

Solving Problems with Ambiguous Subgoal Ordering: Preschoolers' Performance

David Klahr

Carnegie-Mellon University

KLAHR, DAVID. *Solving Problems with Ambiguous Subgoal Ordering: Preschoolers' Performance*. CHILD DEVELOPMENT, 1985, 56, 940-952. Recent evidence has shown that although there are large adult-child differences in overall problem-solving performance, even preschoolers have rudimentary forms of strategies such as means-ends analysis that rely on the use of subgoals. However, in many situations, means-ends analysis is not applicable, and in order to solve a problem, some other method must be used. This study further explores preschoolers' repertoire of problem-solving methods. We use problems in which it is difficult to order subgoals, and in which, therefore, some method other than means-ends analysis must be used. 40 children between 45 and 70 months of age were presented with problems having ambiguous subgoal ordering. Although individual scores varied widely, none of several indices of performance were reliably correlated with age. A detailed analysis of move sequences revealed that preschoolers (a) tend to avoid backup, (b) are sensitive to incremental progress toward the goal, and (c) search 2 or 3 moves ahead for a goal. All of these component skills were combined into a "hill-climbing" method that explains 70% of the variance in problem difficulty.

It is well known that adults and children differ substantially in their ability to solve problems, and several explanations have been proposed to account for these differences. One possibility is that adults' immense knowledge base is likely to include information relevant to just about any task. Evidence supporting this position comes from Chi's (1978) demonstration that, when provided with task-specific knowledge greater than that of adults, children perform at higher levels than do adults. Another potential source of adult-child differences is the capacity of the underlying information-processing system. For example, Case (1980) argues for a trade-off between speed of processing and short-term memory capacity that might underlie the constant improvement in performance with age that he has found in a wide range of tasks.

In this paper we focus on yet another possible source of the difference in adult and child problem solving: the availability of a repertoire of general procedures, known in the cognitive science literature as "weak methods." Although usually inefficient compared to the problem-specific methods that

might be used by a problem solver who was familiar with a domain, the weak methods are extremely general, and they often provide the only basis for intelligent action (Laird, 1984; Laird & Newell, 1983; Newell, 1980).

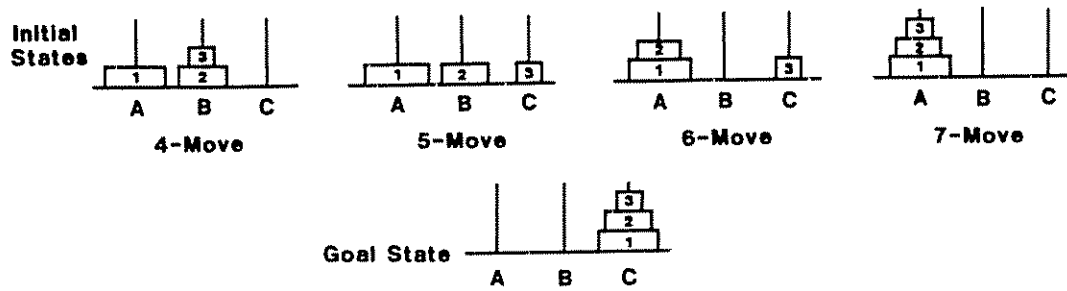
In order to understand the mechanisms responsible for the development of weak methods, we must first assess the extent to which the different methods are used at different ages, and the sequence in which the full repertoire of methods becomes available. The ability of young children to use some of these procedures has been demonstrated by previous investigators (Klahr & Robinson, 1981; Spitz & Borys, 1984; Spitz, Webster, & Borys, 1982). In particular, preschool children spontaneously use methods requiring the use of subgoals, such as means-ends analysis.

Ambiguous and Unambiguous Subgoal Ordering

In this paper, we study preschoolers' ability to solve problems in which means-ends analysis is ineffective. By using a problem that precludes the use of subgoals, we

This research was supported in part by grants from the National Science Foundation (BSN81-12743) and the Spencer Foundation. A preliminary report on this study was presented at the Sixth Annual Conference of the Cognitive Science Society, June 1984, Boulder, CO. Thanks to the parents, the teachers, and particularly the children of the CMU Children's School for their enthusiastic cooperation, and to Margaret Kinsky for her artistry, advice, and assistance in creating materials, running the experiments, and transcribing the results. Thanks also to Sharon Carver, Brian MacWhinney, and especially Bob Siegler for comments on an earlier draft. Herman Spitz, Robert Kail, Robbie Case, and one anonymous reviewer all made important suggestions for improvement. Author's address: Department of Psychology, Carnegie-Mellon University, Pittsburgh, PA 15213

a) Tower Ending



b) Flat Ending

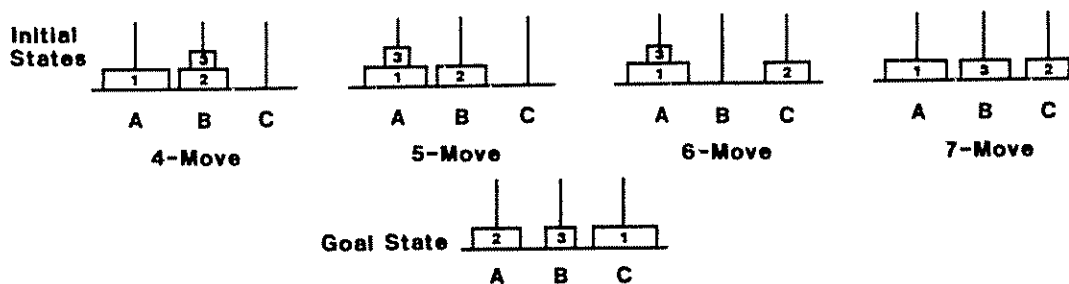


FIG. 1.—Two kinds of Tower of Hanoi Problems: *a*, Tower-ending problems requiring 4, 5, 6, and 7 moves to solution. These problems have an obvious subgoal ordering. *b*, Flat-ending problems requiring 4, 5, 6, and 7 moves to solution. These problems have a nonobvious subgoal ordering.

can assess the extent to which children use some of the other weak methods listed above. Before describing the puzzle used in this study, we discuss a more familiar one in order to make a distinction between problems that have easily ordered subgoals and those that do not.

The "standard" Tower of Hanoi problem, shown at the right side of Figure 1*a*, can be decomposed into a set of subgoals that, if achieved in the appropriate order, will lead to a solution. A subject using means-ends analysis would notice that there are three differences between the initial state and the goal state: each disk is on the wrong peg. Then the subject would have to decide which of these differences to eliminate. In this case, the most important difference is the one associated with the most constrained disk—disk 1—and a subgoal would be established to move disk 1 to the goal peg. But this move cannot be made until disk 1 is free to move, so means-ends analysis sets up a new subgoal to move the obstructing two-disk stack from Peg A to Peg B.¹ The new subgoal would, in turn, gen-

erate its own subgoals, until finally a move could be made directly.

Klahr and Robinson (1981) found that children's performance on the Tower of Hanoi was strongly influenced by both the number of subgoals generated and the ease of ordering them. In general, preschoolers were unable to keep more than two subgoals in mind without becoming confused. But, for the purposes of the present investigation, the most important effect was the performance decline when subgoal ordering was not self-evident.

The two variants of the Tower of Hanoi used by Klahr and Robinson differed in the form of the goal state. On the *tower-ending* problems, in which all the objects are stacked on a single peg (Fig. 1*a*), half of the 6-year-old subjects could solve 6-move problems, and even 5-year-olds were able to solve 4-move problems most of the time. On problems having *flat-ending* goal states, in which each peg has one object on it (Fig. 1*b*), there was a substantial decline in performance. The pro-

¹ For a complete analysis of this and several other strategies on the Tower of Hanoi, see Simon (1975)

portion of 5- and 6-year-olds who could reliably plan at least 4 moves ahead dropped from 81% to 40%. Spitz et al. (1982) explored the effects of non-tower-ending configurations by using towers, flats, and partial towers as goal states. They found that both retarded adults (mean mental age of 10 years) and non-retarded children (from 6 to 11 years old) scored significantly higher on tower-ending problems.

Tower-ending and flat-ending problems differ in the obviousness of the ordering of their subgoals. For tower-ending problems, it is clear that the bottom-most object must reach the goal peg before the second from the bottom, and so on. This subgoal sequence is immediately apparent, even though the exact move sequence necessary to achieve it may not be. In contrast, flat-ending problems do not have an obvious order in which disks reach their goal pegs. The evidence cited above suggests that when the surface form of the problem does not suggest an unambiguous ordering of subgoals, children have a difficult time applying mean-ends analysis. Instead, they must use an even weaker one of the weak methods—one that does not rely on subgoal ordering.

In this study, we further investigate how preschool children behave when confronted with problems having *ambiguous subgoal* ordering. One extreme possibility is that when children cannot use subgoals, they move haphazardly in an unconstrained trial-and-error fashion (see Piaget, 1976). On the other hand, we might discover that they have rudimentary forms of the weak methods that enable them to do some or all of the following: avoid unnecessary backup; evaluate the "goodness" of a move; advance directly toward a goal, once it is "visible." Our problems are designed to reveal the extent to which preschool children can respond intelligently in contexts where means-ends analysis is not applicable. More specifically, we seek evidence that they can use a method known as "hill climbing," in which candidate moves are chosen with respect to improvements in an evaluation function.

The Dog-Cat-Mouse Puzzle

The Dog-Cat-Mouse (DCM) puzzle consists of three toy animals (a dog, cat, and

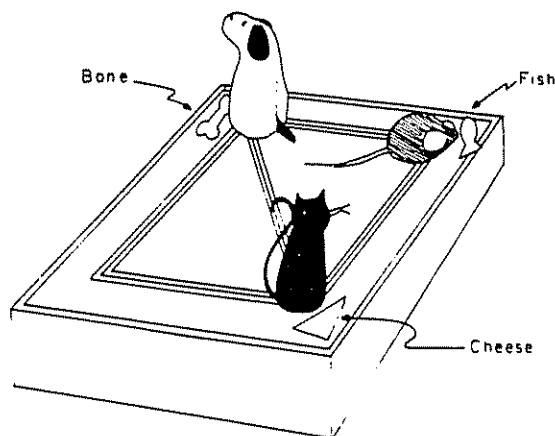


FIG. 2.—The apparatus for the Dog-Cat-Mouse problem. Each animal must be moved to its favorite food: the dog to the bone, the cat to the fish, and the mouse to the cheese.

mouse) and three toy foods that "belong" to the animals (a bone, a fish, and a piece of cheese). The animals and the foods are arranged on a 26 × 26 × 6-cm game board illustrated in Figure 2. The board has four grooves running parallel to each side of the square, and a diagonal groove between the upper left and lower right corners of the square formed by the four outside grooves. The animals can move along the grooves, but they cannot be removed from the board. The foods can be fastened to and unfastened from small patches of Velcro glued to each of the four corners. A problem consists of an initial state—indicated by the placement of each animal in a corner of the puzzle—and a final state—indicated by some arrangement of the bone, fish, and cheese. The goal of the problem is to move each animal to its corresponding food.

This puzzle was chosen for several reasons. First, and most important, it has ambiguous subgoal ordering: the order in which the animals will reach their foods is not at all obvious. Second, it has easily remembered rules and a natural way to represent the goal state. Third, the puzzle has a sufficiently wide range of levels of difficulty. Finally, it is novel,² and children are unlikely to have encountered similar puzzles.

Problem Set

Before describing the particular problems used in this study, it is necessary to de-

² The DCM puzzle is formally identical to the "depth-of-search" puzzle invented by Borys (1984) and first described by Spitz and Borys (1984). The puzzle is the simplest version of a class of puzzles in which T tiles must be arranged in an $M \times N$ array ($T = M \times N - 1$). Examples include the "15-puzzle" studied by Newell and Simon (1972), the "8-puzzle" used by Ericsson (1976), and the "5-puzzle" illustrated by Wicklegren (1974). In these terms, the DCM is a "3-puzzle."

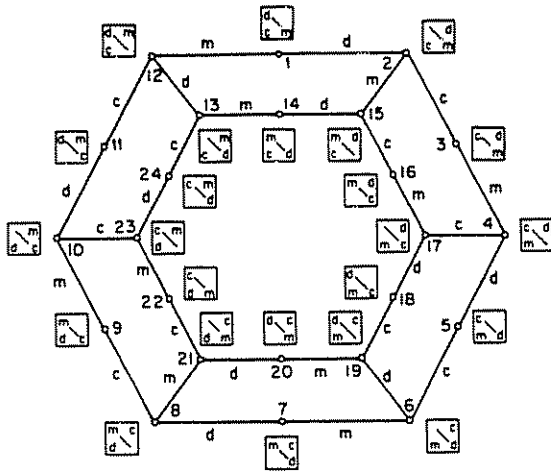


FIG. 3.—State space for the DCM problem. Each node represents a unique configuration of the three animals. Each arc is labeled with the piece that is moved to change states.

scribe some of the formal properties of the DCM puzzle. The state space is illustrated in Figure 3. Each of the 24 nodes represents one of the legal configurations of the three animals at one of the four corners of the game board. Each arc label corresponds to the animal that was moved to get from one state to its neighbor. For example, traversing the state space between nodes 1 and 2 requires a move of the dog. All problems are defined in terms of their initial states (determined by the arrangement of the animals) and their final states (determined by the arrangement of the foods). For example, the problem shown in Figure 2 starts at node 11 and ends at node 20.

Several properties of the state space are relevant to our subsequent discussion:

1. Problems can be characterized in terms of their *path length*: the minimum number of moves required to get from the initial state to the final state. The 552 distinct problems vary in path length from 1 move to 7 moves (example: 1–20 has a path length of 7).

2. If we abstract over the specific identity of the pieces, then there are only two types of nodes: those with *open* diagonals, having three adjacent states (e.g., 2, 4, 19, 21), and those with *closed* diagonals, having two adjacent states (e.g., 1, 3, 18, 20). At an open node, there are 3 possible moves; at a closed node, there are 2 possible moves.

3. *Rotation* problems have both initial and final states on the same hexagon—either the inner or the outer. They have minimum paths that do not use the diagonal of the game board (examples: 1–5, 23–17). All rotation

problems have path lengths less than 7. Those with path lengths less than 6 have unique solution paths (e.g., 1–5). Rotation problems with path lengths equal to 6 have two minimum path solutions (e.g., 1–7).

4. *Permutation* problems have initial and final states on different hexagons, and require the use of the diagonal. These problems start and end with different permutations of the three animals (i.e., D-C-M-D . . . vs. D-M-C-D . . .), and the permutation order can be changed only by using the diagonal (examples: 1–15, 22–3). Permutation problems generally have several minimum paths, for at every open node of the state space, there is the option of using the diagonal. For example, the minimum path from node 1 to node 19 could cross from the outer to the inner loop at nodes 2, 4, or 6.

The effect of problems that differ along these attributes depends on the processes that subjects use to solve them. If they use a lot of forward search, then on average, longer problems should be more difficult, and for equal-length problems, those starting with open nodes (3 possible first moves) should, on average, be more difficult than those with closed nodes (2 possible first moves). Permutation problems should be easier for two reasons: first, they usually have several minimum paths, and if subjects are moving randomly they are more likely to find one; and second, if subjects *are* able to formulate subgoals, then a very useful one would be to fix the permutation (i.e., use the diagonal) and then rotate to the goal.

Eight problems varying in path length (from 4 to 7), type of initial node (open or closed diagonal), and problem type (permutation or rotation) were used. They are listed in the bottom section of Table 1. In order to avoid fatigue, we used only eight problems rather than the full set possible with 4 (path lengths) \times 2 (initial node) \times 2 (problem type). (There are no 7-move rotation problems, so the full set would have 14 problems.) In addition, four 3-move training problems were used to familiarize the children with the rules of the game. They are shown at the top of Table 1.

Method

Subjects

Forty subjects, ranging in age from 45 to 70 months old ($M = 57.6$, $SD = 6.3$), participated. They were all attending the Carnegie-Mellon University Children's School, which has a predominantly, but not exclusively,

TABLE 1
PROBLEM SET

Problem No.	Initial State	Goal State	Path Length	Initial Node	Problem Type
Training set:					
T1	1	4	3	Closed	Rotation
T2	7	22	3	Closed	Permutation
T3	12	9	3	Open	Rotation
T4	2	17	3	Open	Permutation
Problem set:					
1	17	21	4	Open	Rotation
2	18	8	4	Closed	Permutation
3 ^a	11	20	5	Closed	Permutation
4	10	5	5	Open	Rotation
5	13	19	6	Open	Rotation
6	24	18	6	Closed	Rotation
7	14	7	7	Closed	Permutation
8	15	8	7	Open	Permutation

^aThis problem is illustrated in Figure 1.

white, middle-class population. (Data for one subject were lost, so results are reported for 39 subjects.)

Procedure

Children were tested in a small playroom, adjacent to their regular classrooms, that was equipped with videotape recording facilities. The experimenter was a 27-year-old white female who had interacted with all the children as a teaching aide throughout the previous several months.

After being brought into the room, the children were presented with the DCM puzzle in the context of the following cover story.

This is a game about three hungry animals, and your job in the game will be to make sure that each animal gets its favorite food. I have a dog here who loves to chew on bones—would you please give the dog his bone? I have a cat who loves to eat fish, and I have a mouse who loves cheese [subject distributes food]. In this game I will mix up the animals and the food and you will have to move each animal to its favorite food. There are three important rules about how you can move the animals:

First of all, the animals always sit in the corners next to these circles. They can move along these blue lines—around the outside or up the middle backwards or forwards—but they always have to stop in a corner by a circle. That means they can never stop in the middle of a line like this.

The second rule is that only one animal can be in a corner at a time. This is because my mouse is afraid of my cat, my cat is afraid of my dog, and, believe it or not, this big dog is afraid of mice. So they never sit together in one place, and you must never move an animal into a corner where another one is already sitting.

The third rule is easy to remember—they always move one at a time. While the dog moves, the cat and the mouse wait, and while the mouse moves, the dog and the cat wait.

Let's start with a couple of easy ones and then they will get harder.

Children were not explicitly instructed to minimize the number of moves; nevertheless, here, as in many other studies (e.g., DeLoache, Sugarman, & Brown, in press; Karmiloff-Smith, 1979; Klahr & Robinson, 1981), they appear to adopt efficiency as an implicit constraint.

Problems were presented in the order shown in Table 1: the four training problems first, followed by problems 1–8. Children were given two chances to produce a minimum path solution to each problem. If a problem was solved in the minimum number of moves, then the next problem in the sequence was presented. If it was solved in more than the minimum number, or if it had not been solved after twice the minimum number of moves had been made, or if the subject gave up, then the same problem was presented a second time. Regardless of whether the second trial produced the minimum path, a longer solution path, or no solution, the next problem in the sequence was then presented.

As each problem was presented, the children were reminded to rearrange the animals such that each animal would get its favorite food. The children were allowed to make their own moves; if they attempted an illegal move, they were reminded of the rules. The

most common illegal moves were moving an animal only halfway between two corners, moving two animals to the same corner, or attempting to rearrange the foods rather than the animals. However, illegalities occurred on fewer than 5% of trials and tended to occur only on the training problems. All problem-solving sessions were videotaped for subsequent analysis. Sessions took about 15 min to complete.

Scoring

For each problem, subjects were assigned a 1/0 score based on whether or not they found a minimum path solution by the second presentation of the problem. Each subject was assigned a score based on the proportion of passes (1's) across the eight problems. Each problem was assigned a score based on the proportion of subjects passing it. (We also tried an alternative scoring procedure: for each problem, a score of 2 was assigned if a minimum path solution was achieved on the first presentation, a score of 1 if achieved on the second, and 0 otherwise. The correlation between the alternative score and the simpler [0/1] score was $r[38] = .98$; none of the results to be reported would be changed by the use of the alternative scoring.)

Results

In this section, we first present aggregate results across subjects and problems. Then we present a detailed analysis that bears on the components of weak methods listed earlier: backup, partial evaluation, and search for goal states. Following that, we turn to an analysis of relative problem difficulty based on some of the formal properties of the problems that were used in designing them. Then, in the next section, we describe a model that attempts to incorporate all of the components into a single strategy.

Subjects' performance varied widely: the highest-performing subject solved all but one problem (mean score = .875), while three subjects failed all but one (mean scores = .125). Problem difficulty also varied widely, from nearly all subjects passing the easiest problem to over 80% failing the hardest problem. The top-ranked subjects tended to fail only the harder problems, and the lowest-scoring subjects passed the easier rather than the harder problems. However, there were many discrepancies from a Guttman (1950) scale (coefficient of reproducibility = .89).

Although ages were fairly uniformly distributed between 50 and 65 months, the correlation between age and proportion correct

was surprisingly low, $r(38) = .27$, $p < .10$. Furthermore, none of the more detailed analyses (to be described subsequently) produced an age-performance correlation any higher than this.

Backup Analysis

One of the most rudimentary forms of efficiency in problem solving is avoidance of unnecessary moves. In the DCM puzzle, moving the same piece in succession always results in a two-move sequence having no effect: the puzzle returns to the state occupied at the start of the sequence. If moves are made at random, without regard to this "no-backup" constraint, then we would expect 33% of moves at open nodes and 50% of those at closed nodes, or 42% of all moves, to be double moves. Double moves were rare: out of a total 3,350 moves, only 10% were doubles. Individual subjects made double moves on from 1% to 25% of their moves, but proportion of double moves was uncorrelated either with age (as mentioned above) or with overall score.

Not all double moves are inefficient. If a subject realizes that the current path is leading away from the goal, then a double move may be the best way to start to head toward it. Double moves were further analyzed to determine the frequency with which they were helpful. If a double move was the most efficient way to get back on a minimum path, then it was scored as acceptable. For each subject, we computed the proportion of double moves that were acceptable. Forty-four percent of double moves were acceptable, but this is no better than might be expected by a random decision to make a double move. On the rare occasions when children make double moves, they do not make them for any obvious reason.

One final comment on avoiding backup. In the DCM problem, no backup is equivalent to a prohibition on moving the same animal twice in succession, and either one of two general principles ("don't back up" vs. "take turns") could serve as the source of the constraint. In other words, rather than avoiding backup, children might be conforming to a simple turn-taking convention that predisposed them to move the other two pieces before moving the one just moved. But this convention must be violated once on all permutation problems, so that if it were operating here we would expect permutation problems to be more difficult than nonpermutation problems. But they are not: there is no significant difference in the mean scores of rotation and permutation problems. We con-

clude that children are not adhering to a turn-taking convention; instead, they *are* avoiding unnecessary backup.

Partial Evaluation

Another rudimentary ability inherent in several of the weak methods is evaluation of the quality of a proposed move. The simplest evaluation is binary: a state either matches the goal or it does not. (For example, in searching for a key on a key ring, one either has the right key or the wrong key.) Much more useful is the ability to make a *partial evaluation* that gives some measure of how well the current state matches the goal state. Consider an evaluation function— $EV(x, y)$ —that can compute how many of the pieces in state x are in the same positions in state y . For example, $EV(1, 7) = 0$, because none of the pieces are in matching positions, whereas $EV(24, 5) = 2$, because both the cat and the dog are positioned the same way in the two states.

If the children were using such an evaluation function, then we should see two kinds of biases in their move patterns. One bias would show up as a tendency to favor moves that increase the number of pieces in their goal locations. For example, in Problem 2 (18 → 8), a first move of the cat increases the evaluation function, while moving the dog does not. The dog is also off the minimum path. Over all trials and all subjects, on this problem, the cat was moved 81% of the time. Even more revealing are the “garden path” problems, in which the evaluation function produces a local improvement for moves off the minimum path. In Problem 4 (10 → 5), the minimum path move is the mouse, which does not increase the evaluation function. Only the cat increases the partial evaluation function, and it is preferred on 66% of the trials, even though it is off the minimum path. Similarly, on Problem 5 (13 → 19), the non-minimum move of the dog is preferred on 61% of the trials.

The other bias would be a reluctance to remove pieces from their goal locations—to reduce the value of a partial evaluation function. This can be assessed on Problem 3 (11 → 20), where the minimum path sequence requires that the dog be temporarily removed from its goal position. On 65% of all trials with Problem 3, subjects preferred to move the cat rather than the dog, even though this took them off the minimum path.

For each subject, we computed an *evaluation sensitivity score*: the proportion of trials on which, if such an evaluation function pre-

ferred one move to another, then the subject chose (one of) the preferred alternative(s). Of course, for many moves, the evaluation function yields no preferred alternative, and these situations are excluded from the computation.

All subjects showed a sensitivity to partial evaluation. Evaluation sensitivity scores ranged from .60 to .90 (mean = .69, SD = .05). However, as noted in the comments about garden path problems, this sensitivity to local evaluation is not necessarily beneficial. Such “hill climbing” strategies can lead to local maxima that are isolated from the goal. Indeed, evaluation sensitivity scores are *negatively* correlated with overall performance, $r(38) = -.48$, $p < .001$, suggesting that, on this set of DCM problems, exclusive reliance on partial evaluation was dysfunctional.

Goal Detection

Instead of a partial evaluation function, a problem solver could function with only a binary evaluation in conjunction with some forward search capacity. As an extreme example, consider a strategy that searched as far forward as necessary, testing each state for whether or not it was a goal state. Once having detected the goal, the problem solver would simply follow the path that led to it. If a subject lacked any partial evaluation capacity, but had the ability to search n moves ahead for the goal, then we should see perfect performance (i.e., no deviations from a minimum path) from n steps away. Overall, the proportion of minimum path solutions 1, 2, 3, 4, and 5 moves distant from a goal was .99, .96, .88, .66, and .29, respectively. A random model would vary from .40 to .20 over the same range of distances.

Each subject was assigned a *goal detection score* based on the distance from the goal that he or she could reach directly 100% of the time. For example, if a subject produced minimum path solutions every time he or she was 2 moves from the goal state, but only 55% of the occasions on which he or she was 3 moves away, then the subject would get a goal detection score of 2. The distribution of subjects at each level of the goal detection score was 0:4, 1:9, 2:11, 3:13, and 4:2. That is, two-thirds of the subjects could produce perfect solutions on all trials from at least 2 moves from goal states, and one-third of them could even do it from 3 moves away. None could reliably find minimum path solutions more than 4 moves distant. We can grant subjects a 2–3-move capacity to search for the goal state.

Relative Problem Difficulty

The mean problem difficulty is shown in the second column of Table 2. Path length (shown in the third column) is a poor predictor of problem difficulty. The two easiest problems (1 and 2) are also the two shortest, but even though they both have a path length of 4, there is a 30% difference in the proportion of subjects passing them. The next two easiest problems (7 and 8) are the two longest (7 moves). The four hardest problems are intermediate in path length, and within that set, there is a large difference between the pairs with the same path length. Overall, the correlation between path length and problem difficulty is not significant, $r(7) = -.42$, $p > .10$. Neither of the two other independent variables had a reliable effect on problem difficulty. For problem type, $t = -1.75$ (N.S.), and for starting node, $t = -.63$ (N.S.).

Another possible index of problem difficulty is what Spitz and Borys (1984) call subgoal length or "depth of search" (Borys, 1984). It is defined as one less than the number of steps on the minimum path before the first object reaches its final position. Spitz and Borys found that when second- and third-grade children solved Tower of Hanoi problems, subgoal length was a good predictor of problem difficulty. Thus defined, subgoal length (see col. 4 in Table 2) explains only 43% of the variance in problem difficulty. The marginally significant correlation between subgoal length and problem difficulty is further evidence that the DCM has ambiguous subgoals.

Path length, node type, problem type, and subgoal length are all *structural variables*: features of the problem rather than of the problem-solving process. Even if they are good predictors of difficulty, they leave unstated the underlying processes that they affect. For example, the subgoal length calculation used above is based on a tenuous assumption: that children can determine which animal to focus on first in choosing subgoals. As noted earlier, this is precisely what makes the DCM puzzle difficult.

The general point is that structural variables alone do not cause behavior directly: they are mediated by underlying processes. In some cases, the process model is so obvious that it need not be made explicit. But in situations of even modest complexity, such as the DCM puzzle, there are several plausible processes—or components of weak methods—and their interactions can only be understood by formulation of an explicit process model. Thus far, we have presented evidence for sensitivity to partial evaluation, for a no-backup constraint, and for a 2- or 3-move search for goal states. In the next section, we combine all of these into a single model.

Strategic Analysis

How might children attempt to solve these problems? In this section we will propose a model of the strategy that children use when they face problems, such as the DCM, in which subgoal ordering is ambiguous. The presentation has three parts. First, we describe a basic model (Model A) that is insensi-

TABLE 2
SUBJECT PERFORMANCE, STRUCTURAL VARIABLES, AND MODEL PERFORMANCE

Problem No	Subject Means ^a	Path Length	Subgoal Length ^b	Model A		Model B (2 Trials) ^c
				(1 Trial) ^d	(2 Trials) ^d	
1	640	4	2	333	556	541
2	950	4	1	500	750	869
3	436	5	3	375	609	510
4	179	5	3	167	310	278
5	184	6	4	333	556	399
6	263	6	4	250	440	352
7	436	7	5	500	750	740
8	590	7	4	500	750	800

^a Proportion of subjects finding minimum path solution by second attempt

^b Number of moves before first piece reaches final goal position

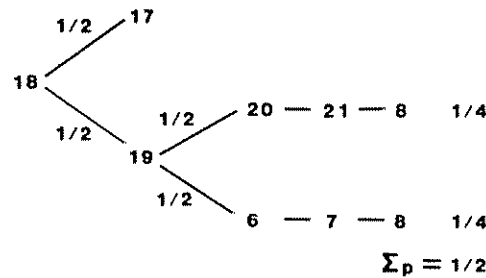
^c Proportion of minimum path solutions found by Model A

^d Computed probability of Model A finding solution with two attempts: $P = p + (1 - p) \times$

p

^e Probability of Model B finding minimum path within two attempts

a) Problem 2 (18-8)



b) Problem 7 (14-7)

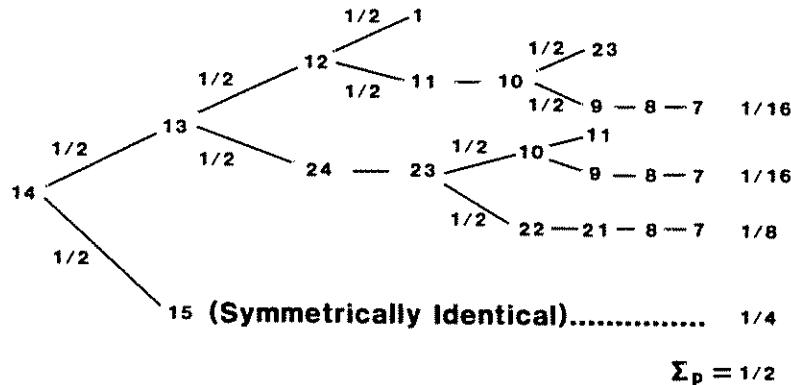


FIG 4.—Minimum solution paths for two of the problems used, with probabilities that Model A would traverse a particular path.

tive to partial solutions and very rigid in adhering to its own rules.³ Then we propose several plausible alternative models, and show that none of them fit the data very well. Finally, we describe an augmented model (Model B) that more accurately captures some of the fine structure of subjects' behavior.

Model A

Consider the following procedure for making moves in the DCM state space:

- 1 If there is a sequence of two or fewer moves that can reach the goal state, then make it. Otherwise.
- 2 Generate all candidate moves (all legal moves, except the piece just moved).
- 3 If there is more than one candidate, choose randomly.
- 4 Go to step 1

This is a simple generate-and-test strategy, with two constraints: (a) 2-move look-ahead for the goal state (the look-ahead has a binary

evaluation function: the state is either the goal or it is not; (b) no immediate backup.

We can determine the probability that Model A would discover a minimum path solution for each problem by computing the compound probabilities that it will stay on a minimum path. Two examples are presented in Figure 4. For problem 2 (Fig 4a), there are 2 possible first moves to nodes 17 or 19, but only node 19 is on the minimum path. At 19, there are 3 legal moves, but the return to state 18 is ruled out by the no-backup constraint, leaving only 2 moves, both of which are on the minimum path. Reaching either nodes 6 or 20 leads directly to the goal via the 2-move look-ahead. The fractions on each branch of the figure show the probability of that move, and the fractions at the goal nodes show the probability that Model A will reach the goal along a minimum path. Figure 4b shows a similar analysis for problem 7.

As Figure 4 illustrates, for problem 2, and for problem 7 as well, there is a .5 proba-

³ The strategies to be described represent first-order approximations to how the typical child in this age range solves this puzzle. Although it is likely that individual children are actually performing according to variants of these models, the data are not sufficiently dense to enable us to identify individual children with specific variants, as did Klahr and Robinson (1981).

bility that Model A will find a minimum path solution, even though they have path lengths of 4 and 7, respectively. The informal explanation for this result is that Model A is likely to solve this relatively short problem because it does not have to go very far to do so, and it is equally likely to solve a relatively long problem because of the symmetry of the state space. In particular, for problem 7, there is no possibility of a "wrong move" until after the first two have been made. Furthermore, for this problem, there are many minimum path solutions.

By applying this analysis to each of the eight problems, we can compare the probability that Model A would pass each problem with the subjects' actual performance. The fifth and sixth columns of Table 2 list, respectively, the probability that Model A would find a minimum path on one trial, and within two trials.⁴ The correlation between the mean problem difficulty and the probability that Model A would pass by the second trial is .70 ($p < .05$, $df = 7$). In other words, Model A explains about half of the variance in problem difficulty.

The generate-and-test method of Model A has three important features: no immediate backup, all-or-none evaluation, and a 2-step look-ahead for the goal state. We can ask two questions about these constraints. First, how well do subjects adhere to them? Second, how important are they? That is, if we modified them, how close would we come to matching subjects' performance? The next few sections address these questions.

The no-backup constraint.—Completely removing the no-backup constraint substantially reduces the probability of solution, for it adds an extra branch to each of the nodes in the analysis shown in Figure 4. That is, all the two-way branches become three-way branches, and all the direct connections (e.g., 11–10 in problem 7) become binary nodes. Even with 2-move look-ahead, such a model would perform far below the average solution rates for our subjects. However, even though children tend not to back up, they do it occasionally (on 10% of their moves), and one important modification to Model A would be a stochastic element that reflected this fact.

Evaluation.—Model A uses an all-or-none evaluation of goal states. As shown earlier, subjects are sensitive to partial evaluations on more than two-thirds of the instances

where such evaluations make a difference. Another modification to Model A would be the inclusion of a mechanism exhibiting the same sensitivity.

Depth of look-ahead.—Model A's 2-move look-ahead predicts perfect performance from up to 2 moves away from a goal, and then a sharp decline. As noted earlier, subject performance is indeed quite good at 2 moves away, but it remains high (nearly 90%) for 3 moves away, rather than dropping as predicted. In fact, as noted in the discussion of goal detection scores, nearly 40% of the subjects exhibited perfect performance once they were 3 moves away from the goal.

Given this relatively good performance from 3 moves away, it is reasonable to consider an alternative to Model A that differs only in having 3-move, rather than 2-move, look-ahead to the goal. But such a model would produce very high likelihoods of success within two trials, ranging from .97 and .94 for problems 8 and 7, to lows of .56 for problems 1, 4, and 5. Not only does this 3-move look-ahead produce unacceptably high solution rates, but also, it only explains about 5% of the variance in subjects' solution rates.

If we degrade the 2-move look-ahead to a 1-move look-ahead, then two things happen. First, the absolute level of performance drops, which is to be expected, since 1-move look-ahead will often branch off the minimum path when it is only 2 moves away from solution, whereas Model A would not. Second, the model now explains only 33% of the variance.

Summary of Model A.—In summary, Model A explains almost 50% of the variance in problem difficulty. Alternative strategies incorporating variations on the depth of the look-ahead to the goal state do not do as well. Elimination of no-backup from Model A yields unacceptably low solution rates. However, Model A oversimplifies children's performance in two respects. First, the children appear to be capable of partial evaluation, whereas Model A is not. Second, children do back up occasionally. In the next section, we describe and evaluate a model that incorporates both of these facts.

Model B

Model B makes moves according to the following rules:

1. If there is a 2-move sequence that can reach the goal state, then make it. Otherwise,

⁴ The probability of success by the second trial is: $P = p + (1 - p) \times p$, where p = probability of success on any single trial

2. Generate all candidate moves. On all but $P_1\%$ of trials, delete the piece just moved from the candidate set (e.g., backup is allowed $P_1\%$ of the time).

3. If there is more than one candidate, then compute EV between each candidate node and the goal node. Choose the move with the maximum EV on $P_2\%$ of trials. Or, if all EVs are equal, choose randomly.

4. Go to step 1.

The simple generate-and-test strategy of Model A has been changed into an imperfect hillclimber to reflect the findings that (a) children do back up, and (b) they are affected by partial evaluations.

The model has two parameters: P_1 is the double-move probability, and P_2 is the probability of being affected by EV. The values for these parameters were empirically derived from the analysis of double moves and all-or-none evaluation described earlier: P_1 was set to .10, and P_2 was set to .69. Recall that 10% of all moves were double moves. In setting P_1 to this value, we are ignoring any possible context-specific variation in the frequencies of double moves. The value of P_2 is based on two considerations. First, it is equal to the mean of the partial evaluation sensitivities described earlier. Second, it was empirically determined (through several simulations) that $P_2 = .69$, provided a better fit than three other values (.65, .75, .80). As in the case of P_1 , selecting a single value for P_2 ignores the fine structure of different contexts and individual differences.

With the two new parameters and the partial evaluation function, it is difficult to compute Model B's probabilities of success analytically. Therefore, we wrote a compu-

ter simulation program that embodied its method. Each problem was presented to the program 400 times, and each solution path was scored as a 1 or 0 by the same criteria used for subjects' performance. Then, the proportion of minimum path solutions (out of the 400) was computed, and this was converted to a probability of solution by the second attempt.

The results are shown in the final column in Table 2. Notice that of the three problems with initially misleading partial evaluations; two of them (3 and 5) are somewhat harder for Model B than for Model A, corresponding more closely to subjects' performance. Problem 2, which is aided by partial evaluation, is easier for Model B than for Model A. As a result of the reordering caused by these changes, Model B accounts for over 70% of the variance in subject performance, $r(7) = .845$, $p < .01$.

Even if subjects are using a Model B strategy, there are many things that might go wrong during its execution, and the longer a problem is, the more likely is their occurrence. As shown in Figure 5, Model B outperforms subjects on the longer problems. The path-length variable picks up these unknown sources of error. We noted earlier that path length alone accounted for only 18% of the variation in problem difficulty. However, path length is uncorrelated with Model B performance, and the multiple regression of both Model B performance and path length against problem difficulty accounts for 97% of the variance.

Discussion

The present study was designed to extend our knowledge about the development

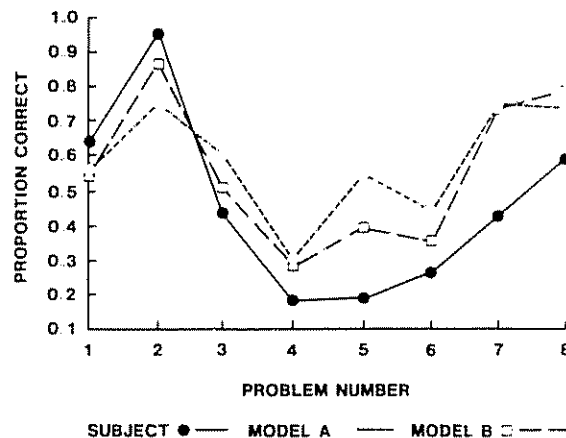


FIG 5—Actual versus predicted problem difficulty. Solid circles show proportion of all subjects passing by second trial. Open circles show probability that Model A would pass by second trial. Open squares show proportion of 400 cases that Model B passed by second trial.

of problem-solving skills. By presenting preschoolers with problems having ambiguous subgoal ordering, we were able to discover what weak methods they could invoke when means-ends analysis was not useful. One extreme possibility is that they resort to random trial-and-error. The other is that they use other, more appropriate, weak methods. In several respects, the results of this study support the latter alternative.

First, as described earlier, even the random component of Model B is highly constrained. The avoidance of double moves suggests a rudimentary knowledge about thoroughly useless actions that is not conveyed by a "trial-and-error" view of young children. The second important finding is that solutions are not simply arrived at by chance, since there is a look-ahead to the goal state, and little deviation from the minimum path once it is in sight. Third, children use partial evaluations of nearly correct states to guide their choice of moves. This sensitivity to incremental progress may actually degrade children's performance (as in the garden path problems). Nevertheless, it is reasonable for children to attempt to use such information.

The full repertoire of weak methods could follow a course of either strict sequential development, in which each method evolves out of its simpler predecessors, or broad-gauged parallel development, in which rudimentary forms of many methods develop simultaneously and relatively independently. The results of this study are consistent with the second view. They show that 4-year-olds already have rudimentary forms of several components of the weak methods. Furthermore, the fact that there were no age-related effects over about a year and a half span suggests wide individual variation in the course of these acquisitions. Of course, with sufficient age variation, there is an effect: Borys (1984), using her original version of this puzzle with 7-, 9-, and 11-year-olds, found a systematic improvement in performance with age. *Something* must be developing, even if not the repertoire of weak methods as such. The likely candidates for "What develops" in problem solving were mentioned at the outset: problem-specific knowledge and a trade-off between efficiency of execution and demands on immediate memory capacity. The results of Borys's (1984) investigation provide some evidence consistent with this view. As for development of weak methods prior to preschool age, not much is known, even though that may be where the most important changes occur (DeLoache & Brown, 1984; DeLoache et al., in press; Sugarman, 1982).

References

- Borys, S. V. (1984). *Depth of search capacity: A developmental study*. Doctoral dissertation, Graduate School of Education, Rutgers, the State University of New Jersey.
- Case, R. (1980). The underlying mechanism of intellectual development. In J. R. Kirby & J. B. Biggs (Eds.), *Cognition, development, and instruction*. New York: Academic Press.
- Chi, M. T. H. (1978). Knowledge structures and memory development. In R. Siegler (Ed.), *Children's thinking: What develops?* (pp. 73-96). Hillsdale, NJ: Erlbaum.
- DeLoache, J. S., & Brown, A. L. (1984). Intelligent searching by very young children. *Developmental Psychology*, 20, 37-44.
- DeLoache, J. S., Sugarman, S., & Brown, A. L. (in press). The development of error correction strategies in young children's manipulative play. *Child Development*.
- Ericsson, K. A. (1976). *Approaches to descriptions and analyses of problem solving processes: The 8-puzzle* (Tech. Rep., Supplement 32). Stockholm: Department of Psychology, University of Stockholm.
- Guttman, L. (1950). Problems of reliability. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarfield, S. A. Star, & J. A. Clausen (Eds.), *Measurement and predictions*. Princeton, NJ: Princeton University Press.
- Karniloff-Smith, A. (1979). Micro- and macro-developmental changes in language acquisition and other representational systems. *Cognitive Science*, 3, 91-118.
- Klahr, D., & Robinson, M. (1981). Formal assessment of problem solving and planning processes in preschool children. *Cognitive Psychology*, 13, 113-148.
- Laird, J. E. (1984). *Universal subgoaling*. Unpublished doctoral dissertation, Computer Science Department, Carnegie-Mellon University, 1984.
- Laird, J. E., & Newell, A. (1983). *A universal weak method* (Tech. Rep. No. CMU-CS-83-141). Pittsburgh: Computer Science Department, Carnegie-Mellon University.
- Newell, A. (1980). Physical symbol systems. *Cognitive Science*, 4, 135-183.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Piaget, J. (1976). *The grasp of consciousness*. Cambridge, MA: Harvard University Press.
- Simon, H. A. (1975). The functional equivalence of problem-solving skills. *Cognitive Psychology*, 7, 268-288.
- Spitz, H. H., & Borys, S. V. (1984). Depth of search: How far can the retarded search through an internally represented problem space? In P. H. Brooks, R. Sperber, & C. McCauley (Eds.),

952 Child Development

- Learning and cognition in the mentally retarded* (pp. 333–358). Hillsdale, NJ: Erlbaum.
- Spitz, H. H., Webster, N. A., & Borys, S. V. (1982). Further studies of the Tower of Hanoi problem-solving performance of retarded young adults and nonretarded children. *Developmental Psychology*, 18(6), 922–930.
- Sugarman, S. (1982). Developmental change in early representational intelligence: Evidence from spatial classification strategies and related verbal expressions. *Cognitive Psychology*, 14, 410–449.
- Wickelgren, W. A. (1974). *How to solve problems*. San Francisco: W. H. Freeman.