SESSION III POTPOURRI

JOHN W. COTTON, *University of California, Presider*

Theory of learning mathematics with computer-assisted instruction: Hierarchical and nonhierarchical options

JOHN W. COTTON, SANDRA P. MARSHALL, STANLEY VARNHAGEN,
and JOHN P. GALLAGHER
University of California, Santa Barbara, California 93106

The learning of arithmetic problems is assumed to be a Markov process involving conditioning of a set of k subskills, each consisting of one or more productions. An axiom set is provided, with the choice between two options for one axiom controlling which of two models results. Model 1, which assumes that every subskill is attempted on every occurrence of a given problem, is a nonhierarchical model. Model 2, which assumes that every subskill in the temporal sequence for a problem is attempted until one is failed, is a somewhat unconventional hierarchical model: It is hierarchical in the sense that conditioning or chance success at one level is prerequisite to performance of the next subskill (next level) in the problem. As the value of the guessing parameter, g , declines, Model 2 leads to less efficient learning than Model 1.

As computer-assisted instruction (CAI) systems increasingly come under the influence of artificial intelligence research (Brown & Burton, 1975; Collins, Warnock, Aiello, & Miller, 1975), the need for new learning models appropriate to those systems increases. These new models will be required to meet traditional goals of parsimony and predictive accuracy; they may prove to be similar to models developed in important series of CAI studies such as Stanford University's arithmetic (Suppes & Morningstar, 1972) and reading (Atkinson, 1974) projects. But a fundamental question must be how to incorporate learning assumptions into computer representations of tasks to be learned and into computer representations of a student's store of knowledge and procedures for performing those tasks.

This paper was prepared under Contract N00123-77-C-0087 from the Navy Personnel Research and Development Center, San Diego, California. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the policies, either expressed or implied, of the U.S. government. The authors thank James G. Greeno, Learning Research and Development Center, University of Pittsburgh, for consultation during the development of this article. All responsibility for the statements herein must lie on the shoulders of the authors, however. Reprints should be requested from John W. Cotton, Department of Education, University of California, Santa Barbara, California 93106. John P. Gallagher is now at Cleveland State University, Cleveland, Ohio 44102.

Suppes (1973) has provided one answer to this question in a recent discussion of the learning and performance of simple arithmetic. Performance is determined by the main programs and subroutines of theoretical devices called register machines, which correspond closely to digital computers. Learning is a Markov process for the development of additional subroutines or program steps. Suppes shows how a hierarchical organization of arithmetic skills can be modeled within the overall system of theory construction that he recommends.

An alternate approach is given by Waterman (1975), who states production systems for such tasks as addition in a form preventing solution until one production rule is called that generates new production rules as needed. Such a production system is called an adaptive production system and has learning capability built directly into its computer program. Waterman was not trying to simulate human arithmetic performance, however. If applied to humans, his addition model would prove deficient because, with Probability 1, it produces learning on the first occasion when a new production rule is needed. It may be desirable to introduce stochastic processes into Waterman's adaptive production systems for mathematics to make them behave as imperfectly as human learners. Or gradual improvement in performance may be generated within a deterministic system of production rules by replacing

unsuccessful productions by new possibilities on successive trials, as was done by Waterman (1976) in developing production systems for serial pattern acquisition. Because of the convenience of using stochastic processes to generate imperfect learning and performance by humans and because of an interest in the hierarchical learning processes discussed by Suppes (1973) and others, we have chosen the Suppes approach in attempting to state a general Markov model of learning simple mathematics. The alternative of building the learning process around adaptive production systems or some other deterministic learning computer model, such as Anderson's (1977) language acquisition system (LAS), is attractive and deserves continued research.

The model to be described here is intended to apply to remedial learning of arithmetic and algebra under a CAI system developed by Cotton, Gallagher, Hopkins, and Marshall (1978). Two special features required of that model are the capability to change predictions as a function of special events on certain training trials, and the capability to predict behavior on complex tasks composed of several subtasks.

The special events just mentioned are such things as presentation of one or more hints to a problem when requested by a student in certain variants of the CAI system. A complex task and its subtasks are illustrated as follows: Add two fractions by (1) finding a common denominator or probably the least common denominator, (2) converting each fraction to a new fraction with that denominator, (3) adding the numerators of the two new fractions and dividing by the new denominator, and (if necessary) (4) reducing the answer to a proper fraction or mixed number having the simplest possible form. The subskills required for performing these subtasks combine into an overall skill of performing the total task, that is, adding two fractions. Methods of identifying the components to be learned in a given situation are discussed in Cotton, Marshall, Varnhagen, and Gallagher (Note 1).

We present a set of axioms from which two closely related Markov models can be developed for the learning of mathematics tasks involving several subskills such as those just discussed. Motivation, interpretation, and application of these axioms follow.

Axiom 1. There are exactly k subskills to be learned in each problem from a class of problems in a training unit. The same k subskills apply to each problem, but the problems differ in the numerical values or (in the case of word and story problems) other arguments of the problems. For example, " $1/3 + 1/2 =$ " and " $2/3 + 2/5 =$ " would be in the same subclass if both were judged to require learning the same subskills, even though different numbers are involved. This judgment depends in part on the specific prerequisite skills assumed present at the beginning of a training period.

Axiom 2. All problems presented to students in one phase of an experiment are members of the class to be

mastered. Or any problems presented during one phase of an experiment that are not in the problem class specified in Axiom 1 have no effect upon the learning of the k target subskills (other than possible distribution of practice effects).

Axiom 3. In attempting to solve a problem, students proceed in the order Subskill₁, Subskill₂, . . . , Subskill_k. If they recognize irreparable failure on any subskill except Subskill_k, they nonetheless continue the problem by trying to perform the next subskill, inserting any arguments required to make remaining steps meaningful.

Alternate Axiom 3. In attempting to solve a problem, students proceed in the order Subskill₁, Subskill₂, . . . , Subskill_k. If they recognize irreparable failure on any subskill, they stop work on the entire problem and wait for the next trial of the experiment.

Conditioning axiom. There is a probability c_i of conditioning Subskill i on any trial at the beginning of which it is unconditioned and in which that subskill is attempted. Once conditioned, a subskill remains conditioned throughout successive trials of an experiment.

Response axiom. If a subskill has been conditioned prior to problem presentation, it will be performed correctly when a subproblem requiring it is encountered. Otherwise, such an encounter will lead to correct performance with probability g_i .

Independence axiom. All conditioning of and responding to different subskills actually attempted on a problem are independent of each other.

This axiom set has been chosen to permit prediction of the learning of complex tasks under fairly general conditions. Notice that Axiom 3 and Alternate Axiom 3 provide the basis for two quite different learning models. Axiom 3 is more traditional in assuming that every subskill is attempted on every trial, but Alternate Axiom 3 is appropriate to any situation in which success on certain subtasks is essential to success on later ones. This alternate axiom is intended to apply whenever the temporal ordering of subtasks and the structure of the overall task make it unclear what to do next if any subtask except the final one is failed. Educational psychologists will recognize these conditions as similar to but not identical to those defining learning hierarchies (Gagné, 1968).

Empirical as well as rational analysis is required to distinguish the conditions under which Axiom 3 is preferable to Alternate Axiom 3. Table 1 presents two temporally ordered tasks that might differ in this regard.

We would expect the learning of the Australian twist tennis service (Task 1) to be in accordance with Axiom 3 because failure to perform an early step in the service does not preclude attempting the next step. In contrast, finding a hypotenuse length (Task 2) is such that Alternate Axiom 3 might be required: If a student could not remember to square the lengths of the legs (Step 1), he or she would probably not be able to attempt to add the results together (Step 2). If unable

Table 1
Comparison of Two Tasks Differing in the Degree
to Which Failure in an Early Subtask Impedes
Performance of Later Subtasks

Task 1: Australian Twist Service in Tennis
Step 1—Change your grip of the tennis racket from the Eastern forehand grip by rotating the racket one quarter turn clockwise.
Step 2—Toss the ball up, somewhat more above your left shoulder than for a flat service.
Step 3—Strike the ball with your racket strings almost horizontal above the ball, moving the racket frontward and to the right.
Task 2: Finding the Hypotenuse Length of a Right Triangle from the Length of Its Legs
Step 1—Square the length of each leg in any convenient order.
Step 2—Add the results of Step 1.
Step 3—Take the square root of the result from Step 2.

to remember Step 2, he or she probably would not be able to attempt to find the square root of the result (Step 3). This difference is partly structural: One cannot find the square root of a nonexistent number. But it is also partly cultural: Little premium is placed by teachers or employers on solving later steps of mathematics problems by using contrived numbers or stating equations at points where the overall calculation has broken down.

Like all theoretical analysis, our comparison of Tasks 1 and 2 has been oversimplified. It is possible to think of procedures for teaching Task 2 with which Axiom 3 might seem more appropriate than Alternate Axiom 3. Therefore, it seems desirable to present a relatively full development of alternate models using the axioms given above and then to discuss experimental procedures that seem most likely to conform to alternate predictions.

Four simplifying assumptions will be made with each model: (1) The number of subskills, k , in each model is exactly three. (2) The learning parameter, c_i , will equal c , a constant for all i values. (3) The guessing parameter, g_i , will equal g , a constant for all i values. (4) All students start an experiment in a completely unconditioned state, U , with respect to these k subskills, as shown in Equation 3 of the next section. Relaxation of these special assumptions can be made without changing the general approach displayed below.

MODEL 1: INDEPENDENTLY ATTEMPTED SUBSKILLS

Model 1 uses Axiom 3 and thus assumes that success or failure on one subskill has no effect on attempts to perform any other subskill. Therefore, there is no effect on the learning or performance of any other subskill. Equation 1 (shown in Table 2) presents the transition matrix T (for the case of three subskills) implied by the axioms presented above when Axiom 3 rather than Alternate Axiom 3 is chosen. The state descriptions in this matrix may be interpreted as follows: C_{123} means that Subskills 1, 2, and 3 have all been conditioned. C_{ij} means that Subskills i and j have been conditioned and that the remaining subskill is still in the unconditioned state. C_i means that Subskill i has been conditioned and that both other subskills are still unconditioned. U means that all three subskills are still unconditioned. As is usual in such matrices, T has rows representing states on one trial (Trial n) and columns representing states on the next trial (Trial $n+1$). Thus the upper left-hand cell of T gives a probability of 1 of moving from C_{123} on Trial n to C_{123} on Trial $n+1$. The quantity c is the learning parameter from the conditioning axiom, under the new assumption that c_i is constant over all subskills.

Consider a complex task such as the addition of two fractions, discussed earlier in this article. When the student's progress on specific subskills is of interest, mathematical analysis based on Equation 1 is necessary. For the moment, however, we will focus on predicting the number of subskills conditioned, rather than their identities. By methods developed by Burke and Rosenblatt (1958) and summarized in Cotton (1971, p. 340), the eight-state model represented in Equation 1 reduces to a four-state model with the following transition matrix T^* and vector P_1 :

$$\begin{array}{ccccc}
 & C'_3 & C'_2 & C'_1 & U \\
 \begin{array}{c} C'_3 \\ C'_2 \\ C'_1 \\ U \end{array} & \begin{array}{c} 1 \\ c \\ c^2 \\ c^3 \end{array} & \begin{array}{c} 0 \\ 1-c \\ 2c(1-c) \\ 3c^2(1-c) \end{array} & \begin{array}{c} 0 \\ 0 \\ (1-c)^2 \\ 3c(1-c)^2 \end{array} & \begin{array}{c} 0 \\ 0 \\ 0 \\ (1-c)^3 \end{array}
 \end{array} \quad (2)$$

$$P_1 = \begin{array}{ccccc} & C'_3 & C'_2 & C'_1 & U \\ \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} & & & & \end{array} \quad (3)$$

Table 2
Matrix T (Equation 1)

	C_{123}	C_{23}	C_{13}	C_{12}	C_3	C_2	C_1	U
C_{123}	1	0	0	0	0	0	0	0
C_{23}	c	$1-c$	0	0	0	0	0	0
C_{13}	c	0	$1-c$	0	0	0	0	0
C_{12}	c	0	0	$1-c$	0	0	0	0
C_3	c^2	$c(1-c)$	$c(1-c)$	0	$(1-c)^2$	0	0	0
C_2	c^2	$c(1-c)$	0	$c(1-c)$	0	$(1-c)^2$	0	0
C_1	c^2	0	$c(1-c)$	$c(1-c)$	0	0	$(1-c)^2$	0
U	c^3	$c^2(1-c)$	$c^2(1-c)$	$c^2(1-c)$	$c(1-c)^2$	$c(1-c)^2$	$c(1-c)^2$	$(1-c)^3$

Let Equation 3 be the initial vector stating the probability of being in each of the possible states in the collapsed model (Equation 2). Here, C'_j means that any j subskills have been conditioned and any remaining subskills are still unconditioned. As before, U means that all subskills in the overall task are unconditioned. If T^* applies for every trial, as would be expected if a constant procedure is used, the state vector for Trial $n+1$ will be given by Equation 4:

$$P_{n+1} = P_n T^*, \quad (4)$$

and the recursion in Equation 4 implies that

$$P_{n+1} = P_1 (T^*)^n. \quad (5)$$

Equation 5, together with the fact that

$$P(\text{Correct}|\text{State}) = \begin{matrix} C'_3 & C'_2 & C'_1 & U \\ [1 & g & g^2 & g^3] \end{matrix} \quad (6)$$

can be used to predict the probability of a correct overall response on Trial n , once c and g (the guessing parameter of the response axiom under an assumption of constancy for all Subskills i) are known.

Methods presented by Kemeny and Snell (1976, Chapter 3) allow the prediction of the average number of trials spent in each state en route to C'_3 , the state in which all subskills have been conditioned. The fundamental matrix giving these entries as a function of rows representing states in which a subject might have started the experiment is given by Equation 7 for N :

$$\begin{matrix} C'_2 & C'_1 & U \\ C'_3 & \begin{bmatrix} [1-(1-c)^2][1-(1-c)^3] & 0 & 0 \\ 2c(1-c)[1-(1-c)^3] & c[1-(1-c)^3] & 0 \\ 3c^2(1-c)[(1-c)^2+1] & 3c^2(1-c)^2 & c[1-(1-c)^3] \end{bmatrix} \\ U & \end{matrix} \quad (7)$$

An important property of a fundamental matrix is that each of its rows gives the mean number of trials expected to be spent in each state represented by a column label before reaching the final state, C'_3 , assuming that the student began an experiment in the state for that row. For example, after division of each Row U entry by the denominator of Equation 7, one

finds that students are predicted to spend an average of $3c(1-c)^2/[1-(1-c)^2][1-(1-c)^3]$ trials in C'_1 , assuming a start in State U as implied by Equation 3 earlier. All students begin in State U , so that only the bottom row of N is needed. However, modification of Equation 3 would require use of additional rows of the fundamental matrix, N .

This concludes presentation of Model 1 in equation form. It is a nonhierarchical model whose implications will be compared with a hierarchical model, Model 2, after that model is outlined below. State notation for Model 2 is much like that for Model 1, with the addition that "+" or "-" indicates whether the last subskill response was correct or incorrect, respectively.

MODEL 2: HIERARCHICALLY SUCCESSFUL SUBSKILLS

Model 2 differs from Model 1 only in having Alternate Axiom 3 apply in place of Axiom 3. Since this axiom requires students to stop a problem as soon as an error is made on a subproblem, predictions from Model 2 very much depend on the analysis of behavior on specific subproblems. Consider, for example, the possibility of starting in State U (all subskills unlearned), conditioning one subskill, and being correct only on the first two subskills of a three-subskill problem. The probability of this combination of events is $2c(1-c)^2 \times g^2(1-g)$, because $2c(1-c)^2$ gives the probability of conditioning on the first and not on the next two subskills or conditioning on the second but not on the other subskills, while $g^2(1-g)$ gives the probability of guessing correctly on the first two subskills and incorrectly on the third. Remember that conditioning the first subskill on one trial did not guarantee a correct response before the next trial. Therefore, guessing correctly had to occur to produce a correct response.

It would be possible to represent the processes of Model 2 using an 8 by 8 transition matrix. However, since the next transition does not depend upon whether a correct or incorrect response is made on Trial n , we omit differentiating rows on the basis of that information. Table 3 (Equation 8) presents a 4 by 8 transition matrix that can be checked fairly easily by drawing one tree of possible conditioning and responding on successive subskills for each starting

Table 3
Abbreviated Transition Matrix T for Model 2: Theory Assuming Hierarchically Successful Subskills (Equation 8)

	C'_3+	C'_3-	C'_2+	C'_2-	C'_1+	C'_1-	$U'+$	$U'-$
C'_3	1	0	0	0	0	0	0	0
C'_2	cg	$c(1-g)$	$(1-c)g$	$(1-c)(1-g)$	0	0	0	0
C'_1	c^2g^2	$c^2g(1-g)$	$2c(1-c)g^2$	$2c(1-c)g(1-g)$	$(1-c)^2g^2$	$(1-c)(1-g)$	0	0
U	c^3g^3	$c^3g^2(1-g)$	$3c^2g^3(1-c)$	$3c^2(1-c)g^2(1-g)$	$3c(1-c)^2g^3$	$3c(1-c)^2g^2(1-g)$	$(1-c)^3g^3$	$(1-c)^3g^2(1-g)$
				$+c^2g(1-g)$		$+2c(1-c)g(1-g)$		$+(1-c)^2g(1-g)$
						$+c(1-g)$		$+(1-c)(1-g)$

state, calculating the probabilities of reaching each leaf of the tree, and combining probabilities as needed.

Methods like those used in developing Equation 2 turn Equation 8 into Equation 9 for T^* below:

$$\begin{array}{ccccccc}
 & C'_3 & C'_2 & C'_1 & U & & \\
 C'_3 & 1 & 0 & 0 & 0 & & \\
 C'_2 & c & 1-c & 0 & 0 & & \\
 C'_1 & c^2g & c(g-2cg+1) & (1-c)(1-cg) & 0 & & \\
 U & c^3g^2 & c^2g(2g-3cg+1) & c^2g^2(3c-4)+cg(g+1-2c)+c & c^2g^2(2-c)+cg(c-g-1)+1-c & (9) &
 \end{array}$$

As before, C'_j is the state in which any j of the three subskills are conditioned and all others are unconditioned. Equations 3-6 also apply to Model 2, permitting prediction of the probability of a correct response on any trial from P_1 or P_n and the present T^* .

Rather than showing an algebraic expression for the fundamental matrix of Model 2, we present the arithmetic value for the matrix, given the specific parameter values, $c = .1$ and $g = .5$. Otherwise, the approach to its solution is the same as mentioned in connection with Equation 7, yielding:

$$\begin{array}{ccccccc}
 & N & C'_2 & C'_1 & U & & \\
 N = & C'_2 & 10.00 & 0 & 0 & & \\
 & C'_1 & 9.66 & 6.90 & 0 & & \\
 & U & 9.66 & 6.50 & 6.05 & (10) &
 \end{array}$$

COMPARISONS OF MODELS 1 AND 2

An imperfect analogy from sports makes Model 1 like football, a game whose duration is unaffected by the performance of the two teams, and Model 2 like baseball, a game in which early failure at bat ends a team's side of the inning. Model 2 is called a model with hierarchically successful subskills because later subskills cannot be successfully performed unless earlier ones have been successful also. This makes Model 2 similar to the alternate route hierarchy (quasihierarchy) discussed by Cotton, Gallagher, and Marshall (1977). Note, however, that Model 2 is a subclass of such quasihierarchies only if no correct responses can be made by chance (i.e., only if $g = 0$). In such a case a student leaving State U can learn either Subskill₁ and Subskills 1 and 2, or Subskills 1, 2, and 3 in a single trial; a student leaving State C'_1 can learn either Subskill₂ or Subskills 2 and 3 in a single trial. No student can make an error on an early subskill and be correct on a later subskill. If $g \neq 0$, the learning of subskills still has the quasihierarchical property that no later subskill can be learned before an earlier one, but performance is also quasihierarchical, in contrast to the Cotton et al. (1977) model, which permits any sequence of correct and incorrect subskills to occur by chance.

By Equation 6 the probability of correct performance of the overall task is a function only of the state and the guessing parameter, g , regardless of whether

Model 1 or Model 2 is applicable. A comparison of performance under the two models may be based on use of Equation 5 with both models or on the basis of other statistics such as those in the fundamental matrix. For example, suppose we assumed that Model 1 held and that an average of 22.21 trials was required to move from State U to State C'_3 . Comparison with the fundamental matrix Equation 12 for Model 1 yields the implication,

$$\begin{aligned}
 & 22.21 \, c[1 - (1 - c)^2][1 - (1 - c)^3] \\
 & = 3c^2(1 - c)[(1 - c)^2 + 1] + 3c^2(1 - c)^2 \\
 & + c[1 - (1 - c)^2].
 \end{aligned} \quad (11)$$

Solution of Equation 11 yields

$$c = .08098. \quad (12)$$

Notice that in selecting a mean number of trials of 22.21 for Model 1 we equated the number of trials to that for Model 2 when $c = .1$ and $g = .5$. Equation 12 shows that Model 1 is more efficient than Model 2 in that a c only about 81% as large as that for Model 2 is required to learn in the same number of trials. Another way of making the same point is to note that 17.90 trials are required to learn with Model 1 if c is set at .1, the same value as just used with Model 2. The efficiency difference between the two models disappears if g is set at 1 so that Model 2 students never stop a problem before performing all three subtasks.

What happens to the efficiency difference with other changes in the value of g ? We state without proof that the general N matrix derivable from Equation 9 yields the following expected number of trials for passage from State U to State C'_3 (sum of all third-row entries in the N matrix):

Expected number

$$\begin{aligned}
 & = [(-3c^3g^3 + 6c^2g^3 + 6c^2g^2 - 4cg^3 - 9cg^2 \\
 & - 6cg + g^3 + 3g^2 + 4g + 3)c^{-1}] \\
 & \times [(-c^3g^3 + 3c^2g^2 + 2c^2g^2 - 3cg^3 - 4cg^2 \\
 & - 2cg + cg^3 + 2g^2 + 2g + 1)^{-1}]
 \end{aligned} \quad (13)$$

The sign of the partial derivative of this expected number as a function of g is not obvious once differentiation has taken place. However, search from $c = .01$ in steps of .02 to .43 and from $g = 0$ to 1 in steps of .05 shows monotonic declines in the expected number as a function of g for any fixed c in the range studied. Since c values larger than .43 are unlikely experimentally, we conclude that the efficiency of learning with Model 2

is, for all practical purposes, a monotonically increasing function of g .

IMPLICATIONS FOR SPECIFIC EXPERIMENTAL SITUATIONS

We now indicate how Models 1 and 2 might be tested with data from the remedial mathematics CAI system developed by Cotton et al. (1978). First, we ask the appropriate response level for use with these models; that is, how complex should be the subskills identified for use with these models? Cotton et al. (1978) have five modules operative: for negative numbers, factors and prime numbers, fractions, inequalities, and simple linear equations. Each module includes two or three levels of numerical problems, whose complexity increases with level number, plus word and story problems.

Two extremes of response definition would be to (1) let the response unit be activation of an individual production rule such as those of Cotton et al. (1978, Table 3, Figure 2), or (2) the occurrence of an overall answer to a problem, evoking feedback of "correct" or "incorrect" to the student making the response. The first option is almost surely too molecular, since many individual productions are limited in the size of their effects and could be replaced by other productions or groups of productions having equivalent effects. The second option may be appropriate in some situations. However, system-generated information (such as hint, compute, and solve), as well as a student's own examination of problem steps as performed, call attention to features of a problem that are intermediate between the two options. A workable unit of analysis may be the size of a single hint or solve step for a given problem, as defined in the information-structure-oriented (ISO) condition of the system.

Let us consider the problem " $1/2 \div (2/3 + 1-1/6) =$." This is a Level 2 fractions problem in that it has one nested term, $2/3 + 1-1/6$. A student who has adequate skills to attempt this problem can already solve the nested term. What parts of the solution to this problem are neither part of solving the nested term nor skills needed by persons who can solve Level 1 problems? Consider the following extracts from the SOLVE routine for this problem (Cotton, Gallagher, Hopkins, & Marshall, Note 2, Table F2): "[G1] This problem contains a nested term, $2/3 + 1-1/6$. It will be necessary to solve it first. . . . [G2] Because the numerator of $11/6$ is larger than the denominator, we could convert this fraction to a 'mixed' form. However, because we will need to operate on more terms in this problem, we will leave it in its 'improper' form. . . . [S1-2] We have reduced the problem from $1/2/(2/3 + 1-1/6) =$ to two terms. We may now solve $1/2/11/6$ directly. [G3] In order to divide $1/2$ by $11/6$, we must apply the 'invert and multiply' rule. That is, the second term must be

changed such that the numerator and denominator switch places. Then the two terms are multiplied. [S3] So, $1/2/11/6$ becomes $1/2 \times 6/11$."

These extracts have been labeled (in the present document only) G1 through G3 for goal-setting and S1-2 or S3 for goal-satisfying operations. It is our claim that the goal-satisfaction operations of reducing the problem to $1/2 \div 11/6$ and of reducing the problem further to $1/2 \times 6/11$ require no new skills once the goal-setting operations have been learned. Accordingly, the present problem may be thought of as requiring three goal-setting operations to be learned, assuming competence at Level 1.

Given this analysis, we are in a position to say that Models 1 and 2 could be applied directly to the class of problems from which this sample problem was drawn. For this class (which is presumably somewhat smaller than the class of all Level 2 fraction problems), k , the number of subskills to be learned, is 3; Axiom 2 permits analysis of data from this class as selected out from a broader set of problems actually presented to the students involved. Strictly speaking, one should say that a model with a much larger k is employed but with all but three of the subskills assumed conditioned before the current training.

LEARNING PARAMETER VALUES AS A FUNCTION OF TREATMENT

Heretofore, we have ignored experimental features that may affect the values of c , the learning parameter, or possibly of g , the guessing parameter. It is hypothesized that adding new features in addition to feedback as to the correctness or incorrectness of an answer will change c for that problem or problem part. For example, calling SOLVE in the ISO condition of Cotton et al. (1978) might increase c for each subskill being taught. Calling SOLVE in the ad hoc frame-oriented condition (AFO) should increase c a bit less because the information provided then is of general usefulness but does not apply exactly to the current problem. Note that c for SOLVE could actually be smaller than in the reference conditioning task if overt responses are more important than information presentation. It should be mentioned that conditioning is specific to a class of problems; although transfer may be expected from numerical problems to story problems, perfect transfer is not expected. Nor does the theory as stated tell how to predict the degree of transfer of this kind.

The foregoing remarks imply that c may vary from trial to trial, depending upon the treatment for that trial. If the treatments do differ from trial to trial, Models 1 and 2 must be modified as follows: Equation 4 uses different transition matrices, depending on the treatment. For example,

$$P_{n+1} = P_n T^*_{\text{solve}} \quad (4a)$$

for a trial in which SOLVE is called, or

$$P_{n+1} = P_n T^*_{\text{answer}} \quad (4b)$$

for a trial on which the specific answer is requested and obtained. If a student receives three trials, the first with conventional feedback, the second with SOLVE, and the third with ANSWER, the state vector for Trial 4 would result from successive applications of Equations 4, 4a, and 4b:

$$P_4 = P_1 T^*_{\text{solve}} T^*_{\text{answer}} \quad (13)$$

The effect of providing a hint is somewhat different from that of providing an answer or a detailed solution. The hint is typically a specification of a goal-setting production and therefore should strengthen this subskill represented by that production. For that one subskill, Subskill_i, the conditioning parameter, c_i , should be larger than for the other subskills of the problem. This takes us back to the general version of the conditioning axiom, allowing different c values for different subskills. Models 1 and 2 can be adapted to this condition, using methods like those just employed in state Equation 13. Notice that, if several hints are presented on one trial, several c_i values will be increased for that trial. Different T matrices for different training trials will differ only in the values of the c_i values employed.

It is tentatively assumed that the COMPUTE command does not directly affect the learning parameter. Using this command may speed learning by permitting performance of more problems in a fixed amount of time. However, it is assumed that proper use of COMPUTE reflects proper goal setting and is thus more an index of learning than an aid to learning.

Effects of Other Student Aids

The other aids available to students are expected to have indirect effects on learning. They are (1) skipping an item, (2) selecting the level of problems to be presented within a module, (3) selecting a new module, and (4) quitting a session prematurely. Skipping an item sets all c s for that item to zero for that trial and permits attempted learning of some other item instead. Whether or not this facilitates overall learning depends on whether the skipped item has already been learned, in which case a benefit results, or whether it was difficult and of such a nature that prior training of other items might cause learning of some of its components, increasing the c for the original problem type overall when it is next presented. The effects of quitting a session prematurely are assumed simply to be reduced learning because of a reduction in number of learning trials. However, a motivational effect of such quitting is also a possibility.

WHEN WOULD MODEL 1 BE MORE APPLICABLE THAN MODEL 2 OR VICE VERSA?

The conditions favoring Model 1 or Model 2 must be identified experimentally. However, anything that leaves later stages of solution in doubt following failure in early stages may be expected to favor Model 2. If SOLVE were completely removed from the CAI system in question, this would reduce the likelihood of success on such later stages. Exclusion of HINT would similarly favor Model 2. It also seems reasonable to assume that mathematical sophistication or special training in problem solving would favor Model 1 by making students more able to predicate later parts of a solution upon hypothetical results of applying some early subskill that they find they do not possess. For example, a sophisticated student asked to compute $Y = 3 \log X + .2$ for integer values of $X = 1-20$ might make a graph of $3X' + .2$ as a function of X' over some arbitrary range of X' , explaining that if someone would give him or her values of $X' = \log X$, the problem could be completed.

It may be noted that each model assumes that students behave quite unimaginatively. Expansion of these models might well do such things as allow a change of solution strategy attempted when one step of an initial strategy fails.

Finally, the close interrelation of subskills in mathematics predisposes students to behave according to Model 2, stopping a problem after perceived failure on any subproblem. Both teaching efficiency and conformity to Model 1 might be increased by developing a more intelligent HINT routine than that of Cotton et al. (1978), such that the computer could track the progress of a student on a problem and provide the required hint at any point of student difficulty. Such a modification of the present CAI system seems entirely feasible.

REFERENCE NOTES

1. Cotton, J. W., Marshall, S. P., Varnhagen, S., & Gallagher, J. P. *Theory relevant to a CAI system for remedial mathematics*. Draft of Navy Personnel Research and Development Center Technical Report, October 1978. (Under review.)
2. Cotton, J. W., Gallagher, J. P., Hopkins, M., & Marshall, S. P. *A multimethod generative CAI system for remedial mathematics*. (Tech. Note) San Diego, Calif: Personnel Research and Development Center, in press.

REFERENCES

- ANDERSON, J. R. Induction of augmented transition networks. *Cognitive Science*, 1977, 1, 125-157.
- ATKINSON, R. C. Teaching children to read using a computer. *American Psychologist*, 1974, 29, 169-178.
- BROWN, J. S., & BURTON, R. R. Multiple representations of knowledge. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding*. New York: Academic Press, 1975.

- BURKE, C. J., & ROSENBLATT, M. A Markovian function of a Markov chain. *Annals of Mathematical Statistics*, 1958, **29**, 1112-1122.
- COLLINS, A., WARNOCK, E. H., AIELLO, N., & MILLER, M. L. Reasoning from incomplete knowledge. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding*. New York: Academic Press, 1975.
- COTTON, J. W. A sequence-specific concept identification model: Infrastructure for the Bower and Trabasso theory. *Journal of Mathematical Psychology*, 1971, **8**, 333-369.
- COTTON, J. W., GALLAGHER, J. P., HOPKINS, M., & MARSHALL, S. P. A multimethod generative CAI system for remedial mathematics. *Behavior Research Methods & Instrumentation*, 1978, **10**, 203-212.
- COTTON, J. W., GALLAGHER, J. P., & MARSHALL, S. P. The identification and decomposition of hierarchical tasks. *American Educational Research Journal*, 1977, **14**, 189-212.
- GAGNÉ, R. M. Learning hierarchies. *Educational Psychologist*, 1968, **6**, 1-9.
- KEMENY, J. G., & SNELL, J. L. *Finite Markov chains*. New York: Springer-Verlag, 1976.
- SUPPES, P. Facts and fantasies of education. In M. C. Wittrock (Ed.), *Changing education. Alternatives from educational research*. Englewood Cliffs, N.J.: Prentice-Hall, 1973.
- SUPPES, P., & MORNINGSTAR, M. *Computer-assisted instruction at Stanford, 1966-1968*. New York: Academic Press, 1972.
- WATERMAN, D. A. Adaptive production systems. In *Fourth International Joint Conference on Artificial Intelligence, Proceedings*. Cambridge, Mass: MIT, Artificial Intelligence Laboratory, 1975.
- WATERMAN, D. A. Serial pattern acquisition: A production system approach. In C. H. Chen (Ed.), *Pattern recognition and artificial intelligence*. New York: Academic Press, 1976.