveals no clear prediction, then a second encoding is effected, this time in terms of big distance or weight. Then the rest of the process follows exactly as in the case of a single encoding.

Figure 5 contains a trace of the model working on one of the problems from Appendix B. The trace shows the state of working memory at the start of each cycle, as well as which production fired. Conflicts are shown when they occur, as are the results of the encoding operators.

The system starts with an element in WM (PREDICT) indicating that it has a goal of making a prediction, and another element representing the current value of the criterion. Since there is no element representing weight or distance, the only production whose conditions are completely satisfied is P8, which tests for (PREDICT) and the absence (ABS) of a weight or distance element (DISTANCE.1). ATTEND, P8's only action, is an encoding operator that is modeled only up to the point of its input/output specifications. In this case the input is presumed to be the physical arrangement of disks on pegs in the configuration (0003/0020), and the outputs, as shown in the trace, are two comparative symbols indicating more weight on the left and more distance on the right. They are directly provided by the model builder.

Thus, at the beginning of Cycle 2, WM contains four elements, and these elements satisfy both P4 and P2 (see Fig. 4). P4 is a special case of P2, so it fires. It recognizes that neither weights nor distances are equal, so it attempts a second encoding (FIND.BIG) to determine some absolute amounts of distance and/or weight. Once again, an unmodeled encoding operator is assumed to produce two elements, indicating a big distance on the right and a big weight on the left. The results are shown at the start of the third cycle.

Five productions are satisfied by the elements now in WM. P2 and P4 are still satisfied since none of the elements that satisfied them on the previous cycle have been changed. P5, P6, and P7 are satisfied because they test for either big weight or big distance. Since P4 is a special case of P2, and P5 of P6 and P7, the special case order principle leaves P4 and P5 in the conflict set. But the elements that match P5 are newer than those that match P4, so WM order selects P5 to fire.

P5 matches whatever the current value of the criterion is (in this case, it is weight) with the corresponding "big" element (in this case [WEIGHT BIG LEFT]) and then uses the value of the directional variable (LEFT) to form its expectation (EXCEPT LEFT DOWN) and to "say" its prediction.

What the system knows at this particular moment is revealed by the contents of WM at the start of the fourth cycle. It knows that:

It expects the left side to go down (EXPECT LEFT DOWN);
It already made a prediction (MADE (PREDICT));
The current criterion is weight (CRITERION WEIGHT);
And it knows the encodings (WEIGHT BIG LEFT) (DISTANCE BIG RIGHT), (WEIGHT MORE LEFT), and (DISTANCE MORE RIGHT).

The rest of the trace is straightforward. During Cycle 4, the system seeks an encoding of what the scale actually did, and it sees that the right side went down. On Cycle 5, it recognizes that what it saw is discrepant with what it expected (E3), so it knows that it got the problem wrong. Finally, on Cycle 6, it recognizes that it was wrong while using the weight criterion, so it changes it to distance.

G. EVALUATION OF REPRESENTATIONS FOR JAN'S KNOWLEDGE

The decision tree in Fig. 3 and the production system in Fig. 4 are logically equivalent: Both account for all but the last of Jan's predictions during the training series. As described above, they differ from the representations in Figs. 1 and 2 in that they model that subject's response to feedback, and because they both represent idiosyncratic encodings of the stimulus. Thus, both models have certain advantages over the previous ones.

However, the models are not equivalent in all respects, and the psychological properties of the production system—properties previously just alluded to—can now be clarified. The production system, since it embodies a general model of the human-information processing system, forces us to form very explicit hypotheses about things that the decision tree lets us finesse. There is no separation of control information from data in a production system. Every relevant piece of information is explicitly represented in WM, and all task-specific knowledge for acting on that information is represented by productions. As indicated by the final list of elements, we are postulating a sizable amount of

material floating around in WM. It is clear that the size of WM is well beyond the estimated short-term memory capacity of from seven (Miller, 1956) to as little as three or four (Broadbent, 1975) items, or the "M-space" estimates (Pascual-Leone, 1970) in the same range. However, it is unclear how a system that did not have immediate access to all of this momentary knowledge could ever do the task. Questions about the amount of control information sufficient to perform the task are not addressed by the decision tree representation.

For all their emphasis on the importance of the outputs from the encoding operators, however, the production system models do not describe the encoding process itself. Neither do they indicate precisely what sort of encoding deficit might affect response to instruction. A remedy to the former limitations would take the form of a model of encoding, and we leave that for future investigation. The second issue, that of the nature and effect of encoding deficits, is directly related to Questions 4 and 5 of our initial set. The specific questions are (1) whether encoding deficits are in fact typical of the younger children, and (2) if such deficits exist, whether they account for the younger children's inability to benefit from instruction on conflict problems. These questions were investigated in Experiment 3.

VIII. Experiment 3: Encoding Hypothesis

Recall that the results of Experiment 2 indicated that older and younger children, equated for initial task-specific knowledge about the balance scale, responded quite differently to the training sequences. This finding motivated a shift in the representation and in the level, or grain, of our analysis of what was going on during training. Lisa's protocol analysis revealed her difficulty in determining the appropriate encoding of the two relevant dimensions, and the analysis of Jan's responses during training led to a production system which incorporated two levels of encoding—one relative, one absolute (big/not big)—for both dimensions. Analysis of other protocols revealed many such stimulus misencodings. This, together with the sizable literature on the development of attentional strategies (cf. Pick, Frankel, & Hess, 1975; Zeaman & House, 1963) suggested to us that differential encoding might be the cause of the differential responsiveness to instruction.

Siegler (1976) described three steps that are necessary to test this hypothesis rigorously. (1) Assess encoding independently of predictive performance and establish the hypothesized encoding differences. (2) Show that the appropriate manipulation can eliminate or at least reduce encoding differences. (3) Demonstrate that when the difference on the explanatory variable—encoding—is eliminated, the initially observed difference on the to-be-explained variable—responsiveness to instruction—is also eliminated. In summary, then, the goal was to show that in a group of older and younger children who were all using Model I initially, there would be a consistent encoding deficit in the younger children, then to eliminate this deficit, and finally to expose both groups to the training sequence and to produce identical learning in both age groups.

Attempting to do this at the fine-grained level of the preceding section would have led to a mass of detailed variation that would make it very difficult to verify the general properties of encoding differences; it also would have been prohibitively expensive in terms of time and effort. Therefore, in this section, we move back up to a more aggregated level of analyses.

A. RECONSTRUCTION PARADIGM

Chase and Simon (1973) utilized a reconstruction paradigm to study the differential ability of chess masters and nonmasters to extract meaningful information from briefly presented board configurations. This procedure suggested to us a means by which differences between older and younger children's encoding of balance scale configurations could be assessed independently of their predictions about the effect of these configurations on the scale's behavior.

In the third experiment in this series, 5- and 8-year-old children were presented with various configurations of weights on a balance scale for a few seconds (the scale was not free to tip). Then the scale was removed from view, and the children were required to reconstruct the initial configuration as accurately as possible on an empty scale. Note that this procedure allowed independent assessment of encoding on both weight and distance dimensions. For example, when given an initial configuration (0300/0200) the child might reconstruct it, for example, as (0300/0200), or (0030/2000), or (0200/0100), or (0010/0003), revealing, respectively, no misencoding, distance only misencoding, weight only misencoding, and both weight and distance misencodings. Our protocol analyses led us to expect that the older children would be accurate on both dimensions, while the younger children would do well on weight but poorly on distance.

1. *Basic Procedure*

The same basic procedure was followed in all phases of Experiment 3, and the full details are given in Siegler (1976). Here we will only describe the major features. Overall, 40 kindergartners ("5-year-olds") and 30 third-graders ("8-year-olds") from two public schools in Pittsburgh participated in Experiment 3.

Two identical balance scales were used. They were slightly different from the one used previously, having seven rather than four pegs on each side of the fulcrum, and having a built-in lever rather than wood blocks to keep the scale from tipping until the experimenter released it. A large Styrofoam board was used to hide one of the balance scales during the reconstruction phases.

The encoding test included 16 problems, on each of which there were from three to five weights on each side, all located on either the third, fourth, or fifth peg from the fulcrum. On any given problem, only one peg on each side was occupied.

Children were tested individually in a vacant room in their school. Each child was presented the encoding test first, and then presented the same 24-item predictions task (without feedback) used in the Experiment 2 posttest. For the encoding test, the 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 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. Just put the weights on the pegs so it's just like the problem you saw on my scale.

After the first 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 your scale in the same way." Children were allowed 10 seconds to observe the initial configurations, and then they were allowed to reconstruct the arrangement immediately on the other scale. There was no time limit for reconstruction, although children usually finished quickly.

Following the last encoding trial, children were told that they were to play another game, and instructions similar to the previous predictions trials were given. The encoding and predictions tasks were given in a single session lasting about 25 minutes.

There were several variations on this basic procedure. We will describe each variation and its results in sequence. The results from all phases are shown in three forms. Table VII shows the percentage of correct distance and weight encodings for both age groups. A strict criterion of perfect reconstruction of both sides of the scale was used for both weight and distance scoring. Table VIII shows the percentage correct predictions for each type of problem, and Table IX shows the classification by model type for each age group in each phase of Experiment 3.

2. *Experiment 3a*

Ten children from each age level participated in the experiment exactly as described above. As shown in Table VII, the results were consistent with the encoding hypothesis. The younger children showed a great disparity between their ability to reproduce weight and their ability to reproduce distance while the older children did not show a significant difference. This pattern held for individual subjects in each age group, and is not the result of aggregating over subjects (see Siegler, 1976, for more extensive statistical analyses of these results). Notice that these encoding differences between older and younger children were not accompanied by a corresponding difference in ability to predict how the balance scale would behave. As shown in Tables VIII and IX, there was virtually no difference in the percentage of different types of problems passed or in the distribution of children using each model.

3. *Experiment 3b*

In this variant, 10 5-year-olds were given 15 rather than 10 seconds to view the initial configuration during the encoding tests. This was done to explore the possibility that the younger children were simply a bit slower than the older ones in encoding the configurations. If they were attempting to encode both dimensions, and had a preferred noticing order of weight first, then giving them more time would be expected to improve their distance scores. As shown in Table VII, this "insufficient time" explanation is unsupported by the results.

4. *Experiment 3c*

Perhaps the younger children did not understand what was meant by "make the same problem." In this variant, the children were told explicitly *what* to encode, and what constituted the experimenter's criterion for the "same" problem. Ten children of each age level participated. The instructions for the encoding task 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 pegs—how many there are on each side

TABLE VII
Percentage 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

TABLE VIII
Percentage Correct Predictions—Experiment 3

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

TABLE IX

Number of Children Using Different Models—Experiment 3

Experiment	Age (years)	Models I	II	III	Unclassifiable
3a	5	9	0	0	1
	8	8	1	0	1
3b	5	7	0	0	3
3c	5	7	0	0	3
	8	6	2	1	1
3d	5	6	0	0	4
	8	6	1	0	3
3e	5	1	3	4	2
	8	0	3	7	0

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 presented the instructions again 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, with a 15-second viewing period.

Once again, as shown in Tables VII, VIII, and IX, the results differed hardly at all from those of Experiments 3a and 3b. Telling children what to encode did not reduce the discrepancy between their encoding of weight and distance, nor did it improve their performance on the predictions task.

5. *Experiment 3d*

This time, children were told not only *what* to encode, but also *how* to encode it. If the problem lay in the inability of the younger children to correctly encode distance, or to handle two dimensions simultaneously, then perhaps direct instruction might help them. Ten children of each age group were given the following additional instructions during the encoding trials:

> You do it like this. First you count the number of weights on this side—one, two, three, four. 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 third peg." Then you would put the right number of weights on the right pegs on each side. Let's practice one.

This was followed by seven practice trials on which the child received feedback on the correct counting of weights and distances. This procedure was expected to reduce or eliminate the weight-distance discrepancy for the younger children, but since the older children presumably already knew how and what to encode, it was not expected to affect their performance. No effect was expected on the predictions performance of either group. All of these expectations were confirmed. Table VII shows that the younger children performed equally well on weight and distance, and that the older children performed better overall, but with no weight–distance discrepancy. Tables VIII and IX show that the predictions performance of both groups was indistinguishable from previous results.

6. *Experiment 3e*

Having finally eliminated the younger children's encoding deficit, we next asked whether that deficit really was the cause of the initial differential responsiveness to instruction. In this final experiment, the same children who participated in Experiment 3d were given the conflict training sequence used in Experiment 2 a few days after they completed Experiment 3d. According to the encoding hypothesis, both older and younger children should now benefit from experience with conflict problems that previously had benefited only the older children. Following the training sequence, the predictions test (without feedback) was again given to the two groups. The results of this posttraining predictions test are shown in Row 3e of Tables VIII and IX. Note that Rows 3d and 3e are based on the same set of subjects at different times. The sequence of manipulations and their corresponding results were: (1) instructions about what and how to encode; (2) encoding task (Table VII, Row 3d); (3) predictions task (Tables VIII and IX, Row 3d); (4) conflict training with feedback, a few days later; and (5) repeat of predictions task (Tables VIII and IX, Row 3e).

Comparison of Rows 3d and 3e in Tables VIII and IX shows that training now aided both age groups. Although there appears to be a slight advantage overall for the older children, there were no significant effects for either age alone, and no age–problem type interaction. It seems clear that the qualitative differences in responsiveness to training were eliminated by prior training in encoding. Although the younger children did not benefit as much as the older, it should be remembered that their encoding performance also did not reach the level of the older children.

B. SUMMARY

The results of Experiment 3 provide strong support for the encoding hypothesis: Younger children clearly do not tend to encode the distance dimension in this task. Without such encoding, they can derive little benefit from the instruction series. However, if given careful and explicit instruction on encoding, they do begin to do it correctly, and such improvement subsequently enables them to benefit spontaneously from a training sequence.

IX. Discussion: Some Answers and Some Further Questions

In this final section we will briefly summarize the work reported thus far, and give some indication of possible future efforts.

A. WHERE WE HAVE BEEN

Let us summarize where this series of investigations has taken us in our attempt to address the five questions posed at the outset.

Question 1: What are the differences in knowledge that underlie different levels of task performance?

The results of Experiment 1 indicated that the four models, in either decision tree or production system formulation, could accurately represent different kinds of knowledge that underlie distinct behavior patterns.

Question 2: What are the alternative strategies that might result in any given level of task performance?

Our analysis defined levels in terms of four modal forms of rule systems. Thus, neither representation could account for alternative means by which a subject might be generating the pattern of responses that led to his classification according to the models. However, the comparison of predictions and explanations for the Model III children suggested that such variations were indeed occurring. Detailed analysis of Lisa's protocol further indicated the need to account for individual variations, and with the construction of Model IIIA for Jan we began to demonstrate how these representations could account for highly idiosyncratic processes underlying Model III response patterns. As we argued earlier, the advantage of the production system representation for this individual level lay in its explicit set of psychological assumptions, assumptions that are consistent with a developing general view of some of the properties of the human information-processing system.

Question 3: For a given level of performance, what is the optimal level of difficulty for an instructional sequence?

The results of Experiment 2 provided a specific example of the general view that "near" training is better than "far" training. Although the particular definitions of "near" and "far" were clearly derived from the underlying representation for knowledge on the task, we did not do enough parametric variation

to be able to claim true optimality. However, the dimensions along which an investigation of such variation could take place are very clear.

Question 4: What are the critical features of an instructional sequence that enable it to have any effect?

The results of Experiment 2 demonstrated that experience with particular types of problems is critically important for improving subsequent performance. This sensitivity was derivable from the initial modal forms of the models. In addition, the role of negative feedback was specified in Jan's model, which is responsive to a mismatch between expectations and the actual outcome of each trial.

Question 5: When and why will two learners at the same initial performance level learn differently from the same instructional sequence?

Most of the development of the protocol analysis and the final production system model for Jan was stimulated by our attempt to answer this question. In Experiment 2 we detected the differential response to experience; Lisa's protocol suggested the encoding hypothesis; Jan's production system indicated the potential complexity and importance of encoding operators and suggested the operational form of the encoding hypothesis, for which the results of Experiment 3 provided strong support.

B. WHERE WE ARE GOING

These investigations have suggested further explorations in two interacting domains: conceptualization of models and experimental studies.

1. Types of Knowledge in the Human Information-Processing System

Our exploration of the issues surrounding the evaluation of different representations for knowledge has revealed that it is possible to distinguish between several different types of knowledge. The suitability of a representation depends upon the particular type of knowledge in which we are interested. In this section, we will briefly indicate what appear to us to be distinctly different kinds of knowledge. The order in which they are described corresponds roughly to their degree of permanence and stability in the human information-processing system.

a. K1: Knowledge About the Momentary State of Affairs. This is the knowledge represented by the elements in WM in a production system, or in the more general concept of "active memory" in other cognitive theories. In a production system, all the productions are continually attempting to recognize familiar elements of K1, and to act upon it through modification. K1 represents what is "going on" from one moment to the next. It contains information about the environment—information that has been produced by encoding operations and by the actions of satisfied productions. It constitutes a record of the system's immediate past.

b. K2: Knowledge About How to Do a Task or Solve a Problem. This type of knowledge is represented by decision trees of the sort used in Figs. 1 and 3, or by the productions in a production system. The knowledge in K2 typically consists of tests for the type of knowledge represented by K1. A production system provides a convenient and flexible representation for K1 and K2. However, many of the explicit assumptions in such representations have no particular psychological relevance, while others, although important, may not be amenable to independent experimental verification. Thus, the evaluation of production system representations for K2 rests upon multiple-level converging empirical measures, including global responses, protocols, and reaction times. Although many particular assumptions may be unverifiable, the integrated behavior of the total system can be observed and evaluated.

c. K3: Knowledge About How to Describe K2. A frequently discussed issue in developmental psychology concerns the relative validity of explanations versus performance (cf. controversey between Braine, 1964; Smedslund, 1965). In our experiments we found that the two forms of measurement were highly correlated, although they did reveal some interesting differences in the Model III children. It would seem that all explanations tasks require that the child have a type of knowledge that is distinct from performance knowledge as such, although it is rarely modeled or represented explicitly in psychological theories.

d. K4: Knowledge About How to Modify K2 and K3. This is the knowledge required for both learning in the long term, and immediate self-modification according to task demands. Some of the general properties of representations for this sort of knowledge have been discussed by Klahr (1976a, 1976b) and Newell (1972a). The content of K4 would be a theory of learning, and it would be premature to even imply that such a full-blown theory is near at hand. However, recent and ongoing work with self-modifying production systems suggests that this is a very promising form of representation for K4 (Waterman, 1974).

e. K5; Knowledge About How to Interpret the Knowledge Stated at the Other Levels. This knowledge would include, in the case of a production system representation, all the rules that the interpreter has to have in order to run a production system. It is the base level knowledge in the system, presumed to be functionally equivalent to the basic system architecture. It is probably inaccessible to introspection, or to instruction, although it may undergo development.

Answers to the questions that are asked about cognitive development may depend, in part, on which of the knowledge types described above are being considered. For example, studies of the development of short-term memory

capacity refer to changes in the amount and type of K1 (Chi, 1976; Huttenlocher & Burke, 1976), while studies of "metacognition" (Flavell, 1976) would appear to be primarily addressed to the development of the interaction between K2 and K3. Questions about "readiness" refer to the K2–K4 interaction and issues related to the competence–performance distinction seem to involve the interplay between K5 and K3, that is, between the "deep" interpretive capacity of the system and particular task-specific knowledge.

2. Experimental Extensions

The results of our experiments suggest a number of directions for further research. One would be to examine problem isomorphs—tasks similar to the balance scale in formal properties but differing in specific characteristics. Efforts in this direction have already been made. Decision tree models have been formulated and tested on Inhelder and Piaget's (1958) projection of shadows task, Bruner and Kenney's (1966) fullness of a water jar task, and Chapman's (1975) probability learning task. In each case the models have been found to accurately represent children's predictive performance. The experiments have also revealed a rich variety of reactions to feedback and encoding strategies (Siegler & Vago, in press). Another approach would be to construct and test more detailed production system models of exactly how children encode balance scale and other problems: Detailed analyses of reaction times might provide the appropriate test for such models. Finally, a host of instructional issues might be examined such as: When should tasks be taught directly and when should appropriate encoding strategies be taught first? Can procedures be devised to teach effective encoding on a variety of problems, or must instruction in encoding proceed on a task by task basis? Do differences in encoding account for individual differences among children of a given age, as well as developmental differences? Pursuing these problems will almost certainly lead to new insights about the five questions posed at the outset of this paper, and also to new questions.

X. Conclusion

Representation of children's knowledge requires that we make testable assertions about both the basic encoding of the environment and the processes that operate on those encodings. Cognitive development and instructional procedures involve changes in both the encoding operators and the rule systems. Instruction will tend to be ineffective if the instructional situation is encoded by the learner in a manner that is unexpected by the instructor. In a limited domain, we have demonstrated that such misencoding was indeed occurring, that we could locate the point of difficulty, eliminate it, and have instruction proceed as we expected it to.

From a broader viewpoint, we have tried to show that the appropriate representation for knowledge depends upon the goals of the scientific endeavor. Different kinds of knowledge are best represented by different formalisms, and are best investigated by different empirical procedures. This pluralistic view of knowledge representation may facilitate our understanding of, and influence upon, what it is that children know.

Appendix A
Protocol Excerpts from Lisa, a 5-Year-Old, on Training Sequence

T1 0001/2000 4:46

E. Okay, Let's put these two here, and this one here.
S. This side will go down (points left).
E. Which side? Touch the side that will go down.
S. (Touches left side)
E. Okay. Let's see if you were right. (Removes blocks. Scale tips left–up; right–down.) Were you right?
S. (Nods no)
E. Which side went down?
S. (Points right)
E. Okay. Why do you think that was? Why did you think before this side would go down?
S. 'Cause that one (points left) didn't have as much as that one (points right).
E. Uh-huh. But what actually happened?
S. This side went down because that one's heavier (points right).

T3 0100/1000 6:10

E. Okay. What do you think will happen this time?
S. They will both stay up.
E. Why do you think that?
S. 'Cause they are both the same.
E. Let's see if you are right. (Removes blocks. Scale tips left–down.) Were you right?
S. (Nods yes)
E. You were? Look. Do they both. . . Are they balanced? Is it like it was before?
S. Well, why are they both the same thing and one's up and one's down?
E. Why do you think that is?
S. I don't know.

T4 0020/0020 7:00

E. Okay. What do you think will happen this time?
S. The same again.
E. They will stay the same again. Why do you think that?
S. 'Cause. Wait a minute. It won't.
E. It won't?
S. 'Cause this one (points left) is closer to this one (points to fulcrum). And this one (points right) is closer to this one (points to fulcrum).
E. So what will happen?
S. This side (points right) will go up.
E. This side will go up?
S. Uh-huh.
E. Okay. What do you mean by "up"? Point which way it will go.
S. I think. . .
E. Which?
S. . . . it will go down.
E. This side will go down (points left)?
S. Uh-huh (nods yes).
E. And this side . . . and so it will be like this? (Tilts balance manually left–down; right–up.)
S. Uh-huh.
E. Is that right?
S. Uh-huh.
E. Okay, let's see if you are right. (Removes blocks, Scale tips right–down.) Were you right?
S. (Nods no)
E. What happened?
S. This went down (points right).
E. Why do you think that is?
S. I don't know!
E. Well, think about it.
S. Ummmm.
E. Okay.
S. I just don't know why.
E. You just don't know why.
S. Uh-huh.
E. Well, we'll keep on working and maybe you'll figure it out.

T6 0102/2010 9:43

E. Okay, what will happen on this one?
S. Yes. This side (points right) will both stay the same (points to both sides).
E. Let's see if . . . Why do you think that?
S. Because they both look the same. One is empty in the middle and one is empty in the middle.
E. Okay (scale balances). . . . Were you right?
S. (Nods yes)
E. Uh-huh. You were. That's right.
S. I was right!

E. That's right.

T7 0200/2000 10:25

E. All righty. Now, what will happen this time?
S. This side (points left) is farther away from this (indicates fulcrum), and this side (points right) is closer to this (indicates fulcrum). So, I think this side (points left) will go up. And this side will go down (points right).

E. Let's see if you are right. (Removes blocks left-down.) Were you right?
S. Uh-uh (no).
E. What happened?
S. (Points left, center, right, and up)
E. What?
S. This went up (points right) and this went down (points left). I thought this would go up (points left).
E. Okay. Try to figure out what's happening.
S. Hm. I don't know why.
E. Well you just keep on trying to figure out.

E7 0003/0004 26:46

S. I'm ready. Ha, wait a minute, I forgot. I did it wrong. I gotta think . . . (can't hear). This is four. This is three. This one will go down (points right).

E. What do you think would happen if we put one more here? What do you think would happen? [0004/0004]
S. Both stay the same. This one is crooked a little bit (adjusts right weights).

E. What do you think'll happen?
S. Stay the same.
E. Yeah? Let's see if you're right. (Removes blocks. Scale tips right-down, with sharp rap as it hits table.) Did they?
S. No!
E. No? No, they didn't. Did they?
S. Plunk. Plunk.
E. Plunk! Why do you think that was?
S. I don't know. They both had four. See, one, two (counts left): one, two, three, four; (counts right): one, two, three, four.
E. They both have four. Is that what made this side go down so much and this side go up so much?
S. No.
E. What do you think it was?
S. I don't know.
E. Think about it. What could it be?
S. I just don't know.
E. Just don't know. Look at it for a moment and try to figure out what it could be. Real carefully.
S. This one is far away (points right) and this one is close (points left).

E. Okay. Have any other ideas?
S. Uh-uh (no).

E8. 0003/3000 28:40

E. Okay. Now, want to make up another problem?
S. Uh-huh.
E. Okay.
S. This one is gonna be a good one. Stay the same.
E. You think so?
S. Uh-huh.
E. Okay, let's see if you're right. (Removes blocks. Scale balances.) Were you right?

S. Uh-huh!
E. Yeah, you were.
S. I'm being right and right and right, but one time I was wrong.

E9 0003/0003 29:26

E. (Requests information on how scale works, and about what would happen on this trial.)

S. There's three, and this side (points right) would go down, I guess.
E. That side would go down?
S. And this side would go up (points left).
E. Why?
S. Because this is far away (points right) and this is close (points left). So I think it would.

E. Think so?
S. Uh-huh.
E. Let's see if you're right.
S. Ohh! Right!

E. What would you do to make it balance, now? (S. starts to move scale manually to balance position.) No. . . . I mean by moving the little . . . little circles around. What could you do to make that balance?

S. This three here (points to right, first peg), and this three stay here (points left). [0003/3000]
E. Let's see if that's right, what you do.
S. I have to hold this up. (Lifts right side and moves weights.) [0003/3000]
E. Are you right?

S. (Nods yes)

E. What would you do to make the other side go down?
S. Whoops. [3000/3000] That side will go up (points right). Whoops, there.
E. Why does that happen?
S. Because that one's far away (points left) and that one's close (points right).

E. I see.

S. But if both had them far away. [3000/0003] Both sides would go down (giggle). They balance.

Appendix B

Trace of production system for Model IIIA on four problems from Table VI. Assume that at start of Problem 5, criterion is set to distance, and retain final criterion value when moving on to next problem. Start with Problem 5. (Notation: Production system written in a special language called PSG. Numbers shown are counts of actions taken since start of each problem. Abbreviations are used for terms such as weight, distance, etc.)

Problem 5 (Ø2ØØ/Ø4ØØ)

Ø. WM: ((PRED) (CRITERION DST))
  Fire P8: ((PRED) (D1) ABS ——→ ATTEND)
    ATTENDING—INPUT NEXT STIMULUS ⇒ (wgt more right) (dst more left)
1. WM: ((DST MORE LEFT) (WGT MORE RIGHT) (PRED) (CRITERION DST))
    CONFLICT.SET: (P2 P4)
  Fire P4: ((PRED) (WGT MORE) (DST MORE) ——→ FIND.BIG)
    ATTENDING—INPUT NEXT STIMULUS ⇒ (wgt big right) (dst big left)
2. WM: ((DST BIG LEFT) (WGT BIG RIGHT) (PRED) (WGT MORE RIGHT) (DST MORE LEFT) (CRITERION DST))
    CONFLICT.SET: (P2 P4 P5 P6 P7)
    CONFLICT.SET: (P4 P5) AFTER SPECIAL.CASE.ORDER
    CONFLICT.SET: (P5) AFTER WM.ORDER
  Fire P5: ((PRED) (CRITERION D1) (D1 BIG X1) (D2 BIG X2) ——→ (MADE **) (EXPECT X1 DOWN) SAY.D)

********** LEFT down

6. WM: ((EXPECT LEFT DOWN) (MADE (PRED)) (CRITERION DST) (DST BIG LEFT) (WGT BIG RIGHT) (WGT MORE RIGHT) (DST MORE LEFT))
  Fire E1: ((EXPECT) ——→ LOOK)
    ATTENDING—INPUT NEXT STIMULUS = (see right down)

7. WM: ((SEE RIGHT DOWN) (EXPECT LEFT DOWN) (MADE PRED)) (CRITERION DST) (DST BIG LEFT) (WGT BIG RIGHT) (WGT MORE RIGHT) (DST MORE LEFT))
    CONFLICT.SET: (E1 E3)
    CONFLICT.SET: (E3) AFTER SPECIAL.CASE.ORDER
  Fire E3: ((EXPECT X1 X2) (SEE X1 X2) ABS (SEE) ——→ (DID **) (SEE ==⇒ SAW) (RESULT WRONG))

1Ø. WM: ((RESULT WRONG) (DID EXPECT LEFT DOWN (SAW RIGHT DOWN) (MADE (PRED)) (CRITERION DST) (DST BIG LEFT) (WGT BIG RIGHT) (WGT MORE RIGHT) (DST MORE LEFT))
  Fire SW1: ((RESULT WRONG) (CRITERION DST) ——→ (OLD **) (DST ==⇒ WGT))

12. WM: ((OLD (RESULT WRONG)) (CRITERION WGT) (DID (EXPECT LEFT DOWN)) (SAW RIGHT DOWN) (MADE (PRED)) (DST BIG LEFT) (WGT BIG RIGHT) (WGT MORE RIGHT) (DST MORE LEFT))

Now do Problem 6. Keep criterion.

TE: ((Ø1Ø212Ø1Ø))

Ø. WM: ((PRED) (CRITERION WGT))
    CONFLICT.SET: (P8)
  Fire P8: ((PRED) (D1) ABS ——→ ATTEND)
    ATTENDING—INPUT NEXT STIMULUS ⇒ (dst same) (wgt same)
1. WM: (WGT SAME) (DST SAME) (PRED) (CRITERION WGT))
  Fire P1: ((PRED) (WGT SAME) ——→ (MADE **) (EXPECT BALANCE EVEN) SAY.B)

********** balance

5. WM: (EXPECT BALANCE EVEN) (MADE (PRED)) (WGT SAME) (DST SAME) (CRITERION WGT))
  Fire E1: ((EXPECT) ——→ LOOK)
    ATTENDING—INPUT NEXT STIMULUS ⇒ (see balance even)

6. WM: ((SEE BALANCE EVEN) (EXPECT BALANCE EVEN) (MADE (PRED)) (WGT SAME) (DST SAME) (CRITERION WGT))
    CONFLICT.SET: (E1 E2)
    CONFLICT.SET: (E2) AFTER SPECIAL.CASE.ORDER
  Fire E2: ((EXPECT X1 X2) (SEE X1 X2) ——→ (DID **) (SEE ==⇒ SAW) (RESULT CORRECT))

9. WM: ((RESULT CORRECT) (DID (EXPECT BALANCE EVEN)) (SAW BALANCE EVEN) (MADE (PRED)) (WGT SAME) (DST SAME) (CRITERION WGT))
Fire SW3

1Ø. WM: ((OLD (RESULT CORRECT)) (CRITERION WGT) (DID (EXPECT BALANCE EVEN)) (SAW BALANCE EVEN) (MADE (PRED)) (WGT SAME) (DST SAME))

Problem 7 (0003/0020)

0. WM: ((PRED) (CRITERION WGT))
ATTENDING—INPUT NEXT STIMULUS ⇒ (wgt more left) (dst more right)

1. WM: ((DST MORE RIGHT) (WGT MORE LEFT) (PRED) (CRITERION WGT))
CONFLICT.SET: (P2 P4)
CONFLICT.SET: (P4) AFTER SPECIAL.CASE.ORDER
Fire P4: ((PRED) (WGT MORE) (DST MORE) ——→ FIND.BIG)
ATTENDING—INPUT NEXT STIMULUS ⇒ (dst big right) (wgt big left)

2. WM: ((WGT BIG LEFT) (DST BIG RIGHT) (PRED) (WGT MORE LEFT) (DST MORE RIGHT) (CRITERION WGT))
CONFLICT.SET: (P2 P4 P5 P6 P7)
CONFLICT.SET: (P4 P5) AFTER SPECIAL.CASE.ORDER
CONFLICT.SET: (P5) AFTER WM.ORDER
Fire P5: ((PRED) (CRITERION D1) (D1 BIG X1) (D2 BIG X2) ——→ (MADE **) (EXPECT X1 DOWN) SAY.D)

**********LEFT down

6. WM: ((EXPECT LEFT DOWN) (MADE (PRED)) (CRITERION WGT) (WGT BIG LEFT) (DST BIG RIGHT) (WGT MORE LEFT) (DST MORE RIGHT))
CONFLICT.SET: (E1)
Fire E1: ((EXPECT) ——→ LOOK)
ATTENDING—INPUT NEXT STIMULUS ⇒ (see right down)

7. WM: ((SEE RIGHT DOWN) (EXPECT LEFT DOWN) (MADE (PRED)) (CRITERION WGT) (WGT BIG LEFT) (DST BIG RIGHT) (WGT MORE LEFT) (DST MORE RIGHT))
CONFLICT.SET: (E1 E3)
CONFLICT.SET: (E3) AFTER SPECIAL.CASE.ORDER
Fire E3: ((EXPECT X1 X2) (SEE X1 X2) ABS (SEE) ——→ (DID **) (SEE ===⇒ SAW) (RESULT WRONG))

10. WM: ((RESULT WRONG) (DID (EXPECT LEFT DOWN)) (SAW RIGHT DOWN) (MADE PRED)) (CRITERION WGT) (WGT BIG LEFT) (DST BIG RIGHT) (WGT MORE LEFT) (DST MORE RIGHT))
Fire SW2: ((RESULT WRONG) (CRITERION WGT) ——→ (OLD **) (WGT ===⇒ DST))

12. WM: ((OLD (RESULT WRONG)) (CRITERION DST) (DID (EXPECT LEFT DOWN)) (SAW RIGHT DOWN) (MADE (PRED)) (WGT BIG LEFT) (DST BIG RIGHT) (WGT MORE LEFT) (DST MORE RIGHT))

Problem 8 (0100/0200) [notice, this has big dst, but not big wgt]

0. WM: ((PRED) (CRITERION DST))
Fire P8: ((PRED) (D1) ABS ——→ ATTEND)
ATTENDING—INPUT NEXT STIMULUS ⇒ (wgt more right) (dst more left)
TE: ((0100/0200))

1. WM: ((DST MORE LEFT) (WGT MORE RIGHT) (PRED) (CRITERION DST))
CONFLICT.SET: (P2 P4)
CONFLICT.SET: (P4) AFTER SPECIAL.CASE.ORDER
Fire P4: ((PRED) (WGT MORE) (DST MORE) ——→ FIND.BIG)
ATTENDING—INPUT NEXT STIMULUS ⇒ (dst big left)

2. WM: ((DST BIG LEFT) (PRED) (WGT MORE RIGHT) (DST MORE LEFT) (CRITERION DST))
CONFLICT.SET: (P2 P4 P7)
CONFLICT.SET: (P4 P7) AFTER SPECIAL.CASE.ORDER
CONFLICT.SET: (P7) AFTER WM.ORDER
Fire P7: ((PRED) (DST BIG X1) ——→ (MADE **) (EXPECT X1 DOWN) SAY.D)

********** LEFT down

6. WM: ((EXPECT LEFT DOWN) (MADE PRED)) (DST BIG LEFT) (WGT MORE RIGHT) (DST MORE LEFT) (CRITERION DST))
Fire E1: ((EXPECT) ——→ LOOK)
ATTENDING—INPUT NEXT STIMULUS ⇒ (see right down)

7. WM: ((SEE RIGHT DOWN) (EXPECT LEFT DOWN) (MADE (PRED)) (CRITERION WGT) (WGT MORE RIGHT) (DST MORE LEFT) (CRITERION DST))
CONFLICT.SET: (E1 E3)
CONFLICT.SET: (E3) AFTER SPECIAL.CASE.ORDER
Fire E3: ((EXPECT X1 X2) (SEE X1 X2) ABS (SEE) ——→ (DID**) (DID**) (SEE || ||==⇒ SAW) (RESULT WRONG))

10. WM: ((RESULT WRONG) (DID (EXPECT LEFT DOWN)) (SAW RIGHT DOWN) (MADE PRED)) (DST BIG LEFT) (WGT MORE RIGHT) (DST MORE LEFT) (CRITERION DST))
Fire SW1: ((RESULT WRONG) (CRITERION DST) ——→ (OLD **) (DST ===⇒ WGT))

12. WM: ((OLD (RESULT WRONG)) (CRITERION WGT) (DID (EXPECT LEFT DOWN)) (SAW RIGHT DOWN) (MADE (PRED)) (DST BIG LEFT) (WGT MORE RIGHT) (DST MORE LEFT))

ACKNOWLEDGMENTS

This study was supported by grants from The Spencer Foundation, by Public Health Service Grant MH-07722 from the National Institute of Mental Health, and by a grant from the Sloan Foundation. Thanks to S. Farnham-Diggory, C. Glenn, and A. Newell for comments on an earlier version.

REFERENCES

Baylor, G. W., & Gascon, J. An information processing theory of aspects of the development of weight seriation in children. *Cognitive Psychology*, 1974, **6**, 1–40.

Becker, J. D. Reflections on the formal description of behavior. In L. Bobrow & A. Collins (Eds.), *Representation and learning*. New York: Academic Press, 1975.

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, 1971.

Bobrow, D. G. Dimensions of representation. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding*. New York: Academic Press, 1975.

Bobrow, D. G., & Collins, A. (Eds.), *Representation and understanding*. New York: Academic Press, 1975.

Bower, G. H. Cognitive psychology: An introduction. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 1). Hillsdale, N.J.: Lawrence Erlbaum Associates, 1975.

Braine, M. D. S. Development of a grasp of transitivity of length: A reply to Smedslund. *Child Development*, 1964, **35**, 799-810.

Broadbent, D. E. The magic number seven after fifteen years. In R. A. Kennedy & A. Wilkes (Eds.), *Studies in long term memory*. New York: Wiley, 1975.

Bruner, J. S., & Kenney, H. On relational concepts, In J. S. Bruner, R. R. Olver, & P. M. Greenfield (Eds.), *Studies in cognitive growth*. New York: Wiley, 1966.

Chapman, R. H. The development of children's understanding of proportions. *Child Development*, 1975, **46**, 141-148.

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. C. Short term memory limitations in children: Capacity or processing deficits? *Memory and Cognition*, 1976, **4**, 559-572.

Flavell, J. H. Metacognitive aspects of problem-solving. In L. B. Resnick (Ed.), *The nature of intelligence*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Gelman, R. Logical capacity of very young children: Number invariance rules. *Child Development*, 1972, **43**, 75-90. (a)

Gelman, R. The nature and development of early number concepts. In H. Reese (Ed.), *Advances in child development* (Vol. 7). New York: Academic Press, 1972. (b)

Glaser, R. *Evaluation of instruction and changing educational models* (Center for the Study of Evaluation of Instructional Programs, Occasional Report No. 13). Los Angeles, Calif.: University of California, 1968.

Greeno, J. G. Cognitive objectives of instruction: Theory of knowledge for solving problems and answering questions. In D. Klahr (Ed.), *Cognition and instruction*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Hunt, E. What kind of computer is Man? *Cognitive Psychology*, 1971, **2**, 57-98.

Huttenlocher, J., & Burke, D. Why does memory span increase with age? *Cognitive Psychology*, 1976, **8**, 1-31.

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.

Klahr, D. Designing a learner: Some questions. In D. Klahr (Ed.), *Cognition and instruction*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976. (a)

Klahr, D. Steps toward the simulation of intellectual development. In L. B. Resnick (Ed.), *The nature of intelligence*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976. (b)

Klahr, D., & Wallace, J. G. Development of serial completion strategies: An information processing analysis. *British Journal of Psychology*, 1970, **61**, 243-257.

Klahr, D., & Wallace, J. G. The role of quantification operators in the development of conservation of quantity. *Cognitive Psychology*, 1973, **4**, 301-327.

Klahr, D., & Wallace, J. G. *Cognitive development: An information-processing view*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

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

Lee, L. C. The concommitant 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.

Mager, R. F. *Preparing instructional objectives*. Palo Alto, Calif.: Fearon Publishers, 1962.

McDermott, J. & Forgy, C. Production system conflict resolution strategies. In D. A. Waterman and F. Hayes-Roth (Eds.), *Pattern-directed inference systems*. New York: Academic Press (in press).

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, **63**, 81-97.

Moore, J., & Newell, A. How can MERLIN understand? In L. W. Gregg (Ed.), *Knowledge and cognition*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1974.

Newell, A. A note on process-structure distinctions in developmental psychology. In S. Farnham-Diggory (Ed.), *Information processing in children*. New York: Academic Press, 1972. (a)

Newell, A. A theoretical exploration of mechanisms for coding the stimulus. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*. Washington, D.C.: Winston, 1972. (b)

Newell, A. Production systems: Models of control structures. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press, 1973.

Newell, A., & McDermott, J. *PSG Manual*. Pittsburgh: Carnegie-Mellon University, Department of Computer Science, 1975.

Newell, A., & Simon, H. A. *Human problem solving*. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Pascual-Leone, J. A mathematical model for the transition rule in Piaget's developmental stages. *Acta Psychologica*, 1970, **32**, 301-345.

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

Pick, A. D., Frankel, D. G., & Hess, V. L. *Children's attention: The development of selectivity*. Chicago: University of Chicago Press, 1975.

Reddy, R., & Newell, A. Knowledge and its representation in a speech understanding system. In L. W. Gregg (Ed.), *Knowledge and cognition*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1974.

Resnick, L. B. Task analysis in instructional designs: Some cases from mathematics. In D. Klahr (Ed.), *Cognition and instruction*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Rychener, M. D. *Introduction to Psnlst*. Pittsburgh: Carnegie-Mellon University, Department of Computer Science, 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. Three aspects of cognitive development. *Cognitive Psychology*, 1976, 8, 481-520.

Siegler, R. S. (Ed.). *Children's Thinking: What Develops?* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

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., & Vago, S. The development of proportionality concepts: Judging relative fullness. *Journal of Experimental Child Psychology* (in press, 1978).

Smedslund, J. The development of transitivity of length: A comment on Braine's reply. *Child Development*, 1965, **36**, 577-580.

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.

Waterman, D. A. *Adaptive production systems* (CIP Working Paper No. 285). Pittsburgh: Carnegie-Mellon University, Department of Psychology, 1974.
Waugh, N. C., & Norman, D. A. Primary memory. *Psychological Review*, 1965, **72**, 89–104.
Young, R. M. *Children's seriation behavior: A production-system analysis*. Unpublished doctoral dissertation, Carnegie-Mellon University, 1973.
Zeaman, D., & House, B. J. The role of attention in retardate discrimination learning. In N. R. Ellis (Ed.), *Handbook of mental deficiency*. New York: McGraw-Hill, 1963. Pp. 159–223.

CHROMATIC VISION IN INFANCY

Marc H. Bornstein
PRINCETON UNIVERSITY

I. Introduction

A. COLOR VISION AND PATTERN VISION

Students of sensation and perception agree that vision predominates among the senses (e.g., Geldard, 1972; Marks, 1974). Specific examples from several different perspectives abound. The largest proportion of sensory neurons in the central nervous system is devoted to visual system function (e.g., 540 million, as compared to 100 million for the auditory system, according to Sinsheimer, 1971). Careful studies of human sensory dominance (e.g., Colavita, 1974; Pick, Warren, & Hay, 1969) and chronometric analyses of sensory information processing (e.g., Posner, Nissen, & Klein, 1976) agree that vision is prepotent. Vision is certainly the richest of the sense departments; the highly complex