DOCUMENT RESUME

ED 027 754 EM 007 186

By-Smallwood, Richard D.
Optimum Policy Regions for Computer-Directed Teaching Systems.
Pub Date [68]
Note-33p.
EDRS Price MF-$0.25 HC-$1.75
Descriptors-Branching, *Computer Assisted Instruction, Computer Oriented Programs, *Computer Science, Decision Making, Educational Research, *Individualized Instruction, Instructional Technology, Learning Processes, *Mathematical Models, Program Descriptions, *Program Design, Programed Instruction, Programed Tutoring, Program Improvement, Programing, Program Planning

The development of computer-directed instruction in which the learning protocol is tailored to each student on the basis of his learning history requires a means by which the many different trajectories open to a student can be resolved. Such an optimization procedure can be constructed to reduce the long and costly calculations associated with straight-forward decision tree optimization calculation. In this procedure the decision logic acts on the basis of the student's history, including his most recent response. A quantitative representation of the purpose of the instruction and the costs of alternative routes is weighted by the probability of the student's following that route. This defines the maximum expected total utility for a student with a given history, the optimization equation. By then using this technique with general models of learning behavior and a branching network design, the optimum alternative at each level in the branching network can be stated. Using such an optimization policy, the total expected instructional cost is sixty to eighty per cent higher than the optimum policy. (BB)

OPTIMUM POLICY REGIONS

FOR

COMPUTER-DIRECTED TEACHING SYSTEMS

Richard D. Smallwood
Dept. of Engineering-
Economic Systems
Stanford University
Stanford, California

OPTIMUM POLICY REGIONS FOR COMPUTER-DIRECTED TEACHING SYSTEMS

I. INTRODUCTION

In the past few years several significant advances in computer-aided instruction have opened the way for an evolution toward more sophisticated educational systems. Perhaps this is the time for some consideration and reappraisal of the direction for this evolution. As I see it, the primary direction for much of the current work in computer-aided instruction is toward the provision of tools that will permit the implementation of essentially classical teaching heuristics. The end result of this line of research will be a set of tools that allow the construction of computer-based teaching systems that provide a faithful mimicry of classical teaching methods.

On the other hand, we might look to the physical sciences and technologies for another possible tack to take in this evolution of educational systems. For example, we could view the past advances in CAI as individual contributions to an expanding technology; that is, as incremental advances to a cohesive set of theoretical concepts, experimental methodologies, and practical tools that together add up to an educational technology. Thus, we can view the evolution of educational systems as centered about an educational technology with each new contribution having for its foundation the distilled essence of previous contributions, and in turn adding its own contribution to the state of the technology. If this technology is to grow and flourish, then it must be founded upon a quantitive science similar in substance to the scientific bases of other technologies. This implies the development and utilization of classes of mathematical models, optimization techniques, and other theoretical structures necessary for the growth of the technology.

The development of a technology typically arises from the repeated application of the experimental-theoretical cycle. That is, experiments to test and extend the present theoretical status of the technology are planned and conducted; and then these are followed by modifications to the theoretical structure based upon the results of the experiments. This paper will present a potential contribution to the theoretical side of this educational technology.

One of the crucial questions to be considered in the future development of computer-aided teaching systems is the extent to which the latent computational power of the machine can be used to make rational decisions on the course of the instruction. A system that included such a decision process in its operation might be termed "computer-directed" rather than "computer-aided" instruction. This discussion will focus on the development of a method for implementing such a decision process in a teaching system.

One of the discouraging problems encountered in a theoretical formulation of the decision process in a computer-directed teaching system has been the excessive computation required. If the decision process is to consider any significant number of future trajectories that the student might experience, then the computation time can become a significant limitation in operating the system. This paper describes a technique that involves a very small amount of computation time for implementing a truly optimum decision policy in a computer-directed teaching system. Furthermore, the results are applicable to a very large class of models of human learning. As mentioned above, these results represent only a theoretical contribution to the educational technology. The experimental testing and validation of these theoretical results are equally important -- and much more difficult to achieve; this, then, represents only the first step toward the solution of the problem.

II. DECISION MAKING IN CAI -- A FORMULATION

This section presents a general formulation of the decision problem in a tutorial computer-directed computer system. This formulation is not new; it has been described in one form or another previously[2,6,7]. It is presented here to provide a general perspective for viewing the results of succeeding sections.

The first question we should ask is: Why should decision processes be incorporated into CAI systems? The answer to this question follows from the natural desire to develop a teaching system that will detect and respond to the differences exhibited by individual students. Thus, we should like to design a decision logic (sometimes called a branching logic) into our CAI system so that the available past history of the student can be used in some meaningful way to influence the future course of the student's instruction. To begin then, let us imagine a hypothetical student with a particular history for whom a decision policy is required. This decision policy will be encoded into the computer teaching system, and will prescribe for the system what alternative instruction should be provided for this student and for other students with different past histories. The role of the past history in the decision process is extremely critical; for this quantity represents a parameterization of the available information about the student that will determine how well the system adapts to the individual learning characteristics of the student. We shall denote the past history of our hypothetical student by h and shall have more to say on this subject later.

The existence of the decision process within the teaching system implies that there must be a set of alternative courses of action available for dealing with our hypothetical student. Since this set of instructional alternatives will typically be dependent upon the student's current status, and furthermore since we have agreed to represent the student's current status in the form of his past history, h , then the set of available instructional alternatives for the student will be denoted by $A(h)$.

For each of the alternative presentations of the material there will typically be a question or set of questions to test the student's comprehension of the material. The student's responses to these questions provide additional information that we must incorporate into his past history to guide the future course of the instruction. Since each of the possible responses that might be elicited from the student will have a different impact upon the student's updated history, it is necessary to consider all possible responses explicitly. Thus we shall assume that for each possible instructional alternative there is a finite set of possible student responses, and we shall denote this set of responses by $R(a)$ where a is the particular instructional alternative from $A(h)$ that has been presented to the student.

With this representation of the instructional alternatives and student responses we can illustrate the complete decision problem with the decision tree shown in Fig. 1. As shown in that figure, for a student with a past history, h , there is a set of instructional alternatives each of which may produce a sample from the set of possible student responses. At the conclusion of this response there will be a new past history, h' , and

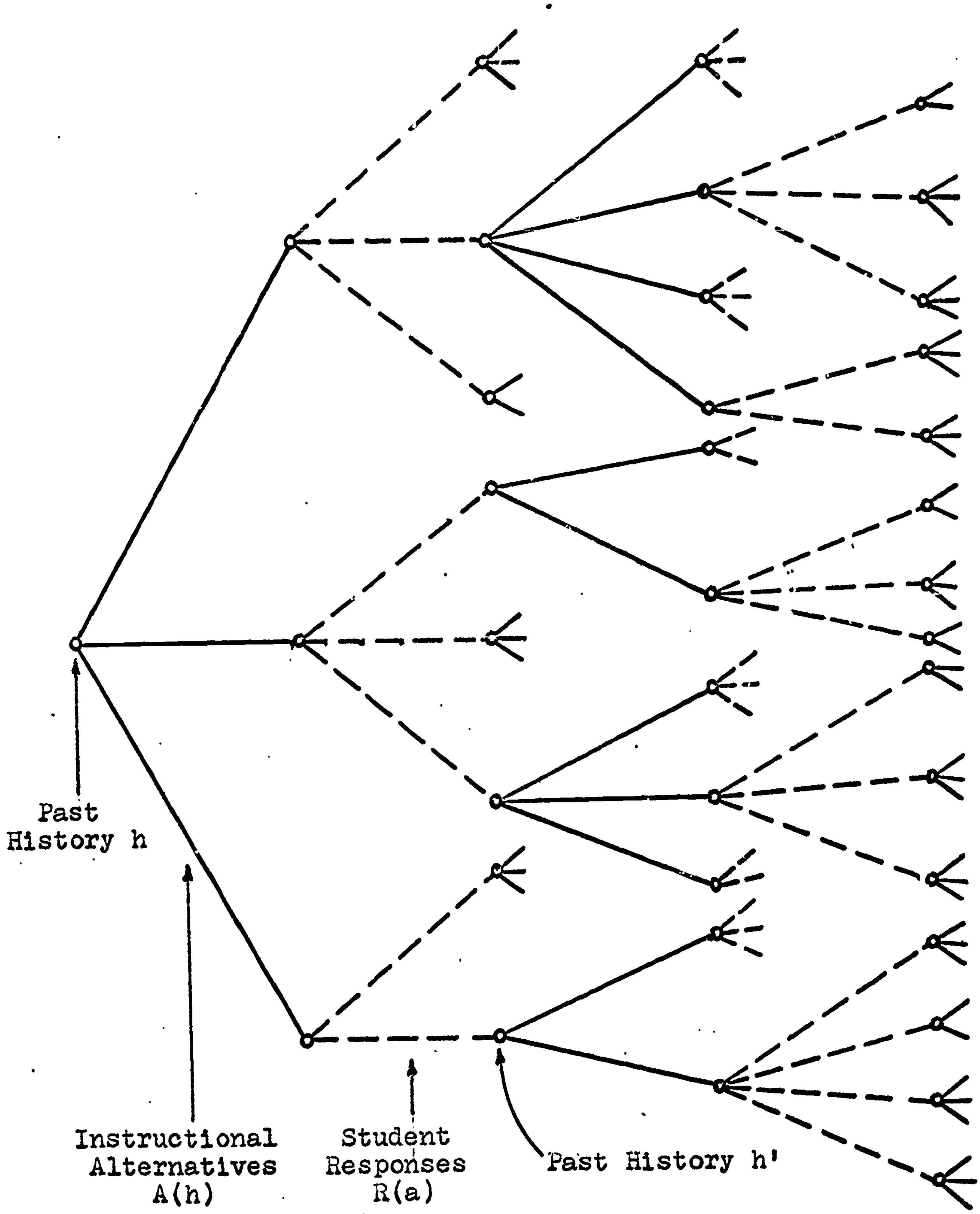

Figure 1 The Decision Tree

a new decision to be made; and this decision-response cycle may extend a considerable distance into the future until the instruction is terminated. The problem then is to calculate the optimum instructional alternative at each decision node taking into account the possible effects that this may have on the future course of the instruction. A brief look at Fig. 1 will show the tremendous number of possible student trajectories through the decision tree that must be considered if all possible paths are to be accounted for in the calculation. For example, if there are five instructional alternatives at each decision node in the tree and if there are two possible responses by the student for each instructional alternative, and if we desire to calculate the optimum instructional alternative based upon those paths by the student that extend ten presentations into the future, then this will require the consideration of ten billion possible student trajectories for each decision. This is clearly an infeasible solution to the problem. This paper will propose an alternative way of viewing this decision problem that will eliminate all but the most trivial of calculations for each decision in the course of a student's instruction.

To continue with the formulation of the decision process, the selection of one of the instructional alternatives at a decision point requires a criterion for appraising the relative value of each alternative. More explicitly, we shall need a quantitative representation of the purposes of the instruction as well as the relative costs of alternative presentations. For our purposes here we shall assume the existence of an utility function, $u_a(k,h)$ that specifies the immediate value that is accrued if alternative a is presented to a student with past history h and the k^{th} response is elicited. This function describes the immediate rewards (or costs) that are associated with each particular stage in the decision tree of Fig. 1. There is the additional question of the terminal rewards (or costs) that are accrued by terminating the instruction with the student in a particular status. For this purpose we assume the existence of a terminal utility function, $u_o(h)$, that describes the utility associated with terminating the instruction for a student with a past history h . A particular example of such a utility structure will be illustrated in Section IV.

In considering the different possible student trajectories in Fig. 1 we must weight the utilities associated with each trajectory by the probability that the student will in fact traverse that path in the decision tree. This requires a model of student behavior that allows the calculation of the probability that a student will produce a particular response to the presentation of an instructional alternative. Thus we assume the existence of a mathematical model for calculating the probability, $p(k|h,a)$, that a student with past history h who has been presented instructional alternative, a , will respond with the k^{th} response (where $k \in R(a)$ and $a \in A(h)$).

These definitions lead directly to an equation that defines the maximum expected total utility $v(h)$ that can be achieved for a student with a particular past history, h . To write this equation, consider all possible responses that a student might produce for a particular instructional alternative. For each such response there will be an immediate utility that is accrued plus the contribution from all future instruction that will follow with the updated past history, h' . Thus a recursive equation for the expected total utility is:

$$v(h) = \max_{a \in A(h)} \left[\sum_{k \in R(a)} p(k|h)[u_a(k,h) + v(h')] \right] \qquad (1)$$

In this equation we assume the existence of a rule for updating the student's past history; that is, h' represents the past history associated with the student who had a past history, h, was given alternative a and who responded with the k^{th} response.

The formulation of the decision problem represented by Eq. (1) is a typical dynamic programming recursive equation. Previous works have formulated the decision process in a computer-directed teaching system in a similar way.[2,7] In particular, the reader is referred to the excellent review article by Groen and Atkinson[2].

The implementation of an actual teaching system with a decision process based upon this formulation was attempted in 1961[6]. For this simple system the number of instructional alternatives at each decision point ranged from one to four, while the number of possible responses ranged from two to five. In terms of the decision tree in Fig. 1, the calculation of the optimum alternative at each decision node was carried out by extending the calculations in Eq. (1) three stages into the future. The weakest component in that early system was the mathematical model used for the calculation of student response probabilities. Also, the particular choice of past history parameterization for the student was very simple and did not realize the full capabilities of the system. In the next section, we shall consider a very general class of models that might be used for describing student learning behavior. The incorporation of this class of models into the decision process will alleviate many of the shortcomings of that earlier system.

III. A CLASS OF MODELS

The first step in formulating a model is to attempt an explicit description of our intuitive understanding of the phenomenon. One such description of the instructional process defines it as the systematic attempt to change the student's internal state of knowledge about the material being presented. Suppose now that it were possible to describe these internal knowledge states as a finite number of entities each of which represents one possible internal state of knowledge that a student may occupy during his course of instruction on the subject material. We shall refer to these entities as states, and it seems reasonable at this point to assume that they are mutually exclusive and exhaustive.

Within the limits of this representation, the instructional process can be viewed as the selection of alternative mechanisms for causing a student to make transitions from one internal state to another. These transitions will seldom be deterministic; that is, a particular instructional alternative will generally only cause a transition from one state to another state with a certain probability. Thus, we define as a parameter of the model the quantity $t_{ij}(a)$; this is the probability that a student occupying the i^{th} state will take the transition to the j^{th} state if he is presented with the material associated with instructional alternative a .

With this description for the influence of instructional material upon a student's internal state of knowledge, the question arises: How can we gain access to information concerning the internal state of the student? The mechanism for accomplishing this, of course, is to ask the student questions, the answers to which will depend upon the student's internal state of knowledge about the material. Thus if we assume that there is a discrete set of responses that a student will give for a particular instructional alternative, a , then we can model the relationship between the student's internal state and his response. For this purpose we define the probability

$r_{jk}(a)$ that a student who is presented instructional alternative a and who is presently occupying the j^{th} internal state will give the k^{th} response to the question associated with the presentation of the material. There is an explicit assumption in this definition that the student's response is dependent only upon his internal knowledge state. Figure 2 is a graphical representation of this class of models.

Now we must consider what additional parameters of the student's past history should be incorporated into the decision process as a result of this model. If we somehow were given access to information concerning the true state of the student, then this would be a very valuable component in the parameterization of the student's past history. Since we seldom, if ever, have perfect information about the student's state, the logical component for the student's past history is the current state of information about the student's internal knowledge state. We can represent this state of information as a set of probabilities, $[\pi_1, \pi_2, \ldots]$, where π_i is the probability that the student presently occupies the i^{th} state. If this set of probabilities is included as a parameter of the student's past history, then we can visualize this set of numbers changing as the student is presented with various instructional alternatives throughout the course of his instruction and as his responses to various questions are used to update the state of information about his progress.

If a model of the type presented here is to be used in the decision process in a teaching system, that is, if a model of this type is to be used in calculating $v(h)$ in Eq. 1, then two analytical results are required from the model. The first of these is a procedure for calculating the response probability, $p(k|h,a)$, and the second is the mechanism for

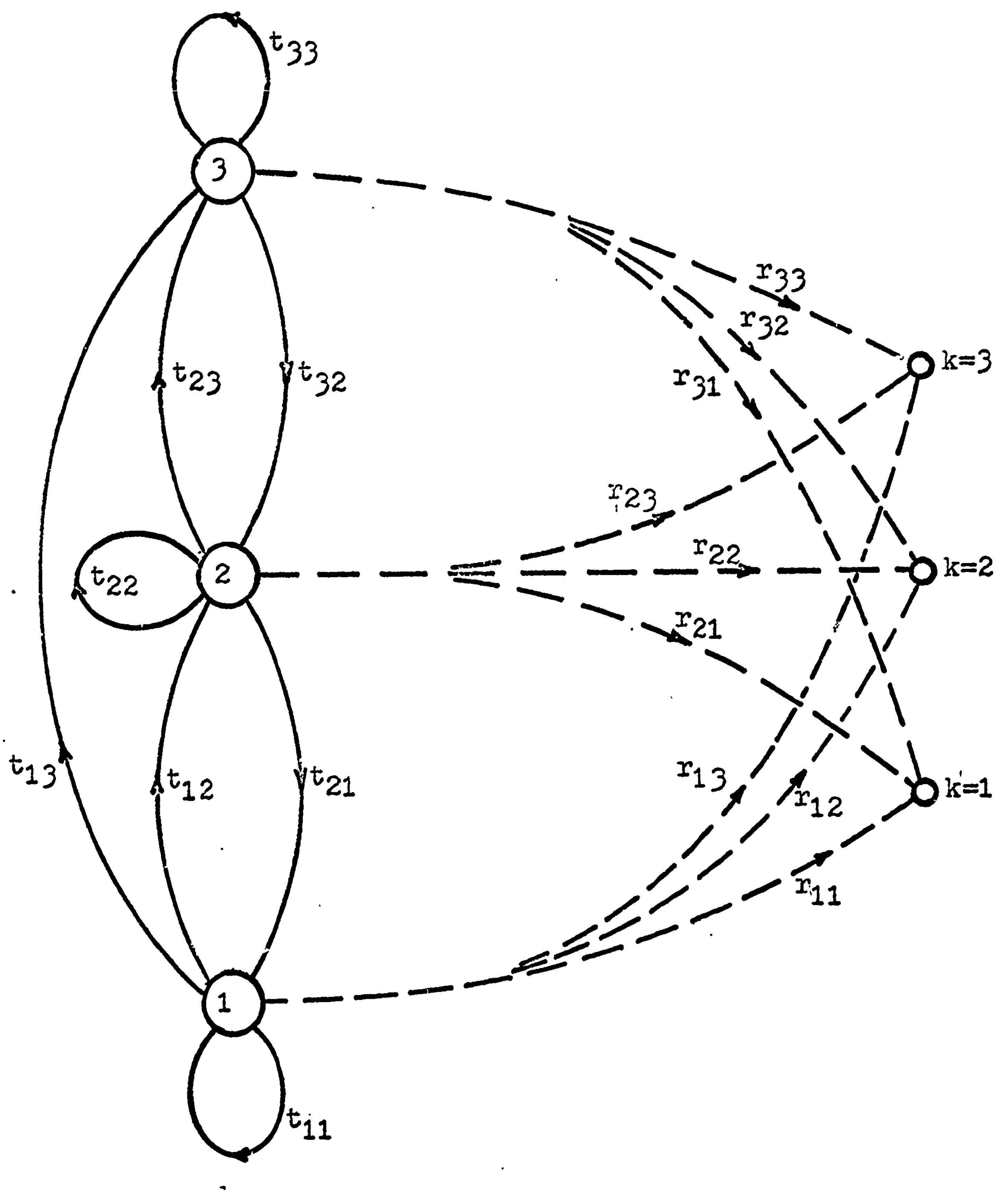

Figure 2 A Three State-Three Response Model

updating the past history $h = [\pi_1, \pi_2, \ldots]$ as a result of presenting an ins'.ructional alternative and observing a particular student response.

To describe the answers to these two demands let us assume that for a particular student with past history $h = [\pi_1, \pi_2, \ldots]$ we are considering presenting instructional alternative a. This instructional alternative will consist of some simple textual material followed by a question to test the student's comprehension of the material. We shall further assume that any transitions of the student's internal state occur prior to his response*.

We shall consider the response probability, $p(k|h,a)$, first. This quantity can be easily calculated by considering all possible states that the student might occupy after presentation of the textual material. The application of el.mentary probability operations yields for this quantity:

$$p(k|h,a) = \sum_i \sum_j \Pr\{\text{prior state} = i, \text{ succeeding state} = j, k^{\text{th}} \text{ response} | h, \text{ give alternative } a\}$$

$$= \sum_i \sum_j \pi_i t_{ij}(a)\, r_{jk}(a) \tag{2}$$

The procedure for calculating the updated state probabilities, $[\pi_1', \pi_2', \ldots]$, can be derived in a somewhat analogous way. Let us suppose that a particular student with a past history $h = [\pi_1, \pi_2, \ldots]$ has been given instructional alternative a and has given the k^{th} response tc the

---

*This is the so called pre-response transition case(7). Similar results can be easily calculated for the post-response transition case in which state transitions occur after the student's response.

question associated with that alternative. The updated state probability, π'_j , can be written through a simple application of Bayes' rule plus some elementary probability operations as:

$$\pi'_j = \sum_i \Pr\{\text{prior state} = i,\ \text{succeeding state} = j \mid k^{th} \text{ response},\ h,\ \text{give alternative } a\}$$

$$= \frac{\sum_i \Pr\{\text{prior state} = i,\ \text{succeeding state} = j,\ k^{th} \text{ response} \mid h,\ \text{give alternative } a\}}{p(k \mid h, a)}$$

$$= \frac{\sum_i \pi_i t_{ij}(a)\ r_{jk}(a)}{\sum_{i,j} \pi_i t_{ij}(a)\ r_{jk}(a)} \tag{3}$$

Thus this class of models provides a very simple mechanism for calculating the response probabilities as well as updating the past history. In the next section we shall show how this model can be easily incorporated into the optimum decision calculation of Section II.

IV. THE OPTIMIZATION PROBLEM

In a tutorial computer-aided teaching system it is often desirable that each student be exposed to certain basic information even though the actual presentation of this information may take on many forms. The general branching network shown in Fig. 3 illustrates a very general and flexible technique for achieving this result. In the general branching network each student starts the instruction at the first level. On the basis of the initial evaluation of the student's past history, one of the instructional

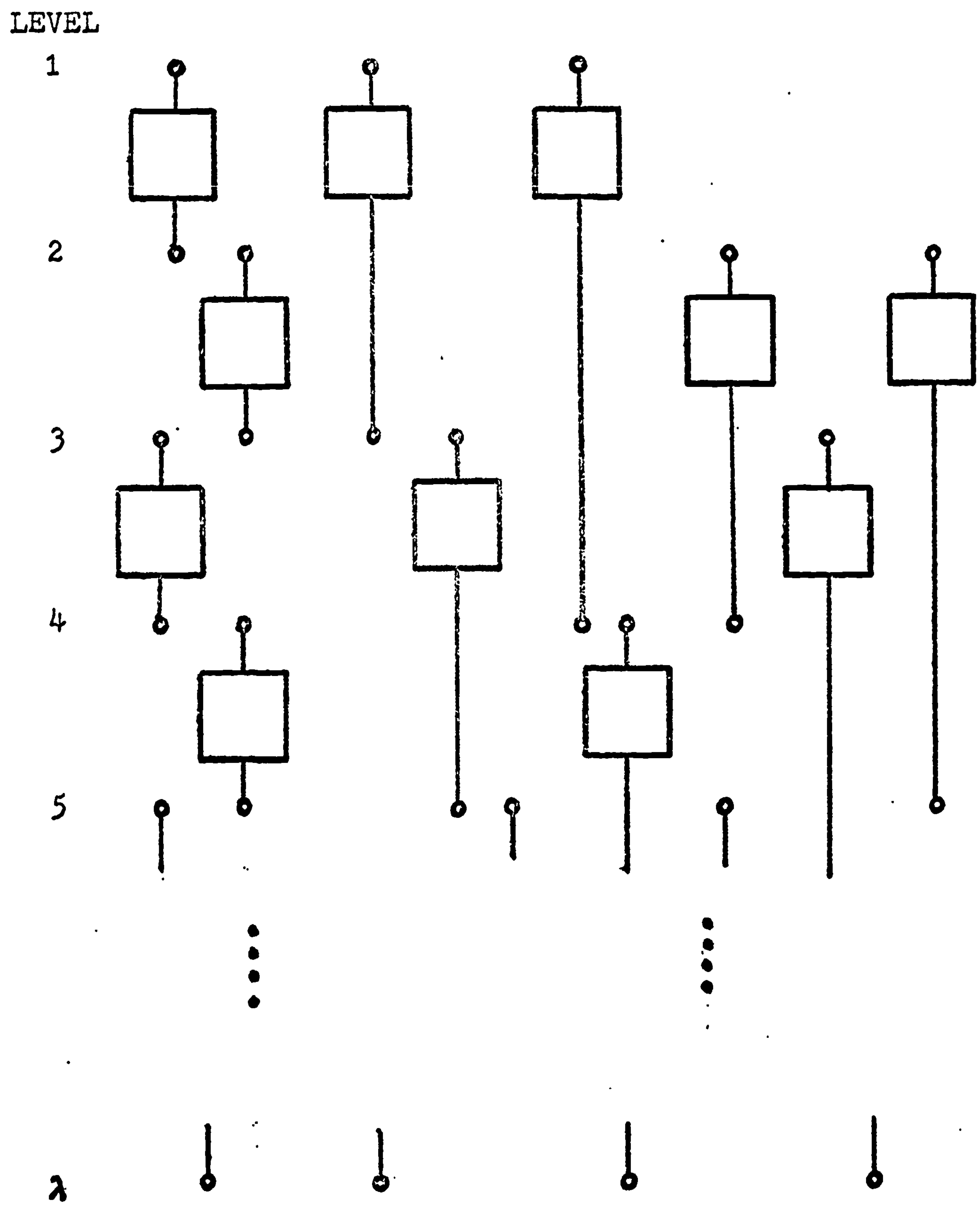

Figure 3 The General Branching Network

alternatives leaving the first level is presented to the student. Each of these instructional alternatives will be assumed to consist of a presentation of some textual material followed by a question designed to test the student's comprehension of the material. If the student responds with the correct answer to this question then he is placed at the final level corresponding to that alternative. On the other hand, if the student's response is not correct, then we shall assume the existence of some assignment rule that places the student at some level appropriate to that response. In other words, we shall assume the existence of a function $\ell(m,n,k)$ that determines the next level for a student who responded with the k^{th} response to the n^{th} instructional alternative leaving the m^{th} level. Once this student has been assigned this new level, then of course a new decision calculation must be carried out to determine which of the instructional alternatives leaving the student's level should be presented next.

Given such a general branching network for a set of subject materials, it seems feasible that one of the models discussed in Section III might very well describe the student's learning dynamics while progressing through the instruction defined by the various alternative presentations in the branching network. Thus, let us assume that such a model does indeed exist and that there is a set of transition probabilities, $t_{ij}(a)$, and response probabilities, $r_{jk}(a)$, for each of the instructional alternatives in the branching network -- that is, for each of the blocks in Fig. 3. The problem then is to use the optimization procedure in Eq. 1 with this formal structure to calculate, on the basis of the student's past history, the optimum alternative at each level in the branching network.

To accomplish this task we must define the student's past history. For the general branching network of Fig. 3 and a mathematical model of the form in Section III the appropriate parameterization of the student's past history is his current level in the general branching network and the current state probabilities. In other words, we let $h = [m, \pi_1, \pi_2, \ldots]$ where m is the student's current level in the general branching network. For a student at the m^{th} level who has been presented the n^{th} instructional alternative leaving that level and who has responded with the k^{th} response, the updated past history is $h' = [\ell(m,n,k), \pi_1', \pi_2', \ldots]$ where π_j' is calculated according to Eq. 3.

The one remaining component for the optimization is the utility structure. One reasonable description of a utility structure, and the one that will be used here, defines a presentation cost for each of the blocks in the general branching network and also defines a terminal cost that is dependent upon the student's terminal state when he finishes the instruction at the last level. Thus, we define the presentation cost for the n^{th} instructional alternative leaving the m^{th} level as c_{mn}.* The terminal cost at the conclusion of the instruction is just $\sum_i \gamma_i \pi_i$ where γ_i is the cost of terminating the instruction with the student in the i^{th} state. Since this utility structure has been postulated in terms of cost rather than values we must transform the value formulation of Eq. 1 into a cost formulation. This is easily done by multiplying that equation by (-1) and replacing the "max" by "min". For this cost formulation we can define the quantity $w_m(\Pi)$ as the total expected optimum cost for a student who

*This presentation cost can also be made dependent upon the student's response with no loss in applicability of the results. This generality will not be included in this section for the sake of notational convenience.

is at the m^{th} level and whose vector of state probabilities is $\Pi = [\pi_1, \pi_2, \ldots]$. The substitution of these definitions into the general formulation of Eq. 1 yields the following recursive equation for this more specific problem:

$$w_m(\Pi) = \min_n \left[\sum_k p(k|h,n)[c_{mn} + w_\ell(\Pi')] \right]$$

$$= \min_n \left[c_{mn} + \sum_k p(k|h,n) w_\ell(\Pi') \right] \quad (4)$$

In Eq. 4 the subscript ℓ is the assignment function $\ell(m,n,k)$ and the elements of the updated probability vector Π' are calculated from Eq. 3. The cost associated with the terminal level in the branching network is of course just

$$w_\lambda(\Pi) = \sum_i \gamma_i \pi_i \quad (5)$$

where λ is the last level in the branching network.

Appendix A uses the formulation in Eqs. 4 and 5 to show that the quantity $w_m(\Pi)$ can have the following relatively simple form

$$w_m(\Pi) = \min_n \min_i \left[\sum_j \alpha_{nij}^{(m)} \pi_j \right] \quad (6)$$

where n ranges over the set of instructional alternatives leaving the m^{th} level and i is simply an integer valued index for each instructional alternative. With this simple expression for the minimum expected cost, the optimum decision policy for all student past histories can be written very simply:

Select the instructional alternative, n, for which the quantity $\min_i \left[\sum_j \alpha_{nij}^{(m)} \pi_j \right]$ is minimum . (7)

Once the values for $\alpha_{nij}^{(m)}$ have been calculated, the implementation of this decision policy is very simple. The extensive searches throughout the decision tree have been eliminated through the prior calculation of a set of optimum policy regions that uniquely determine the optimum policy as a function of the student's past history.

Appendix B describes an iterative technique for calculating the values of the α coefficients in Eq. 6.

To test out these ideas a simple but non trivial example was constructed and the iterative technique of Appendix B was used to calculate the optimum policy regions. The mathematical model that was used is the simple two state model shown in Fig. 4. As can be seen, this model has only two parameters associated with it, the single transition probability, t, and the single response probability, r. This is the simple one element model that has been considered so extensively in the literature(1,3,4,5,7). The "zero" state in this model is generally associated with the unconditioned or unlearned state, and the "one" state with the conditioned or learned state. There are two parameters for this model; the transition probability t is the probability that a student in the "zero" state will make the transition to the "one" state on a particular presentation of the instructional alternative,

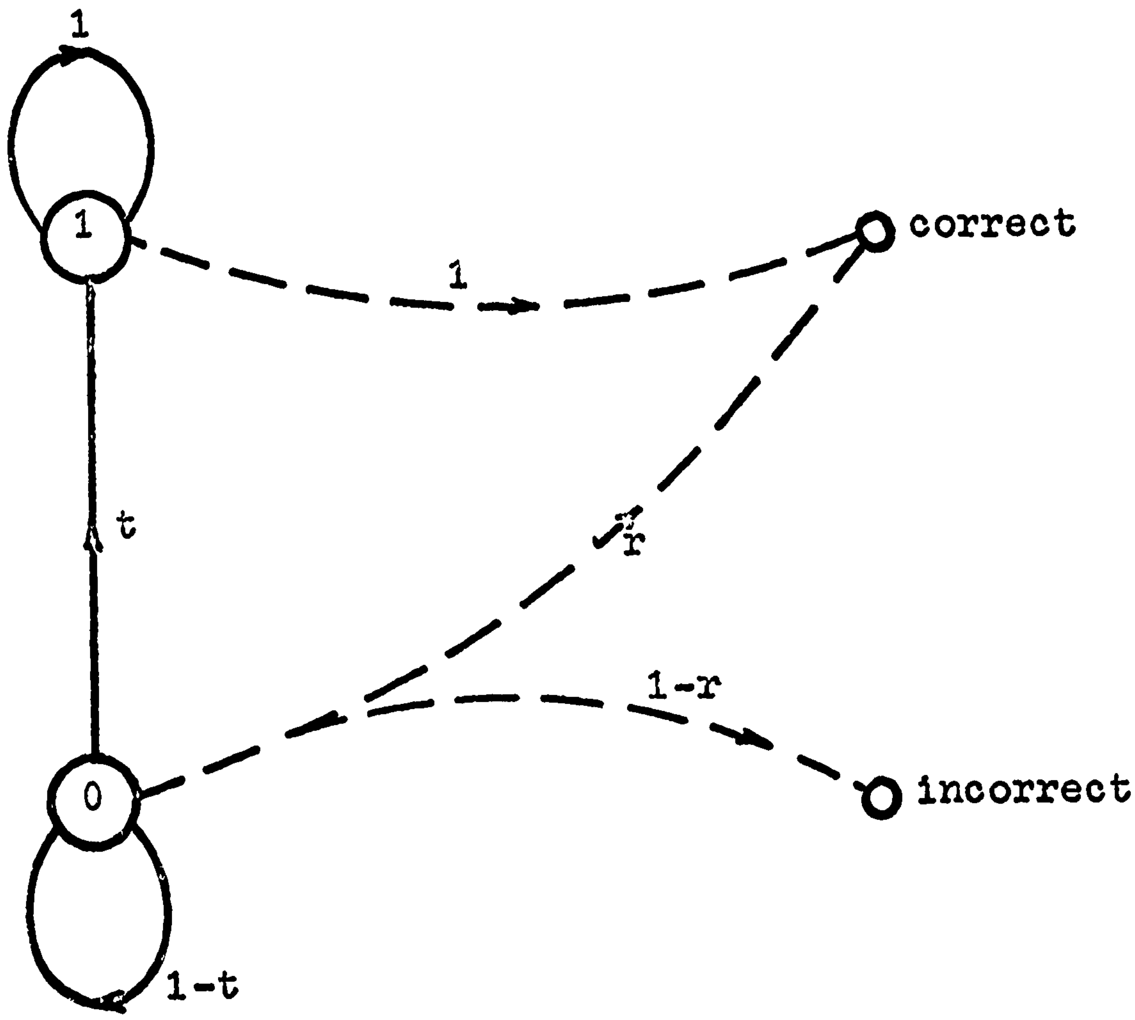

Figure 4 The One Element Model

and the response probability, r , is the probability that a student in the zero state will still respond with the correct answer (this is often referred to as the "guessing probability").

Figure 5 shows the sixteen level branching network that was used for the example. In this figure, the values for the transition probability, t , and the presentation cost c_{mn} are shown within the rectangle representing that instructional alternative. The outputs from each block that exit from the side of the rectangle represent the level assignment function for incorrect responses to the question associated with that instructional alternative. The response probability, r , was equal to 0.2 for all of the alternatives. The terminal costs, γ_0 and γ_1 , were set equal to 30 and 0, respectively. (There is an interesting physical interpretation for the quantity γ_0 in this formulation of the problem. This quantity is simply the maximum amount that we are willing to pay in order to achieve the transition of a student from the zero state to the one state.)

When the iterative procedure described in Appendix B was applied to this problem, approximately 11 iterations were necessary for convergence of the optimum policy regions. This optimum policy is shown in Fig. 6. The optimum policy region for each of the instructional alternatives is plotted as a function of the state probability, π_1 . Some typical trajectories that students might take through the general branching network are also plotted.

It is interesting to consider the speed of convergence of the iterative process. Figure 7 shows the total expected instructional cost starting at the first level for several of the decision policies that were calculated during the 11 iterations. As can be seen the iterative process converges

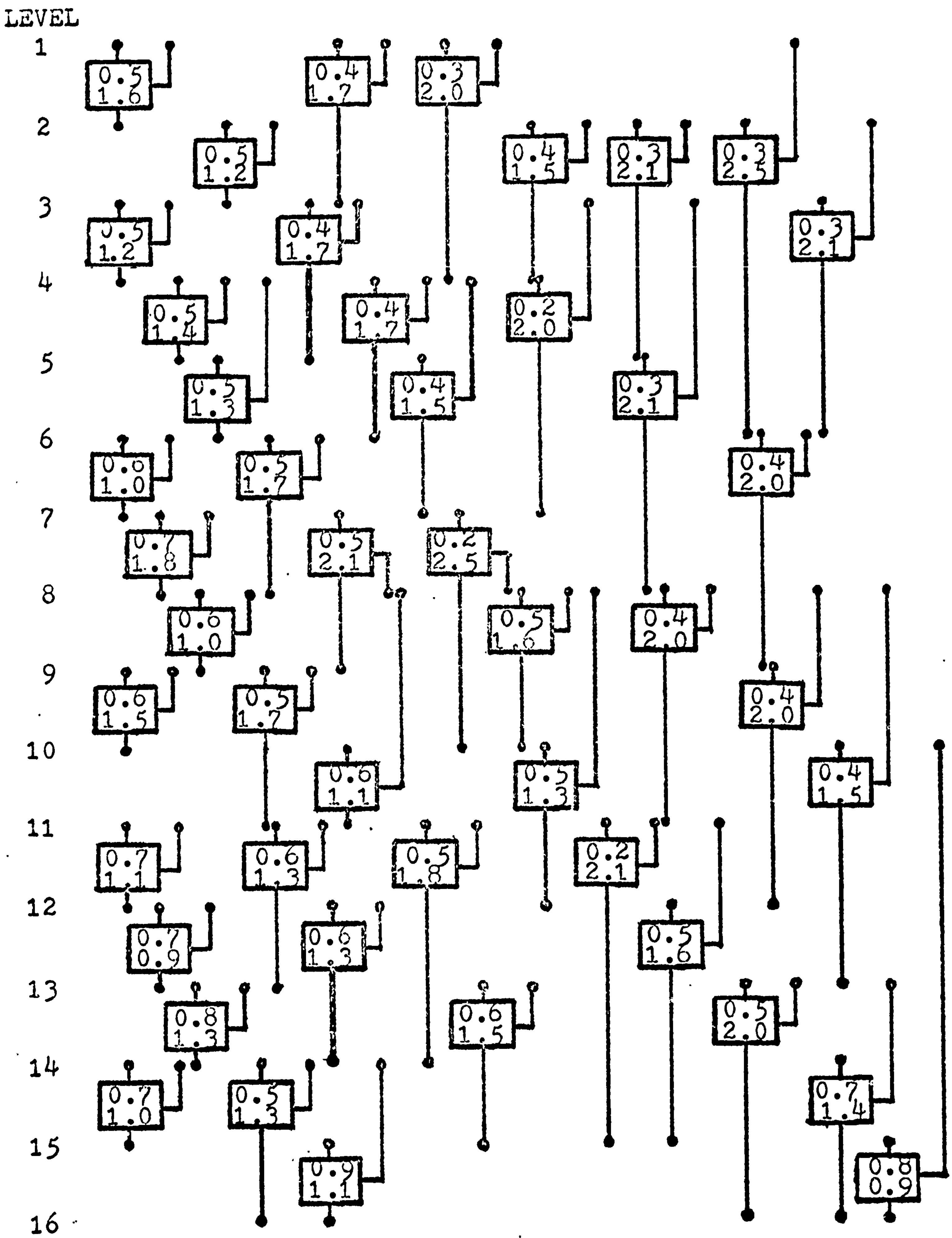

Figure 5 The Branching Network for the Example

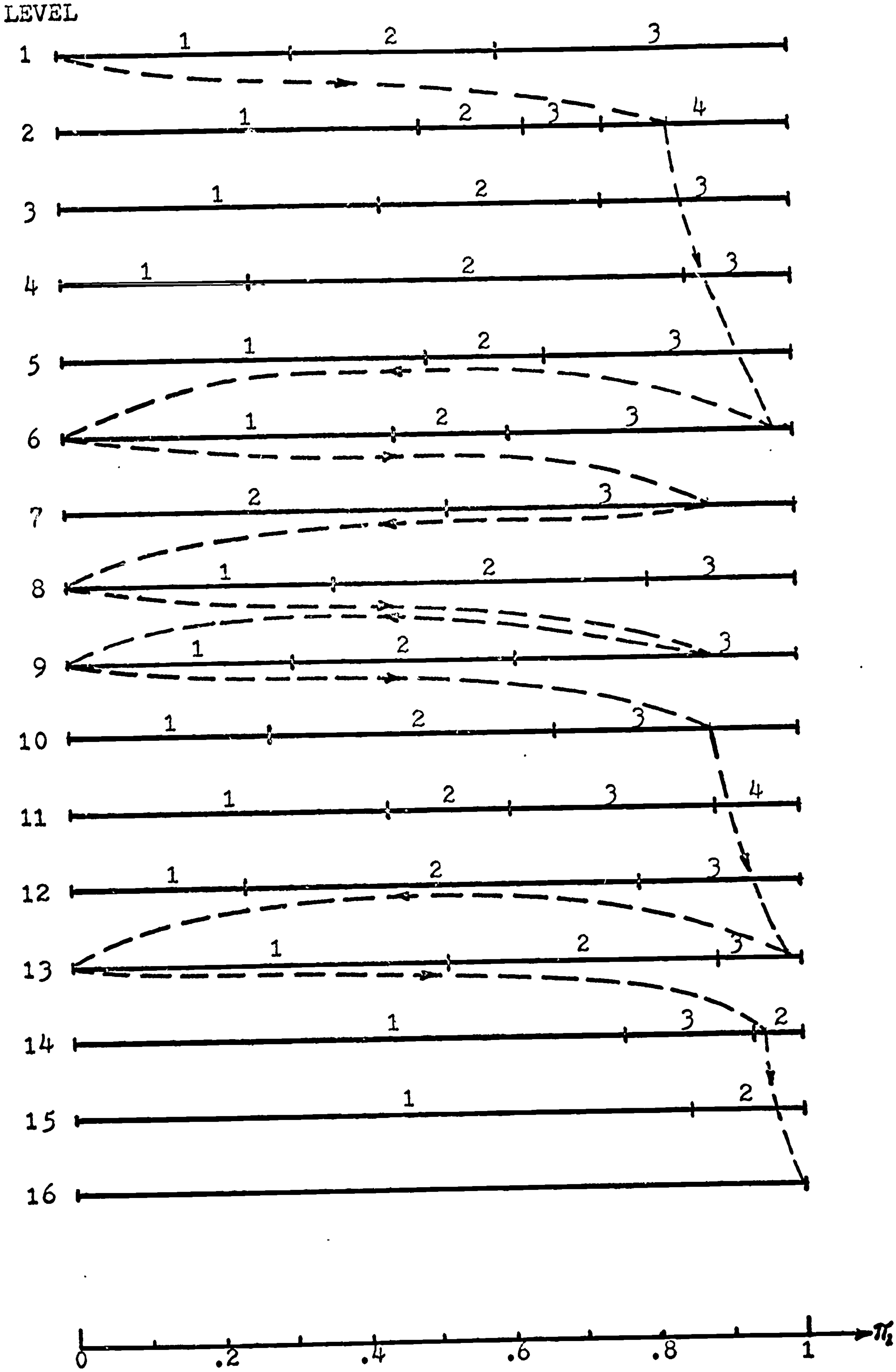

Figure 6 Optimum Policy Regions for the Example

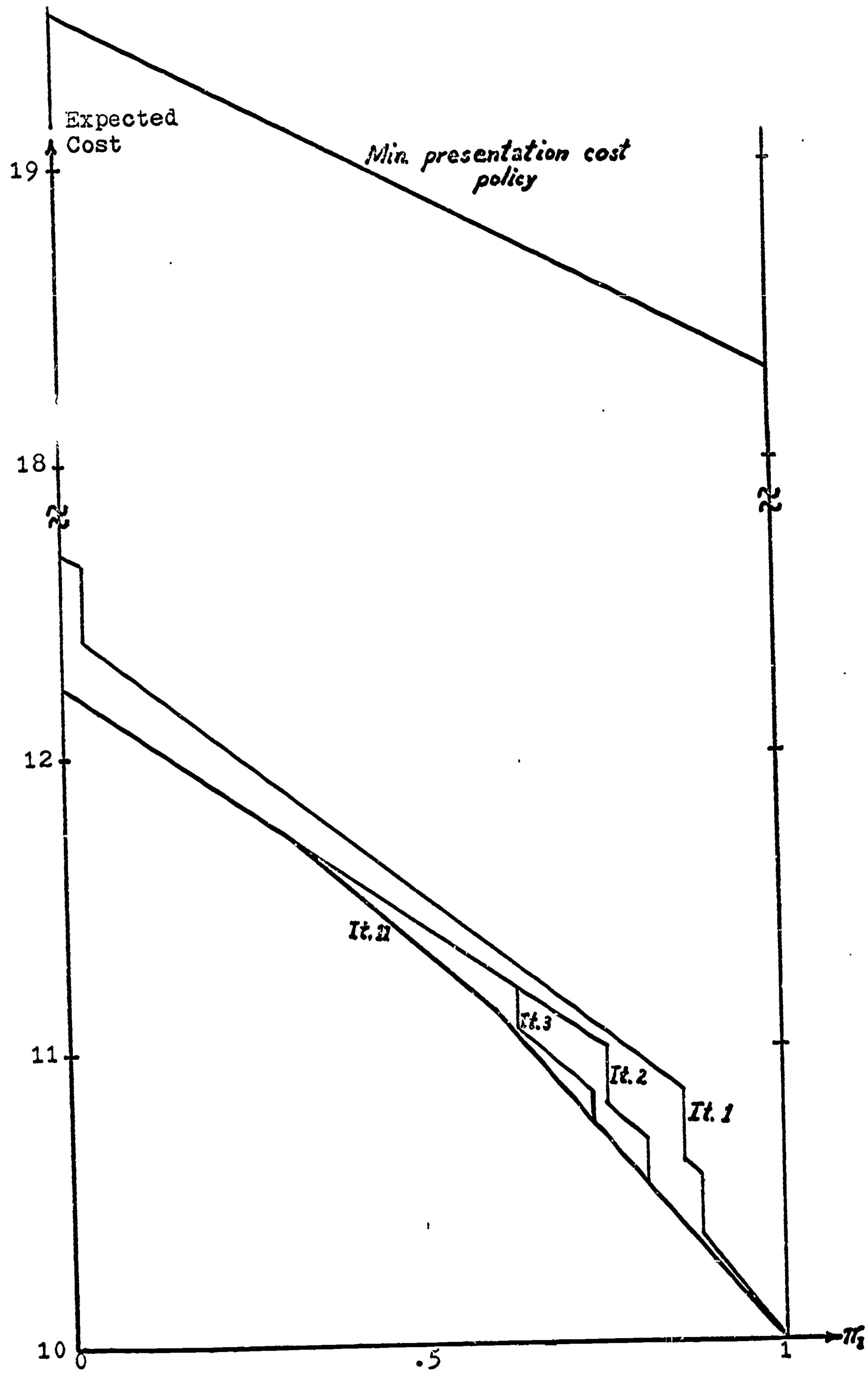

Figure 7 Expected First Level Instructional Cost for Several Iterated Decision Policies

quite rapidly in terms of the total expected cost function. As an illustration of the efficacy of such an optimization procedure, Fig. 7 also shows the total expected cost for a student for whom the minimum presentation cost alternative is always chosen. As illustrated in Fig. 7 this policy results in a total expected instructional cost that is sixty to eighty per cent higher than the optimum policy.

V. SUMMARY AND CONCLUSIONS

As indicated in the introduction to this paper, the results presented here only represent the first (and probably the easiest) step in an evolutionary sequence of theoretical-experimental advances to the educational technology. This paper presents an optimization procedure for a general class of learning models; the procedure essentially eliminates the tedious costly calculations associated with a straight-forward decision tree optimization calculation. Hopefully, later contributions to the educational technology will explore some of the experimental implications of these results. Specifically, much work remains to be done on the validation of models and more experiments must be conducted to test the efficacy of optimum decision processes in computer-directed teaching systems. The potential benefits of educational systems that truly adapt to the individual learning characteristics of the students will justify the allocation of future research resources toward these goals.

VI. REFERENCES

1. Atkinson, R.C., Bower, G.H. & Crothers, E.J., An Introduction to Mathematical Learning Theory, New York, Wiley, 1965, Chapter 3.

2. Groen, G.J. & Atkinson, R.C., "Models for Optimizing the Learning Process", Psychological Bulletin, 66:309-320, Oct. 1966.

3. Karush, W. & Dear, R.E., "Optimum Stimulus Presentation Strategy for a Stimulus Sampling Model of Learning," Journal of Mathematical Psychology, 3:19-47, 1966.

4. Karush, W. & Dear, R.E., "Optimal Strategy for Item Presentation in a Learning Process", Management Science, 13:773-785, July 1967.

5. Matheson, J.E., "Optimal Teaching Procedures Derived from Mathematical Learning Models", Report CCS-2, Dept. of Engineering-Economic Systems, Stanford University, Stanford, California, 1964.

6. Smallwood, R.D., A Decision Structure for Teaching Machines, M.I.T. Press, Cambridge, Massachusetts, 1962.

7. Smallwood, R.D., "Quantitative Methods in Computer-Directed Teaching Systems", Dept. of Engineering-Economic Systems, Stanford University, Stanford, Calif., March 15, 1967.

8. Suppes, P., "Problems of Optimization in Learning a List of Simple Items", In M.W. Shelly & G.L. Bryan (Eds.), Human Judgement and Optimality, New York, Wiley, 1964, pp 116-126.

Appendix A: THE OPTIMUM POLICY COST FUNCTION

In Section IV the following recursive equation was derived for the optimum policy cost function for a student with past history $h = [m, \pi_1, \pi_2, \ldots]$:

$$w_m(\Pi) = \min_n \left[c_{mn} + \sum_k p(k|h,n)\, w_\ell(\Pi') \right] \tag{4}$$

where n is the number of the instructional alternative leaving the m^{th} level. This appendix shall show that Eq. 4 is consistent with a solution of the form:

$$w_m(\Pi) = \min_n \min_g \left[\sum_j \alpha^{(m)}_{ngj}\, \pi_j \right] \tag{8}$$

First of all, the two quantities $p(k|h,n)$ and π'_j in Eq. 4 can be written directly from Eqs. 2 and 3 in Section III:

$$p(k|h,n) = \sum_i \sum_j \pi_i t_{ij}(a)\, r_{jk}(a) \tag{9}$$

$$\pi'_j = \frac{\sum_i \pi_i t_{ij}(a)\, r_{jk}(a)}{\sum_{i,j} \pi_i t_{ij}(a)\, r_{jk}(a)} = \frac{1}{p(k|h,n)} \sum_i \pi_i t_{ij}(a)\, r_{jk}(a) \tag{10}$$

where a represents the n^{th} instructional alternative leaving the m^{th} level. Now if we assume that $w_\ell(\pi')$ on the right side of Eq. 4 is of the form shown in Eq. 8 then the substitution of π'_j and $w_\ell(\pi')$ into

Eq. 4 yields:

$$w_m(\Pi) = \min_n \left[c_{mn} + \sum_k p(k|h,n) \min_{n'} \min_g \sum_j \alpha^{(\ell)}_{n'gj} \frac{\sum_i \pi_i t_{ij}(a)\ r_{jk}(a)}{p(k|h,n)} \right] \quad (11)$$

where n' refers to the n'^{th} instructional alternative leaving the new level $\ell(m,n,k)$. Now since the response probability $p(k|h,n)$ is independent of n' and g, this quantity can be canceled in the final term of Eq. 11 to give:

$$w_m(\Pi) = \min_n \left[c_{mn} + \sum_k \min_{n'} \min_g \sum_j \alpha^{(\ell)}_{n'gj} \sum_i \pi_i t_{ij}(a)\ r_{jk}(a) \right]$$

$$= \min_n \left[c_{mn} + \sum_k \min_{n'} \min_g \sum_i \pi_i \sum_j \alpha^{(\ell)}_{n'gj} t_{ij}(a)\ r_{jk}(a) \right] \quad (12)$$

For each set of state probabilities $\Pi = [\pi_1, \pi_2, \ldots]$ there will be a set of indices, $S(\Pi) = [n'_1, g_1, n'_2, \ldots n'_k, g_k, \ldots]$ that satisfy the last two minimizations in Eq. 12. In other words, for each value of Π we define n'_k and g_k as the two indices that satisfy the minimizations in Eq. 12 for the k^{th} response, and $S(\Pi)$ is the set of these indices. Furthermore, the number of possible such index sets will be finite; and so if we were to investigate the space of possible values for the state probabilities, we would find this space divided into regions each with its own value for the index set $S(\Pi)$. For the sake of this development, we shall define an index over these regions; that is we shall let h denote the h^{th} region in the space of state probabilities and $S_h(\Pi)$ is the set of indices corresponding to this region.

Now by the definition of this index set, we can rewrite Eq. 12 as:

$$w_m(\Pi) = \min_n \left[c_{mn} + \min_h \sum_i \pi_i \sum_k \sum_j \alpha^{(\ell)}_{n_k' g_k j} t_{ij}(a)\ r_{jk}(a) \right] \tag{13}$$

where n_k' and g_k are elements of the h^{th} index set and h ranges over the possible index sets, S_h, corresponding to the various regions in the space of state probabilities.

And finally by using the fact that the sum of the state probabilities must be unity we can move c_{mn} inside the summation to give:

$$w_m(\Pi) = \min_n \min_h \left[\sum_i \pi_i \left(c_{mn} + \sum_k \sum_j \alpha^{(\ell)}_{n_k' g_k j} t_{ij}(a)\ r_{jk}(a) \right) \right] \tag{14}$$

Eq. 4 is of the same form as Eq. 8 with

$$\alpha^{(m)}_{nhi} = c_{mn} + \sum_k \sum_j \alpha^{(\ell)}_{n_k' g_k j} t_{ij}(a)\ r_{jk}(a) \tag{15}$$

Thus, we have shown that an optimum policy cost function of the form shown in Eq. 8 is consistent with the recursive equation of Eq. 4. Of course, the terminal cost function in Eq. 5 is also in this form, and so the argument is complete.

Since the optimum instructional alternative is just the one that minimizes the cost function, it follows that the α's that determine $w_m(\Pi)$ can also be used to prescribe the optimum policy as described in Eq. 7.

Appendix B: THE CALCULATION OF OPTIMUM POLICY REGIONS

This appendix describes an iterative scheme for calculating the optimum policy cost function, $w_m(\Pi)$. The basic equation defining the iterative process is very similar to Eq. 4; if $w_m^{(z)}(\Pi)$ is the optimum policy cost function after the z^{th} iteration, then we define the process by:

$$w_m^{(z+1)}(\Pi) = \min_n \left[c_{mn} + \sum_k p(k|h,n)\, w_\ell^{(z)}(\Pi') \right] \tag{16}$$

for $z \geq 0$. The process is started by assuming an initial value for $w_m^{(0)}(\Pi)$:

$$w_m^{(0)}(\Pi) = \sum_j \alpha_{oj}^{(m)} \pi_j \tag{17}$$

where the $\alpha_{oj}^{(m)}$'s are to be specified later. Of course, the terminal cost function will always be equal to

$$w_\lambda^{(z)}(\Pi) = \sum_i \gamma_i \pi_i \quad \text{for all } z \tag{18}$$

where λ is the terminal level of the branching network.

According to the form for $w_m(\Pi)$ shown in Eq. 6, a convenient method for specifying the function is by several sets of α's -- one set for each possible combination of n and g in Eq. 6. Each iteration then amounts to using the previous sets of α's to calculate new sets of α's for each level. The complete iteration process thus consists of the following steps:

1. Set up the initial values of the α's for each level.
2. For each level m, search through the space of possible state probabilities, Π, and find all those sets of α's that, on the basis of the α's calculated on the previous iteration, determine the value of $w_m^{(z)}(\Pi)$.
3. Check to see if the new values of the α's are sufficiently close to the previous ones to justify stopping the iteration process; if not, return to Step 2.

One possible method for carrying out Step 2 is first to find the sets of α's at several points throughout the space of state probabilities; e.g. at the points defined by $\pi_1 = 1, \pi_2 = 1, \pi_3 = 1, \ldots$. The intersection of the hyperplanes defined by these sets of α's, $\sum_i \alpha_{ngi}^{(m)} \pi_i$, will generally determine one or more additional points in the space of state probabilities, and the α's for these additional points can be added to the list of α's for the level under consideratic ı. This process continues until there are no intersections of the hyperplanes that yield a new set of α's for the level under consideration.

This process of finding a new set of α's for a particular point, Π, in the state probability space is not a difficult one. Equation 12 can be used to find the appropriate values of n, n_k', and g_k and then Eq. 15 can be used in the actual calculation of the α's.

In practice, there is a slight modification of Step 2 in the iterative process that yields somewhat faster convergence. For this modified version of Step 2, we start with the next to last level $(\lambda - 1)$ and work backwards. In addition in the calculations of $w_m(\Pi)$ we use the values of the α's already calculated during the present iteration when calculating $w_\ell(\Pi')$ for any ℓ greater than m.

The proof of convergence for this iteration process proceeds by induction. Suppose that $w_m^{(z)}(\Pi) < w_m^{(z-1)}(\Pi)$ for all $m<\lambda$, and all Π. Then since $p(k|h,n)\geq 0$, from Eq.16 we have:

$$w_m^{(z+1)}(\Pi) < \min_n \left[c_{mn} + \sum_k p(k|h,n)\ w_m^{(z-1)}(\Pi')\right] = w_m^{(z)}(\Pi) \qquad (19)$$

Thus, if we can find an initial set of α's such that $w_m^{(1)}(\Pi) < w_m^{(0)}(\Pi)$, then the sequence of iterations will yield a monotonically decreasing sequence $[w_m^{(0)}(\Pi),\ w_m^{(2)}(\Pi),\ldots]$ bounded below by $\min_n [c_{mn}]$; and this will prove convergence.

The first iteration of the process yields

$$w_m^{(1)}(\Pi) = \min_n \left[c_{mn} + \sum_k p(k|h,n) \sum_j \alpha_{oj}^{(\ell)}\ \pi_j'\right] \quad \text{for} \quad 1 \leq m < \lambda \qquad (20)$$

where we have substituted Eq.17 into Eq.16 with z=0. The problem now is to find a set of values for the α_{oj}'s such that the expression in Eq.20 is less than $\sum_i \alpha_{oi}^{(m)} \pi_i$ for all Π.

If we substitute π_j' from Eq.10 into Eq.20 we have:

$$w_m^{(1)}(\Pi) = \min_n \left[c_{mn} + \sum_i \pi_i \sum_{j,k} \alpha_{oj}^{(\ell)} t_{ij}(a)\ r_{jk}(a)\right]$$

$$= \min_n \left[\sum_i \pi_i \left(c_{mn} + \sum_{j,k} \alpha_{oj}^{(\ell)}\ t_{ij}(a)\ r_{jk}(a)\right)\right] \qquad (21)$$

Now $w_m^{(1)}(\Pi)$ will be less than $\sum_i \alpha_{oi}^{(m)} \pi_i$ for all Π if there is some instructional alternative, n, for which:

$$\alpha_{oi}^{(m)} = c_{mn} - \epsilon + \sum_{j,k} \alpha_{oj}^{(\ell)} t_{ij}(a) r_{jk}(a) \tag{22}$$

where $\epsilon > 0$. Thus, if Eq. 22 is satisfied for some $\epsilon > 0$, the condition $w_m^{(1)}(\Pi) < w_m^{(0)}(\Pi)$ will be true and convergence of the iterative process is proved.

It can be shown that the set of simultaneous linear equations in Eq. 22 will always have a positive solution as long as the quantities $(c_{mn} - \epsilon)$ are positive. Thus, we can be assured of convergence of the iteration process if we select for each level m an instructional alternative with $c_{mn} > 0$ and then solve Eq. 22 for the starting α's. Of course, in most practical situations the solution of these equations will not be necessary; some reasonable set of initial α's will usually suffice.