August 1977

Fixed Structure Automata in a Multi-Teacher Environment

Daniel E. Koditschek
University of Pennsylvania, kod@seas.upenn.edu

Kumpati S. Narendra
Yale University

Follow this and additional works at: http://repository.upenn.edu/ese_papers

Recommended Citation


This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the University of Pennsylvania's products or services. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to pubs-permissions@ieee.org. By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

NOTE: At the time of publication, author Daniel Koditschek was affiliated with Yale University. Currently, he is a faculty member in the Department of Electrical and Systems Engineering at the University of Pennsylvania.
Fixed Structure Automata in a Multi-Teacher Environment

Abstract
The concept of an automaton operating in a multi-teacher environment is introduced, and several interesting questions that arise in this context are examined. In particular, we concentrate on the consequences of adding a new teacher to an existing n-teacher set as it affects the choice of a switching strategy. The effect of this choice on expediency and speed of convergence is presented for a specific automaton structure.

Comments

This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the University of Pennsylvania's products or services. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to pubs-permissions@ieee.org. By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

NOTE: At the time of publication, author Daniel Koditschek was affiliated with Yale University. Currently, he is a faculty member in the Department of Electrical and Systems Engineering at the University of Pennsylvania.
shown stable behavior, and it proved easy to apply. Although the test problems have been of modest size \( (d \leq 6) \), the model should be applicable to considerably larger problems \( (d \approx 40) \).

REFERENCES


Fixed Structure Automata in a Multi-Teacher Environment

DANIEL E. KODITSCHEK AND KUMPATI S. NARENDRA, SENIOR MEMBER, IEEE

Abstract—The concept of an automaton operating in a multi-teacher environment is introduced, and several interesting questions that arise in this context are examined. In particular, we concentrate on the consequences of adding a new teacher to an existing \( n \)-teacher set as it affects the choice of a switching strategy. The effect of this choice on expediency and speed of convergence is presented for a specific automaton structure.

I. INTRODUCTION

THE THEORY of learning automata was introduced in 1961 by Tsetlin [1], and is now studied quite extensively in both the Soviet Union and the United States. Learning, in this field, is investigated through the formal paradigm of an automaton interacting with a generalized environment. The automaton, finding its behavior repertoire confronted by a set of environmental consequences which it preferentially orders, learns to choose a behavior which produces the most favorable response. Narendra and Thathachar survey this theory and review some fundamental questions in [2]. The power and versatility of automata schemes is being slowly recognized, and several applications have recently been suggested which involve learning in complex situations [2], [3].

In spite of these advances, almost all schemes developed have dealt with a single automaton operating in a single environment. Yet much of what we generally call learning takes place outside this context. While some of our actions are better than others, it is often impossible to compare them linearly. Students are exposed to a variety of teachers; businesses are confronted with conflicting expertise; babies (usually) learn from two parents. In all cases learning seems to be effected in spite of the vector character of the teachers' responses. This paper is concerned with further generalizing the concept of environment to include these common situations. More formally, we shall undertake to modify the theory of learning automata as it currently exists by replacing a totally ordered input set to the automaton with a partially ordered set and study the effects upon optimal learning strategies. The modification will be accomplished by exposing automata to several single-teacher environments.

There is an ambiguity in learning automata theory be-
between the notion of a teacher and an environment. While this has been unimportant in the past, it becomes crucial in the context of the problem discussed in this paper. We will distinguish between an environment which rewards or penalizes the automaton directly and a teacher which conveys information about an implicit environment by approving or disapproving the automaton’s behavior. This distinction arises naturally when we address the question whether “learning” is manifested by the capacity to gain approval from conflicting judges or by the tendency to perform one “correct” action more often than others. Our interest lies primarily in the latter situation: we postulate the existence of an underlying environment within which there exists a correct behavior pattern which is inaccessible to the automaton. A set of teachers which attempts to describe this environment forms the “multi-teacher environment” with which the automaton interacts. The objective of the automaton then is to interrogate the teachers and determine the particular behavior pattern that the underlying environment treats as the best. To make the problem meaningful we postulate, regardless of other differences, that the teachers “agree” with the underlying environment on the ordering of the automaton’s actions.

For a more complete understanding of the problems that may arise in multiple-teacher environments, it is necessary to introduce some distinctions between teachers who satisfy the postulate stated above. For discussion purposes we distinguish between benign, harsh, and good teachers. While the first two are defined only qualitatively, the last concept will be defined as a relation over the set of all teachers. Using these, an investigation of the nature of “improvement” in a multi-teacher environment is undertaken in the body of the paper.

We shall first formulate the problem a little more precisely in Section II and then address the following questions in the succeeding sections. It seems intuitively clear that two good teachers must provide more information than one. Is it true that two teachers are better than either one alone, or that \( n + 1 \) teachers are better than any \( n \) of them? If more teachers improve the automaton’s accuracy, do they necessarily decrease the rate of the learning process? Is it sometimes preferable to ignore a new teacher’s response? Are there different effective learning strategies in the multi-teacher environment, or is the automaton limited to one optimal strategy?

It will be shown in this paper that the switching strategy chosen almost completely affects the answers to the foregoing questions. There exists a strategy by which we can always assure that \( n + 1 \) teachers are “better” than any \( n \) of them. However, such a strategy affects the speed of learning adversely and may consequently be not of great practical interest. If convergence rate is an important consideration, then the automaton may do better to choose strategies whose responses are worsened by the addition of the new teacher. Whether a new teacher is to be ignored or included depends on the performance criterion as well as the old strategy used and the information that is available regarding the new teacher.

The multi-teacher environment considered in this paper will be confined to the context of a fixed structure automaton. We describe the \( P \)-model learning automaton briefly in this section. For a fuller discussion of this topic and for further information regarding other learning models, the reader is referred to [4].

A learning automaton is an automaton-teacher pair \((\mathcal{A}, E)\) as shown in Fig. 1. The automaton \(\mathcal{A}\) is described by the triple \(\{A, B, F\}\), where \(A\) is an action set \(\{a, b\}\), \(B\) is the input set \(\{0,1\}\), and \(F: A \times B \rightarrow A\) is the switching strategy. The teacher is defined by the set of probabilities \(\{c_i, \bar{c}_i\}\), where \(c_i = \Pr\{\beta = 1 | x\}\) and \(\bar{c}_i = \Pr\{\beta = 0 | \bar{x}\}\) (note: \(\beta \in B\)). The automaton performs one of the actions \(a\) or \(\bar{a}\), receives a reward (\(\beta = 0\)) or penalty (\(\beta = 1\)) from the teacher, and uses the information in its updating strategy to determine a new action. At any time \(t\), the automaton will perform \(a\) with total probability \(\pi(t)\) and \(\bar{a}\) with total probability \(\bar{\pi}(t)\), where the total probability vector

\[
\pi(t) = \begin{bmatrix} \pi(t) \\ \bar{\pi}(t) \end{bmatrix}
\]

is a Markov sequence. An expedient strategy is defined as [4]

\[
\lim_{t \to \infty} M(t) = E[\beta(t) | \pi(t)] = c_1 \pi + \bar{c}_1 \bar{\pi} \leq \frac{c_1 + \bar{c}_1}{2}.
\]

Expedience results in the expected penalty being less than chance. It also corresponds to the better action being chosen with higher probability than the worse action (i.e., \(\pi > \bar{\pi}\)).

The single-teacher environment Tsetlin automaton discussed above will be denoted \(\mathcal{A}(1)\). To facilitate discussion of higher order environments, we introduce the measure space associated with \(\mathcal{A}(1), \{B, \mathcal{P}(B), p\}\), where \(B\) is the quotient set \([0,1]/4\) and \(p\) is the measure on \(\mathcal{P}(B)\) assigning \(0)/\mathcal{P}(B) \rightarrow 1 - c_i, \{1, \mathcal{P}(B) \rightarrow \bar{c}_i, \{0})/\mathcal{P}(B) \rightarrow 1 - \bar{c}_i, \{1})/\mathcal{P}(B) \rightarrow \bar{c}_i.

Fig. 2 presents the general \(n\)-teacher environment \(\mathcal{A}(n)\). This is the pair \((\mathcal{A}, \{E_i\}, i = 1, 2, \cdots, n)\) where \(\{E_i\}\) represents a set of \(n\) teachers. The input set to the automaton is now the set of all \(n\)-tuples of zeros and ones and is denoted \(B^n\) where

\[
B^n = \prod_{i=1}^{n} [0,1].
\]
The measure space \((\mathcal{B}, \mathcal{P}(\mathcal{B})), P_n\) corresponds to the environment \(\mathcal{A}(n)\), where \(P_n\) is the measure on \(\mathcal{P}(\mathcal{B})\) induced by the Cartesian product of the individual teacher’s probabilities making the assignments:

\[
(0, 0, \cdots, 0)/2 \leftrightarrow \prod_{i=1}^{\infty} (1 - c_i)
\]

\[
(1, 1, \cdots, 1)/2 \leftrightarrow \prod_{i=1}^{\infty} c_i
\]

\[
(0, 0, 0, \cdots, 0)/2 \leftrightarrow \prod_{i=1}^{\infty} (i - c_i)
\]

\[
(1, 1, \cdots, 1)/2 \leftrightarrow \prod_{i=1}^{\infty} c_i.
\]

The distinction between a multi-teacher environment and a joint exposure to many environments may now be clarified. In the former case we postulate an underlying environment \(E\), which is inaccessible to \(\mathcal{A}\) and has \(c = 1\), \(\bar{c} = 0\). Thus \(E\) defines \(\bar{A}\) as the “correct” behavior. It is this information that \(\mathcal{A}\) must learn from the multi-teacher environment. As stated in the introduction, we postulate that all teachers “agree” with \(E\), i.e., all of the teachers agree that \(\bar{A}\) is the correct behavior so that \(c_i > \bar{c}\). The set \(D\) of acceptable teachers is defined by

\[
D = \{ E_i \in [0,1] \times [0,1] | c_i > \bar{c}_i \}
\]

and is shown in Fig. 3. Qualitatively, if \(c_i\) is small we will call \(E_i\) “benign,” and if \(\bar{c}_i\) is large we will call the teacher “harsh.” A benign teacher will tend to call both actions good, while a harsh teacher will tend to call both actions bad most of the time.

In the case of the \(n\)-teacher environment, the former simple situation no longer prevails. The set \(\mathcal{B}\) is unordered, and an output cannot be classified as a reward or a penalty in a straightforward manner. Intuitively, a response \((1, 1, \cdots, 1)\) must be a penalty and \((0, 0, \cdots, 0)\) must be a reward. But there are \(2^n - 2\) other actions which must be categorized before a switching strategy can be developed. In this context it no longer makes sense to talk about the minimum expected penalty, since penalty is undefined. Consequently, the obvious interpretation that the expectation of \(a\) in the input should be minimized is not meaningful. The automaton’s primary objective in this case is to maximize the final probability of choosing the correct action \(\bar{A}\), and the outputs of the environments are interpreted as rewards or penalties to aid in achieving this objective. Hence, we will define expedience by the condition

\[
\lim_{t \to \infty} \#(t) = \bar{r} > \frac{1}{2}
\]

and if \(\mathcal{A}_1\) and \(\mathcal{A}_2\) evolve total probabilities \(\bar{r}_1\) and \(\bar{r}_2\), we shall say \(\mathcal{A}_1\) is more expedient than \(\mathcal{A}_2\) if and only if \(\bar{r}_1 > \bar{r}_2\). Thus expedience has been redefined as a total ordering over the set of all automata.

A natural comparison of teachers would determine which produces the most expedient \(\mathcal{A}\) if it acts alone. Given two teachers \(E_i\) and \(E_j\) we find that \(\bar{r}_j \geq \bar{r}_i\) if and only if

\[
\frac{\bar{c}_j}{\bar{c}_i} \geq \frac{\bar{c}_j}{\bar{c}_i}.
\]

Hence, we define a “good teacher” \(E_j\) as one whose ratio \(\bar{c}_j/\bar{c}_i\) is low, and we say that it is “better” than \(E_i\), if condition (3) holds. By this definition, “goodness” is a total ordering over the set \(D\) of all teachers.

III. THE TWO-TEACHER ENVIRONMENT

The two-teacher environment \(\mathcal{A}(2)\) is a prototype of all higher order environments. Strategies and solutions in \(\mathcal{A}(n)\) are motivated by results and considerations in \(\mathcal{A}(2)\). In particular, the interesting questions regarding strategic adjustment from \(\mathcal{A}(n)\) to \(\mathcal{A}(n + 1)\) are addressed in the passage from \(\mathcal{A}(1)\) to this case.

A. A Simple Strategy

Given \((A_i | E_i, E_j)\) as shown in Fig. 2, the probabilities assigned to \(B^2\) by \(P_2\) are the following.
The automaton’s response to these inputs will completely determine the behavior of the overall system. Intuitively, \( \mathcal{A} \) should switch actions on \((1,1)\) and treat \((0,0)\) as a reward. The appropriate response to \((0,1)\) and \((1,0)\) is less obvious. Recalling from [2] that reward-inaction is an absolutely expedient strategy, one may consider the use of an inaction strategy for \(\{(0,1),(1,0)\}\). However, for a fixed structure single-memory-level automaton, inaction is the reward strategy, i.e., in \(\mathcal{A}(1)\) the reward transition matrix

\[
F^0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.
\]

Hence such a strategy will lead to

\[
F^{(0,0)} = F^{(0,1)} = F^{(1,0)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}
\]

\[
F^{(1,1)} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}
\]

yielding the stochastic Markov matrix

\[
P_{21} = \begin{bmatrix} 1 - c_1c_2 & c_1c_2 \\ \bar{c}_1\bar{c}_2 & 1 - \bar{c}_1\bar{c}_2 \end{bmatrix}.
\]

The eigenvector of \(P_{21}\) corresponding to the eigenvalue \(\lambda = 1\) is

\[
\pi_{21} = \left[ \begin{array}{c} \bar{c}_1\bar{c}_1 \\ c_1c_2 + \bar{c}_1\bar{c}_2 \\ c_1c_2 + \bar{c}_1\bar{c}_2 \end{array} \right].
\]

This is clearly expedient since \(c_1c_2 \geq \bar{c}_1\bar{c}_2\). The more interesting subsequent question is whether \(\mathcal{A}(2)\) is more expedient under this strategy than was \(\mathcal{A}(1)\), i.e., is \(\bar{\pi}_1 \leq \pi_{21}\)?

The inequalities

\[
\frac{c_1}{c_1 + \bar{c}_1} \leq \frac{c_1c_2}{c_1c_2 + \bar{c}_1\bar{c}_2} \quad \text{or} \quad \frac{c_2}{c_2 + \bar{c}_2} \leq \frac{c_1c_2}{c_1c_2 + \bar{c}_1\bar{c}_2}
\]

are equivalent to

\[
c_1\bar{c}_1(c_2 - \bar{c}_2) \geq 0 \quad \text{or} \quad c_2\bar{c}_2(c_1 - \bar{c}_1) \geq 0,
\]

hence, they are always true (since \(c_1 \geq \bar{c}_1\) by (1)); agreement.

The Measure Space \(M_{21}\): In the above discussion, if we define \(e^* \triangleq c_1c_2\) and \(\bar{e}^* \triangleq \bar{c}_1\bar{c}_2\), then \(e^*\) and \(\bar{e}^*\) become the “penalty probabilities” corresponding to the actions \(a\) and \(\bar{a}\) of a composite teacher \([E_1, E_2]\). This intuitive realization may be derived more formally as follows. The partition on the input set \(B^2\) imposed by the switching strategy results in a “reward” subset which allows its elements to contain one penalty component. These subsets

\[
G_{21} \triangleq \{(0,0),(0,1),(1,0)\} \quad \text{and} \quad G_{21}^* \triangleq \{(1,1)\}
\]

generate a new measure space \(M_{21}\) imposed by the switching strategy. \(M_{21}\) is defined by the triple

\[
M_{21} \triangleq (B^2, H_{21}, \rho_2)
\]

where \(B^2\) is the entire input space,

\[
H_{21} \triangleq \{(\phi, G_{21}, G_{21}^*, B^2)\},
\]

and the new measure function \(\rho_2\) is the old \(P_2\) summed over the elements of \(H_{21}\):

\[
\rho_2(G) \triangleq \int_G dP_2.
\]

By this formal reasoning we see that the switching strategy has “coarsened” the measure set from the power set of \(B^2\) to \(H_{21}\), so that \(\rho_2(\{G_{21}\}) = c_1c_2\) is the penalty probability of a new composite teacher. We will refer to this strategy in \(\mathcal{A}(2)\) as the automaton (and overall environment) \(\mathcal{A}(2)\).

The expedient strategy in \(\mathcal{A}(1)\) will be denoted \(\mathcal{S}_1\).

Other Strategies: The automaton \(\mathcal{S}_{21}\) described above has two teachers and interprets any input with at most one penalty (the second subscript) as a “reward.” According to this definition, the automaton \(\mathcal{S}_{20}\) with measure space \(M_{20} \triangleq (B^2, H_{20}, \rho_2)\) would allow into the “reward” subset only inputs with no penalty component. Under this strategy the final total probability vector is

\[
\pi_{20} = \left[ \begin{array}{c} \frac{c^*_2}{c^*_2 + \bar{e}^*} \\ \frac{c^*_1}{c^*_1 + \bar{e}^*} \\ \frac{c^*_1}{c^*_1 + \bar{e}^*} \end{array} \right] = \left[ \begin{array}{c} \frac{\bar{c}_1 + \bar{c}_2 - \bar{c}_1\bar{c}_2}{c_1 + c_2 - c_1c_2 + \bar{c}_1 + \bar{c}_2 - \bar{c}_1\bar{c}_2} \\ \frac{c_1 + c_2 - c_1c_2}{c_1 + c_2 - c_1c_2 + \bar{c}_1 + \bar{c}_2 - \bar{c}_1\bar{c}_2} \\ \frac{c_1 + c_2 - c_1c_2}{c_1 + c_2 - c_1c_2 + \bar{c}_1 + \bar{c}_2 - \bar{c}_1\bar{c}_2} \end{array} \right].
\]

Since \(\bar{e}^* < c^*_2\) we know that \(\mathcal{S}_{20}\) is expedient. However, the inequality \(\bar{\pi}_{21} > \pi_{20}\) holds if and only if

\[
c_1c_2 \geq \frac{c_1 + c_2 - c_1c_2}{c_1 + c_2 - c_1c_2 + \bar{c}_1 + \bar{c}_2 - \bar{c}_1\bar{c}_2}
\]

or

\[
c_1\bar{c}_1(c_2 - \bar{c}_2) + c_2\bar{c}_2(c_1 - \bar{c}_1) \geq 0,
\]

which is always true by the “agreement” postulate (1). Hence, we have the result that, independent of the characteristics of the two teachers, \(\mathcal{S}_{20}\) is always less expedient than \(\mathcal{S}_{21}\). Further, the two teachers are not necessarily better than either one taken separately since \(\pi_{20} > \bar{\pi}_{10}\) if and only if

\[
\frac{\bar{c}_2}{c_2} \leq \frac{\bar{c}_1(1 - c_1)}{c_1(1 - \bar{c}_1)}
\]

which is not always true.

We may briefly comment on the only remaining viable deterministic switching strategies in \(\mathcal{A}(2)\), which ignore one or the other of the two teachers, respectively. For example, if the automaton switches on receiving the inputs \((1,1)\) and...
(0,1), then clearly the first teacher \( E_1 \) is being disregarded. Similarly if it switches on (1,1), (1,0), the second teacher is being disregarded. In Section III-B we discuss the conditions under which this policy becomes attractive.

The preceding discussion has demonstrated that the strategy used by an automaton determines whether or not two teachers as a composite entity are going to be unconditionally better than any one teacher. We know \( \mathcal{J}_{21} \) is more expedient than the strategy in \( \mathcal{J}(1) \) using either \( E_1 \) or \( E_2 \). The composite teacher is better than either one used singly. However, \( \mathcal{J}_{20} \) does not enjoy such properties even though it is also expedient: the characteristics of one of the teachers may actually result in a deterioration of the overall performance.

**B. Rate of Convergence and Choice of Strategy**

It is clear from the previous section that \( \mathcal{J}_{21} \) is the most expedient strategy in \( \mathcal{J}(2) \). Why, then, should we even mention any others? Unfortunately expedience may be bought at cost of speed: rate of convergence to steady-state behavior is an equally important attribute. The automaton \( \mathcal{J}_{21} \) does not switch actions unless the outputs of both teachers are penalties. The less frequent this occurrence the fewer the automaton’s “experiments” with alternative behaviors. While the convergence problem in \( \mathcal{J}(2) \) does not necessarily preclude \( \mathcal{J}_{21} \), a large number of teachers will mar the desirability of its higher order analogues precisely because the automaton may continue to perform the same action for long periods of time. In this section we briefly consider the different strategies in a two-teacher environment with regard to convergence rate criteria.

A measure of the convergence rate of an automaton [4] is the distance from unity of the second eigenvalue of the probability transition matrix \( P \). Let \( r_{nm} = |1 - c_{nm} - d_{nm}| \) denote this distance for strategy \( \mathcal{J}_{nm} \) in the \( n \)-environment case. By inspection, in the two-environment case, we have \( e_1^{21} + e_2^{21} < c_1 + c_2 < c_2 + d_2 + c_2^{20} \) all of which values increase monotonically between 0 and 2 as some function of \( \| \langle c_1, c_2, e_1, e_2 \rangle \| \) depicted in Fig. 4. Four regions are constructed in the figure illustrating where the sum of the penalties using the various strategies are greater than unity.

In region 1 it is clear that \( r_{21} \) is lower than either \( r_{21} \) or \( r_{20} \). In this region \( E_1 \) and \( E_2 \) are both harsh enough to bring \( e_1^{21} + e_2^{21} > 1 \). Hence, strategy \( \mathcal{J}_{21} \) is both more expedient and converges more rapidly than either \( \mathcal{J}_{20} \) or \( \mathcal{J}_{1} \) with either \( E_1 \) or \( E_2 \). Outside of this small region, however, we are faced with the choice between speed and accuracy (i.e., between convergence rate and expediency).

Disregarding \( \mathcal{J}_{21} \), we are left to choose between \( \mathcal{J}_{20} \) and \( \mathcal{J}_{i}(1) \) \((i = 1, \ 2)\), the latter notation denoting the case where either teacher \( E_2 \) or \( E_1 \) is ignored, respectively. If the strategy \( \mathcal{J}_{20} \) is chosen, the reward set includes only the element \((0,0)\). While this strategy is expedient, it is more or less expedient than \( \mathcal{J}_{i}(1) \) (ignoring the second teacher) depending on whether (4) is true or not, i.e.,

\[
\begin{align*}
\bar{e}_2 &< \bar{e}_1(1 - c_1) \\
\bar{e}_2 &< \bar{e}_1(1 - \bar{e}_1)
\end{align*}
\]

In other words, condition (4) determines when \( E_2 \) “improves” \( E_1 \) in strategy \( \mathcal{J}_{20} \). Hence, we have defined a new strict partial ordering on the set of all teachers yielding properties:

- i) \( E_1 \) can never improve itself,
- ii) \( E_i \) and \( E_j \) may be found such that neither improves the other.

Fig. 5 illustrates an occurrence of iii).

Fig. 6 magnifies region 2 of Fig. 4, where the inequalities \( c_1 + \bar{e}_1 > 1 \) and \( c_1^{21} + c_2^{21} < 1 \) are satisfied. We consider \( 2a,2b,2c \) defined by \( g \), where \( r_{21} = d_1 = r_1 \), and \( h \), where \( r_{21} = d_2 = r_20 \), as depicted. In region 2a, \( \mathcal{J}_{21} \) is not only more expedient but also faster than the other strategies (as in region 1). In 2c the teachers yield \( r_1 < r_{20} < r_{21} \), while in 2h, \( r_1 < r_{21} < r_{20} \). Consequently, we must choose a performance criterion when the teachers have characteristics of these latter two subregions. This choice entails the following considerations. If speed of convergence is more desirable, then we must ignore the second teacher. If a compromise between speed and accuracy is desired, then we must determine whether \( E_2 \) improves (4) \( E_1 \). If there is no improvement, the choice of \( \mathcal{J}_{i}(1) \) is again immediate. If the second teacher does improve the first, then we will choose \( \mathcal{J}_{20} \) in 2c, and \( \mathcal{J}_{21} \) in 2h.

A similar discussion holds in region 3 of Fig. 4, where \( r_1 \) is always smaller than \( r_{21} \). Again, the choice of strategy depends on the performance criterion and the specific character of \( E_2 \). In region 4, \( \mathcal{J}_{20} \) is always faster than both \( \mathcal{J}_{1} \) and \( \mathcal{J}_{21} \). We note again that in all four regions expedience is entirely independent of the convergence rate: 1) \( \mathcal{J}_{21} \) is the most expedient strategy; 2) \( E_2 \) may or may not improve \( E_1 \), regardless of the region in question; 3) hence, speed and accuracy must be evaluated separately.
there are many situations where estimates of these probabilities are available based on past performance. In such cases the estimates rather than the true values of \( c_i \) are used in the decision process described above.

IV. THE \( n \)-TEACHER ENVIRONMENT

As in the previous section, a central concern of this discussion is the formulation of optimal strategies with respect to the different performance criteria which may arise in the context of varying degrees of information about an \( n \)-teacher environment. There are a greater number and variety of choices in \( \mathcal{A}(n) \) than in \( \mathcal{A}(2) \), but the situation is essentially the same. Mathematical proofs of many statements made in the section are lengthy, and have been relegated to the Appendix of the paper.\(^1\) We intend in this section merely to highlight the salient results of that more formal mathematical investigation.

A. \( \mathcal{A} \) Strategies and Their Measure Spaces

As described in Section II, the response of \( n \) teachers to an input is an \( n \)-dimensional vector of ones and zeros. \( \mathcal{A}_{nm} \) denotes the strategy whose reward subset \( G_{nm} \) contains all vectors with no more than \( m \) penalty components (i.e., nonzero entries). In other words, the automaton switches only when at least \( n-m \) teachers disapprove of an action. As in Section III, the measure space associated with this strategy is \( M_{nm} = (B^<_n, \mu_{nm}(\theta)) \). \( M_{nm} \) is proven to be a true probability space in P.1. All such strategies are shown to be expedient in P.2 using the monotonicity of \( \rho_n(G_{nm}) \) when considered as a function of \( |c_1, \ldots, c_n|_{nm} \).

B. Central Results of the Analysis

Interesting questions arise when we know nothing about the \( n \) teachers and are interested in the relative expedience of two strategies. In P.3 the central result of this paper, that \( \mathcal{A}_{nk} \) is more expedient than \( \mathcal{A}_{ni} \) if and only if \( l + 1 \leq k \), is proved (of course, \( k < n \)). That this seemingly intuitive result requires a relatively complicated proof may be motivated by the following considerations. Allowing the automaton to consider an input vector with more penalty components as a member of the "reward" subset naturally (tautologically) decreases the incidence of what the automaton regards as a "penalty." But this decrease obtains for \( a_{x} \) and \( \bar{a}_{x} \) together. Our definition of expedience (2) reflects the characterization of learning (established in the introduction) as a process of distinguishing between the relative values of two actions. Hence the fact that both yield more "rewards" does not aid the learning process. But the proposition is true. The new composite teacher is not only more benign but is also better: the probability of choosing the correct action increases.

A consequence of this result concerns the effect of adding a new teacher to an \( n \)-teacher set. In P.4 it is shown that \( \mathcal{A}_{n+1,k} \) is more expedient than \( \mathcal{A}_{ni} \) (for any combination of \( n \) teachers out of the \( n+1 \)) if and only if \( k \geq l + 1 \) (again

\(^1\) The facts, propositions and their corollaries in the Appendix are numbered separately. Hence, e.g., P.4, which denotes the result proven as "Proposition 4" occurs before C.1, "Corollary 1."
k < n). In other words, increasing the number of teachers by one results in a more expedient automaton if the strategy is modified to include an extra penalty component in its reward set.

It follows immediately that the most expedient strategy for an n-teacher environment is one where input vector (1, 1, \cdots, 1) is the sole element of the penalty set. A more expedient automaton can only result from the addition of an extra teacher.

C. Convergence Rate in \( \mathcal{A}(n) \)

The extended discussion of convergence rate in Section III-B carries over to \( \mathcal{A}(n) \) with n functions \( \{r_m^i\}_{i=1}^n \) defining \( n + 1 \) regions as they cross unity. Corresponding to region 1 of Fig. 4, if the teachers are very harsh, then the relative frequency of \( (1, 1, 1, \cdots, 1) \) will be sufficient to make \( \mathcal{J}_{n,n-1} \) the fastest strategy. On the other extreme, corresponding to region 4 of that figure, if the teachers are very benign, then penalty components will be very rare in the input vectors, and switching will be too infrequent unless \( \mathcal{J}_{\emptyset} \) is used. In between these extremes are the regions corresponding to 2 and 3 where we require more knowledge of teachers’ values and a predetermined performance criterion as described for \( \mathcal{A}(2) \). We mention again that in practice this “extra” knowledge will obtain from estimates \( \{c_i, \varepsilon_{ij}^i\}_{i=1}^n \) based on previous and accumulating experience.

D. \( \emptyset \) Strategies

An alternative class of strategies in \( \mathcal{A}(n) \) is suggested by the argument used to develop \( \mathcal{J}_{21} \) in the transition from \( \mathcal{A}(1) \) to \( \mathcal{A}(2) \). As has been described in Section III, an \( \mathcal{J}_{n-1,m} \) strategy in \( \mathcal{A}(n-1) \) actually produces a composite teacher \( \{E_{i}^n\}_{i=1}^n \), with penalty probabilities \( \rho_{n-1}(G_{n-1,m}^i|x) \), \( \rho_{n-1}(G_{n-1,m}^i|\bar{x}) \). In this light we may view the addition of the extra teacher \( E_n \) to the \( n-1 \) teacher set as forming \( \mathcal{A}(2) \) out of \( \{E_{i}^n\}_{i=1}^n \) rather than \( \mathcal{A}(n) \) out of \( \{E_{i}^n\}_{i=1}^n \), that is, we will treat \( \{E_{i}^n\}_{i=1}^n \) and \( E_n \) as two separate teachers rather than simply adjoining \( E_n \) into the \( n-1 \) teacher set. A possible strategy would be to switch actions only when the new teacher responds with a 1 and the \( n-1 \) teacher input vector falls into the penalty subset \( G_{n-1,m} \) (i.e., has fewer than \( n-1 - m \) zeros). We will refer to such strategies as \( \mathcal{J}_{n-1,1} \), denoting an \( \mathcal{J}_{21} \) construction of \( E_n \) with \( \{E_{i}^n\}_{i=1}^n \) (with \( \mathcal{J}_{n-1,1} \)). Similarly, \( \mathcal{J}_{n,m} \) will denote an \( \mathcal{J}_{20} \) construction of \( E_n \) with \( \{E_{i}^n\}_{i=1}^n \) (with \( \mathcal{J}_{n-1,1} \)); the automaton will switch strategies unless the new teacher responds with a zero and the \( n-1 \) teacher input vector falls into the reward subset \( G_{n-1,m} \) (i.e., has fewer than \( m \) ones). We note that \( \mathcal{J}_{n,m} \) is more expedient than \( \mathcal{J}_{n-1,m} \) in \( \mathcal{A}(n-1) \) if and only if \( E_n \) improves \( \{E_{i}^n\}_{i=1}^n \) (under the \( \mathcal{J}_{n-1,1} \) strategy). Further, the reader should realize that \( \mathcal{J}_{n,n-1} = \mathcal{J}_{n,n-1} \) and \( \mathcal{J}_{\emptyset} = \mathcal{J}_{\emptyset} \).

E. Transition from \( \mathcal{A}(n-1) \) to \( \mathcal{A}(n) \)

It turns out that expedience comparisons between \( \emptyset \) and \( \mathcal{J} \) strategies provide useful tests of the value of an extra teacher. The interpretation of C.1-C.3 is as follows. We start with an \( n-1 \) teacher set and employ an \( \mathcal{J}_{n-1,m} \) strategy \( (m \leq n-3) \). We are allowed the use of a new teacher \( E_n \) should it be ignored; or, if not, what strategy should be used in \( \mathcal{A}(n) \)? \( E_n \) does not improve the \( n-1 \) teacher set if we know \( \mathcal{J}_{n-1,m+1} \) (ignoring \( E_n \)) is more expedient than either \( \mathcal{J}_{n,m+1} \) or \( \mathcal{J}_{n,m+1} \) (using \( E_n \)). If \( E_n \) does improve the \( n-1 \) teacher set, then \( \mathcal{J}_{n,m+1} \) is more expedient than \( \mathcal{J}_{n-1,m+1} \) (ignoring \( E_n \)); moreover \( \mathcal{J}_{n,m+1} \) yields a more expedient automaton than \( \mathcal{J}_{n,m+1} \), even though both are using \( E_n \). Of course, if \( m = n-2 \), a more expedient strategy in \( \mathcal{A}(n-1) \) does not exist, and we shall always use the new teacher with a \( \mathcal{J}_{n,m+1} = \mathcal{J}_{n,m+1} \) strategy.

Since \( r_{n-1,m+1} \) is always smaller than \( r_{n,m+1} \), and very likely (if \( E_n \) is benign enough) smaller than the function associated with \( \mathcal{J}_{n,m+1} \), ignoring \( E_n \) and increasing the size of the reward set in \( \mathcal{A}(n-1) \) will yield a slower strategy unless \( r_{n-1,m+1} \) is closer to unity than the other two functions. Hence, if no performance cost is placed on an additional teacher, we may use it whether or not the \( n-1 \) teacher set is improved: the difference in accuracy will be made up by speed.

The importance of a performance criterion should be clear from the foregoing discussion. More obviously, all of the comparisons described in this section will require varying degrees of information regarding the teachers.

V. STOCHASTIC STRATEGIES AND MEMORY

Up to this point we have discussed two-state deterministic automata. It is reasonable to suspect that stochastic strategies, allowing a more flexible (than binary) switching choice, will thereby make more effective use of the different “strengths” of approval accorded by an n-teacher set. This, however, is not the case. From [4] we know that memory capacity enhances the automaton’s expedience. The multi-teacher environment will augment this effect.

A. Stochastic \( \mathcal{J} \) Strategies

A stochastic strategy in \( \mathcal{A}(1) \) is characterized by state transition matrices

\[
F^1 = \begin{bmatrix}
\zeta_1 & 1 - \zeta_1 \\
1 - \zeta_1 & \zeta_1 \\
\end{bmatrix}
\]

and

\[
F^0 = \begin{bmatrix}
\zeta_0 & 1 - \zeta_0 \\
1 - \zeta_0 & \zeta_0 \\
\end{bmatrix}
\]

where \( \zeta_i \in [0,1] \). Thus a given input causes a switching response with probability \( 1 - \zeta_1 \). While \( \zeta_0 = 1 \) is necessary for expedience, \( \zeta_1 \) does not affect expedience at all and may increase or decrease convergence rate depending upon whether \( \zeta_1 + \zeta_2 > 1 \) [4]. This uncompelling result may prepare us for equally lackluster consequences in \( \mathcal{A}(n) \).

The obvious extension of stochastic capability in \( \mathcal{A}(n) \) is characterized by transition matrices

\[
F^i = \begin{bmatrix}
\zeta_i & 1 - \zeta_i \\
1 - \zeta_i & \zeta_i \\
\end{bmatrix}
\]

where the n components of \( \beta \) yield

\[
\sum_{j=1}^{n} \beta_j = i.
\]
The \(\{\xi_i\}_{i=1}^n\) express the automaton’s interpretation of the “degree” of disapproval implied by an input which contains \(i\) penalties. By a stochastic \(S_{nm}\) strategy we denote the situation where \(\xi_i \in [0,1] (i \leq m)\) and \(\xi_i = 0\ (m < i \leq n)\): the automaton will switch actions when the input is an element of \(G_{nm}\), and will consider elements of \(G_{nm}\) as rewards only with probability \(1 - \xi_i\) (\(i\) is the number of component penalties). The result of P.5 demonstrates that every deterministic \(S_{nm}\) strategy is just as expedient or more expedient than its stochastic cousins. The most expedient stochastic strategy, \(S_{nm,n-1}\) with \(\xi_i = 1\ (i < n)\) and \(\xi_i \in [0,1]\) (i.e., never switches when input is in \(G_{nm,n-1}\), switches with probability \(1 - \xi_i\) when input is all ones), is slower than \(S_{nm,n-1}^*\) unless \(r_{n,n-1} > 1\), which as mentioned in Sections III-B and IV-C, happens extremely rarely.

B. Stochastic \(\mathcal{G}\) Strategies

Alternatively, we might implement a stochastic strategy that switches actions with probability \(1 - \xi_1\) if the \(n\)th component of the input vector is a penalty, and if the first \(n - 1\) components fall into \(G_{n,n-1}^*\) while considering everything else as a reward. This is a stochastic \(S_{n,n-1}^*\) strategy. It is no more expedient than the deterministic \(S_{n,n-1}^*\), and will be slower unless the associated \(r\) (rate) function is greater than unity.

C. Memory

Allowing the automaton extra states provides the possibility of schemes with memory. In \(\mathcal{A}(1)\) it is well known [4] that a Tsetlin \(m\)-memory level automaton is \(\epsilon\)-optimal if \(\epsilon_1 < \frac{1}{2}\), and a Krinsky model becomes \(\epsilon\)-optimal as memory depth approaches infinity. Placing the automaton in a higher order environment employing the reward–penalty partitions discussed in this report amounts to substituting a particular composite teacher for the single teacher in \(\mathcal{A}(1)\). Hence, if the composite teacher is better, the automaton must be more expedient. Since we know when a higher order environment becomes a better teacher, we know how to improve the performance of these established memory schemes. But, substituting a better teacher does not affect the essential characteristics of memory management. No switching strategies discussed in this report can make a Krinsky or Tsetlin automaton \(\epsilon\)-optimal (unless, of course, in the latter case, \(\epsilon_n^* < \frac{1}{2}\), where \(\epsilon_1 > \frac{1}{2}\) previously).

However, the input sets from a multi-teacher environment suggest many new memory schemes which may very possibly induce stronger expedience characteristics than yet seen in deterministic fixed structure automata. Unfortunately, these schemes entail nonsparse \(2m \times 2m\) (where \(m\) is the memory depth) algebraic matrices whose solutions are rather complicated. Those solutions are now being developed.

VI. CONCLUSION

We have replaced a single environment with a set of environments, and, positing some a priori correct behavior, have investigated fixed structure automata learning schemes that increase the probability of its being performed. By emphasizing the importance of learning a “correct” behavior, we have turned away from the problem of minimizing an expected incidence of penalties, establishing a distinction between teachers and environments, and thereby motivating a strong initial postulate that all teachers agree (as to the correct behavior).

Initial analysis indicates that the relative success of a particular strategy depends upon how it partitions an unordered input set into reward and penalty subsets. Somewhat surprisingly, the maximally expedient partitions evince a single element penalty subset containing the vector with no reward components. Since the incidence of one out of \(2^n\) inputs decreases drastically as \(n\) increases, these maximally expedient strategies will usually be very slow. Exact decisions regarding speed and accuracy may be analytically determined, but require further knowledge of the environment.

Future study of the \(n\)-teacher environment might fruitfully consider variable structure or sample mean automata. An investigation of the expected penalty minimization problem would also be interesting.

APPENDIX

We only present the proofs for the two central propositions of the paper, and simply list all other results (which have been proven in [5]).

We establish the following notational conventions:

\[
B^n \triangleq \prod_{i=1}^{n} [0,1]_i
\]

is the set of \(n\)-dimensional vectors with zero or one entries.

\[
G_{np} \triangleq \left\{ \beta \in B^n \mid \sum_{i=1}^{n} \beta_i \leq p \right\}
\]

is the set of such vectors containing \(p\) or fewer ones. \(\mathcal{M}_n\) is the \(\sigma\)-algebra on \(B^n\) generated by the nested sequence \(\{G_{n,m}\}_{m=1,2,\ldots,n}\).

\[
\rho_n: \mathcal{M}_n \rightarrow \mathbb{R} \text{ by } \rho_n(G) = \int_G dP_n
\]

(where \(P_n\) is the probability induced by the Cartesian product of teachers \(\prod_{i} E_i\)) is a measure on \(\mathcal{M}_n\) \(\rho_n(G_{nm})(G_{nm}|z)\) is the \(\rho\)-measure of the set of vectors with no more than \(m\) penalty components, and hence, the reward probability for \(\mathcal{M}_n\) \(\rho_n(G_{nm})(G_{nm}|z)\) is \(f_{nm}(G_{nm}|z)\) the \(\rho\)-measure of those vectors with exactly \(p\) penalty components.

Finally, let \(x \in [0,1]^n\). Then \(f_{nm}(x), \eta_{nm}(x)\) will denote the appropriate mappings from \([0,1]^n \rightarrow [0,1]\) formed by evaluating \(\rho_n(G_{nm}|z)\) when \(c_i = x_i\) \((i = 1,n)\), and \(f_{nm}(x), \eta_{nm}(x)\) \((x)\) will denote those mappings formed by evaluating \(\rho_n(G_{nm}|z)\) when \(c_i = y_i\) \((i = 1,n)\). For any vector \(u \in [0,1]^n\), the \(n - 1\) dimensional vector \(u_j\) will denote \((u_1,\ldots,u_{j-1},u_{j+1},\ldots,u_n)^T\).

Proposition 1: \((B^n, \mathcal{M}_n, P_n)\) is a probability space.

Proposition 2: For all \(nm\), and \(x \in [0,1]^n\), the reward probability function \(f_{nm}(x)\) is monotonically decreasing in each \(x_i\).

Proposition 3: Given the automaton-environment pair \(\mathcal{A}(nm)\), strategy \(S_{nm}\) is more expedient than \(S_{nm}\) independent of teacher characteristics if and only if \(n > p > q \geq 0\).
Proof: The hypothesis is equivalent to the condition \( \bar{n}_{np} > \bar{n}_{qq} \), which, in turn, is equivalent to the inequality:

\[
(\ast) \quad \frac{\bar{n}_{np}}{\bar{n}_{np}^2} < \frac{\bar{n}_{nq}}{\bar{n}_{nq}^2}.
\]

It will suffice to prove this for \( p = q + 1 \), which is done by induction on \( n \).

i) True for \( n = 2 \): We need only consider the case \( p = 1 = q + 1 \): \((\ast)\) may be rewritten

\[
1 - f_{21} < 1 - f_{20}
\]

which is equivalent to \( c_1 \xi (c_2 - c_2) + c_2 \xi (c_1 - c_1) > 0 \).

This is true by the agreement postulate.

ii) True for \( n \) implies true for \( n + 1 \): We must show, for \( y, x \in [0,1]^n \)

\[
\frac{\bar{n}_{n + 1, n + 1, + 2}(y)}{\bar{n}_{n + 1, n + 1, + 1}(y)} < \frac{\bar{n}_{n + 1, n + 1, + 1}(x)}{\bar{n}_{n + 1, n + 1, + 1}(x)}.
\]

This may be recursively expanded and simplified, for each \( j \leq n + 1 \), as

\[
(1 - x_j)(1 - y_j)[\bar{n}_{n + 1}(x_j)\bar{n}_{nq}(y_j) - \bar{n}_{nq + 1}(y_j)\bar{n}_{nq}(x_j)]
+ x_j y_j[\bar{n}_{nq}(x_j)\bar{n}_{nq - 1}(y_j) - \bar{n}_{nq - 1}(x_j)\bar{n}_{nq}(y_j)]
+ x_j(1 - y_j)[\bar{n}_{nq}(x_j)\bar{n}_{nq}(y_j) - \bar{n}_{nq - 1}(x_j)\bar{n}_{nq}(y_j)]
- y_j(1 - x_j)[\bar{n}_{nq}(x_j)\bar{n}_{nq}(y_j) - \bar{n}_{nq - 1}(y_j)\bar{n}_{nq}(x_j)] > 0.
\]

By inductive hypothesis, there is some \( j \) for which the first and second terms are positive. Note that if \( q = 0 \), the second is, nevertheless, positive since \( \bar{n}_{nq} > \bar{n}_{nq} \) by P.2. It requires a separate proof by induction to show that the third term is positive. Finally, it can be shown that the third term is larger than the fourth since \( x_j(1 - y_j) > y_j(1 - x_j) \) and \( \bar{n}_{nq - 1} > \bar{n}_{nq + 1} = \bar{n}_{nq - 1} + 0 \) by inductive hypothesis. Hence the inequality is true.

Proposition 4: Given the automaton-environment pair \( \mathcal{A}(n) \) and an extra teacher \( E_{n+1} \), strategy \( \mathcal{S}_{n+1,p} \) is more expedient than \( \mathcal{S}_n \) independent of teacher characteristics if and only if \( n \geq p > q \).

Proof: The hypothesis is equivalent to \( \bar{n}_{n+1,p} > \bar{n}_{nq} \) which is true if and only if

\[
\frac{\bar{n}_{n+1,p}}{\bar{n}_{n+1,p} + 1} < \frac{\bar{n}_{nq}}{\bar{n}_{nq} + 1}.
\]

By recursively expanding the left-hand side, and simplifying, this yields the inequality

\[
(\ast) \quad y < \frac{\bar{n}_{nq} \bar{n}_{nq} - \bar{n}_{nq} \bar{n}_{nq}}{\bar{n}_{nq} \bar{n}_{nq} - \bar{n}_{nq} \bar{n}_{nq}} \frac{1}{\bar{n}_{nq} \bar{n}_{nq} - \bar{n}_{nq} \bar{n}_{nq}}.
\]

Define \( D' = \{(x,y) \in [0,1]^2\} \). By the agreement postulate on teachers we must have \( (c_1, \xi) \in D = \{(x,y) \in [0,1]^2 | x > y\} \).

i) Sufficiency: Assume \( p \geq q + 1 - \text{show} \ D \subseteq D' \). By P.3 we know \( \bar{n}_{nq} \bar{n}_{nq} - \bar{n}_{nq} \bar{n}_{nq} \leq \bar{n}_{nq} \bar{n}_{nq} - \bar{n}_{nq} \bar{n}_{nq} \), which may be rewritten as \( \bar{n}_{nq} \bar{n}_{nq} - \bar{n}_{nq} \bar{n}_{nq} \leq \bar{n}_{nq} \bar{n}_{nq} - \bar{n}_{nq} \bar{n}_{nq} \), and this is equivalent to \( \bar{n}_{nq} \bar{n}_{nq} - \bar{n}_{nq} \bar{n}_{nq} \leq \bar{n}_{nq} \bar{n}_{nq} - \bar{n}_{nq} \bar{n}_{nq} \). Since the inequality is made strict by multiplying the left-hand side by \( x < 1 \), we have, rearranging,

\[
x < \frac{\bar{n}_{nq} \bar{n}_{nq} - \bar{n}_{nq} \bar{n}_{nq}}{\bar{n}_{nq} \bar{n}_{nq} - \bar{n}_{nq} \bar{n}_{nq}} \frac{x + 1}{\bar{n}_{nq} \bar{n}_{nq} - \bar{n}_{nq} \bar{n}_{nq}}.
\]

But if \((x,y) \in D', \) then \( x < y \), hence \( (x,y) \in D' \).

ii) Necessity: Assume \( p < q \); show \( D \subseteq D' \). If \( p < q \) choose \( \epsilon > 0 \) such that \( \bar{n}_{nq} \bar{n}_{nq} < \bar{n}_{nq} \bar{n}_{nq} - 2 \epsilon \). Choose

\[
\left( \frac{\epsilon}{\bar{n}_{nq} \bar{n}_{nq} - 2 \epsilon}, \frac{\epsilon}{\bar{n}_{nq} \bar{n}_{nq} - 2 \epsilon} \right) \in D.
\]

This is not an element of \( D' \). If \( p = q \) choose

\[
\left( \frac{x}{\bar{n}_{nq} \bar{n}_{nq} - \bar{n}_{nq} \bar{n}_{nq}}, \frac{\bar{n}_{nq} \bar{n}_{nq} - \bar{n}_{nq} \bar{n}_{nq}}{\bar{n}_{nq} \bar{n}_{nq} - 2 \epsilon} \right) \in D.
\]

This is not an element of \( D' \).

Corollary 1: A necessary and sufficient condition for \( \mathcal{S}_{n+1,k} \) in \( \mathcal{A}(n) \) to be more expedient than \( \mathcal{S}_n \) in \( \mathcal{A}(n) \) where \( \{E_j\}^\ast_{j=1} = \{E_j\}^\ast_{j=1} \) is

\[
(\bar{c}_{n+1} \leq \bar{n}_{nq}(\bar{c}_1, \cdots, \bar{c}_n) \bar{n}_{nq - 1} \cdots \bar{c}_1) \leq \bar{n}_{nq}(\bar{c}_1, \cdots, \bar{c}_n) \bar{n}_{nq - 1} \cdots \bar{c}_1).
\]

The next two corollaries are consequences of the definitions of the \( \mathcal{C} \) strategies and the inequalities in P.4.

Corollary 2: If \( \mathcal{S}_{n+1,k} \) is more expedient than \( \mathcal{S}_{n+1,k} \) in \( \mathcal{A}(n) \) then \( \mathcal{S}_n \) in \( \mathcal{A}(n) \) is more expedient than \( \mathcal{S}_{n+1,k} \).

Corollary 3: If \( \mathcal{S}_{n+1,k} \) in \( \mathcal{A}(n) \) is more expedient than \( \mathcal{S}_n \) in \( \mathcal{A}(n) \), then \( \mathcal{S}_{n+1,k+1} \) is more expedient than \( \mathcal{S}_{n+1,k+1} \) in \( \mathcal{A}(n) \).

The reader is reminded that a more complete exposition of stochastic strategies as well as the proof of the final proposition are to be found in [5].

Proposition 5: For all \( n \), \( \mathcal{S}_{am} \) is as expedient as any stochastic strategy of its type.

REFERENCES