#3

ADAPTIVE DECISION PROCESSES

JACK LEE ROSENFELD

Loan Copy Only

TECHNICAL REPORT 403

SEPTEMBER 27, 1962

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
RESEARCH LABORATORY OF ELECTRONICS
CAMBRIDGE, MASSACHUSETTS

The Research Laboratory of Electronics is an interdepartmental laboratory in which faculty members and graduate students from numerous academic departments conduct research.

The research reported in this document was made possible in part by support extended the Massachusetts Institute of Technology, Research Laboratory of Electronics, jointly by the U.S. Army (Signal Corps), the U.S. Navy (Office of Naval Research), and the U.S. Air Force (Office of Scientific Research) under Signal Corps Contract DA 36-039-sc-78108, Department of the Army Task 3-99-20-001 and Project 3-99-00-000; and in part by Signal Corps Contract DA-SIG-36-039-61-G14.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

RESEARCH LABORATORY OF ELECTRONICS

Technical Report 403 September 27, 1962

ADAPTIVE DECISION PROCESSES

Jack Lee Rosenfeld

This report is based on a thesis submitted to the Department of Electrical Engineering, M.I.T., January 9, 1961, in partial fulfillment of the requirements for the degree of Doctor of Science.

(Revised manuscript received May 1, 1962)

Abstract

General adaptive processes are described. In these processes a measure of performance is increased as the experimenter gathers more information; the actions taken by the experimenter determine both the profit and the type of information gathered.

In particular, the adaptive decision process is a two-person, zero-sum, m × n game with some unknown payoffs. This game is played repeatedly. The true values of the unknown payoffs are learned only during those plays of the game at which the unknown payoffs are received. The players are given a priori probability distributions for the values of the unknown payoffs. A measure of performance is defined for the players of adaptive decision processes.

An optimum strategy for one player is derived for the case in which the opponent uses one mixed strategy, known to the player, repeatedly. Optimum minimax strategies for both players are derived for the case in which the players are given the same information about the unknown payoffs. An optimum strategy, from a restricted class of strategies, is derived for one player when he is playing against nature, which is assumed to be an opponent whose strategy is unknown but is unfavorable to the player.

TABLE OF CONTENTS

I. ADAPTIVE SYSTEMS

Ever since the advent of large stored-program digital computers, engineers have been concerned with the problem of how to exploit fully the capabilities of these machines. Much thought has been directed toward using the basic assets of digital computers — the ability to store large amounts of data and perform arithmetical and logical operations very rapidly — to enable computers to gather data during the performance of some tasks and use the gathered information to improve the performance of the tasks. This type of self-improvement process has been called "adaptive behavior." If the nature of the environment in which a computing system is to operate is known to the system designer, and if the computing system is to operate only in that environment, then the designer can often plan an optimum system. However, if the nature of the environment is unknown, if it changes with time or if a single computer must be designed to work well in a variety of environments, then it may be practical to design the system to gather data about its environment and use that data to change its mode of operation. The goal of the change is a more nearly optimum mode of operation, according to some measure of performance.

During the past ten years much work has been done in the field of adaptive systems. Recently, great interest has been shown in randomly connected networks of logical elements. Two of the important contributions in this field are those of Farley and Clark[1, 2] and of Rosenblatt.[3] In these systems, both of which are simulated by digital computers, inputs are applied to the networks, and outputs are received. If the outputs are judged to be correct by the experimenter, the weights of those logical elements that contributed to the output are increased; if the outputs are not correct, the weights of those logical elements that contributed to the output are decreased. The systems are said to adapt if the ratio of the number of correct outputs to the number of incorrect outputs increases as the system gathers data about the desired performance. Experimental results demonstrate the adaptive behavior of these schemes.

The study of random networks is only one phase of the research in adaptive systems. Other interesting work has been done by Oettinger,[4] Bellman and Kalaba,[5] Widrow,[6] White,[7] Mattson,[8] Widrow and Hoff,[9] and others. Aseltine, Mancini, and Sarture[10] have written a fine summary of the work in the field of adaptive control systems.

The common features of the systems just mentioned are the utilization of data gathered in order to increase the expected return, and the independence of the type of data gathered from the actions of the adaptive systems. A less restricted class of adaptive systems is characterized by a dependence of the type of data gathered upon the action of the system. (To distinguish the more general class from the class just described, the respective adjectives "general" and "special" will be used when necessary.) The behavior of a general adaptive system has a twofold result: it determines what type of data will be gathered, and it determines the present return. The data gathered now generally enable the system to improve its future return.

The problem of forming the research policy for an industrial concern is in the class of general adaptive problems. The company's net profit is a function both of its present technical knowledge and the amount of money funded to research. Each year the policy of the company affects the net profit for that year and also the amount of technical knowledge gained through research. The last quantity should help the company to increase its future net profits. The game of Kriegspiel is another example of a general adaptive process. Kriegspiel is a modified game of chess in which neither player is allowed to see his opponent's moves. A referee watches both playing boards and informs players when pieces are captured, or when a player attempts to make a move that is illegal because the path is blocked by a piece of his opponent. A player can learn much about the disposition of his opponent's pieces by attempting to make an illegal move. As a result, both the amount of information a player gathers about the arrangement of pieces and the amount by which the strength of his position changes depend upon his move.

Bush and Mosteller[11] have developed one of the most widely known general adaptive systems. They made no claims for their "stochastic models for learning" other than that the models are good representations for the outcomes of certain experiments with animals and perhaps can be applied to human behavior. The Bush-Mosteller model supposes that the behavior of an organism can be represented at any time by a probability distribution over the courses of action available to the organism. At each trial the response of the organism and the outcome selected by the experimenter determine what event has occurred. Each possible event is associated with a Markov operator that operates on the probability distribution. This produces a new probability distribution that represents the behavior of the system at the next trial. Some organisms that become better at performing certain tasks as they gain experience can be simulated by Bush-Mosteller models. Furthermore, these models can be classified as general adaptive systems (although this was not the intent of their authors' work) because both the data gathered and the reward received at each trial are dependent upon the system's response at that trial. If the parameters are properly chosen, the ratio of the number of successful to unsuccessful events increases as the system gathers more data.

Robbins[12] has posed an interesting problem: "An experimenter has two coins, coin 1 and coin 2, of respective probabilities of coming up heads equal to $p_1 = 1 - q_1$ and $p_2 = 1 - q_2$, the values of which are unknown to him. He wishes to carry out an infinite sequence of tosses, at each toss using either coin 1 or coin 2, in such a way as to maximize the long-run proportion of heads obtained." (In this paper Robbins gave a good— but not optimum – rule for selecting the coin at each toss. A better rule was suggested by Isbell.[13]) A system (experimenter) that performs this maximization is a general adaptive system. The outcome of a toss is dependent upon which coin is tossed, and this outcome determines both the payoff and the data available to the system.

Several authors have made other excellent contributions to the field of general adaptive systems: Robbins,[14] Flood,[15] Bradt, Johnson, and Karlin,[16] Kochen and Galanter,[17] and Friedberg.[18, 19]

II. ADAPTIVE DECISION PROCESSES

The mathematical systems that we call adaptive decision processes are general adaptive systems. They are more restricted than the sequential decision problems posed by Robbins[14]; but they represent a fairly broad class of general adaptive systems. It is hoped that the solutions derived for these processes will be a step toward the solution of more general types of sequential decision problems.

The following warfare situation is a simple example of the type of "realistic" activity represented by adaptive decision processes. The aggressor, called player B, sends missiles toward the defender, called player A. Two indistinguishable types of missile can be sent by B – an armed rocket or a decoy. Player A can use a thoroughly reliable and accurate antimissile missile if he wishes; furthermore, A can tell whether or not a missile sent by B was armed after it has been destroyed or after it has landed in A's territory. The only unknown quantity is the destructive power of B's armed missile when it is allowed to reach its target. Player A has information from two equally reliable spies. One asserts that A will lose one unit (the units may be megabucks) if he allows a warhead to reach his shores; the other spy says the loss will be four units. Player A assigns probability 1/2 to each of these values. However, once A allows an armed missile to land, he will know from then on whether the true destructiveness is 1 or 4. The only other significant loss occurs if A sends an antimissile missile to destroy an unarmed enemy rocket; the loss for this event is 2 units, because of the needless expense. Since A faces the prospect of enduring B's bombardment for a long time, he considers the advisability of learning, by sad experience, the loss that is due to a live missile that is allowed to reach its target. After A has that information, he can decide upon the desirability of using antimissile missiles. In order to make a scientific decision, A constructs the 2 × 2 payoff matrix shown in Fig. 1. The entry in row i and column j,

$$\begin{array}{ccc} & & \begin{array}{cc} & B \\ \text{armed} & \text{decoy} \end{array} \\ A & \begin{array}{l} \text{no defense} \\ \\ \text{defense} \end{array} & \begin{bmatrix} a_{11} & 0 \\ & \\ 0 & -2 \end{bmatrix} \end{array} \qquad \begin{array}{l} \Pr(a_{11}=-4) = 1/2 \\ \Pr(a_{11}=-1) = 1/2 \end{array}$$

Fig. 1. Payoff matrix for warfare example.

a_{ij}, represents the expected return to A and the expected loss to B is a selects the alternative corresponding to row i and B selects the alternative corresponding to column j. For example, a_{12} equals 0 because A receives no return when he sends no antimissile missile against a decoy; however, a_{22} equals −2 because A gains −2 units (loses 2 units) and B loses −2 units (gains 2 units) when A sends an antimissile

missile to destroy a decoy.

In this report we present derivations for optimum strategies for player A, based upon certain assumptions about player B. Decision processes in which the payoff matrix is only partially specified at the beginning of an infinite sequence of decisions are studied.

A brief introduction to the theory of games is given in Appendix I. A reader who has no knowledge of the subject will find this introduction adequate to carry him through all but the most detailed of the following arguments. Other references are also suggested.[21-25]

2.1 Definition of Adaptive Decision Processes

An adaptive decision process consists of an m × n, two-person, zero-sum game that is to be played an infinite number of times. After each step ("Step" implies a single play of the m × n game.) the payoff is made and each player is told what alternative has been selected by his opponent. The payoff matrix is not completely specified in advance. The nature of the uncertain specification of the matrix and the process by which the uncertainty can be resolved are the heart of adaptive decision processes. Unknown payoffs are selected initially according to a priori probability distributions, and the players are told only these probability distributions. If a_{ij} is one of the unknown payoffs, the players do not learn the true value of a_{ij} until, at some step of the infinite process, player A (the maximizing player) uses alternative i and player B (the minimizing player) uses alternative j. At this step both players are told the true value of a_{ij}, so that it is no longer unknown; A receives a_{ij} and B loses a_{ij}. Of course, when all of the unknown payoffs have been received, the process is reduced to the repeated play of a conventional m × n, two-person, zero-sum game.

One can visualize a large stack of matrices, each with all of its payoffs permanently recorded. Some of these payoffs are hidden by opaque covers on each matrix. A probability is assigned to each matrix in the stack. The players know this probability distribution. One of the matrices is chosen, according to the probability distribution, by a neutral referee, and that matrix is shown to the players with the opaque covers in place. The game is played repeatedly until the pair of alternatives corresponding to one of the covered payoffs is used. The cover is then removed, and the play resumes until the next cover must be removed, and so on. This process continues until no cover remains. The completely specified game is then repeated indefinitely.

Three basic types of adaptive decision process are discussed in this report. Adaptive Bayes decision is covered in Section III. This case is the situation in which player B is nature, and the probabilities of occurrence of the n states of nature are known and are the same at each step of the process. For example, in the problem illustrated by Fig. 1 if player B announced that half of his missiles were duds, and that there was no correlation between the alternatives that he had selected from one step to the next, then A could use the results given in Section III of this report to determine an optimum strategy.

Adaptive decision under uncertainty is discussed in Section V. In this case player A knows nothing about B's strategy. The analysis is based upon the assumption that A uses the same probability distribution over his alternatives at each step of the process until he receives one of the unknown payoffs, after which he changes to another repeated distribution, and so on. Furthermore, A is assumed to adopt the conservative attitude that he should use the strategy that maximizes his return when B selects a strategy that minimizes A's return. Referring to the problem associated with Fig. 1, we see that adaptive decision under uncertainty implies that aggressor B knows both the true loss associated with a hit by an armed missile and also what repeated probability distribution defender A will use. If B always uses this information to minimize A's expected return, then A must select a distribution to maximize the minimum return. Other ways of approaching the problem of adaptive decision under uncertainty are discussed.

A third case, considered in Section IV, covers adaptive competitive decision. Player B is assumed to be an intelligent player attempting to minimize the return to player A. There are two subclasses of adaptive competitive decision: the equal information case, in which A and B are given the same a priori knowledge about the unknown payoffs; and the unequal information case, in which the a priori data are different. The former subclass corresponds to the situation shown in Fig. 1 when both players make the same evaluation of the probabilities for payoff a_{11}; one example of the latter subclass is the situation in which B knows the true value of a_{11}, but A does not.

2.2 Measure of Performance

The phrase "maximize the return" is not precise enough to form a basis for further analysis. It is certainly true that A should play so as to receive a large payoff, learn the payoffs that are unknown to him, and prevent B from learning the payoffs that are unknown to B. Also, A should extract what information he can about the payoffs unknown to him by observing the alternatives that B has chosen during previous steps, and not divulging to B, by the alternatives A chooses, any information about the payoffs unknown to B. It is necessary to find a quantitative measure of performance that will incorporate all of these aims and then to select a strategy for A that optimizes this measure.

The measure that first occurs to us is the expected sum of the payoffs at each step of the game. Player A should attempt to maximize this sum, and player B to minimize it. However, since this measure is generally infinite, the maximization or minimization of the infinite quantity would be meaningless. The difficulty of having to deal with an infinite quantity can be solved by dividing the expected sum of the payoffs for the first N steps by N in order to get the expected payoff per step. The limit of the expected payoff per step can be taken as N approaches infinity. The difficulty with this measure of performance is illustrated by a simple example. Consider two different strategies of player A for which all of the unknown payoffs are learned before the thousandth step of the game and the appropriate minimax strategy of conventional game theory is repeated from the thousandth step on. Both strategies will have the same limit of expected payoff per step, which is equal to the minimax value of the payoff matrix. Essentially, this is true because in the limit the contribution of the first thousand payoffs is negligible. This measure of performance was rejected because it neglects the effects of the data-gathering process. A measure of performance that does not discriminate among the many strategies for which all the unknown payoffs are learned in a finite number of steps is not useful.

A more useful measure of performance is the expected sum of the discounted payoffs at each step. This measure of performance has been used successfully by Arrow, Harris and Marschak,[26] and by Gillette[27] for handling infinite processes. The discounted payoff is especially pertinent to economic situations. For example, if the steps of an adaptive decision process are made annually and the payoff is invested at 3 per cent interest (compounded annually), then $100 payoff received now will be worth $103 one year from now. Also $100 received a year from now is equivalent to $100/1.03 invested now. Therefore the present worth of all of the discounted payoffs is the expected sum of the current payoff plus 1/1.03 times the payoff that will be received a year from now, plus $(1/1.03)^2$ times the payoff that will be received two years hence, and so on. The expected sum of the discounted payoffs converges. This measure of performance places more emphasis upon present return than future return. Therefore it overcomes the objection raised against the limit of expected payoff per step. Nevertheless, the possibility exists that with one of two strategies all the unknown payoffs are learned within a finite number of steps, while with the other they are not. Yet the expected sum of the discounted payoffs for the former strategy may exceed the sum for the latter strategy for some values of the discounting factor and may be less for other values. This is a reasonable objection to the use of the expected sum of discounted payoffs.

Before some notation is introduced for the purpose of defining the mean loss measure of performance, a basic theorem will be stated explicitly. This theorem states that the players of adaptive decision processes lose no flexibility by restricting their strategies to a class called behavior strategies. A player is said to be using a behavior strategy if, at each step in an adaptive decision process, he selects a probability distribution over his m (or n) alternatives and uses that distribution to select an alternative. The distribution that he chooses may be dependent upon his knowledge of the history of the process (alternatives selected by both players and payoffs received at all preceding steps). Since behavior strategies are completely general for adaptive decision processes, in the following discussions it will be assumed that players do use behavior strategies. This theorem is an obvious extension of Kuhn's results.[28]

Some notation can be introduced. The probability distribution used by player A at the k^{th} step is denoted $p^k \equiv \left(p_1^k, p_2^k, \ldots, p_m^k\right)$, where p_i^k is the probability that A selects alternative i at the k^{th} step. Hence

$$\sum_{i=1}^{m} p_i^k = 1.$$

Similarly, the probability distribution used by B at the k^{th} step is denoted $q^k \equiv \left(q_1^k, q_2^k, \ldots, q_n^k\right)$. Note that k is a superscript – not a power. In general, p^k and q^k depend upon the past history of the process. The value of the payoff matrix is denoted v; it represents the maximum expected return that A could guarantee himself (by an appropriate selection of a probability distribution) at one step, if all of the unknown payoffs were uncovered. Of course, v is a function of the values of the unknown payoffs. The precise meaning of the value of v for the three cases of adaptive decision processes will be discussed in the appropriate sections of this report. The expected return to player A at the k^{th} step, when A uses p^k and B uses q^k, is denoted r^k.

$$r^k \equiv \sum_{i=1}^{m} \sum_{j=1}^{n} p_i^k q_j^k a_{ij}.$$

Since some of the quantities a_{ij} represent unknown payoffs, r^k is a function of p^k, q^k, and of the values of the unknown payoffs. The term $L^k = v - r^k$ is called the single-step loss at the k^{th} step; it is the difference between the expected payoff that A could guarantee himself if the values of the unknown payoffs were known and the expected payoff A does receive. If the limit as N approaches infinity of the sum of single step losses for the first k steps exists, it is called the total loss L. The values $+\infty$ and $-\infty$ are allowable limits.

$$L = \sum_{k=1}^{\infty} L^k.$$

The expected value of the total loss L, with respect to the probability distributions for the unknown payoffs, is called the mean loss. It is denoted

$$\overline{L} = \int L(\text{unknown payoffs})\, dP(\text{unknown payoffs}),$$

where P(unknown payoffs) denotes the cumulative probability distribution function for the unknown payoffs. The derivations of the mean loss for the three cases of adaptive decision will also be covered.

The single-step loss, L^k, represents the expected loss to A at the k^{th} step because of his lack of data for the unknown payoffs. L^k is similar to the "regret" or "loss" function defined by Savage,[29] except that regret is defined as the difference between what a player could receive if he knew his opponent's choice of alternative, and what he does receive. L^k is the loss for a single step of the game. When the losses for all of the steps are summed, the total is L, which is a function of the values of the unknown payoffs and p^1, q^1, p^2, q^2, ... L is the total loss that A sustains because of his ignorance of the true values of the unknown payoffs. $\overline{L}$, the expected value of L (with respect to the probability distributions for the unknown payoffs), is the measure of performance used in this report.

Player A should play so as to minimize the mean loss $\overline{L}$; whereas, B should play to maximize this quantity. If A plays wisely, L^k will be smaller, in general, than if he plays foolishly, and as a result $\overline{L}$ will also be smaller. Two other factors suggest that the mean loss is a good measure of performance. First, it will be demonstrated that there exist strategies for the players that make $\overline{L}$ finite. Second, if we consider the definitions for L and $\overline{L}$ applied to an N-truncated adaptive decision process (a process that terminates after N steps), we arrive at conclusions that seem reasonable. The following relationships are clearly true:

$$L_N = Nv - \sum_{k=1}^{N} r^k$$

$$\overline{L_N} = N\overline{v} - \sum_{k=1}^{N} \overline{r^k},$$

where L_N is the total loss, and $\overline{L_N}$, $\overline{v}$, and $\overline{r^k}$ are the mean values of L_N, v, and r^k, respectively. Since $\overline{v}$ is not dependent upon the strategies used, the mean N-truncated loss, $\overline{L_N}$, is minimized when player A maximizes the expected sum of his returns at the first N steps, and $\overline{L_N}$ is maximized when B minimizes that sum. Hence, by using the mean N-truncated loss measure of performance, we arrive at the same optimum strategies for A and B as we would when we apply the measure of performance that was suggested first to the truncated process (the expected sum of the payoff at each step).

Note that assumptions of linear utility preference, independence of utility with time and absence of intrapersonal variations in utility, have been tacitly made. They enter implicitly into the definition of the mean loss.

2.3 Summary of Results

Through adaptive Bayes decision it has been demonstrated that an optimum strategy for player A consists of the repeated use of one probability distribution over A's alternatives until one of the unknown payoffs is received, and then the use of another distribution until another payoff is received, and so on, until all of the payoffs are known. Then A must repeat another probability distribution indefinitely. Such a procedure is called a piecewise-stationary strategy. The probability distributions in the optimum piecewise-stationary strategy assign probability 1 to one of A's alternatives and 0 to the others. In general, A should attempt to learn the unknown payoffs as soon as possible. A technique is presented for reducing the computational effort required to determine the optimum piecewise-stationary strategy. This simple method eliminates a large amount of computation.

The analysis of adaptive decision under uncertainty is based upon the assumption that player A uses a piecewise-stationary strategy. The significant result is that A can guarantee that the mean loss is finite if he selects a probability distribution that lies inside a certain linear constraint space. That is, if the components, p_1, p_2, ..., p_m, of the p vector satisfy a given set of linear inequalities, then $\overline{L}$ must be finite. When only one payoff is unknown, the optimum p vector is calculated by a simple algorithm. No simple rule has been derived for the case in which there are two or more unknown payoffs.

The principal results for adaptive competitive decision are: (a) a minimax solution exists for the case in which there is equal information, and piecewise-stationary strategies are optimum for both players; and (b) in general, piecewise-stationary strategies are not optimum for unequal information. A straightforward technique for computing the minimax strategies in problems of equal information will be developed in Section V, but no solution is available for the unequal information case. Another result demonstrates that competitive processes with unequal information can be given meaning as infinite game theory problems.

III. ADAPTIVE BAYES DECISION

If player B uses the same probability distribution at each step of the decision process, and if the alternatives are selected independently at each step, then B is said to be using a stationary strategy. In the adaptive Bayes decision process player B is assumed to use a stationary strategy, which is known to player A. This corresponds to situations in which the alternatives of player B represent possible states of nature, the probability distribution over the states of nature is known, and the state of nature is independently determined at each step of the process. The payoff a_{ij} represents the award to player A when he uses alternative i and state of nature j occurs. An example of the situation illustrated by Fig. 1 would be a problem of adaptive Bayes decision if the defender (player A) learned through his spies that the aggressor intended to send a certain fraction, q_1, of his missiles with warheads and $q_2 = 1 - q_1$ without warheads.

It is also assumed that none of B's alternatives occurs with zero probability:

$$q_j > 0 \quad \text{for } j = 1, \dots, n.$$

This eliminates from consideration extraneous columns of the payoff matrix.

It is demonstrated in Appendix II that the piecewise-stationary strategy for which the mean loss is smallest is actually the optimum strategy

$$\min \bar{L}(S') = \min \bar{L}(S),$$

where S' represents the class of all piecewise-stationary strategies, and S is the class of all possible strategies. This result is the one that intuition leads us to expect. Since no new data are gathered until one of the unknown payoffs is received, it does not seem likely that the probability distribution for each step of the optimum strategy should change between the times when unknown payoffs are learned. The reader is advised to defer the reading of Appendix II until he has read section 3.1. The result given in Appendix II, however, is used hereafter.

3.1 Single Unknown Payoff

The value, v, of the payoff matrix in the adaptive Bayes decision process represents the largest expected return that A could guarantee himself, for a single step of the process, if he knew the true values of the unknown payoffs

$$v = \max_p \left(\sum_{i=1}^{m} \sum_{j=1}^{n} p_i q_j a_{ij} \right) = \max_p \sum_{i=1}^{m} p_i \sum_{j=1}^{n} q_j a_{ij} , \qquad (1)$$

where p represents the set of all probability distributions over the alternatives of A. The expected return when alternative i is used will be denoted

$$E(\text{row } i) \equiv \sum_{j=1}^{n} q_j a_{ij} .$$

When the notation E(row i) is used in Eq. 1, we have

$$v = \max_p \sum_{i=1}^{m} p_i E(\text{row } i) = \max_i E(\text{row } i) \qquad (i=1, \ldots, m).$$

The case for a single unknown payoff is considered first. It is completely general, in order to let a_{11} be the unknown payoff. The single-step loss at the first step and at each succeeding step, until a_{11} is received, is $L^1 = v - r$. Because of the stationariness of the strategies of A and B, it is true that

$$p = p^1 = p^2 = p^3 = \ldots$$

$$q = q^1 = q^2 = q^3 = \ldots$$

until a_{11} is received. Therefore, no superscript is applied to the expected return, r.

After a_{11} is received, A is assumed to use an optimum strategy for the succeeding steps of the process. This is a fundamental assumption that will be used many times in this report. In general, for the purpose of calculating optimum strategies at any step of an adaptive process, the assumption is made that the players use optimum strategies for the process that remains after the next unknown is discovered. This assumption may be made only for situations in which the techniques of finding the optimum strategies have been developed. Once a_{11} is discovered, A has all the data available to determine the alternative(s) for which the expected return equals the value of the payoff matrix. As a result, by repeatedly using an optimum alternative, A can play so that the single-step loss is zero after a_{11} is received.

It follows that

$$L^k = (1-p_1q_1)^{k-1}L^1$$

because the single-step loss at the kth step equals the probability that a_{11} is not discovered before the kth step $[(1-p_1q_1)^{k-1}]$ times the single-step loss when a_{11} is not known $[L^1]$ (plus the probability that a_{11} is discovered before the kth step times the single-step loss when a_{11} is known, which equals zero). Therefore, the total loss is

$$L = \sum_{k=1}^{\infty} L^k = \sum_{k=1}^{\infty} (1-p_1q_1)^{k-1}L^1,$$

and the mean loss is

$$\overline{L} = \sum_{k=1}^{\infty} (1-p_1q_1)^{k-1}\overline{L^1}, \qquad (2)$$

where $\overline{L^1}$ is the mean value of $L^1(a_{11})$ with respect to the unknown payoff. If $P(a_{11})$ is the cumulative probability distribution function of unknown payoff a_{11}, then

$$\overline{L^1} = \int L^1(a_{11})\, dP(a_{11}).$$

The single-step loss L^1 is non-negative for all possible values of a_{11} and all distributions P because the value v of the payoff matrix represents the maximum value of r. Consequently, $\overline{L^1}$ is always non-negative. As a result, $\overline{L}$ exists and is a non-negative number or $+\infty$.

It may be possible to select a distribution, p, for which $\overline{L^1}$ equals zero. This is true if r = v for all possible values of a_{11}; that is, if

$$\sum_{i=1}^{m} p_i \; E(\text{row } i) = \max_i E(\text{row } i)$$

for all possible values of a_{11}. This is equivalent to stating that there exists an alternative i_o for which E(row i_o) $\geqslant$ E(row i) for all possible values of a_{11} and all $i \neq i_o$. This is true if either of these cases holds:

(a) $E_{\min}$(row 1) $\geqslant$ E(row i) for all $i \neq 1$.

(b) E(row i_o) $\geqslant E_{\max}$(row 1), and E(row i_o) $\geqslant$ E(row i) for all $i \neq 1$ or i_o.

$E_{\max}$(row 1) and $E_{\min}$(row 1) denote, respectively, the maximum and minimum possible values of E(row 1):

$$E_{\max}(\text{row } 1) = q_1 a_{11\ \max} + \sum_{j=2}^{n} q_j a_{1j}$$

$$E_{\min}(\text{row } 1) = q_1 a_{11\ \min} + \sum_{j=2}^{n} q_j a_{1j},$$

where $a_{11\ \max}$ and $a_{11\ \min}$ are, respectively, the maximum and minimum possible values of a_{11}. If we introduce the following notation, the results just derived can be expressed more concisely:

$$\overline{E}(\text{row } 1) = q_1 \int a_{11} dP(a_{11}) + \sum_{j=2}^{n} q_j a_{1j}\, .$$

If case (a) holds, $\overline{v} = \overline{E}$(row 1). Thus $\overline{L}$ equals zero if A uses alternative 1 repeatedly. If case (b) holds, $\overline{v}$ = E(row i_o), so $\overline{L}$ equals zero if A uses alternative i_o repeatedly. Case (a) implies that the expected return for alternative 1 is at least as large as the expected return for any other alternative, irrespective of the true value of a_{11}; case (b) implies that the expected return for alternative i_o is at least as large as the expected return for any other alternative, irrespective of the true value of a_{11}.

Once the cases for which $\overline{v} = \overline{r}$ have been dealt with, it is possible to consider the remaining cases, for which $\overline{v} > \overline{r}$. (The quantity $\overline{r}$ is the mean value of r: $\overline{r} = \int r(a_{11})\, dP(a_{11}) = p_1 q_1 \int a_{11}\, dP(a_{11}) + \sum_{(i,j)\neq(1,1)} p_i q_j a_{ij}$.) Equation 2 implies that

$$\overline{L} = \frac{\overline{v} - \overline{r}}{p_1 q_1}. \tag{3}$$

It is demonstrated in Appendix III that $\frac{\overline{v} - \overline{r}}{p_1 q_1}$ assumes its minimum value for some distribution, p, with one component equal to 1 and the remaining components equal to 0; therefore, Eq. 4 follows from Eq. 3.

$$\overline{L}_{\min} \equiv \min_p \overline{L} = \min_p \left[\frac{\overline{v} - \overline{r}}{p_1 q_1}\right] = \min\left[\frac{\overline{v} - \overline{E}(\text{row } 1)}{q_1}, \frac{\overline{v} - E(\text{row } 2)}{0}, \ldots, \frac{\overline{v} - E(\text{row } m)}{0}\right]. \tag{4}$$

Because $\overline{v} > \overline{r}$ for all distributions p and because $q_1 > 0$, it follows that

$$\overline{L}_{\min} = \frac{\overline{v} - \overline{E}(\text{row } 1)}{q_1}.$$

The optimum strategy for A is to use alternative 1 repeatedly.

All of the preceding results can be summarized by saying that

$$L_{\min} = \begin{cases} 0 & \text{if } \overline{v} = E(\text{row } i_o) \text{ for any } i_o = 2, \ldots, m \\ \dfrac{\overline{v} - \overline{E}(\text{row } 1)}{q_1} & \text{otherwise.} \end{cases}$$

Therefore, the minimum mean loss, $\overline{L}_{\min}$, is bounded if all positive payoffs are bounded. If a distribution, p, with its i^{th} component, p_i, equal to one is denoted by e_i, then we can say

$$p_{\text{opt}} = \begin{cases} e_{i_o} & \text{if } \overline{v} = E(\text{row } i_o) \text{ for any } i_o = 2, \ldots, m \\ e_1 & \text{otherwise.} \end{cases}$$

The meaning of this result is clear. The logic behind the cases in which $\overline{v} = E(\text{row } i)$ or $\overline{v} = \overline{E}(\text{row } 1)$ has been discussed already. The mean loss is 0, for in these cases player A has no reason to wish to know a_{11}, since no matter what value the unknown assumes, alternative i dominates all others. Therefore, A's strategy involves no attempt to learn the true value of a_{11}. However, in the case for which there is no uniformly best alternative, A's optimum strategy is to use alternative 1 repeatedly (in order to discover the true value of unknown payoff a_{11} as soon as possible), and after a_{11} has been received to use an optimum alternative repeatedly for the conventional Bayes decision process that results. In this case the mean loss equals the mean single-step loss,

$$\overline{L^1} = \overline{v} - \overline{E}(\text{row } 1) > 0,$$

times the expected number of steps before a_{11} is discovered, which is $1/q_1$.

If $q = (1/2, 1/2)$ for the missile defense problem of Fig. 1, the following quantities are easily calculated:

$$E(\text{row } 1) = \begin{cases} -2 & \text{if } a_{11} = -4 \\ -1/2 & \text{if } a_{11} = -1, \end{cases} \qquad v(a_{11}) = \begin{cases} -1 & \text{if } a_{11} = -4 \\ -1/2 & \text{if } a_{11} = -1, \end{cases}$$

$$E(\text{row } 2) = -1.$$

The preceding analysis indicates that $\overline{L}_{\min}$ equals +1:

$$\overline{L}_{\min} = \frac{-3/4 - (-5/4)}{1/2} .$$

A's optimum strategy is to use alternative 1 repeatedly until a_{11} is received, after which he should use alternative $\left\{\begin{matrix}1\\2\end{matrix}\right.$ repeatedly if $a_{11} = \left\{\begin{matrix}-1\\-4\end{matrix}\right.$.

3.2 Multiple Unknown Payoffs

The following discussion is for the purpose of determining the optimum probability distribution, p_{opt}, for the first segment of A's optimum piecewise-stationary strategy when two payoffs, a_{11} and a_{22}, are unknown. The special cases in which both unknown payoffs are in the same row or column of the payoff matrix will also be mentioned. After one step of the process has occurred, either a_{11} has been received (with probability p_1q_1), a_{22} has been received (with probability p_2q_2) or neither has been received (with probability $(1-p_1q_1-p_2q_2)$). Player A can play in an optimum fashion after he has discovered a_{11} or a_{22}, since the optimum strategy for cases with a single unknown payoff has been derived. Let the total loss sustained by A if he uses the optimum strategy for the process with a single unknown payoff, a_{22}, be denoted $L_{\min}|a_{11}$. The optimum strategy is a function of the value of a_{11}; therefore $L_{\min}|a_{11}$ is a function of both a_{11} and a_{22}. $L_{\min}|a_{22}$ is defined analogously.

Since player A uses a piecewise-stationary strategy, the single-step loss at each step is L^1 until a_{11} or a_{22} is received, after which the loss is $L_{\min}|a_{11}$ or $L_{\min}|a_{22}$. Therefore, the total loss is

$$\begin{aligned} L &= L^1 + p_1q_1L_{\min}|a_{11} + p_2q_2L_{\min}|a_{22} \\ &\quad + (1-p_1q_1-p_2q_2)\Big[L^1+p_1q_1L_{\min}|a_{11}+p_2q_2L_{\min}|a_{22} \\ &\quad + (1-p_1q_1-p_2q_2)[L^1+\dots]\Big] \\ &= \left(L^1+p_1q_1L_{\min}|a_{11}+p_2q_2L_{\min}|a_{22}\right)\sum_{k=0}^{\infty}(1-p_1q_1-p_2q_2)^k; \qquad (5) \end{aligned}$$

and the mean loss is

$$\overline{L} = \left(\overline{L^1} + p_1q_1\overline{L_{min}|a_{11}} + p_2q_2\overline{L_{min}|a_{22}}\right)\sum_{k=0}^{\infty}(1-p_1q_1-p_2q_2)^k, \tag{6}$$

where

$$\overline{L_{min}|a_{11}} \equiv \int \overline{L}_{min}(a_{11})\,dP(a_{11}).$$

$\overline{L}_{min}(a_{11})$ is the minimum mean loss for the process with a single unknown payoff a_{22}, as a function of a_{11}. $\overline{L_{min}|a_{22}}$ is defined similarly.

It may be possible to select a distribution, p, for which the mean loss, $\overline{L}$, equals zero. Because L^1, $L_{min}|a_{11}$, and $L_{min}|a_{22}$ are non-negative for all possible values of the unknown payoffs and for all distributions, Eq. 6 implies that the mean loss is zero if and only if $\overline{L^1} = \overline{L_{min}|a_{11}} = \overline{L_{min}|a_{22}} = 0$. The mean single-step loss, $\overline{L^1}$, equals zero when r = v for all possible values of (a_{11}, a_{22}); that restriction also implies that both $\overline{L_{min}|a_{11}}$ and $\overline{L_{min}|a_{22}}$ equal zero. Conditions (similar to cases (a) and (b) for the single unknown payoff process) that must be satisfied if the preceding restriction is to hold, are easily derived. These are cases in which $\overline{v}$ equals E(row i_o), $\overline{E}$(row 1), or $\overline{E}$(row 2). Player A has no need to learn the true values of the unknown payoffs. If these cases are eliminated first, then the situation in which $\overline{L^1}$ is positive may be handled. Equation 6 leads to the following expression for the mean loss.

$$\overline{L} = \frac{\overline{L^1} + p_1q_1\overline{L_{min}|a_{11}} + p_2q_2\overline{L_{min}|a_{22}}}{p_1q_1 + p_2q_2}.$$

A result of Appendix III implies that $\overline{L}$ assumes its minimum value for some distribution $p = e_i$:

$$\overline{L}_{min} = \min\left[\frac{\overline{v} - \overline{E}(\text{row }1) + q_1\overline{L_{min}|a_{11}}}{q_1}, \frac{\overline{v} - \overline{E}(\text{row }2) + q_2\overline{L_{min}|a_{22}}}{q_2}, \frac{\overline{v} - E(\text{row }3)}{0}, \ldots, \frac{v - E(\text{row }m)}{0}\right].$$

If $\overline{v} > \overline{r}$ for all p, the following relation is true:

$$\overline{L}_{min} = \min\left[\frac{\overline{v} - \overline{E}(\text{row }1) + q_1\overline{L_{min}|a_{11}}}{q_1}, \frac{\overline{v} - \overline{E}(\text{row }2) + q_2\overline{L_{min}|a_{22}}}{q_2}\right],$$

where the definition of $\overline{E}$(row 2) is similar to that of $\overline{E}$(row 1). All of the preceding results can be summarized in the following form:

$$\overline{L}_{\min} = \begin{cases} 0 & \text{if } \overline{v} = E(\text{row } i_o) \text{ for any } i_o = 3, \ldots, m \\ \min \left[\dfrac{\overline{v} - \overline{E}(\text{row } 1) + q_1 \overline{L_{\min} | a_{11}}}{q_1}, \dfrac{\overline{v} - \overline{E}(\text{row } 2) + q_2 \overline{L_{\min} | a_{22}}}{q_2} \right] & \text{otherwise} \end{cases}$$

($\overline{L}_{\min}$ is bounded if all possible payoffs are bounded.)

$$p_{\text{opt}} = \begin{cases} e_{i_o} & \text{if } v = E(\text{row } i_o) \text{ for any } i_o = 3, \ldots, m \\ e_1 & \text{otherwise, if } \dfrac{\overline{v} - \overline{E}(\text{row } 1) + q_1 \overline{L_{\min} | a_{11}}}{q_1} \leqslant \dfrac{\overline{v} - \overline{E}(\text{row } 2) + q_2 \overline{L_{\min} | a_{22}}}{q_2} \\ e_2 & \text{otherwise.} \end{cases}$$

Once again, the solution shows that the optimum strategy for player A is to use alternative 1 or 2 repeatedly in order to find out unknown payoff a_{11} or a_{22} as soon as possible, unless the expected return of row i is greater than the expected returns for all of the other rows, irrespective of the true values of a_{11} and a_{22}. (In the last case, A is not interested in learning the true values of a_{11} and a_{22}, so he uses alternative i repeatedly.) After A learns the value of a_{11} or a_{22}, he should use the optimum strategy for the process with a single unknown payoff that remains.

We may ask the questions: If player A must learn both a_{11} and a_{22} eventually, what difference does it make which he tries to learn first? Why should there be any difference between the mean losses when we use $p = e_1$ or $p = e_2$? These questions are answered with the help of the mathematical manipulation included in Appendix IV. The results in Appendix IV may enable player A to determine his optimum strategy by means of very simple calculations. This method of determining p_{opt} is called the abbreviated method, and is valid when a_{11} and a_{22} are statistically independent. Three possible situations can arise:

(i) $E_{\max}(\text{row } 1) > E_{\max}(\text{row } 2)$

(ii) $E_{\max}(\text{row } 1) < E_{\max}(\text{row } 2)$

(iii) $E_{\max}(\text{row } 1) = E_{\max}(\text{row } 2)$.

In the first case, the optimum strategy for A is to use $p = e_1$. The reason is that when player A uses e_1 and discovers the true value of a_{11}, it is possible that $E(\text{row } 1) \geqslant E_{\max}(\text{row } 2)$. (The expected return for alternative 1 is at least as large as the expected return for alternative 2 – irrespective of the true value of a_{22}.) Thus, after discovering a_{11}, A may never wish to learn the true value of a_{22}. On the other hand, if player A starts the two-unknown payoff process by using e_2, he must always learn a_{11} after he discovers the true value of unknown payoff a_{22} because it is impossible, by the definition of case (i), to find that E(row 2) > E_{max}(row 1) for any value of a_{22}. Therefore, if A must always use e_1 at some part of his piecewise-stationary strategy until he discovers a_{11}, his optimum strategy is to do this first and then use e_2 only if it is necessary. In case (i) it is not true that A "must" learn both a_{11} and a_{22} eventually. In case (ii) $p_{opt} = e_2$; analogous reasoning demonstrates the validity of this result.

There are four subcases of case (iii)

(a) $Pr(a_{11}=a_{11\ max}) = 0$, and $Pr(a_{22}=a_{22\ max}) = 0$,

(b) $Pr(a_{11}=a_{11\ max}) > 0$, and $Pr(a_{22}=a_{22\ max}) = 0$,

(c) $Pr(a_{11}=a_{11\ max}) = 0$, and $Pr(a_{22}=a_{22\ max}) > 0$,

(d) $Pr(a_{11}=a_{11\ max}) > 0$, and $Pr(a_{22}=a_{22\ max}) > 0$.

Subcase (a) is the situation in which the random variables a_{11} and a_{22} have probability distribution functions with probability zero of actually attaining the maximum values, or else they have infinite maxima. The solution for subcase (a), according to the results of Appendix IV, is that the mean losses resulting from the use of distribution e_1 or e_2 first are the same, so both strategies are equally optimum. The reason is that after learning a_{11}, it will be necessary with probability one for A to learn a_{22} in order to discover the optimum strategy for the payoff matrix, and vice versa. The solution for subcase (b) states that $p_{opt} = e_1$, since if e_2 is used first, it will be necessary with probability one to use $p = e_1$ to discover a_{11}; however, if e_1 is used first, it will be necessary only with probability $Pr(a_{11} < a_{11\ max})$ to use e_2 in order to discover a_{22}. Subcase (c) is the converse of subcase (b): $p_{opt} = e_2$. Subcase (d) is not as simple as the others, and involves the comparison of the following expressions:

$$\frac{E_{max}(\text{row 1}) - \bar{E}(\text{row 1})}{q_1} Pr(a_{22}=a_{22\ max}),$$

$$\frac{E_{max}(\text{row 2}) - \bar{E}(\text{row 2})}{q_2} Pr(a_{11}=a_{11\ max}).$$

If the former is smaller, $p_{opt} = e_1$; if the latter is smaller, $p_{opt} = e_2$; if the two terms are equal, both e_1 and e_2 are optimum. It is difficult to read any significance into this result.

Because of the abbreviated method it is possible to derive the optimum strategy from a few easily calculated quantities. An example is worked out in Appendix V by both the regular and abbreviated methods, to illustrate the concepts just derived. This example is a dramatic demonstration of the power of the abbreviated method.

The special cases in which the two unknown payoffs are in the same row or column must be considered now. If the unknown payoffs are in the same column, the preceding results apply with extremely minor modifications. It is obvious that the preceding results can also be specialized to handle the case in which both unknown payoffs are in the same row. Assume that a_{11} and a_{12} are not known. Some simple manipulations lead to the following conclusions:

$$\bar{L}_{min} = \begin{cases} 0 & \text{if } \bar{v} = E(\text{row } i_o) \text{ for any } i_o = 2, \ldots, m \\ \dfrac{\bar{v} - \bar{E}(\text{row } 1) + q_1 \overline{L_{min}|a_{11}} + q_2 \overline{L_{min}|a_{12}}}{q_1 + q_2} & \text{otherwise} \end{cases}$$

and

$$p_{opt} = \begin{cases} e_{i_o} & \text{if } \bar{v} = E(\text{row } i_o) \text{ for any } i_o = 2, \ldots, m \\ e_1 & \text{otherwise.} \end{cases}$$

The case with three unknown payoffs, a_{11}, a_{22}, and a_{33}, is handled just as the case of two unknown payoffs:

$$\bar{L}_{min} = \begin{cases} 0 & \text{if } \bar{v} = E(\text{row } i_o) \text{ for any } i_o = 4, \ldots, m \\ \min\limits_{i=1,2,3} \left[\dfrac{\bar{v} - \bar{E}(\text{row } i) + q_i \overline{L_{min}|a_{ii}}}{q_i} \right] & \text{otherwise.} \end{cases}$$

Here, for example, $\overline{L_{min}|a_{11}} = \int \bar{L}_{min}(a_{11})\, dP(a_{11})$, and $\bar{L}_{min}(a_{11})$ is the minimum mean loss for the case with the two unknown payoffs a_{22} and a_{33} as a function of a_{11}.

The reader will appreciate the difficulties in notation that arise when an attempt is made to write a general expression for cases of more than two unknown payoffs with all possible locations of the unknown payoffs taken into account. Nevertheless, the principles that have been described are still valid for more than two unknown payoffs. A general principle that deserves attention is that $\bar{L}_{min}$ is bounded whenever all possible payoffs are bounded.

Algorithms that take into account all possible situations that arise can be constructed for the purpose of machine computation of optimum strategies. The computations for k unknown payoffs depend upon computations for the k cases of k-1 unknown payoffs, each of which, in turn, depends upon the k-1 calculations for processes with k-2 unknown payoffs, and so on. The reader who has ventured into Appendix V will realize how very rapidly the magnitude of the computational effort grows with the number of unknown payoffs.

It is regrettable that the complexity of the calculations for three or more statistically independent unknown payoffs prevents an extension of the type of analysis for the abbreviated method which was performed in Appendix IV with two statistically independent unknown payoffs. Nevertheless, the arguments presented above in support of the analytic results are valid, so the abbreviated method can be extended to the cases in which there are more than two independent unknown payoffs. The essence of the method is, first, to check for the cases in which $\bar{v} = E(\text{row } i)$ or $\bar{v} = \bar{E}(\text{row } i)$ and for the cases in which the minimum expected return for some alternative exceeds the maximum expected return for another alternative. After these situations are dealt with in the appropriate manners (if $\bar{v} = E(\text{row } i)$ or $\bar{v} = \bar{E}(\text{row } i)$, $p_{opt} = e_i$; if alternative i is dominated, eliminate it from consideration), a comparison is made of $E_{max}(\text{row } i)$ for all alternatives associated with unknown payoffs. If there is a single maximum term, then $p_{opt} = e_i$, where i is the index of the maximum alternative. If the maximum is assumed for two or more alternatives but the probability is zero that the expected return for any of these alternatives assumes its maximum value, then $p_{opt} = e_i$, where i corresponds to any one of the maximum alternatives. The case in which several alternatives have the same maximum value of expected return but only one has a finite probability of assuming the maximum implies that $p_{opt} = e_i$, i corresponding to the unique row. Because the few remaining cases have proved too complex to understand, it is necessary to return to the standard method in order to calculate the optimum strategies when several alternatives have positive probability of assuming the same maximum value of expected return.

IV. ADAPTIVE DECISION UNDER UNCERTAINTY

4.1 The Meaning of Adaptive Decision Under Uncertainty

When player A is making decisions in the face of uncertainty, he knows that at any step of the decision process one of n states of nature exists. The uncertainty about which one exists, and the uncertainty about the process that selects the state of nature are the problems player A must face. If the probability distribution of the states of nature is stationary, and if A knows what this distribution is, he should use the optimum strategies developed in Section III for adaptive Bayes decision processes. Under other circumstances he must resort to different techniques.

If nature uses a stationary strategy but A does not know the repeated distribution, he is forced to make some assumption that will make the problem amenable to solution. The validity of the assumption depends upon the nature of the process. For example, A may assume the existence of an a priori probability distribution over the possible probability distributions of nature's stationary strategy. (A common a priori distribution is the one for which all of nature's distributions are equally likely.) After each step of the process player A can derive an a posteriori probability distribution of nature's distributions, which is a function of the a priori distribution and the alternative used by nature. When A has learned all of the unknown payoffs, his problem is not completely solved. Because he does not know B's strategy, he does not know which of his own strategies is optimum. The problem A faces when all of the payoffs are known is an example of a special adaptive process, since the information gathered about B's strategy is independent of A's strategy. The problem is a generalization of a problem discussed by White.[7] The optimum strategy for player A is to use, at each step, the alternative for which the expected return, at that step, is maximized. The correct alternative is easily determined. The problem that A faces before he knows the entire payoff matrix is a general adaptive process, since the information that he gains about the unknown payoffs depends upon the alternative he selects, so the interesting question is, How should A play when some payoffs are unknown? The mean-loss measure of performance can be applied to this form of the adaptive decision under uncertainty problem. A reasonable definition for v, the value of the payoff matrix, is the expected return player A could guarantee himself if he knew both the true values of the unknown payoffs and the distribution used by nature. In this case the single-step loss does not equal zero when all of the payoffs are known, as it does in the adaptive Bayes decision process. Since it is not known whether the mean loss can be finite for any strategy of A, it may be necessary to use a different measure of performance.

Player A faces a more difficult task when it is not reasonable to assume an a priori distribution for the distribution of nature's stationary strategy. It must be realized that the simpler problem of how to play a game against nature when nature's strategy is unknown – but all the payoffs are known – has not been solved yet. One conservative approach advises player A to use the minimax distribution for the payoff matrix repeatedly. This strategy guarantees A an expected return of at least v at each step. Another approach advises A to make use of his knowledge of the alternatives selected by nature at preceding steps in order to estimate nature's strategy. A paper by Hannon[30] deals with this technique. However, there is no generally accepted solution to the problem. Because of the difficulty in finding a satisfactory solution for the special adaptive process under uncertainty when all of the payoffs are known, the general adaptive process of repeated decision under uncertainty when some payoffs are unknown appears to be a monumental problem.

When the assumption that nature uses a stationary strategy is not valid, the problem is even more difficult. A very cautious approach suggests that A assume that nature's strategy is chosen to maximize A's loss. That is, whatever strategy A uses, nature selects the worst (from A's viewpoint) possible strategy. Therefore, A should select a strategy that will minimize the maximum loss. Then he can guarantee that his loss never exceeds the minimax value, irrespective of the actual strategy used by nature. (Because this is an infinite process, the minimax loss does not necessarily equal the maximin loss.) The problem handled in sections 4.2 and 4.3 is closely related to the minimax formulation. The true minimax problem is a problem of adaptive competitive decision with unequal information – the case in which player B knows all the payoffs. The general solution for this problem has not been found; the problem discussed in sections 4.2 and 4.3 is the minimax solution when player A is restricted to the use of piecewise-stationary strategies. Player B is assumed to use the strategy that maximizes the total loss; this maximizing strategy is a function of both the true values of the unknown payoffs and A's strategy. Player A's optimum piecewise-stationary strategy is the one that minimizes the mean value of the maximum total loss.

The calculation of minimum mean loss to be given presently is an upper bound to the loss sustained by player A in problems of adaptive competitive decision with unequal information. If A uses a better strategy than the optimum piecewise-stationary strategy, the mean loss will be smaller than the quantity calculated here; however, the strategy derived below will be a fairly good mode of play for both the problem of competitive decision and adaptive decision under uncertainty. Furthermore, some very interesting concepts are brought to light by this study.

The problem of Fig. 1 represents a case of decision under uncertainty if it is known that the aggressor sends only live missiles, but the warheads are unreliable and may or may not explode. Player B is assumed to be a capricious gremlin who determines whether each missile will explode. The first alternative of B represents explosion; the second alternative, nonexplosion. (The payoffs in Fig. 1 ought to be modified in order to take into account the cost to B of armed missiles that fail; however, the purposes of this exposition will not be furthered by a change in Fig. 1.) Player A, being very cautious, assumes that the gremlin knows both the piecewise-stationary strategy that A will use and the true value of a_{11}, and will use this information to maximize A's total loss. Therefore, A should select a strategy that minimizes the mean value of the maximum total loss.

4.2 Single Unknown Payoff

The value of the payoff matrix, v, represents the largest expected return player A can guarantee for one step of the process when the payoff matrix is completely known. Since it is assumed that B selects a strategy to minimize the return, v equals the minimax value of the payoff matrix, which is a function of the unknown payoffs:

$$v = \max_{p} \min_{q} \sum_{i=1}^{m} \sum_{j=1}^{n} p_i q_i a_{ij} = \min_{q} \max_{p} \sum_{i=1}^{m} \sum_{j=1}^{n} p_i q_j a_{ij} .$$

Assume that only payoff a_{11} is not known. Because A is assumed to use a piecewise-stationary strategy known to B, it can be shown that player B maximizes the total loss by using his optimum piecewise-stationary strategy. The arguments of Appendix II apply almost directly to this case. Because both players are assumed to use the minimax distributions repeatedly after payoff a_{11} has been received, the single-step loss equals zero after that step. This result allows us to prove that to every nonstationary strategy of B there corresponds a memoryless sequence of distributions for which the total loss to player A is no smaller than the loss for the nonstationary strategy. The total loss can be written

$$L = \sum_{k=1}^{\infty} (v-r)^k \prod_{t=0}^{k-1} \left(1-p_1 q_1^t\right),$$

where $1 - p_1 q_1^o$ is defined to be equal to one. No matter what probability distribution A uses, B can select a distribution for which the single-step loss is non-negative; therefore, the least upper bound to the total loss, with respect to B's strategies, is non-negative. Let the least upper bound of L be denoted L_o. If $L_o = 0$, player B can attain a total loss of zero by using a stationary strategy. If $0 < L_o < +\infty$, the logic of Appendix II can be used to show that the optimum piecewise-stationary strategy for B ensures a total loss of L_o. It is easily shown that if $L_o = +\infty$, B can attain a total loss of $+\infty$ by using a stationary strategy.

It is possible, therefore, to write an expression for the total loss as a function of the piecewise-stationary strategies of A and B (and of a_{11}):

$$L = \sum_{k=1}^{\infty} (1-p_1 q_1)^{k-1} (v-r). \tag{7}$$

Three cases can arise:

(a) the maximum value of v−r (with respect to q) equals zero;

(b) the maximum value of v−r is positive, and $p_1 = 0$; and

(c) the maximum value of $v - r$ is positive, and $p_1 > 0$.

In case (a), $L = 0$ when B sets $q = e_j$ for any j for which E(col j) = v. At least one such alternative j exists. Then the maximum total loss equals zero. The quantity E(col j) represents the expected return when B selects alternative j:

$$E(\text{col } j) \equiv \sum_{i=1}^{m} p_i a_{ij} .$$

In case (b), $L = +\infty$ if B sets $q = e_j$ for any j for which $v > E(\text{col } j)$. There must be at least one such alternative. The maximum total loss for case (c) must be positive. This expression for the total loss follows from Eq. 7:

$$L = \begin{cases} \dfrac{v - r}{p_1 q_1} & \text{if } v \neq r \\ 0 & \text{if } v = r. \end{cases} \tag{8}$$

In Appendix III it is shown that an expression of the form $(v-r)/p_1 q_1$ assumes its maximum value with respect to q when $q = e_j$ for some $j = 1, \ldots, n$. If the largest of the n terms equals a positive number c and occurs for $q = e_j$ with $j \neq 1$, and if $v = E(\text{col } j)$, then

$$\max_q \frac{v - r}{p_1 q_1} = \frac{0}{0} = c.$$

But Eq. 8 implies that $L = 0$ when $v = r$. This difficulty can be avoided if B chooses q close — but not equal — to e_j, so that L is as close to c as desired. Therefore, it is possible to write

$$L_{\max} \equiv \max_q L = \max \left[\frac{v-E(\text{col } 1)}{p_1}, \frac{v-E(\text{col } 2)}{0}, \ldots, \frac{v-E(\text{col } n)}{0} \right]. \tag{9}$$

(This expression is admittedly meaningless and is to be accepted only as a convenient notation for the preceding description of the quantity $L_{\max}$.) Notice that cases a and b are also included in the notation of Eq. 9 for $L_{\max}$.

Because $L_{\max}$ is non-negative for all possible values of p and a_{11}, the mean loss,

$$\overline{L}(p) \equiv \int L_{\max}(p, a_{11})\, dP(a_{11}),$$

is infinite if $L_{\max} = +\infty$ for any values of a_{11} that have finite probability. In order to ensure that $\overline{L}(p)$ is finite, A must select p so that the following relations are satisfied for all possible values of a_{11}:

$$E(\text{col } 1) \geqslant v \quad \text{if } p_1 = 0$$

$$E(\text{col } j) \geqslant v \quad \text{for } j = 2, \ldots, n.$$

These inequalities follow from Eq. 9. If they are satisfied for the largest possible value of v, they are satisfied for all other possible values. Since v is a monotonic, non-decreasing function of a_{11}, it assumes its maximum possible value, v_{max}, when the unknown payoff assumes its maximum possible value, $a_{11\,max}$: $v_{max} \equiv v(a_{11\,max})$. Therefore, A guarantees that $\overline{L}(p)$ is finite if he selects a probability distribution, p, that satisfies these inequalities:

$$\begin{aligned}
&a)\ \sum_{i=1}^{m} p_i a_{ij} \geqslant v_{max} && \text{for } j = 2, \dots, n, \\
&b)\ \sum_{i=1}^{m} p_i a_{i1} \geqslant v_{max} && \text{if } p_1 = 0 \\
&c)\ \sum_{i=1}^{m} p_i = 1 && \\
&d)\ p_i \geqslant 0 && \text{for all } i = 1, \dots, m.
\end{aligned} \tag{10}$$

This is an important result. No matter what piecewise-stationary strategy A selects, and no matter what value the unknown payoff assumes, player B can select a piecewise-stationary strategy (that depends upon p and a_{11}) for which the total loss is non-negative. If A uses a distribution p that does not satisfy inequalities (10a), then the expected return for some alternative $j \neq 1$ of player B is less than the value of the payoff matrix for some possible value(s) of a_{11}. B's optimum strategy for that value(s) is to select alternative j repeatedly. In this case the single step loss is positive at each step:

$$L^k = v - E(\text{col } j) > 0 \qquad \text{for } k = 1, 2, \dots .$$

Because $j \neq 1$, player A never learns the true value of the unknown payoff. As a result, the total loss is infinite, so the mean loss is infinite also. If inequality (10b) is not satisfied, then for some possible value(s) of a_{11}

$$L^k = v - E(\text{col } 1) > 0 \qquad \text{for } k = 1, 2, \dots ,$$

if B selects alternative 1 repeatedly. Because A never receives a_{11} when $p_1 = 0$, the total loss and the mean loss are infinite. Relationships (10c) and (10d) are the restrictions imposed by the fact that p is a probability distribution.

Inequalities (10a), (10c) and (10d) describe a closed convex polyhedron in m space. The coordinates of this m space are $p_1, \dots, p_m$. (Actually, the inequalities determine a closed convex polygon in one hyperplane of m space.) This polyhedron will be called a constraint space for player A.

There exists at least one distribution that lies in the constraint space. Let $a_{11\,max}$

be substituted for a_{11} in the payoff matrix. A minimax strategy of player A for one play of this game is a probability distribution that satisfies inequalities (10a), (10c) and (10d), as well as (10b). Therefore, there exists at least one distribution, p, inside the constraint space for which $\overline{L}(p)$ is finite. This is a major conclusion. It demonstrates that $\overline{L}(p_{opt})$ is always finite.

The next step in the solution is the selection of a distribution that lies inside the constraint space and for which the mean loss is minimized. The case in which the probability equals zero that v actually attains its maximum value is considered before the more involved cases: $Pr(v=v_{max}) = 0$. This condition means that a_{11} attains its maximum value with probability zero (the cumulative probability distribution function $P(a_{11})$ is continuous at $a_{11\ max}$) and that $v(a_{11})$ has a positive derivative at $a_{11\ max}$. In this case when p lies in the constraint space, the following relationship is true with probability 1:

$$\sum_{i=1}^{m} p_i a_{ij} \geqslant v_{max} > v \qquad \text{for } j = 2, \ldots, n.$$

By the definition of v, it is impossible that

$$\sum_{i=1}^{m} p_i a_{ij} > v \qquad \text{for all } j = 1, \ldots, n;$$

therefore,

$$E(\text{col } 1) = \sum_{i=1}^{m} p_i a_{i1} \leqslant v$$

with probability 1. As a result, B must use alternative 1 repeatedly in order to maximize the total loss for all possible values of a_{11}. Equation 11 follows immediately.

$$L_{max} = \begin{cases} \dfrac{v - E(\text{col } 1)}{p_1} & \text{if } v > E(\text{col } 1) \\ 0 & \text{if } v = E(\text{col } 1), \end{cases} \tag{11}$$

and

$$\overline{L}(p) = \begin{cases} \dfrac{\overline{v} - \overline{E}(\text{col } 1)}{p_1} & \text{if } \overline{v} > \overline{E}(\text{col } 1) \\ 0 & \text{if } \overline{v} = \overline{E}(\text{col } 1) \end{cases} \tag{12}$$

where $\overline{E}$(col 1) represents the mean value of the expected return for alternative 1 with respect to unknown payoff a_{11}. Note again that when A selects p inside the constraint space, he forces B to use alternative 1 repeatedly, with probability 1.

According to a result given in Appendix III, the quantity $(\overline{v}-\overline{E}(\text{col } 1))/p_1$, assumes its minimum value at one of the vertices of the convex polyhedron described by inequalities (10). Also the case in which $\overline{v} = \overline{E}(\text{col } 1)$ occurs at a vertex, if it occurs at all. Because the polyhedron has a finite number of vertices, $\overline{L}(p)$ can be calculated for each of the vertices. The smallest of these numbers is $\overline{L}_{min}$:

$$\overline{L}_{min} \equiv \min_{p} \overline{L}(p) = \overline{L}(p_{opt}).$$

The techniques for finding the vertices of a polyhedron in m space should be familiar to those who are acquainted with the linear programming problem. It is admitted that finding all of the vertices and calculating $\overline{L}(p)$ for each of them can be a lengthy process.

The following example illustrates the procedure outlined above for finding $\overline{L}_{min}$ and p_{opt}.

$$A \quad \begin{array}{c|ccc} & & B & \\ & 1 & 2 & 3 \\ 1 & a_{11} & 0 & 0 \\ 2 & 0 & -1 & 0 \\ 3 & 0 & 0 & -1 \end{array} \qquad P(a_{11}) = \begin{cases} 0 & \text{if } a_{11} < -2 \\ 2 + a_{11} & \text{if } -2 \leqslant a_{11} < -1 \\ 1 & \text{if } a_{11} \geqslant -1 \end{cases}$$

Here, $P(a_{11})$ is the cumulative distribution function for a_{11}, corresponding to a flat probability density between -2 and -1, and probability 0 elsewhere. The following quantities are easily calculated:

$$v(a_{11}) = \begin{cases} \dfrac{a_{11}}{1 - 2a_{11}} & \text{if } a_{11} \leqslant 0, \quad a_{11\,max} = -1, \\ \\ 0 & \text{if } a_{11} \geqslant 0, \quad v_{max} = -1/3. \end{cases}$$

The constraint space is defined by these relationships:

$$\left.\begin{array}{l} 0p_1 - 1p_2 + 0p_3 \geqslant -1/3, \\ 0p_1 + 0p_2 - 1p_3 \geqslant -1/3, \end{array}\right\} \qquad \text{corresponding to (10a)}$$

$$p_1 + p_2 + p_3 = 1, \qquad \text{corresponding to (10c)}$$

$$p_1 \geqslant 0, \quad p_2 \geqslant 0, \quad p_3 \geqslant 0, \qquad \text{corresponding to (10d).}$$

The four vertices of this polyhedron can be easily derived:

$$\begin{array}{ll} p_a = (1/3,\ 1/3,\ 1/3), & p_c = (2/3,\ 1/3,\ 0), \\ p_b = (2/3,\ 0,\ 1/3), & p_d = (\ 1,\ \ 0,\ \ 0). \end{array} \qquad (13)$$

The following quantities are needed to evaluate $\overline{L}(p)$:

$$\overline{v} = \int_{-2}^{-1} \frac{a_{11}}{1-2a_{11}}\, da_{11} = -.372, \qquad \overline{a}_{11} = -1.5.$$

Therefore, if p lies in the constraint space

$$\overline{L}(p) = \begin{cases} \dfrac{-.372-(-1.5p_1+0p_2+0p_3)}{p_1} = \dfrac{1.5p_1-.372}{p_1} & \text{if } 1.5p_1-.372>0 \\ 0 & \text{if } 1.5p_1-.372=0. \end{cases}$$

Thus

$$\overline{L}_{\min} = \overline{L}(p_a) = .384, \qquad p_{\text{opt}} = (1/3,\ 1/3,\ 1/3).$$

This example has illustrated the straightforward procedure that is followed for finding $\overline{L}_{\min}$ in case v attains its maximum value with zero probability. There are two other cases that can occur. The possible situations that arise are illustrated graphically in Appendix VI. When there is a finite probability that v attains its maximum value, the problem becomes more complicated. Let $P(a_{11})$ for the preceding example be changed from the function already given to

$$P_1(a_{11}) = \begin{cases} 0 & \text{if } a_{11} < -2 \\ \frac{1}{4}(2+a_{11}) & \text{if } -2 \leqslant a_{11} < -1 \\ 1 & \text{if } a_{11} \geqslant -1 \end{cases}$$

This means that the random variable a_{11} has a flat probability density of magnitude 1/4 from −2 to −1; the probability that a_{11} equals −1 is 3/4; and a_{11} has probability 0 elsewhere. It is easily established that $v_{\max} = -1/3$, and $\Pr(v=v_{\max}) = 3/4$. Another case in which $\Pr(v=v_{\max}) > 0$ occurs for the following cumulative distribution function:

$$P_2(a_{11}) = \begin{cases} 0 & \text{if } a_{11} < -2 \\ \frac{1}{4}(2+a_{11}) & \text{if } -2 \leqslant a_{11} < +2 \\ 1 & \text{if } a_{11} \geqslant +2 \end{cases}$$

Here a_{11} has a flat probability density from −2 to +2 and has probability 0 elsewhere. Because v = 0 for all a_{11} greater than zero, it follows that $v_{\max} = 0$ and $\Pr(v=v_{\max}) = 1/2$. The example will be solved for both P_1 and P_2.

Consider the case in which $v = v_{\max}$, which was not encountered in the preceding discussion. It is possible that p lies in the constraint space and that $E(\text{col } 1) > v = v_{\max}$. If this is so, $E(\text{col } j) \geqslant v = v_{\max}$ for all j = 1, . . . , n. Then case (a) is true, because the maximum value of v − r equals zero, $L_{\max} = 0$, and Eq. 11 is not a valid expression for L_{max}. However, if E(col 1) = v = v_{max}, then $L_{max} = 0$, but Eq. 11 is still valid. In order to test whether or not E(col 1) > v_{max} for any distribution in the constraint space and any value of a_{11}, it is simply necessary to test whether or not E_{max}(col 1) > v_{max} at any of the vertices.

$$\left(E_{max}(\text{col } 1) \equiv p_1 a_{11\ max} + \sum_{i=2}^{m} p_i a_{i1}\right).$$

When cumulative distribution function $P_1(a_{11})$ is applied to the preceding example, Eq. 11 is valid. This can be seen very easily. The constraint space for P_1 is the same as for the original probability distribution, because $v_{max} = -1/3$ for both distributions. Therefore, the vertices of the constraint space are given by Eq. 13. In particular, $p_1 \geqslant 1/3$. Furthermore, E(col 1) = $-p_1$. Therefore, E(col 1) $\leqslant v_{max}$, for all possible values of a_{11} and for all points in the constraint space. As a result, Eq. 11 is valid when $v = v_{max}$, and Eq. 12 is valid. Since

$$\bar{v} = \frac{1}{4}\int_{-2}^{-1} \frac{a_{11}}{1 - 2a_{11}}\, da_{11} + \frac{3}{4}\frac{-1}{3} = -.343 \qquad \text{and } \bar{a}_{11} = -1.125,$$

$$\bar{L}(p) = \frac{\bar{v} - \bar{p}_1\bar{a}_{11}}{p_1} = \frac{-.343 - (-1.125p_1)}{p_1},$$

and

$$\bar{L}_{min} = \bar{\bar{L}}(p_a) = .096.$$

The problem of Fig. 1, when viewed as a problem of adaptive decision under uncertainty, falls into this category in which $Pr(v=v_{max}) > 0$, but E_{max}(col 1) $\leqslant v_{max}$ for all p in the constraint space. The reader should be able to verify the fact that $\bar{L}_{min} = 1$ and $p_{opt} = (2/3, 1/3)$.

The situation becomes more complicated if $P_2(a_{11})$ is assumed for the preceding example. Since v_{max} equals zero, the constraints are:

$$0p_1 - 1p_2 + 0p_3 \geqslant 0$$

$$0p_1 + 0p_2 - 1p_3 \geqslant 0$$

$$p_1 + p_2 + p_3 = 1$$

$$p_1 \geqslant 0, \quad p_2 \geqslant 0, \quad p_3 \geqslant 0.$$

There is only one p vector that satisfies this set of constraints: $p_o = (1, 0, 0)$. Therefore, this unique vector must be the optimum distribution for player A. But E(col 1) = $p_1 a_{11} > 0 = v_{max}$ when a_{11} is greater than zero. Equation 11 is valid only for $a_{11} \leqslant 0$, since $L_{max} = 0$ for $a_{11} > 0$. Therefore, the mean loss is

$$\bar{L}(p) = \int_{-2}^{0} \frac{v - E(\text{col } 1)}{p_1} dP(a_{11}) + \int_{0}^{2} 0 \, dP(a_{11})$$

$$= 1/4 \int_{-2}^{0} \left(\frac{a_{11}}{1 - 2a_{11}} - a_{11} \right) da_{11} = .351.$$

The preceding discussion can be summarized easily. In order to calculate $\bar{L}$ when $Pr(v=v_{max}) = 0$, use the expression

$$\frac{\bar{v} - \bar{E}(\text{col } 1)}{p_1}. \tag{14}$$

If $Pr(v=v_{max}) > 0$ and a unique p vector satisfies the constraints, that vector is optimum, and

$$\bar{L}(p) = \int_{a_{11}=-\infty}^{a_{11}=a_o} \frac{v - E(\text{col } 1)}{p_1} dP(a_{11}),$$

where a_o is a critical value of a_{11}. If $a_{11} < a_o$, it is true that $E(\text{col } 1) < v_{max}$; and if $a_{11} > a_o$, then $E(\text{col } 1) > v_{max}$. The critical value a_o is defined by the equation

$$p_1 a_o + \sum_{i=2}^{m} p_i a_{i1} = v_{max}.$$

If there is more than one p vector in the constraint space, then a check must be made to see whether or not $E_{max}(\text{col } 1) \leqslant v_{max}$ for all p in the constraint space. If this is true, then Eq. 14 should be used.

The remaining possibility is that $Pr(v=v_{max}) > 0$; there are several p vectors in the constraint polyhedron; and $E_{max}(\text{col } 1) > v_{max}$ for at least one of the vertices; then it is said that an unstable condition occurs. The following example illustrates this instability.

$$\begin{array}{cc} & \text{B} \\ \text{A} & \begin{array}{c|cc} & 1 & 2 \\ \hline 1 & a_{11} & 2 \\ 2 & 1 & 2 \end{array} \end{array} \qquad P(a_{11}) = \begin{cases} 0 & \text{if } a_{11} < 0 \\ a_{11}/3 & \text{if } 0 \leqslant a_{11} < 3 \\ 1 & \text{if } a_{11} \geqslant 3 \end{cases}$$

(a_{11} has a flat probability density function of magnitude 1/3 from 0 to 3, and has probability 0 elsewhere.)

$$v(a_{11}) = \begin{cases} 1 & \text{if } a_{11} \leqslant 1 \\ a_{11} & \text{if } 1 \leqslant a_{11} \leqslant 2 \\ 2 & \text{if } a_{11} \geqslant 2, \end{cases} \qquad \begin{array}{l} v_{max} = 2 \\ \Pr(v=v_{max}) = 1/3. \end{array}$$

The constraint equations become

$$2p_1 + 2p_2 \geqslant 2$$

$$p_1 + p_2 = 1$$

$$p_1 \geqslant 0, \quad p_2 \geqslant 0.$$

As a result, p_1 can lie anywhere between 0 and 1. The two vertices of the constraint space are

$$p_a = (1,0), \quad p_b = (0,1).$$

The check to see if $E_{max}(\text{col } 1) > v_{max}$ yields a positive answer when $p = p_a$, since $E_{max}(\text{col } 1) = 3 > v_{max} = 2$. Because of the relative simplicity of the problem, the calculation of $\bar{\bar{L}}_{min}$ is feasible. This rather tedious solution is not given here. The end product of the calculation is that $\bar{L}_{min} = 1/6$, and $p_{opt} = (1,0)$.

Now let the payoff matrix be modified slightly:

$$\begin{bmatrix} a_{11} & 2 \\ 1 & 2+\epsilon \end{bmatrix}$$

where $\epsilon > 0$.

$$v = \begin{cases} 1 & \text{if } a_{11} \leqslant 1, \\ a_{11} & \text{if } 1 \leqslant a_{11} \leqslant 2, \\ \dfrac{a_{11}(2+\epsilon) - 2}{a_{11} - 1 + \epsilon} & \text{if } a_{11} \geqslant 2, \end{cases} \qquad v_{max} = \frac{4 + 3\epsilon}{2 + \epsilon}$$

This is now a problem in which $\Pr(v=v_{max}) = 0$. The solution is

$$\bar{\bar{L}}_{min} \approx 0.5, \quad p_{opt} = \left(\frac{1+\epsilon}{2+\epsilon}, \frac{1}{2+\epsilon}\right) \approx (1/2, 1/2),$$

for $0 < \epsilon \ll 1$.

The significant point brought out by the example is that an infinitesimal perturbation of the payoff matrix has caused a large change in both the optimum strategy, p_{opt}, and the minimum mean loss, $\bar{L}_{min}$. The solutions given for both matrices are correct; however, when such an instability exists, the validity of the problem itself should be questioned. The fact that the entries of the payoff matrix are not accurately known in many practical situations would render worthless a problem statement for which the solution was critically dependent upon the precise values of the entries. Furthermore, since numbers are not stored in digital computers without roundoff, the actual values of payoffs are subject to small perturbations when adaptive processes are solved by digital computers. The solutions found in unstable cases would depend strongly upon the rule used to round off the numbers. This is obviously an undesirable situation.

The solution of unstable cases has not been investigated. The preceding problem illustrated both the difficulty of finding a solution when $Pr(v{=}v_{max}) > 0$ and $E_{max}(\text{col } 1) > v_{max}$ for several points in the constraint space, and how the payoff matrix can be perturbed to change the problem into one for which $Pr(v{=}v_{max}) = 0$. The matrix can also be modified by adding ϵ to a_{12}; this changes the problem to one for which $Pr(v{=}v_{max}) > 0$ and $E_{max}(\text{col } 1) > v_{max}$, but for which there is a single point in the constraint space ($p_{opt} = (1, 0)$). The discussion in Appendix VI shows the relationships among the three cases just mentioned. It points out that the unstable case is the boundary case between the other two situations.

The fundamental result that has been obtained here, aside from complete techniques for finding optimum strategies for all but the unstable cases, is the demonstration that if p is located inside the constraint space, $\bar{L}(p)$ is bounded (unless, perhaps $p_1=0$), and $\bar{L}_{min}$ is always bounded if all possible payoffs are bounded.

4.3 Multiple Unknown Payoffs

Let a_{11} and a_{22} be the unknown payoffs in a problem of adaptive decision under uncertainty. Players A and B are assumed to select the optimum strategies for the problem of adaptive decision with a single unknown payoff that remains after one of the two unknown payoffs is received. $L_{min}|a_{11}$ represents a loss function for the decision process that remains after a_{11} has been received. It is a function of both a_{11} and a_{22}. Player A is assumed to select the optimum piecewise-stationary strategy for the problem with the single unknown payoff a_{22}; that is, A selects p_{opt}, the value of p that minimizes $\bar{L}(p)$ for this problem. p_{opt} is a function of a_{11}. Player B selects q to maximize $L(p_{opt})$. Therefore, the loss function $L_{min}|a_{11}$ is identical to the quantity that was denoted $L_{max}(p_{opt})$ in section 4.2. The quantity $L_{min}|a_{22}$ is defined in an analogous fashion.

Arguments similar to those employed in section 4.2 can be used to demonstrate that B's optimum strategy is piecewise-stationary. If it is possible for A to select p so that r = v for all possible values of (a_{11}, a_{22}), then $L_{max} = 0$. Otherwise, $L_{max} > 0$, and it is necessary to proceed with the following analysis. It is not difficult to show that the total loss can be written

$$L = \begin{cases} \dfrac{v - r + p_1q_1L_{min}|a_{11} + p_2q_2L_{min}|a_{22}}{p_1q_1 + p_2q_2} & \text{if the numerator} \neq 0 \\[2em] 0 & \text{if the numerator} = 0. \end{cases}$$

A result of Appendix III leads to this expression for the maximum value of the total loss:

$$L_{max} = \max\left[\frac{v - E(\text{col } 1) + p_1 L_{min}|a_{11}}{p_1}, \frac{v - E(\text{col } 2) + p_2 L_{min}|a_{22}}{p_2}, \frac{v - E(\text{col } 3)}{0}, \ldots, \frac{v - E(\text{col } n)}{0}\right],$$

where the same meaning is given to this expression as was given to Eq. 9. In order to guarantee that the mean value of L_{max} is finite, player A must select a p vector that satisfies the following inequalities:

$$\text{a)} \quad \sum_{i=1}^{m} p_i a_{ij} \geqslant v_{max} \qquad \text{for } j = 3, \ldots, n,$$

$$\text{b)} \quad \sum_{i=1}^{m} p_i a_{i1} \geqslant v_{max} \qquad \text{if } p_1 = 0,$$

$$\text{c)} \quad \sum_{i=1}^{m} p_i a_{i2} \geqslant v_{max} \qquad \text{if } p_2 = 0, \tag{15}$$

$$\text{d)} \quad \sum_{i=1}^{m} p_i = 1,$$

$$\text{e)} \quad p_i \geqslant 0 \qquad \text{for } i = 1, \ldots, m,$$

where $v_{max} = \max_{a_{11}, a_{22}} v(a_{11}, a_{22})$. Inequalities (15a), (15d), and (15e) describe the constraint space for this problem. There exists at least one p vector in the constraint space for which L_{max} is finite for all possible payoffs. Thus the mean value of L_{max} is finite for some point in the constraint space. If p lies in the constraint space and if $Pr(v{=}v_{max}) = 0$ or $Pr(v{=}v_{max}) > 0$ but either $v_{max} \geqslant E(\text{col } 1)$ or $v_{max} \geqslant E(\text{col } 2)$, then

$$L_{max} = \max\left[\frac{v - E(\text{col } 1) + p_1 L_{min}|a_{11}}{p_1}, \frac{v - E(\text{col } 2) + p_2 L_{min}|a_{22}}{p_2}\right] \tag{16}$$

for all possible values of (a_{11}, a_{22}). The mean value of L_{max}, with respect to a_{11} and a_{22}, is the mean loss:

$$\bar{L}(p) \equiv \int L_{max}\, dP(a_{11}, a_{22}).$$

This cannot be expressed in a neat form like that of Eq. 12. The minimum value of $\bar{L}$

with respect to p is denoted $\bar{L}_{\min}$:

$$\bar{L}_{\min} \equiv \min_{p} \bar{L}(p) \equiv \bar{L}(p_{\text{opt}}).$$

When $\Pr(v=v_{\max}) = 0$ or $\Pr(v=v_{\max}) > 0$ but either $v_{\max} \geqslant E(\text{col } 1)$ or $v_{\max} \geqslant E(\text{col } 2)$ at all vertices of the constraint space and for all (a_{11}, a_{22}) for which $v(a_{11}, a_{22}) = v_{\max}$, then Eq. 16 may be used to calculate L(p). The example worked out in Appendix VII is a situation for which $\Pr(v=v_{\max}) > 0$, but Eq. 16 can be used; it also illustrates some concepts mentioned briefly in this section.

This example is for a case in which the mean loss is smaller at some point inside the constraint space than it is at any of the vertices of the space; therefore, the minimum mean loss occurs for some vector p_{opt} that does not lie at one of the vertices. Because it is not possible to find p_{opt} and $\bar{L}_{\min}$ by calculating $\bar{L}(p)$ at the vertices of the constraint polyhedron, it may not be possible to find p_{opt} and $\bar{L}_{\min}$ by means of a finite number of computations, as it was in the case of a single unknown payoff. Because of the relatively complex form of the expression for $L_{\max}$ in Eq. 16, it is still not possible to find a simple algorithm that can be used to locate p_{opt} with two unknown payoffs. It is always possible to follow some minimum-seeking technique and search the constraint space, calculating $\bar{L}(p)$ at each examined point. However, the lengthy computation required to find $\bar{L}(p)$ for one point may condemn a minimum-hunting method that involves a large number of trial calculations. One means of compromise is to select any point in the constraint space (for which inequalities (15b) and (15c) are also satisfied) until a_{11} or a_{22} is received and to use the optimum strategy for the process with a single unknown payoff from then on. This, of course, is not optimum, but it guarantees a finite loss with a reasonable amount of computation.

The cases for which Eq. 16 may not be a valid expression for $L_{\max}$ have not been considered. These are cases in which $\Pr(v=v_{\max}) > 0$. When $v = v_{\max}$ it is possible that $v_{\max}$ is less than both E(col 1) and E(col 2) at some vertex of the constraint space. Then $L_{\max} = 0$, and Eq. 16 is invalid. Nevertheless, computation of $\bar{L}(p)$ in these cases is not much more difficult than for those cases in which Eq. 16 can be used.

The preceding discussion concerned unknown payoffs in different columns of the payoff matrix. In the following analysis of several unknown payoffs in the same column, many details have been omitted in order to present an interesting sidelight in compact form. If a_{11} and a_{21} are the unknown payoffs, it can be shown that

$$L = \frac{v - r + p_1 q_1 L_{\min}|a_{11} + p_2 q_1 L_{\min}|a_{21}}{(p_1+p_2)\, q_1},$$

$$L_{\max} = \max \left[\frac{v - E(\text{col } 1) + p_1 L_{\min}|a_{11} + p_2 L_{\min}|a_{21}}{(p_1+p_2) q_1}, \frac{v - E(\text{col } 2)}{0}, \ldots, \frac{v - E(\text{col } n)}{0} \right].$$

If p lies in the constraint space defined by

$$\sum_{i=1}^{m} p_i a_{ij} \geqslant v_{\max} \qquad \text{for } j = 2, \ldots, n,$$

$$\sum_{i=1}^{m} p_i = 1,$$

$$p_i \geqslant 0 \qquad \text{for } i = 1, \ldots, m,$$

then

$$L_{\max} = \frac{v - E(\text{col } 1) + p_1 L_{\min}|a_{11} + p_2 L_{\min}|a_{21}}{p_1 + p_2}.$$

Furthermore, the mean loss is the average of $L_{\max}$ with respect to a_{11} and a_{21}:

$$\bar{L} = \frac{\bar{v} - \bar{E}(\text{col } 1) + p_1 \overline{L_{\min}|a_{11}} + p_2 \overline{L_{\min}|a_{21}}}{p_1 + p_2}, \tag{17}$$

where

$$\overline{L_{\min}|a_{i1}} = \int L_{\min}|a_{i1} \, dP(a_{11}, a_{21}).$$

Equation 17 indicates that $\bar{L}$ assumes its minimum value at one of the vertices of the constraint space. Therefore, this case is not essentially different from the case with a single unknown payoff, and p_{opt} and $\bar{L}_{\min}$ can be found by using the techniques described before. Furthermore, this situation generalizes to any number of unknown payoffs in one column of the payoff matrix. If the unknown payoffs are all in the same row, then the problem does not differ essentially from the case with two unknown payoffs in different rows and different columns. In the case of three or more unknown payoffs there exists a constraint space of the same form as that for two unknown payoffs, except that inequalities (15a) are applied only to the columns of the payoff matrix in which no unknown payoffs are located. If A selects a p vector that lies inside the constraint space and that satisfies inequalities analogous to (15b) and (15c), then he guarantees that the mean loss is finite. When all columns contain an unknown payoff, the constraint space consists of all legitimate probability vectors.

An essential result of this work is the concept of a constraint space. This space is a function of the maximum values of the unknown payoffs. If player A selects a piecewise-stationary strategy corresponding to a point in the constraint space, he guarantees that his mean loss will be finite. (Other minor constraints must also be satisfied.) That is, he guarantees that either he learns one of the unknown payoffs in a finite number of steps, or his total loss is less than or equal to zero. If player B were an intelligent player and A promised to select a p vector in the constraint space, then, in general, B's optimum strategy would be to select repeatedly the alternative corresponding to one of the columns of the payoff matrix that contain unknown payoffs. As a result, the minimum mean loss $\bar{L}_{min}$ found for a problem of decision under uncertainty represents an upper bound to the loss that player A must sustain. It has been assumed for this problem that nature knows all of the unknown quantities and selects her strategy to maximize A's total loss. If A is opposed by an intelligent opponent who knows all of the payoffs, then $\bar{L}_{min}$ for the adaptive decision under uncertainty problem is an upper bound to the mean loss that player A sustains if he plays intelligently; and if A is opposed by a player who does not know all of the payoffs, $\bar{L}_{min}$ is an upper bound to A's mean loss.

V. ADAPTIVE COMPETITIVE DECISION

There are two subclasses of the adaptive competitive decision problem: equal information, and unequal information. Two players playing the game illustrated in Fig. 1 with equal information would both be given the same probability distribution for the value of the unknown payoff. This case could arise if the aggressor had an untested warhead about which the defender knew as much as the aggressor. The equal-information case also applies to situations involving several payoffs unknown to both players. In the unequal-information case the players are given different probability distributions for the values of the unknown payoffs, or else the sets of payoffs unknown to the two players are different. (The latter condition is actually a special case of the former condition.) When the aggressor knows the true value of a_{11}, the process in Fig. 1 is an unequal-information situation. The meaning of unequal-information situations and a discussion of how they can arise are considered in section 5.2. The problem of adaptive competitive decision with equal information is solved first.

5.1 Equal Information

It will be demonstrated that by using an optimum piecewise-stationary strategy player A can guarantee that no matter how B plays, A's mean loss is no greater than a finite number $\bar{L}_{opt}$; and B has an optimum piecewise-stationary strategy, the use of which ensures that no matter what strategy A uses, A's mean loss is at least $\bar{L}_{opt}$. (The preceding statement is not rigorously true; nevertheless, the detailed proof that follows demonstrates that the statement is essentially correct.) That is,

$$\min_{S_A} \max_{S_B} \bar{L} = \max_{S_B} \min_{S_A} \bar{L} = \bar{L}_{opt},$$

where S_A and S_B represent all possible mixed strategies (not necessarily stationary) of A and B, respectively. $\bar{L}_{opt}$ may be negative as well as positive or zero.

That a minimax strategy exists for adaptive competitive decision processes with equal information, although not unexpected, is significant. Any adaptive competitive decision process can be formulated as an infinite, two-person, zero-sum game, begun with a random move by the referee and continued with moves by players A and B alternated ad infinitum. This aspect of the problem is discussed in section 5.2. Infinite games, however, do not always possess minimax solutions. (See, for example, the infinite game discussed in Appendix I.) Nevertheless, there exists a group of problems involving the repeated play of certain game matrices; these problems are infinite games, and they have minimax solutions.[20, 27, 32] Adaptive competitive decision processes are related to these infinite games.

The definition of the single-step loss function at the k^{th} step is

$$L^k \equiv v - r^k,$$

where v is the minimax value of the payoff matrix. The loss for N steps of the decision process, L_N, is

$$L_N \equiv \sum_{k=1}^{N} L^k,$$

and the mean loss for N steps is

$$\overline{L}_N \equiv \int L_N(\text{unknown payoffs})\, dP(\text{unknown payoffs}).$$

The limit of $\overline{L}_N$ as N approaches infinity, if it exists, is the measure of performance that player A attempts to minimize and B to maximize. This measure, called the mean loss, is slightly different from the measures of performance used in Sections III and IV. The fundamental reason for this departure is that the mean loss defined there may not exist in the competitive process. The total loss is normally defined as

$$L = \sum_{k=1}^{\infty} L^k; \tag{18}$$

however, because L^k may be positive or negative in competitive situations, the sum in Eq. 18 may not converge to a finite number or to infinity. Thus, the mean value of L may not exist. (L^k is non-negative in adaptive Bayes decision or adaptive decision under uncertainty, so in these cases L always converges to a non-negative number or $+\infty$.) However, it is shown in Appendix VIII that the limit of $\overline{L}_N$ is a useful measure of performance. The results of Appendix VIII will be discussed after the following intuitive argument, which presents the essence of that material.

Consider, again, the matrix of Fig. 1. In this situation $\overline{v} = -1$, and $\overline{a}_{11} = -2.5$. Consider the auxiliary game defined by the matrix

$$\begin{bmatrix} \overline{v} - \overline{a}_{11} & \overline{v} - a_{12} + \overline{L} \\ \overline{v} - a_{21} + \overline{L} & \overline{v} - a_{22} + \overline{L} \end{bmatrix} = \begin{bmatrix} 1.5 & \overline{L} - 1 \\ \overline{L} - 1 & \overline{L} + 1 \end{bmatrix} \tag{19}$$

What relationship does this matrix have to the original process? Let $\overline{L}$ be the mean loss associated with the competitive decision process, if a mean loss exists. When one step of the process is played, the mean loss for that step is $\overline{v} - \overline{a}_{11}$ if a_{11} is received, and $\overline{v} - a_{ij}$ if a_{11} is not received. In the former case, the process terminates with no further loss to player A because both players should use optimum strategies from the second step on. If a_{11} is not received, the mean loss from the second step on is $\overline{L}$ because the situation faced by the players is identical to the situation faced by them before the first step. This reasoning leads to matrix (19). Player A should use a mixed strategy (which is a function of $\overline{L}$) that minimizes the expected payoff from matrix (19), and B should maximize the expected payoff because the payoff represents a loss to A and a gain to B. The minimax value for matrix (19) when A is the minimizing player and B the maximizing player is

$$v(\bar{L}) = \begin{cases} \dfrac{-(\bar{L})^2 + 3.5\bar{L} + .5}{-\bar{L} + 4.5} & \text{if } \bar{L} \leqslant 2.5 \\ \\ \bar{L} - 1 & \text{if } \bar{L} \geqslant 2.5 \end{cases} \tag{20}$$

If (19) truly represents the infinite process, then the minimax value of that matrix should equal the mean loss for the process:

$$v(\bar{L}) = \bar{L}. \tag{21}$$

In other words, players A and B should be as willing to make one play of the game specified by (19) as they are to participate in the original decision process, if the value of $\bar{L}$ that satisfies Eq. 21 is used for $\bar{L}$ in matrix (19). That value is called the optimum mean loss, $\bar{L}_{\text{opt}}$. For the value function of Eq. 20, $\bar{L}_{\text{opt}} = 0.5$. When 0.5 is substituted for $\bar{L}$ in matrix (19), the minimax strategies are found to be

$$p_o = (1/2, 1/2),$$

$$q_o = (1/2, 1/2).$$

Because neither player's information about the game changes from one play to the next, it seems reasonable that the optimum strategies for both players are piecewise stationary. Assume that A uses the piecewise-stationary strategy that begins with distribution p_o and continues with the optimum distribution after a_{11} has been received. The arguments of Appendix II can be applied to this situation in order to demonstrate that B's optimum piecewise-stationary strategy yields the maximum mean loss. The mean loss for a piecewise-stationary strategy of B that begins with distribution q is written

$$L = \begin{cases} \dfrac{v - \left(p_{1o}q_1\bar{a}_{11} + \displaystyle\sum_{(i,j)\neq(1,1)} p_{io}q_j a_{ij}\right)}{p_{1o}q_1} & \text{if the numerator} \neq 0 \\ \\ 0 & \text{if the numerator} = 0 \end{cases}$$

$$= \begin{cases} \dfrac{-1 - (-1.25q_1 - 1q_2)}{.5q_1} & \text{if the numerator} \neq 0 \\ \\ 0 & \text{if the numerator} = 0 \end{cases} = \begin{cases} .5 & \text{if } q_1 \neq 0 \\ \\ 0 & \text{if } q_1 = 0. \end{cases}$$

Therefore $\max\limits_{q} \bar{L} = 0.5$. This result agrees with the preceding analysis. Now the same procedure is followed for q_o when A uses a piecewise-stationary strategy:

$$\bar{L} = \begin{cases} \dfrac{\bar{v} - \left(p_1 q_{1o} \bar{a}_{11} + \displaystyle\sum_{(i,j)\neq(1,1)} p_i q_{jo} a_{ij}\right)}{p_1 q_{1o}} & \text{if the numerator} \neq 0 \\ 0 & \text{if the numerator} = 0 \end{cases}$$

$$= \begin{cases} \dfrac{-1 - (-1.25p_1 - p_2)}{.5p_1} & \text{if the numerator} \neq 0 \\ 0 & \text{if the numerator} = 0 \end{cases} = \begin{cases} .5 & \text{if } p_1 \neq 0 \\ 0 & \text{if } p_1 = 0. \end{cases}$$

Therefore $\min_p \bar{L} = 0$. This is a surprising result, since it was expected that player B could guarantee that the mean loss to player A would equal at least .5, by using q_o. But suppose B's piecewise-stationary strategy begins with a small perturbation of q_o:

$$q = (1/2-\epsilon, 1/2+\epsilon),$$

where $0 < \epsilon \ll 1$. In this case the mean loss is

$$\bar{L} = \begin{cases} \dfrac{-1 - [-2.5(.5-\epsilon)p_1 - 2(.5+\epsilon)p_2]}{(.5-\epsilon)p_1} & \text{if the numerator} \neq 0 \\ 0 & \text{if the numerator} = 0. \end{cases}$$

It can be shown that the numerator is always positive, and

$$\min_p \bar{L} = \frac{.25 - 2.5\epsilon}{.5 - \epsilon} \approx .5 - 4\epsilon.$$

Therefore, player B can guarantee that the mean loss is as near .5 as he desires, but he cannot actually guarantee a loss of 0.5.

It is now time to state the complete results of Appendix VIII. The proof is modeled after a proof by H. Everett.[33]

All points on the real axis can be divided into two sets on the basis of the value of the auxiliary game matrix:

$$\begin{aligned} &\bar{L} \text{ is in Set 1 if } \begin{cases} v(\bar{L}) < \bar{L} & \text{and } \bar{L} < 0, \text{ or} \\ v(\bar{L}) \leqslant \bar{L} & \text{and } \bar{L} \geqslant 0; \end{cases} \\ &\bar{L} \text{ is in Set 2 if } \begin{cases} v(\bar{L}) > \bar{L} & \text{and } \bar{L} > 0, \text{ or} \\ v(\bar{L}) \geqslant \bar{L} & \text{and } \bar{L} \leqslant 0. \end{cases} \end{aligned} \qquad (22)$$

(In the preceding example — see Eq. 20 — Set 1 = $[.5, +\infty]$, and Set 2 = $[-\infty, .5)$.) If a quantity $\bar{L}$ lies in Set 1 and a minimax strategy for player A is found when that value of $\bar{L}$ is substituted in the auxiliary matrix, then by using the minimax distribution repeatedly, player A can guarantee that no matter what strategy B uses, either

$$\lim_{N\to\infty} \bar{L}_N \leqslant \bar{L}$$

if the limit exists, or there is an integer N_o that is such that $\bar{L}_N \leqslant \bar{L}$ if $N > N_o$. If a value of $\bar{L}$ is in Set 2 and the minimax strategy for B is found when that value of $\bar{L}$ is substituted in the auxiliary matrix, then by using the minimax distribution repeatedly, Player B can guarantee that no matter what strategy A uses, either

$$\lim_{N\to\infty} \bar{L}_N \geqslant \bar{L}$$

if the limit exists, or there is an integer N_o that is such that $\bar{L}_N \geqslant \bar{L}$ if $N > N_o$. The essence of this result is that A can guarantee that his loss is no greater than $\bar{L}$ if $\bar{L}$ is in Set 1, and B can guarantee that A's loss is no less than $\bar{L}$ if $\bar{L}$ is in Set 2. (In the preceding example, A can guarantee that his mean loss is no greater than $\bar{L}$ for any $\bar{L}$ greater than or equal to .5, and B can guarantee that A's loss is no less than $\bar{L}$ for any $\bar{L}$ less than 0.5. In order to minimize his mean loss, A should choose the minimax distribution corresponding to $\bar{L} = .5$, and B should choose a minimax distribution for some $\bar{L}$ arbitrarily close to (but less than) 0.5. This agrees with the preceding analysis.)

A further result of Appendix VIII is that there always exists a unique value of $\bar{L}_{opt}$ that is such that all points greater than $\bar{L}_{opt}$ are in Set 1 and all points less than $\bar{L}_{opt}$ are in Set 2. Furthermore, $v(\bar{L}_{opt}) = \bar{L}_{opt}$, although there may be more than one value of $\bar{L}$ that satisfies Eq. 21. Figure 2 illustrates a case in which $v(\bar{L}) = \bar{L}$ for all values of $\bar{L}$ between $\bar{L}_a$ and $\bar{L}_b$. In Fig. 2, Set 1 = $[\bar{L}_a, +\infty]$, and Set 2 = $[-\infty, \bar{L}_a)$ because $\bar{L}_a > 0$; therefore, $\bar{L}_{opt} = \bar{L}_a$.

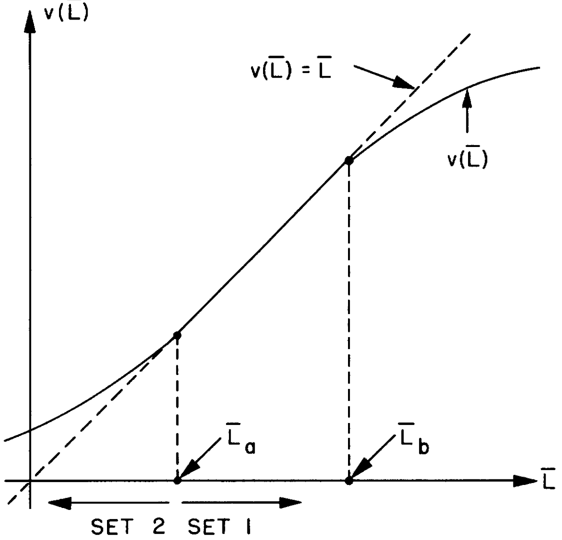

Fig. 2. Typical curve for $v(\bar{L})$ vs $\bar{L}$.

Machine computation of $\bar{L}_{opt}$ is a reasonably straightforward process. Because of the general shape of the curve $v(\bar{L})$, an iterative technique can always be used to find $\bar{L}_{opt}$. For example, in Fig. 2 if some initial point, $\bar{L} > \bar{L}_b$, is chosen, then $\bar{L}' = v(\bar{L})$ lies between $\bar{L}$ and $\bar{L}_b$. Also $\bar{L}'' = v(\bar{L}')$ lies between $\bar{L}'$ and $\bar{L}_b$, and so on. The point $\bar{L}^{(n)}$ converges to $\bar{L}_b$. Furthermore, if $\bar{L} < \bar{L}_a$, then $\bar{L}' = v(\bar{L})$ lies between $\bar{L}$ and $\bar{L}_a$, and $\bar{L}^{(n)}$ converges to $\bar{L}_a$. Once the points $\bar{L}_a$ and $\bar{L}_b$ have been determined, we know that $v(\bar{L}) = \bar{L}$ for all $\bar{L}$ between $\bar{L}_a$ and $\bar{L}_b$; therefore, it is an easy matter to apply Eq. 22 to find what point $\bar{L}_{opt}$ lies on the boundary between Sets 1 and 2. This latter part of the analysis is unnecessary if there is a single point that satisfies Eq. 21, as there was in the preceding example.

The fact that $\bar{L}_{opt}$ is finite follows from the proof that $\bar{L}_{min}$ is finite in the problem of adaptive decision under uncertainty. That is, there exists a piecewise-stationary strategy for player A which guarantees that his mean loss is finite. As a result, $\bar{L} < +\infty$. Similarly, if B is the player of an adaptive decision under uncertainty process, by using the optimum piecewise-stationary strategy, he can guarantee that $\bar{L} > -\infty$. Therefore, $-\infty < \bar{L}_{opt} < +\infty$. This result is based upon the assumption that the possible values of the unknown payoffs are bounded above and below.

In the case of two unknown payoffs, it is assumed that once one of them is received both players play in an optimum fashion; therefore, the remaining mean loss to player A is $\bar{L}_{opt}$ for the resultant adaptive competitive decision problem with a single unknown payoff. The following auxiliary matrix is used to find the optimum mean loss, $\bar{L}_{opt}$, for the two unknown payoff case (in a 3 X 3 process with unknown payoffs a_{11} and a_{22}):

$$\begin{bmatrix} \bar{v} - \bar{a}_{11} + \overline{L_{opt}|a_{11}} & \bar{v} - a_{12} + \bar{L} & \bar{v} - a_{13} + \bar{L} \\ \bar{v} - a_{21} + \bar{L} & \bar{v} - \bar{a}_{22} + \overline{L_{opt}|a_{22}} & \bar{v} - a_{23} + \bar{L} \\ \bar{v} - a_{31} + \bar{L} & \bar{v} - a_{32} + \bar{L} & \bar{v} - a_{33} + \bar{L} \end{bmatrix}.$$

$\bar{L}_{opt}|a_{11}$ denotes the optimum mean loss for the case of a single unknown payoff, a_{22}, when a_{11} is known; $\overline{L_{opt}|a_{11}}$ denotes the mean value of $\bar{L}_{opt}|a_{11}$, with respect to the probability distribution of a_{11}:

$$\overline{L_{opt}|a_{11}} = \int \bar{L}_{opt}|a_{11} \, dP(a_{11}).$$

$\overline{L_{opt}|a_{22}}$ is defined similarly. Sets 1 and 2 can be defined by Eq. 22. Then there is an optimum mean loss, $\bar{L}_{opt}$, which is the greatest lower bound of Set 1 and the least upper bound of Set 2. For any value of $\bar{L}$ in Set 1, player A can use a piecewise-stationary strategy to guarantee that no matter how B plays, either

$$\lim_{N\to\infty} \bar{L}_N \leq \bar{L}$$

if this limit exists, or else there is an integer N_o that is such that $\bar{L}_N \leq \bar{L}$ if $N > N_o$. An analogous statement can be made about values of $\bar{L}$ in Set 2. In other words, all of the conclusions drawn about the single unknown payoff case are true for the two unknown payoff case. The derivation of these results is not given, but it is a very simple extention of Appendix VIII.

When r payoffs are unknown (r > 1), an auxiliary matrix is used with entries $\bar{v} - \bar{a}_{ij} + \overline{L_{opt} | a_{ij}}$ corresponding to the r unknown payoffs, and $\bar{v} - a_{ij} + \bar{L}$ as the remaining entries. $\bar{L}_{opt} | a_{ij}$, which is a function of a_{ij}, is the optimum mean loss for the problem with the r - 1 other unknown payoffs when a_{ij} is known, and $\overline{L_{opt} | a_{ij}}$ is the mean value of $\bar{L}_{opt} | a_{ij}$ with respect to the probability distribution for a_{ij}.

The labor involved in the computation of the optimum mean loss and the optimum strategies becomes forbidding rapidly as r increases. This is also true with adaptive Bayes decision (except when the abbreviated method can be used to eliminate much of the labor) and adaptive decision under uncertainty.

5.2 Unequal Information

The concept of competitive decision with unequal information requires some explanation. It seems reasonable that situations can occur in which some payoffs are unknown to one player, other payoffs are unknown to the other player, and some payoffs are unknown to both players with the same probability distributions given to both. However, cases in which the players are given different probability distributions for the same unknown payoffs may appear unrealistic. The example that follows demonstrates how such a situation can logically arise.

Two biased coins are flipped by the referee. The first has probability p of landing heads up; the second, q. Unknown payoff a_{11} has value +a if both coins land with the same side up and -a if the coins land with different sides up. Player A is shown the outcome of the toss of the first coin, and player B that of the second coin. The complete process may be tabulated as follows:

Outcome	Probability	a_{11}	$Pr_A(a_{11}=+a)$	$Pr_B(a_{11}=+a)$
HH	pq	+a	q	p
HT	p(1-q)	-a	q	1 - p
TH	(1-p)q	-a	1 - q	p
TT	(1-p)(1-q)	+a	1 - q	1 - p,

where $Pr_A(a_{11}=+a)$ denotes A's a posteriori probability that $a_{11} = +a$ after he is told about the outcome of the toss of the first coin. $Pr_B(a_{11}=+a)$ has a similar meaning. Notice that the quantities $Pr_A(a_{11}=+a)$ and $Pr_B(a_{11}=+a)$ can be different from each other for all four outcomes.

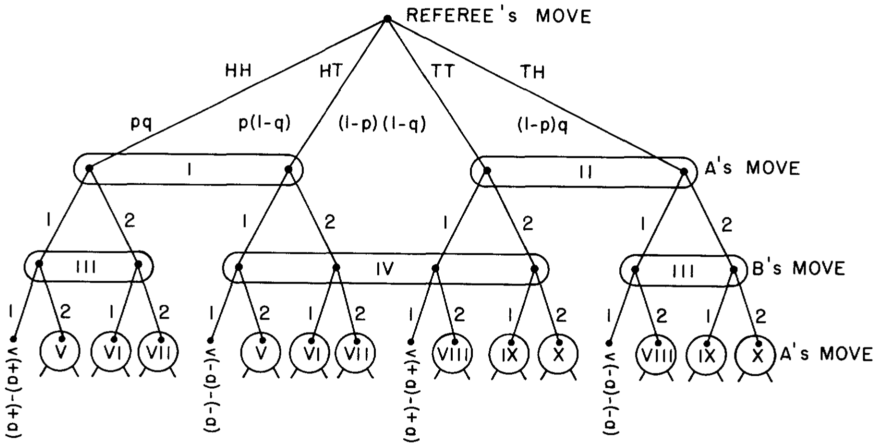

Fig. 3. Part of tree for coin-tossing game with unequal information.

Figure 3 illustrates the first part of the game tree for an adaptive competitive decision process based upon this example. It is assumed, for convenience, that the payoff matrix is 2 × 2 with a single unknown payoff a_{11}. Roman numerals denote information sets; all nodes indexed by the same Roman numeral are indistinguishable. For example, if the second coin lands heads up, player B does not know from which of the four nodes labeled III he is making his choice at his first move. These four nodes are all in the same information set, although in Fig. 3 they are separated into two groups for the sake of appearance. The four terminal nodes represent cases in which the unknown payoff is received and no further loss is sustained. The quantities written beneath these nodes are the total losses sustained by A.

Now consider a four-outcome process that is more general than the coin experiment. Each outcome i = 1, 2, 3 or 4 has an associated probability α_i. Let one of the values +a or −a be associated with each outcome. The outcomes may be divided into more general information sets for A and B than were the information sets of the coin-tossing problem. For every information set x of player A there exists an a posteriori probability $Pr_A(a_{11}=+a|x)$; similarly, $Pr_B(a_{11}=+a|y)$ exists for each information set y of player B. These quantities are all interrelated: $Pr(a_{11}=+a)$; α_i for i = 1, 2, 3, and 4; $Pr_A(a_{11}=+a|x)$ for all information sets x; and $Pr_B(a_{11}=+a|y)$ for all information sets y. We may ask whether $Pr(a_{11}=+a)$; α_i, and information sets for both players can be adjusted so that it is possible to arbitrarily pick four pairs of a posteriori probabilities $Pr_A(a_{11}=+a|i)$ and $Pr_B(a_{11}=+a|i)$ for the four outcomes i. ($Pr_A(a_{11}=+a|i)$ is A's a posteriori probability that a_{11} = +a after A is told of what information set, x, the outcome i that has occurred is a member.) The answer to our question is no; this cannot be done in general. Notice that the following relationship must be satisfied:

$$\sum_{i=1}^{4} \alpha_i Pr_A(a_{11}=+a|i) = \sum_{i=1}^{4} \alpha_i Pr_B(a_{11}=+a|i) = Pr(a_{11}=+a). \qquad (23)$$

This restricts the assignment of distribution pairs. In particular, assume that it is desired to make $\Pr_A(a_{11}=+a\,|x) = q$ for all information sets and $\Pr_B(a_{11}=+a\,|y) = p$ for all information sets. Then Eq. 23 implies

$$\sum_{i=1}^{4} \alpha_i q = \sum_{i=1}^{4} \alpha_i p = \Pr(a_{11}=+a).$$

In other words, p must equal q. The case in which each player has the same a posteriori probability distribution for all information sets, but the distributions of the two players are different is impossible. The only case in which each player can have the same a posteriori distribution for all information sets is the equal information case. The principles discussed in this paragraph can be generalized to apply to referee moves more general than the four-outcome process used for illustration.

The problem of adaptive competitive decision with unequal information is an infinite two-person, zero-sum game, since a game tree can be constructed. At the first move, made by the referee, the unknown payoffs are selected according to some initial probability distribution. If the number of possible outcomes for the referee's move is finite, then a finite number of branches leaves the first node; otherwise, an infinite number leaves. (The question of whether the continuous case can be approximated by an appropriately selected discrete case is not germane to the present discussion.) Let the second move be that of player A. (The choice is arbitrary.) If the payoff matrix for the process is $m \times n$, player A has m alternatives from all information sets. There is one information set for every possible outcome of the referee's random move about which A is informed. The next move belongs to B, and his information sets correspond similarly to the possible outcomes of the random move about which he is told. The game has a terminating node after B's move if all the unknown payoffs have been received, and the payoff equals the total loss that A sustains for the sequence of moves leading to that node. When the game does not terminate, each player has at least one more move. The information sets for A at the first move are subdivided to form the information sets for A's second move. Each subdivision corresponds to one of the $m \cdot n$ possible pairs of alternatives used by A and B at the previous two moves. (Some of these subdivisions are terminating nodes.) Player B's information sets at his first move are similarly subdivided. The construction of the game tree continues indefinitely in the same manner, with moves by A and B alternated.

Although infinite games do not necessarily possess minimax solutions, it was demonstrated in section 5.1 that adaptive competitive decision processes with equal information do have minimax solutions. Equal information situations are special cases of unequal information situations in which the a posteriori probability distributions of the unknown payoffs for both players are identical for all possible outcomes of the referee's move.

The solution to the following example represents the present extent of our knowledge about adaptive competitive decision with unequal information. The details of the solution are contained in Appendix IX. Only the results are presented here. Consider the following matrix in which payoff a_{11} is unknown to player A but known to B:

$$\begin{array}{cc|cc|} & & \multicolumn{2}{c}{B} \\ & & 1 & 2 \\ A & 1 & a_{11} & 0 \\ & 2 & 0 & -1 \end{array} \qquad P(a_{11}) = \begin{cases} 0 & \text{if } a_{11} < -1 \\ \frac{1}{2} & \text{if } -1 \leq a_{11} < +1 \\ 1 & \text{if } a_{11} \geq +1. \end{cases}$$

(a_{11} assumes values +1 and −1 with probability 1/2 each.)

Cumulative probability distribution function $P(a_{11})$ is known to both players, but B also knows the true value of payoff a_{11}.

An optimum strategy for player A is to use alternative 1 repeatedly until a_{11} is received and thereafter to use the correct minimax distribution repeatedly. An optimum strategy for player B is at the first step to use alternative 1 if a_{11} equals −1 or alternative 2 if a_{11} equals +1, and from the second step on to use the correct minimax distribution repeatedly. These are minimax strategies for the competitive process. This means that if A uses the strategy given above, the mean loss that he sustains is guaranteed to be no greater than 1/4 (the minimax mean loss). Similarly, if B uses the strategy given above, the mean loss sustained by A is no less than 1/4, no matter what strategy A uses.

The minimax strategy given for player B is not piecewise-stationary. A further result of Appendix IX is that there exists no piecewise-stationary strategy for B, the use of which guarantees that the mean loss is at least as large as 1/4.

The significant conclusions that can be drawn from this example are: (i) there exist examples of unequal information processes for which minimax solutions exist; and (ii) piecewise-stationary strategies may not be optimum when minimax solutions exist. Nothing more has been learned about the general existence of minimax strategies.

The following discussion is based upon pure hypothesis, with neither proofs to substantiate it nor counterexamples to contradict it. It seems reasonable in the case in which player B knows all of the payoffs unknown to A and their a priori probability distributions that B's optimum distribution at the first step should be a function of both the true values of the unknown payoffs and also their a priori probability distribution, and player A's optimum strategy should be a function of the a priori distributions. Because A knows the functional form of B's optimum strategy, A can calculate an a posteriori probability distribution for the unknown payoffs, based upon the knowledge of the alternative B used at the first step. He assumes that B always uses his optimum strategy. The a posteriori probability distribution may serve as the a priori distribution for the next step of the process. This procedure could be repeated until all of the unknown payoffs have been received. Note that if the optimum strategies are truly determined in the fashion just described, player A's distributions are dependent upon only the sequence of alternatives used by B in the past; and B's distributions are dependent upon only the true values of the unknown payoffs and the past history of his own choices. Neither player's distributions are influenced by the past history of A's choices. We hope that the preceding hypothesis will be verified soon. Once the problem of unequal information in which one player knows all of the unknown payoffs is solved, it should be possible to attack more general unequal information situations.

For want of a better strategy, the player of an unequal information process can use the optimum strategy derived in the section on adaptive decision under uncertainty (Section IV). This strategy may not be optimum for the problem of competitive decision; however, it is a strategy that guarantees a finite loss for the player, irrespective of his opponent's strategy.

VI. TOPICS FOR FURTHER STUDY

Several questions that have been raised in this report have not been answered, and there are related topics that have not been discussed which merit study. Because these questions have been discussed at some length, they will be mentioned very briefly here. First, there is the problem of adaptive decision under uncertainty with more than one unknown payoff. No simple algorithm has been found for determining the optimum strategies in these cases. We doubt that a neat method exists. The second unanswered question concerns adaptive competitive decision with unequal information. There appears to be a good chance that this problem can be solved completely. One problem confronting the investigator is the difficulty of working out examples. For instance, the simple example handled in Appendix IX was solved by taking a wild guess in extrapolating the solution of a two-step process to an infinite process. In order to calculate the optimum strategies for the two-step process, it was necessary to solve a 6 × 36 payoff matrix. This, in itself, was a huge task.

Three major problems to which the theory developed in this report should be extended are: (i) adaptive decision when nature is assumed to use an unknown stationary strategy; (ii) the processes in which the unknown payoffs are stochastic variables with unknown mean values; and (iii) the relationship between infinite processes and truncated processes.

Ways of viewing the problem of adaptive decision when nature uses an unknown stationary strategy were discussed in section 4.1. A paper by Hannan[30] attacks the problem of repeated decision-making when nothing at all is assumed about nature's strategy and all of the payoffs are known. There may be some merit in applying this approach to problems with unknown payoffs. The question of how to define a good measure of performance arises in many formulations of the problem of adaptive decision under uncertainty. Some suggestions to this effect were given in section 4.1.

The second suggestion – that processes in which the unknown payoffs are stochastic variables with unknown mean values be investigated – is a very significant one. Many of the problems of adaptive decision which arise in practice are of this type. For example, the unknown payoff in the problem of Fig. 1 may be a stochastc variable with unknown mean value. Payoff a_{11} represents the mean value of the gain to A if he allows an armed missile to reach his territory. If the warheads are not uniform, player A still does not know the mean value of the destructiveness after one armed missile has been allowed to explode. The more missiles he allows to explode, the more samples he will have from the distribution of the unknown payoff, and the better estimate he will have of the average value of the distribution, a_{11}. The coin-tossing problem of Robbins is a case in which the payoffs are stochastic variables with unknown mean values. Figure 4 is the payoff matrix for this problem, which is really a one-person game. The payoff for alternative 1 may be either 1, with probability p_1, or zero, with probability $1 - p_1$. A similar statement can be made about the payoff for alternative 2. The true

$$\begin{array}{cc} & B \\ A & \begin{array}{c} 1 \\ \\ 2 \end{array}\begin{bmatrix} p_1 \\ \\ p_2 \end{bmatrix} \end{array}$$

Fig. 4. Payoff matrix for Robbins' problem.

values of p_1 and p_2 are never actually learned by A.

The basic difference between the adaptive processes discussed in Sections III-V and processes in which the unknown payoff is a stochastic variable lies in the fact that there it was assumed that once an unknown payoff is received, the value of that payoff is known to both players from then on. If x is a stochastic variable, its mean value will never be truly known by the decision-makers; however, after many receipts of x, a good approximate mean value will be known. This situation is similar to the case in which repeated decisions are made when nature uses an unknown stationary strategy. After many steps of the process, nature's true probability distribution can be approximated very closely. Because the player can never be certain of the true mean value of an unknown payoff, he can never guarantee that his single-step loss is zero. Thus, if we apply the mean-loss measure of performance to stochastic problems, we may find that there exists no strategy for the player for which the mean loss is finite. In this case a better measure of performance must be found.

Since truly infinite processes rarely occur in practice, it is worth while to consider the relationship between the infinite processes discussed here and processes that last a finite number of steps. It would be valuable to know if the optimum strategies for the infinite processes yield "almost optimum" returns when applied to truncated processes of length N, where N is large. Since, in general, optimum strategies are more easily calculated for infinite processes than for truncated processes, it may be useful to apply the optimum strategies for infinite processes to truncated processes. An area of research that deserves attention is the study of how large N must be before the discrepancy (between the optimum loss and the loss when infinite process strategies are used) exceeds allowable limits. Appendix X is a small sample of the kind of work suggested. The optimum strategy is derived for an N-truncated adaptive Bayes decision process with a single unknown payoff. The essence of the result is that if N is large enough, the optimum strategy for the N-truncated process is the same as the optimum strategy for the infinite process.

VII. CONCLUDING REMARKS

In Sections I and II the adaptive decision process was introduced and the mean-loss measure of performance was defined. The concept of behavior strategy was shown to be useful, and it was demonstrated that the players of adaptive decision processes could be assumed to use behavior strategies, without destroying the generality of the problem.

Adaptive Bayes decision was discussed in Section III. Aside from the complete solution of the problem, a major result in that section is that the player's optimum piecewise-stationary strategy is the best of all possible strategies. An abbreviated technique for solving Bayes problems sheds much light on the meaning of the optimum solutions.

In Section IV a method of attack was proposed for the problem of adaptive decision under uncertainty. This approach was also found to be useful for problems of competitive decision with unequal information. The optimum strategy was derived for the case of a single unknown payoff, but no simple rule was found for determining optimum strategies in cases of more than one unknown payoff.

In Section V we demonstrated that a minimax solution exists for adaptive competitive decision problems with equal information. Optimum piecewise-stationary strategies were shown to be the best of all possible strategies. A technique was derived for the complete solution of this problem. The results of section 5.2 demonstrated that piecewise-stationary strategies are not necessarily optimum for adaptive competitive decision with unequal information. Furthermore, a formalization of the problem of unequal information was presented.

It must be admitted that the theory developed in this report does not enable the reader to solve a large class of practical problems. Nevertheless, it is hoped that this work has paved the way for further developments in the area of adaptive decision processes, which will be more generally applicable to practical problems. The ideas introduced here of how to tackle repetitive decision processes when the rewards are not completely known in advance and how to set up a reasonable measure of performance are basic to general adaptive processes, and should provide a foundation for future work.

APPENDIX I

A BRIEF INTRODUCTION TO THE THEORY OF GAMES

In general, any game can be represented by a tree. The tree in Fig. A-1 is drawn for a finite, two-person, zero-sum game. The game is finite because the tree is finite. It is a two-person game for two players and a referee. In this game player A will be considered to be a two-man team. The game is zero-sum because the payoffs at the terminal nodes represent a gain to player A and a loss to player B. For example, the payoff at the rightmost terminal node represents a gain of -2 units (a loss of 2 units) for A and a loss of -2 units (a gain of 2 units) for B. In a nonzero-sum, two-person game a payoff might be written (1, -2), which would mean a gain of one unit for A and a loss of 2 units for B, with the remaining unit absorbed by some neutral agent.

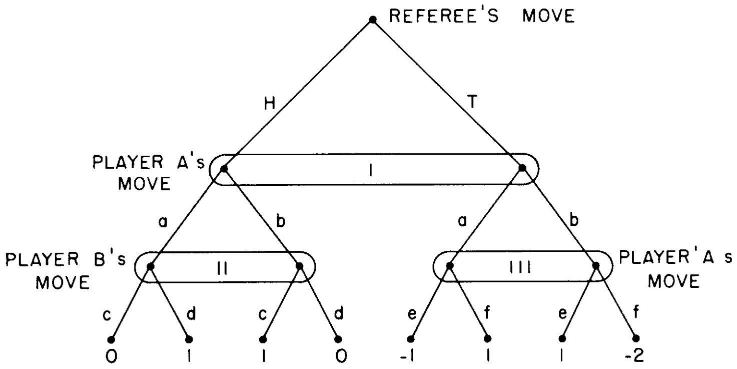

Fig. A-1. Tree for example game.

The play of the game of Fig. A-1 begins when the referee tosses an unbiased coin. The two branches leaving the top node represent this event. The first man of team A must then select alternative a or b. The line which encloses the two nodes that result from the coin toss indicates the fact that the first man of team A is not told about the outcome of the coin toss. Such a group of nodes that cannot be differentiated by a member of team A is called an <u>information set</u> of A; the set labeled I in the diagram is an information set of player A. After the first man of team A makes his choice, either the referee announces to player B that it is his turn to select alternative c or d, or else he tells the second man of team A that he must select alternative e or f. The referee's instructions correspond to the outcome of the coin toss. Information set II implies that B is not told about the outcome of A's preceding choice. Information set III implies that the second man of team A is not allowed to know the outcome of his teammate's preceding choice. The game then terminates, and team A is paid by B the amount indicated at the terminal node.

There are four possible courses of action that team A can take. These correspond to the four possible ways of assigning choices to information sets I and III: (a, e), (a, f), (b, e), and (b, f). These couples are called pure strategies of player A. Strategy (a, f), for example, represents the first man choosing a and the second man choosing f if he is given the opportunity to make a choice. B has two strategies: (c) and (d). A matrix that has an entry for every pair consisting of a pure strategy for A and a pure strategy for B is shown below. Each entry is the expected payoff to A and loss to B if the players use the corresponding pair of pure strategies. (Player A is by convention the maximizing player, and B the minimizing player.) For example, the entry at (a, e), (c) is -1/2, because if the coin lands head up then choices a and c result in a payoff of 0, and if the coin lands tail up the choices a and e result in payoff -1. Because the coin is unbiased, the expected payoff is -1/2. This 4 X 2 matrix is called the payoff matrix for the game of Fig. A-1. A finite payoff matrix can be constructed for any finite game. The tree formulation is called the extensive form of the game, while the equivalent matrix is called the normalized form.

$$\begin{array}{ccc} & & \text{B} \\ & & \begin{array}{cc} (c) & (d) \end{array} \\ \text{A} & \begin{array}{c} (a,e) \\ (a,f) \\ (b,e) \\ (b,f) \end{array} & \left[\begin{array}{cc} -1/2 & 0 \\ 1/2 & 1 \\ 1 & 1/2 \\ -1/2 & -1 \end{array}\right] \end{array}$$

If A decides to use either strategy (a, f) or (b, e), the actual strategy being chosen by a random process which assigns probability 1/2 to each strategy, then he is said to be using a mixed strategy with p = (0, 1/2, 1/2, 0). The p vector represents the probabilities that he assigns to the four pure strategies. Similarly, the mixed strategy that selects pure strategy (c) or (d) with equal likelihood is denoted by q = (1/2, 1/2).

When B uses mixed strategy q = (1/2, 1/2), the expected payoff to A is no greater than 3/4, irrespective of which pure strategy A uses. If A uses p = (0, 1/2, 1/2, 0), his expected return is no less than 3/4 for both pure strategies of B. (The expected payoff actually equals 3/4 for both of B's strategies.) This example illustrates a general theorem for finite, zero-sum, two-person, m X n games:

$$\max_p \min_q \sum_{i=1}^{m} \sum_{j=1}^{n} p_i q_j a_{ij} = \min_q \max_p \sum_{i=1}^{m} \sum_{j=1}^{n} p_i q_j a_{ij} = v, \qquad (A-1)$$

where p is a mixed strategy $(p_1, p_2, \ldots, p_m)$ for A, q is a mixed strategy $(q_1, q_2, \ldots, q_n)$ for B, and a_{ij} is the entry in the payoff matrix for pure strategy pair i and j.

Equation A-1 represents the minimax theorem. The meaning of this theorem is straightforward. If A assumes, when he selects his mixed strategy p, that B knows his selection and will use a mixed strategy q to minimize A's expected return, then the largest expected return A can guarantee is

$$v_1 = \max_p \left[\min_q \sum_{i=1}^{m} \sum_{j=1}^{n} p_i q_j a_{ij} \right].$$

Similarly, if B assumes that A knows the mixed strategy that B will use and maximizes the expected payoff for this strategy, then B can guarantee that A's expected return is no greater than

$$v_2 = \min_q \left[\max_p \sum_{i=1}^{m} \sum_{j=1}^{n} p_i q_j a_{ij} \right].$$

The minimax theorem states that $v_1 = v_2$. This common quantity is called v, the value of the game. Because A and B are assumed to be intelligent players, the assumption that each knows the optimum strategy of his opponent is valid. Therefore, a reasonable strategy for each player is the one that guarantees the minimax value of the payoff matrix. The optimum minimax strategies are accepted by most workers in the area of game theory as good strategies in competitive situations between two intelligent opponents.

Note that in the game of Fig. A-1 player A should never use strategy (a, e), because no matter what strategy B uses, A can do better by using strategy (a, f). This is reflected in the associated payoff matrix by the fact that both payoffs in the row of strategy (a, e) are smaller than the corresponding payoffs of row (a, f). Strategy (a, f) is said to dominate strategy (a, e). Strategy (b, e) dominates strategy (b, f), so A should never use strategy (b, f). There can also be column domination: if all of the payoffs in one column of a matrix are no less than the corresponding payoffs in another column, then the latter column dominates the former, and the strategy associated with the former column should never be used.

In general, infinite games do not have minimax solutions. Consider the game in which each of two players independently selects an integer. The player who selects the larger integer receives one unit of payoff from the player who selects the smaller integer. In case of a tie, the payoff is zero. The game is infinite because each player has an infinite number of pure strategies. There is no minimax value for this game. For any mixed strategy that A uses, there is a strategy for B that causes the expected payoff to be as close to -1 as desired, so v_1 equals -1; and for any mixed strategy used by B there exists a strategy for A that causes the payoff to be as near +1 as desired, so $v_2 = +1$. (In order to state precisely the idea of nonexistence of a minimax solution it is necessary to introduce infinums to replace the minimums in Eq. A-1 and supremums to replace the maximums. This formality has been omitted for the sake of brevity.)

APPENDIX II

PROOF THAT IN THE PROBLEM OF ADAPTIVE BAYES DECISION THE OPTIMUM PIECEWISE-STATIONARY STRATEGY IS THE OPTIMUM STRATEGY

Consider first the case of a single unknown payoff a_{11}. Let player A use a nonstationary behavior strategy. This means that A's sequence of probability distributions can be written $p^1, p^2, p^3, \ldots,$ where distribution p^2 is a function of the alternatives used by A and B at the first step, p^3 is a function of the alternatives used by A and B at the first and second steps, and so on. The mean loss to player A can be written

$$\bar{L} = (\bar{v}-\overline{r^1}) + \sum_{(i,j)\neq(1,1)} p_i^1 q_j \bar{L}(i,j). \tag{A-2}$$

The first term on the right-hand side of (A-2) represents the mean loss for the first step of the process. The second term represents the mean loss from the second step on. The quantity $\bar{L}(i,j)$ will be defined below. Equation A-2 is based upon the assumption that after he has received payoff a_{11}, player A will use an optimum strategy (which is stationary), so that there is no further loss. The second quantity on the right in (A-2) is the sum of mn−1 terms, each of which is the probability of occurrence of one particular pair of alternatives for players A and B at the first step times the mean loss from the second step on. $\bar{L}(i,j)$ represents the mean loss from the second step on for the sequence of probability distributions $p^2, p^3, \ldots,$ which is a function of (i,j), the pair of alternatives that has been used at the first step by A and B.

The first stage in this proof is to find

$$\bar{L}^*(1) = \min_{(i,j)\neq(1,1)} \bar{L}(i,j).$$

The pair of alternatives, (i_o, j_o) for which the minimum of $\bar{L}(i,j)$ occurs leads to a sequence of distributions $p^2, p^3, \ldots.$ In general, a different pair would lead to a different sequence. Suppose that after the first step player A assumes, for the purpose of calculating the sequence from the second step on, that the pair of alternatives, (i_o, j_o), has been used at the first step, no matter what pair actually has been chosen. Then the sequence of distributions, $p^2, p^3, \ldots,$ will be used after the first step that would have been used if the pair (i_o, j_o) had actually occurred. What will be the mean loss from the second step on? If B's distributions depended upon the outcomes of past steps, then the mean loss from the second step on resulting from distributions $p^2, p^3, \ldots$ would be a function of the true alternative pair, (i,j), used at the first step. However, because B's strategy is stationary, the mean loss from the second step on is $\bar{L}^*(1)$. Therefore, A can select a sequence of distributions from the second step on for which p^2 is independent of past history and $p^3, p^4, \ldots$ are dependent only upon the alternatives used after the first step, and for which the mean loss is no greater than the mean loss, $\overline{L}$, for A's original nonstationary strategy. To repeat: For any sequence of distributions for A there is an equally good or better sequence in which the second and later distributions are independent of the alternative pair used at the first step.

The next stage in the proof is to assume that A uses a sequence in which the distributions used at the first N steps are independent of past history and in which the distributions used at later steps are dependent only upon the alternatives actually chosen at the N^{th} step and later. It will then be proved that A can do as well or better by using a sequence in which the first N + 1 distributions are independent of the past history, and the later distributions are dependent upon only the alternatives actually chosen at the $N+1^{th}$ step and later. Since this has been proved already for N = 1, it will follow by induction that the conclusion is true for all N. Thus, in general, A can do as well or better than he can with a general nonstationary strategy by using a distribution sequence in which the probability distribution at any step is independent of past history. Such a sequence is called memoryless.

Note that under the assumption made in the preceding paragraph, the following is an expression of the mean loss:

$$\overline{L} = \sum_{k=1}^{N} (\overline{v}-\overline{r^k}) \prod_{t=0}^{k-1} (1-p_1^t q_1) + \sum_{(i,j)\neq(1,1)} p_i^N q_j \overline{L}(i,j), \qquad \text{(A-3)}$$

where $1 - p_1^o q_1$ is defined to be equal to 1. The first term on the right-hand side of (A-3) represents the mean loss for the first N steps, and the second term represents the mean loss from the $N+1^{th}$ step on when a_{11} has not yet been received and alternative pair (i, j) was chosen at the N^{th} step. Pair (i, j) determines sequence p^{N+1}, p^{N+2}, ... The following term is introduced:

$$\overline{L}^*(N) \equiv \min_{(i,j)\neq(1,1)} \overline{L}(i,j).$$

The arguments employed previously can be used to show that the mean loss sustained after the N^{th} step, if a_{11} has not been received, need not be greater than $\overline{L}^*$ (N). It follows that A can do as well or better than he could by using the original nonstationary sequence, by using a sequence in which all distributions at the first N + 1 steps are independent of past history, and the remaining distributions depend upon only the alternatives chosen at and after the $N+1^{th}$ step. This concludes the proof that to any general nonstationary strategy there corresponds a memoryless strategy for which the mean loss is no greater than the loss for the nonstationary strategy. The mean loss for a memoryless strategy can be written

$$\overline{L} = \sum_{k=1}^{\infty} (\overline{v}-\overline{r^k}) \prod_{t=0}^{k-1} (1-p_1^t q_1). \qquad \text{(A-4)}$$

Now consider the class of all possible memoryless strategies. Because the mean loss can never be less than zero, there exists a greatest lower bound for $\bar{L}$ over the set of all possible strategies. Call this bound L_o. Therefore, given any $\epsilon > 0$, there exists a strategy for which the mean loss is less than $L_o + \epsilon$. L_o is bounded if all possible payoffs are bounded; this fact is a result of the boundedness of $\bar{L}$ for the optimum stationary strategy, which was demonstrated in section 3.1. Either (a) L_o equals zero, or (b) L_o is greater than zero.

Case (a) is easily disposed of. The fact that the mean single-step loss is never less than zero and the assumption that the greatest lower bound of the mean loss is zero imply that the greatest lower bound of the mean single-step loss at the first step is zero. But the mean single-step loss is a continuous function of probability distribution p^1, and the set of all probability distributions is a compact set. Therefore there is some distribution for which the single-step loss is zero. If a stationary strategy that repeats this distribution is used, the mean loss for the infinite process is zero.

Case (b), $(L_o>0)$, is now considered. Given any $\epsilon > 0$, there exists a memoryless sequence p^1, p^2, ... for which the mean loss is less than $L_o + \epsilon$, and not less than L_o:

$$L_o + \epsilon > \bar{L} \geq L_o.$$

It will now be shown that the optimum piecewise-stationary strategy is at least as good as strategy sequence p^1, p^2, ... It was shown in section 3.1 that the optimum stationary strategy, p_{opt}, satisfies the relationship

$$\frac{\bar{v} - \bar{r}(p_{opt})}{p_{1\,opt}\,q_1} \leqslant \frac{\bar{v} - \bar{r}(p)}{p_1 q_1} \tag{A-5}$$

for all distributions p. The case in which it is possible to make r = v has also been covered. If inequality (A-5) is substituted in Eq. A-4, the following inequality results:

$$\bar{L} \geqslant \sum_{k=1}^{\infty} (\bar{v}-\bar{r}(p_{opt})) \frac{p_1^k q_1}{p_{1\,opt}\,q_1} \prod_{t=0}^{k-1} (1-p_1^t q_1)$$

$$= \frac{\bar{v} - \bar{r}(p_{opt})}{p_{1\,opt}\,q_1} \sum_{k=1}^{\infty} p_1^k q_1 \prod_{t=0}^{k-1} (1-p_1^t q_1).$$

The term $(\bar{v}-\bar{r}(p_{opt}))/p_{1\,opt}\,q_1$ equals the mean loss sustained by player A if he uses stationary strategy p_{opt}. The expression $p_1^k q_1 \prod_{t=0}^{k-1} (1-p_1^t q_1)$ equals the probability of receiving a_{11} at the k^{th} step; therefore, the expression $\sum_{k=1}^{\infty} p_1^k q_1 \prod_{t=0}^{k-1} (1-p_1^t q_1)$ is the probability of receiving a_{11} eventually. If it can be shown that the probability of receiving a_{11} eventually equals one, then $\bar{L} \geqslant \frac{\bar{v} - \bar{r}(p_{opt})}{p_{1\,opt}\, q_1}$. Because ϵ can be made as small as desired, it follows that the mean loss sustained by A if he uses the optimum piecewise-stationary strategy equals L_o.

The one unanswered question is, Does the probability equal one that a_{11} is received eventually, or is there a positive probability that a_{11} is never received? Note that the expected loss per step when a_{11} is unknown must be positive — otherwise L_o would be zero. Therefore, if the probability is positive that a_{11} is never known, the mean loss is infinite. But in section 3.1 it was shown that player A can attain a finite mean loss by using the optimum stationary strategy. Therefore, L_o, the greatest lower bound to L, is finite, and there is a contradiction to the assumption that $\bar{L}$ is within ϵ of L_o.

This theorem that the mean loss corresponding to the optimum piecewise-stationary strategy is no greater than the mean loss for any nonstationary strategy must now be generalized to the case of several unknown payoffs. The outline of the solution for the case with the two unknown payoffs a_{11} and a_{22} will illustrate the technique that can be used for the general proof. It is assumed that in order to minimize his loss, player A will always select the optimum piecewise-stationary strategy for the process that exists when only one unknown payoff remains. The mean losses $\overline{L_{min}|a_{11}}$ and $\overline{L_{min}|a_{22}}$ are defined in section 3.2. Then if any general nonstationary strategy is used by A, the mean loss can be written

$$\bar{L} = \bar{v} - \overline{r^1} + p_1^1 q_1 \overline{L_{min}|a_{11}} + p_2^1 q_2 \overline{L_{min}|a_{22}} + \sum p_i^1 q_j \bar{L}(i,j),$$

where the summation of the last term is understood to be over all pairs of alternatives (i, j), except (1, 1) and (2, 2). The reasoning used previously leads to the conclusion that there is a sequence in which distribution p^2 does not depend upon the outcome of the first step, and for which the mean loss is no greater than $\bar{L}$ for the original nonstationary strategy. Then the inductive argument is applied to the expression

$$\bar{L} = \sum_{k=1}^{N} \left(\bar{v} - \overline{r^k} + p_1^k q_1 \overline{L_{min}|a_{11}} + p_2^k q_2 \overline{L_{min}|a_{22}}\right) \prod_{t=0}^{k-1} \left(1 - p_1^t q_1 - p_2^t q_2\right) + \sum p_i^N q_j \bar{L}(i,j),$$

where $1 - p_1^o q_1 - p_2^o q_2$ is defined equal to 1. The reasoning does not differ from the reasoning for a single unknown payoff, and the conclusion follows that for every nonstationary strategy with loss L there exists a memoryless sequence of probability distributions with loss no greater than L. The mean loss for any memoryless strategy can be expressed as

$$\bar{L} = \sum_{k=1}^{\infty} \left(\bar{v} - \overline{r^k} + p_1^k q_1 \overline{L_{min}|a_{11}} + p_2^k q_2 \overline{L_{min}|a_{22}}\right) \prod_{t=0}^{k-1} (1 - p_1^t q_1 - p_2^t q_2). \qquad \text{(A-6)}$$

Next, we note that L_o, the greatest lower bound to $\overline{L}$ over the set of all memoryless sequences, is non-negative. If L_o equals zero, A can attain zero mean loss by using a stationary strategy; if L_o is positive, then there is some memoryless strategy that is such that $L_o + \epsilon > \overline{L} \geqslant L_o$, for any $\epsilon > 0$. The following inequality is true for the optimum stationary strategy, p_{opt}:

$$\frac{v - r(p_{opt}) + p_{1\,opt}\, q_1\, \overline{L_{min} | a_{11}} + p_{2\,opt}\, q_2\, \overline{L_{min} | a_{22}}}{p_{1\,opt}\, q_1 + p_{2\,opt}\, q_2}$$

$$\leqslant \frac{\bar{v} - \bar{r}(p) + p_1 q_1\, \overline{L_{min} | a_{11}} + p_2 q_2\, \overline{L_{min} | a_{22}}}{p_1 q_1 + p_2 q_2} \qquad \text{(A-7)}$$

for all distributions p. If inequality (A-7) is substituted in (A-6), the following inequality results:

$$\overline{L} \geqslant \frac{\bar{v} - \bar{r}(p_{opt}) + p_{1\,opt}\, q_1\, \overline{L_{min} | a_{11}} + p_{2\,opt}\, q_2\, \overline{L_{min} | a_{22}}}{p_{1\,opt}\, q_1 + p_{2\,opt}\, q_2}$$

$$\times \sum_{k=1}^{\infty} \left(p_1^k q_1 + p_2^k q_2\right) \prod_{t=0}^{k-1} \left(1 - p_1^t q_1 - p_2^t q_2\right). \qquad \text{(A-8)}$$

The summation in inequality (A-8) equals the probability that at least one of the two unknown payoffs is received eventually. The proof that this probability equals one is the same as the proof for the case with a single unknown payoff, and it follows that $\overline{L}$ is no less than the mean loss for the optimum stationary strategy. A brief final argument completes the proof of the theorem for unknown payoffs a_{11} and a_{22}.

The proof for more than two unknown payoffs or any two payoffs in the same row or column differs only in minor details from the preceding argument.

APPENDIX III

DETERMINING THE EXTREMA OF CERTAIN LOSS FUNCTIONS

Consider the following function of k variables to be maximized (or minimized) over the domain of a convex polyhedron in k space:

$$f(x_1, \ldots, x_k) = \frac{\sum_{i=1}^{k} a_i x_i + b}{\sum_{i=1}^{k} c_i x_i + d} = \frac{\bar{a} \cdot \bar{x} + b}{\bar{c} \cdot \bar{x} + d} = f(\bar{x}),$$

Here, $\bar{a}$ and $\bar{c}$ are k-dimensional vectors and b and d are constants. Two possibilities exist: either $f(\bar{x})$ assumes its maximum (minimum) value at one of the vertices of the convex polyhedron, or $f(\bar{x})$ does not take on its maximum (minimum) value at any one of the vertices. Assume that the last is true. Select any point in the polyhedron, $\bar{x}_o$, for which $f(\bar{x})$ assumes its maximum (minimum) value. Consider the set of points $\bar{x}_o + \alpha\bar{w}$, where α is a variable parameter and $\bar{w}$ is a k-dimensional, non-null vector. This set of points lies on a straight line that passes through $\bar{x}_o$ in the direction of vector $\bar{w}$. Because $\bar{x}_o$ is not a vertex, there exists some vector $\bar{w}$ for which the line passes through points of the polyhedron on both sides of $\bar{x}_o$. For any point of this line, $f(\bar{x})$ can be written

$$f(x) = \frac{\bar{a} \cdot (\bar{x}_o + \alpha\bar{w}) + b}{\bar{c} \cdot (\bar{x}_o + \alpha\bar{w}) + d} = \frac{\alpha\bar{a} \cdot \bar{w} + \bar{a} \cdot \bar{x}_o + b}{\alpha\bar{c} \cdot \bar{w} + \bar{c} \cdot \bar{x}_o + d} = \frac{\alpha e + f}{\alpha g + h},$$

where e, f, g, and h are constants independent of $\bar{x}$. If the derivative of $f(\bar{x})$ is taken along this line through $\bar{x}_o$, the following result is obtained:

$$\frac{df(\bar{x})}{ds} = \frac{eh - fg}{(\alpha g + h)^2 |\bar{w}|}.$$

The numerator of this expression is independent of $\bar{x}$, so the sign of the derivative is the same along the entire line. As a result, $f(\bar{x})$ is a monotonic function of the position along the entire line (unless the denominator of $f(\bar{x})$ becomes zero somewhere along the line). Therefore, $f(\bar{x}_o)$ cannot be larger (smaller) than $f(\bar{x})$ for all of the other points of the line that lie in the polyhedron. In particular, there exists a point, $\bar{x}_a$, at one of the two intersections of the line through $\bar{x}_o$ and a face of the polyhedron, which is such that $f(x_a)$ is at least as large (small) as $f(\bar{x}_o)$.

Now the entire process may be repeated by drawing a line through point $\bar{x}_a$ that lies in one face of the polyhedron. (The face of the polyhedron is actually a hyperplane of dimension $k - 1$.) Since the dimension of the problem is reduced by one each time a new line is chosen, it can eventually be shown that $f(\bar{x})$ at one of the vertices is at least as large (small) as $f(\bar{x}_0)$ (unless the denominator of $f(\bar{x})$ becomes zero somewhere along one of the auxiliary lines). This contradicts the initial assumption of case (2) that $f(\bar{x})$ does not take on its maximum (minimum) value at any of the vertices. Therefore, $f(\bar{x})$ does assume its maximum and minimum values at vertices of the convex polyhedron. The cases in which the denominator of $f(\bar{x})$ can become zero will be handled separately.

For adaptive Bayes decision, $\bar{x}$ is equivalent to p, and the dimension is m. The convex polyhedron is defined as

$$\sum_{i=1}^{m} p_i = 1,$$

$$p_i \geqslant 0 \qquad \text{for } i = 1, \ldots, m.$$

The vertices are the m points e_i. The function to be minimized is

$$f = \frac{\bar{v} - p_1\bar{\bar{E}}(\text{row } 1) - \sum_{i=2}^{m} p_i E(\text{row } i)}{p_1 q_1}$$

for one unknown payoff, a_{11}. Because this expression is used only when the numerator is known to be positive, and because the denominator is zero only on one of the faces of the polyhedron (where $f = +\infty$) and not at an interior point, the preceding argument demonstrates that f assumes its minimum value at one of the vertices of the polyhedron. The reasoning for cases of more than one unknown payoff is very similar, and the conclusion is the same.

In the maximizing part of the decision under uncertainty problem, x is equivalent to q, and the dimension is n. The convex polyhedron is defined as

$$\sum_{j=1}^{n} q_j = 1$$

$$q_j \geqslant 0 \qquad \text{for } j = 1, \ldots, n.$$

The vertices are the n points e_j. The function to be maximized is

$$f = \frac{v - \sum_{j=1}^{n} q_j E(\text{col } j)}{p_1 q_1}$$

for a single unknown payoff. The cases of interest in section 4.2 are those for which

$p_1 > 0$. The denominator of f can become zero only on one of the faces of the polyhedron and not at an interior point. In the plane $q_1 = 0$, the numerator of f is linear, and so f is a monotonic function of the position along any line in that plane. Consequently, f assumes its maximum value at one of the vertices of the polyhedron. The same conclusion is reached in cases of two or more unknown payoffs.

In the minimizing part of the problem of decision under uncertainty, $\bar{x}$ is equivalent to p, with dimension m. The polyhedron is defined as

$$\sum_{i=1}^{m} p_i a_{ij} \geqslant v_{max} \qquad \text{for } j = 2, \ldots, n$$

$$\sum_{i=1}^{m} p_i = 1$$

$$p_i \geqslant 0 \qquad \text{for } i = 1, \ldots, m.$$

The function to be minimized is

$$f = \frac{\bar{v} - \left(p_1 \bar{a}_{11} + \sum_{i=2}^{m} p_i a_{i1}\right)}{p_1},$$

where $\bar{a}_{11} = \int a_{11}\, dP(a_{11})$. The denominator of f is zero only on one boundary plane ($p_1 = 0$) but not at an interior point of the polyhedron. As a result, the function assumes its minimum value at one of the vertices.

APPENDIX IV

AN ABBREVIATED METHOD FOR FINDING THE OPTIMUM STRATEGY IN AN ADAPTIVE BAYES DECISION PROCESS WITH TWO STATISTICALLY INDEPENDENT UNKNOWN PAYOFFS, a_{11} AND a_{22}

Some assumptions must be made initially to eliminate from consideration the trivial cases in which $\bar{v} = E(\text{row } i)$, or E(row 1) or E(row 2) is smaller than the expected return of another row for all possible values of a_{11} and a_{22} (in which case the dominated alternative should be eliminated):

$$E_{\max}(\text{row } 1) > E(\text{row } i) \qquad \text{for } i = 3, \ldots, m$$

$$E_{\max}(\text{row } 2) > E(\text{row } i) \qquad \text{for } i = 3, \ldots, m$$

$$E_{\max}(\text{row } 1) > E_{\min}(\text{row } 2)$$

$$E_{\max}(\text{row } 2) > E_{\min}(\text{row } 1).$$

Then the results of section 3.2 indicate that

$$\bar{L}_{\min} = \min\left[\frac{\bar{v} - \bar{E}(\text{row } 1)}{q_1} + \overline{L_{\min}|a_{11}}, \ \frac{\bar{v} - \bar{E}(\text{row } 2)}{q_2} + \overline{L_{\min}|a_{22}}\right]. \qquad \text{(A-9)}$$

Because a_{11} and a_{22} are statistically independent, it follows that

$$\bar{L}_{\min}(a_{11}) = \begin{cases} \dfrac{\bar{v}(a_{11}) - \bar{E}(\text{row } 2)}{q_2} & \text{if } E(\text{row } 1) < E_{\max}(\text{row } 2) \ \left(a_{11} < a_{11}^{o}\right) \\ 0 & \text{otherwise} \end{cases} \qquad \text{(A-10)}$$

where

$$\bar{v}(a_{11}) = \int v(a_{11}, a_{22})\, dP(a_{22}).$$

The quantity a_{11}^{o} is defined as

$$q_1 a_{11}^{o} + \sum_{j=2}^{n} q_j a_{1j} = E_{\max}(\text{row } 2).$$

Here, a_{11}^{o} represents a critical value of unknown payoff a_{11}. Therefore the mean value of Eq. A-10 is written

$$\overline{L_{\min}|a_{11}} = \int_{a_{11}<a_{11}^{o}} \frac{\bar{v}(a_{11}) - \bar{\bar{E}}(\text{row } 2)}{q_2}\, dP(a_{11})$$

$$= \frac{\bar{\bar{v}}}{q_2} - \frac{1}{q_2} \int_{a_{11}\geqslant a_{11}^{o}} E(\text{row } 1)\, dP(a_{11}) - \frac{\bar{\bar{E}}(\text{row } 2)}{q_2} \Pr\left(a_{11}<a_{11}^{o}\right), \quad \text{(A-11)}$$

and similarly

$$\overline{L_{\min}|a_{22}} = \frac{\bar{\bar{v}}}{q_1} - \frac{1}{q_1} \int_{a_{22}\geqslant a_{22}^{o}} E(\text{row } 2)\, dP(a_{22}) - \frac{\bar{\bar{E}}(\text{row } 1)}{q_1} \Pr\left(a_{22}<a_{22}^{o}\right), \quad \text{(A-12)}$$

where a_{22}^{o} is defined as

$$q_1 a_{21} + q_2 a_{22} + \sum_{j=3}^{n} q_j a_{2j} = E_{\max}(\text{row } 1).$$

Three possible cases arise:

(1) $E_{\max}(\text{row } 1) > E_{\max}(\text{row } 2)$,

(2) $E_{\max}(\text{row } 1) < E_{\max}(\text{row } 2)$,

(3) $E_{\max}(\text{row } 1) = E_{\max}(\text{row } 2)$.

In case (1) $a_{22}^{o} > a_{22\ \max}$, so it is simple to find $\bar{\bar{L}}_{\min}$. When Eqs. A-11 and A-12 are substituted in Eq. A-9, we find

$$\bar{L}_{\min} = \min\left[\frac{\bar{\bar{v}}}{q_1} - \frac{\bar{\bar{E}}(\text{row } 1)}{q_1} + \frac{\bar{\bar{v}}}{q_2} - \frac{1}{q_2} \int_{a_{11}\geqslant a_{11}^{o}} E(\text{row } 1)\, dP(a_{11}) - \frac{\bar{\bar{E}}(\text{row } 2)}{q_2} \Pr\left(a_{11}<a_{11}^{o}\right),\ \frac{\bar{\bar{v}}}{q_2} - \frac{\bar{\bar{E}}(\text{row } 2)}{q_2} + \frac{\bar{\bar{v}}}{q_1} - 0 - \frac{\bar{\bar{E}}(\text{row } 1)}{q_1}\right].$$

The first term is smaller if

$$\int_{a_{11}\geqslant a_{11}^{o}} E(\text{row } 1)\, dP(a_{11}) > \bar{\bar{E}}(\text{row } 2) \Pr\left(a_{11}\geqslant a_{11}^{o}\right).$$

This is true, since $E(\text{row } 1) > \bar{\bar{E}}(\text{row } 2)$ for all $a_{11} \geqslant a_{11}^{o}$. Therefore, in case (1), $p_{\text{opt}} = e_1$; similarly, in case (2), $p_{\text{opt}} = e_2$. For case (3), $a_{11}^{o} = a_{11\ \max}$, and $a_{22}^{o} = a_{22\ \max}$, so

$$\overline{L}_{\min} = \min\left[\frac{\bar{\bar{v}}}{q_1} - \frac{\bar{\bar{E}}(\text{row }1)}{q_1} + \frac{\bar{v}}{q_2} - \frac{E_{\max}(\text{row }1)}{q_2}\,\Pr(a_{11}=a_{11\ \max}) - \frac{\bar{\bar{E}}(\text{row }2)}{q_2}\,\Pr(a_{11}<a_{11\ \max}),\right.$$

$$\left.\frac{\bar{v}}{q_2} - \frac{\bar{\bar{E}}(\text{row }2)}{q_2} + \frac{\bar{\bar{v}}}{q_1} - \frac{E_{\max}(\text{row }2)}{q_1}\,\Pr(a_{22}=a_{22\ \max}) - \frac{\bar{\bar{E}}(\text{row }1)}{q_1}\,\Pr(a_{22}<a_{22\ \max})\right].$$

The first term is smaller if

$$\frac{E_{\max}(\text{row }1) - \bar{\bar{E}}(\text{row }1)}{q_1}\,\Pr(a_{22}=a_{22\ \max})$$

$$< \frac{E_{\max}(\text{row }2) - \bar{\bar{E}}(\text{row }2)}{q_2}\,\Pr(a_{11}=a_{11\ \max}). \qquad \text{(A-13)}$$

Four subcases arise:

(a) $\Pr(a_{11}=a_{11\ \max}) = 0$, $\Pr(a_{22}=a_{22\ \max}) = 0$

(b) $\Pr(a_{11}=a_{11\ \max}) > 0$, $\Pr(a_{22}=a_{22\ \max}) = 0$

(c) $\Pr(a_{11}=a_{11\ \max}) = 0$, $\Pr(a_{22}=a_{22\ \max}) > 0$

(d) $\Pr(a_{11}=a_{11\ \max}) > 0$, $\Pr(a_{22}=a_{22\ \max}) > 0$

For subcase (a) both terms of inequality (A-13) are equal, so $p_{opt} = e_1$ or e_2. In subcase (b) the first term is smaller, so $p_{opt} = e_1$. Similarly, in subcase (c) $p_{opt} = e_2$. In subcase (d) it is necessary to use the inequality (A-13) as it stands to find p_{opt}.

It is not difficult to find a counterexample that demonstrates that the preceding results do not necessarily hold when a_{11} and a_{22} are statistically dependent.

APPENDIX V

EXAMPLE OF ADAPTIVE BAYES DECISION WITH TWO UNKNOWN PAYOFFS

$$q = (1/3, 1/3, 1/3)$$

$$\begin{array}{cc|ccc|} & & & B & \\ & & 1 & 2 & 3 \\ & 1 & 3 & a_{12} & -1 \\ A & 2 & -2 & 0 & a_{23} \\ & 3 & 1 & -1 & 0 \end{array}$$

$$P(a_{12}) = \begin{cases} 0 & \text{if } a_{12} < -3 \\ \frac{1}{3}(a_{12}+3) & \text{if } -3 \leq a_{12} < 0 \\ 1 & \text{if } 0 \leq a_{12} \end{cases}$$

$$P(a_{23}) = \begin{cases} 0 & \text{if } a_{23} < 0 \\ \frac{1}{3}a_{23} & \text{if } 0 \leq a_{23} < 3 \\ 1 & \text{if } 3 \leq a_{23} \end{cases}$$

Unknown payoffs a_{12} and a_{23} are assumed to be statistically independent random variables. $P(a_{12})$ and $P(a_{23})$ are cumulative distribution functions, corresponding to flat density functions between −3 and 0, and 0 and 3, respectively.

This example will be solved by the standard method, and the results will be checked by the abbreviated method.

The following terms are easily calculated:

$$E(\text{row } 1) = \frac{2 + a_{12}}{3}, \ \bar{E}(\text{row } 1) = \frac{1}{6}, \ E_{\max}(\text{row } 1) = \frac{2}{3}, \ E_{\min}(\text{row } 1) = -\frac{1}{3},$$

$$E(\text{row } 2) = \frac{-2 + a_{23}}{3}, \ \bar{E}(\text{row } 2) = -\frac{1}{6}, \ E_{\max}(\text{row } 2) = \frac{1}{3}, \ E_{\min}(\text{row } 2) = -\frac{2}{3},$$

$$E(\text{row } 3) = 0.$$

First, the process with the single unknown payoff a_{12} is solved, as a function of a_{23}. The preliminary checks are made to see whether $\bar{v} = E(\text{row } i_o)$ for $i_o \neq 1$, or $\bar{v} = \bar{E}(\text{row } 1)$:

$$E_{\max}(\text{row } 1) > \begin{cases} E(\text{row } 3) \\ \\ E(\text{row } 2) \text{ for all possible values of } a_{23}, \text{ and} \end{cases}$$

$$E(\text{row } 3) \quad E_{\min}(\text{row } 1).$$

Therefore, $p_{opt} = e_1$, and

$$\bar{L}_{\min}(a_{23}) = \frac{\bar{v}(a_{23}) - \bar{E}(\text{row } 1)}{q_2} > 0 \tag{A-14}$$

for all possible values of a_{23}. The value of the payoff matrix can be calculated:

$$v(a_{23}) = \left\{ \begin{array}{l} \left.\begin{array}{ll} 0 & \text{if } E(\text{row } 1) \leqslant 0 \\ E(\text{row } 1) & \text{if } E(\text{row } 1) \geqslant 0 \end{array}\right\} \text{ if } E(\text{row } 2) \leqslant 0 \ (a_{23} \leqslant 2) \\ \left.\begin{array}{ll} E(\text{row } 2) & \text{if } E(\text{row } 1) \leqslant E(\text{row } 2) \\ E(\text{row } 1) & \text{if } E(\text{row } 1) \geqslant E(\text{row } 2) \end{array}\right\} \text{ if } E(\text{row } 2) \geqslant 0 \ (a_{23} \geqslant 2), \end{array}\right.$$

and

$$\overline{v}(a_{23}) = \begin{cases} \frac{1}{3} \cdot 0 + \int_{-2}^{0} \left(\frac{2 + a_{12}}{3}\right) \frac{1}{3}\, da_{12} = \frac{2}{9} & \text{if } a_{23} \leqslant 2 \\ \frac{1}{3}(a_{23}-1)\left(\frac{-2 + a_{23}}{3}\right) + \int_{a_{23}-4}^{0} \left(\frac{2 + a_{12}}{3}\right) \frac{1}{3}\, da_{12} & \\ \quad = \frac{1}{18}\left(a_{23}^{2}-2a_{23}+4\right) & \text{if } a_{23} \geqslant 2. \end{cases}$$

Therefore, Eq. A-14 yields

$$\overline{L}_{\min}(a_{23}) = \begin{cases} \frac{1}{6} & \text{if } a_{23} \leqslant 2 \\ \frac{1}{6}\left(a_{23}^{2}-2a_{23}+1\right) & \text{if } a_{23} \geqslant 2, \end{cases}$$

and

$$\overline{L_{\min}|a_{23}} = \frac{2}{3} \cdot \frac{1}{6} + \int_{2}^{3} \frac{1}{6}\left(a_{23}^{2}-2a_{23}+1\right) \frac{1}{3}\, da_{23} = \frac{13}{54}.$$

Now the case for which a_{23} is the single unknown will be considered. The preliminary checks are made to see whether $\overline{v} = E(\text{row } i_o)$ for $i_o \neq 2$, or $\overline{v} = \overline{E}(\text{row } 2)$:

if $a_{12} \geqslant -1$, then $E(\text{row } 1) \geqslant E_{\max}(\text{row } 2) > E(\text{row } 3)$, which implies that $\overline{L}_{\min}|a_{12} = 0$ and $p_{\text{opt}} = e_1$;

if $a_{12} < -1$, then $E_{\max}(\text{row } 2) > \begin{cases} E(\text{row } 3) \\ E(\text{row } 1) \end{cases}$ for all $a_{12} < -1$, in which case $\overline{L}_{\min}|a_{12} > 0$, and $p_{\text{opt}} = e_2$.

These results can be summarized as follows:

$$p_{\text{opt}} = \begin{cases} e_1 & \text{if } a_{12} \geqslant -1 \\ e_2 & \text{if } a_{12} < -1, \end{cases}$$

and

$$\bar{L}_{\min}(a_{12}) = \begin{cases} 0 & \text{if } a_{12} \geqslant -1 \\ \dfrac{\bar{v}(a_{12}) - \bar{E}(\text{row } 2)}{q_3} & \text{if } a_{12} < -1 \end{cases} \qquad \text{(A-15)}$$

The quantity $\bar{v}(a_{12})$ can be calculated without excessive difficulty:

$$\bar{v}(a_{12}) = \begin{cases} \dfrac{1}{18} & \text{if } a_{12} \leqslant -2 \\ \dfrac{1}{18}\left(a_{12}^2+8a_{12}+13\right) & \text{if } -2 \leqslant a_{12} \leqslant -1 \\ \dfrac{2}{3} + \dfrac{a_{12}}{3} & \text{if } a_{12} \geqslant -1. \end{cases}$$

If the expressions for $\bar{v}(a_{12})$ are substituted in Eq. A-15, the following expressions are obtained:

$$\bar{L}_{\min}(a_{12}) = \begin{cases} \dfrac{2}{3} & \text{if } a_{12} \leqslant -2 \\ \dfrac{1}{6}\left(a_{12}^2+8a_{12}+16\right) & \text{if } -2 \leqslant a_{12} < -1 \\ 0 & \text{if } a_{12} \geqslant -1, \end{cases}$$

and

$$\overline{L_{\min}|a_{12}} = \frac{31}{54}$$

$\bar{v}$ is easily calculated from $\bar{v}(a_{12})$ or $\bar{v}(a_{23})$: $\bar{v} = 20/81$. Because $\bar{v} > \bar{E}(\text{row } 1) > \bar{E}(\text{row } 2)$,

$$\bar{L}_{\min} = \min\left[\frac{\frac{20}{81} - \frac{1}{6} + \frac{1}{3}\cdot\frac{31}{54}}{\frac{1}{3}}, \frac{\frac{20}{81} + \frac{1}{6} + \frac{1}{3}\cdot\frac{13}{54}}{\frac{1}{3}}\right] = \frac{22}{27},$$

and $p_{\text{opt}} = e_1$. This completes the solution by the standard method.

Now let us apply the abbreviated method. It is easily determined that $\bar{v} > \bar{r}$ for all p. Because $E_{\max}(\text{row } 1) > E_{\max}(\text{row } 2)$, case (1) applies; therefore, $p_{\text{opt}} = e_1$. Notice the saving in labor! Of course, this method does not yield the value of $\bar{L}_{\min}$; however, the labor involved in the calculation of $\bar{L}_{\min}$ is considerably reduced when p_{opt} is known.

APPENDIX VI

ILLUSTRATIONS OF THE POSSIBLE SITUATIONS THAT ARISE IN ADAPTIVE DECISION UNDER UNCERTAINTY WITH A SINGLE UNKNOWN PAYOFF

The technique of converting a game problem into a linear programming problem, which is used here, has been given by Luce and Raiffa.[34]

Consider a 2 × 3 payoff matrix in which a_{11} represents the unknown payoff. All payoffs are assumed to be positive; this is reasonable, since the addition of the same large positive constant to each payoff of any matrix causes all terms to be positive yet does

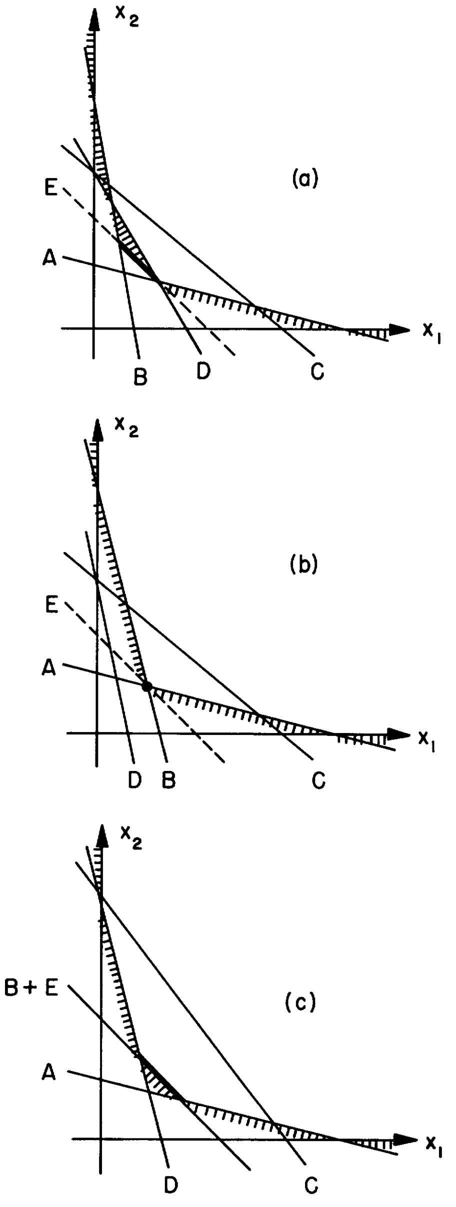

Fig. A-2. Graphical illustration of 2 × 3 decision under uncertainty problem.

not change the optimum strategies. Figure A-2 is used to illustrate the possible significant situations that arise. The following equations are represented by lines in all parts of this figure.

$$a_{12}x_1 + a_{22}x_2 = 1 \quad (A) \qquad a_{11\ \min}x_1 + a_{21}x_2 = 1 \quad (C)$$

$$a_{13}x_1 + a_{23}x_2 = 1 \quad (B) \qquad a_{11\ \max}x_1 + a_{21}x_2 = 1 \quad (D).$$

The dashed lines E represent this equation: $x_1 + x_2 = 1/v_{\max}$. The following argument demonstrates why this is so. Make the substitutions $x_i = p_i/y$, where $y > 0$, in equations D, A, and B. In this case the regions on or above the shaded lines correspond to points satisfying the inequalities

$$\begin{aligned} &a_{11\ \max}p_1 + a_{21}p_2 \geqslant y && (D) \\ &a_{12}p_1 + a_{22}p_2 \geqslant y && (A) \qquad\qquad (A\text{-}16) \\ &a_{13}p_1 + a_{23}p_2 \geqslant y && (B) \end{aligned}$$

with $p_1 \geqslant 0$, $p_2 \geqslant 0$. The further restriction that $p_1 + p_2 = 1$ implies that

$$x_1 + x_2 = \frac{p_1 + p_2}{y} = \frac{1}{y}. \qquad (A\text{-}17)$$

Equation A-17 is the equation of a line with slope −1 that intersects both axes at distance $1/y$ from the origin. The largest possible value of y that satisfies Eqs. A-16 and A-17 equals $v_{\max}$. These equations can be satisfied and y can be maximized at a point lying on or above the shaded lines and on the line with slope −1 that intersects the axes at the smallest possible distance from the origin. Because lines E in all parts of the figure are closest to the origin, they intersect the axes at length $1/v_{\max}$ from the origin. All points on lines E can be described by the equation

$$x_1 + x_2 = \frac{1}{v_{\max}} \qquad (E)\ .$$

This leads to the conclusion that points that lie on or above lines A and B, on or above the x_1 axis, on or to the right of the x_2 axis, and on line E satisfy the inequalities

$$a_{12}p_1 + a_{22}p_2 \geqslant v_{\max}, \qquad p_1 \geqslant 0,\ p_2 \geqslant 0,$$

$$a_{13}p_1 + a_{23}p_2 \geqslant v_{\max}, \qquad p_1 + p_2 = 1.$$

These inequalities define the constraint polyhedron for the payoff matrix.

The case in which $\Pr(v=v_{\max}) = 0$ is illustrated in Fig. A-2a. As a_{11} increases from $a_{11\ \min}$ to $a_{11\ \max}$, the line $a_{11}x_1 + a_{21}x_2 = 1$ moves from line C to line D, pivoting at the x_2 axis, and the line E moves closer to the origin. If the probability that a_{11} equals $a_{11\ \max}$ is zero, then $\Pr(v=v_{\max}) = 0$. The constraint polyhedron is the heavy segment of line E in Fig. A-2a. Although the figure does not demonstrate that E(col 1) < v whenever $v < v_{max}$, a fact that has already been proved does illustrate the situation when $v = v_{max}$. Because the heavy segment of line E lies on or below line D, all points in the constraint polyhedron satisfy the relationship $a_{11\ max}p_1 + a_{21}p_2 \leq v_{max}$, or

$$E_{max}(\text{col } 1) \leq v_{max}. \qquad \text{(A-18)}$$

Figure A-2a also illustrates a case in which $\Pr(v=v_{max}) > 0$. This is true if a_{11} has a positive probability of assuming the value $a_{11\ max}$. Inequality (A-18) is valid in this case also.

Figure A-2b illustrates another case in which $\Pr(v=v_{max}) > 0$. Here the constraint polyhedron consists of the single point on line E at the intersection of lines A and B. Naturally, this point represents p_{opt}. Because the point lies above line D, the following relation is true:

$$E_{max}(\text{col } 1) > v_{max}.$$

In fact, E(col 1) $> v_{max} = v$ for all values of a_{11} from a critical value a_o (that causes the line for equation $a_{11}x_1 + a_{21}x_2 = 1$ to pass through the unique point on line E) to $a_{11\ max}$.

The unstable situation is illustrated in Fig. A-2c. The constraint space corresponds to the heavy segment of line E; however,

$$E_{max}(\text{col } 1) > v_{max}$$

for those points in the constraint space that lie above line D. The reason for the designation "unstable" has to do with the fact that if line B were rotated an infinitesimal amount clockwise, this problem would become one of the type illustrated in Fig. A-2a, and if line B were rotated an infinitesimal amount counterclockwise, the problem would become one of the type illustrated in Fig. A-2a. In general, the two solutions for the perturbed matrices will be considerably different.

Note that although the figures have been drawn, of necessity, for a 2 × 3 matrix, the principles illustrated by these drawings are valid for any m × n payoff matrix.

APPENDIX VII

EXAMPLE OF ADAPTIVE DECISION UNDER UNCERTAINTY WITH TWO UNKNOWN PAYOFFS IN WHICH THE MEAN LOSS IS SMALLER AT SOME POINT INSIDE THE CONSTRAINT SPACE THAN AT ANY OF THE VERTICES

$$\begin{array}{cc} & B \\ & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ A \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \left[\begin{array}{ccc} a_{11} & -1 & 0 \\ 2 & a_{22} & -1 \\ -2 & 0 & 0 \end{array}\right] \end{array} \qquad \begin{array}{l} \Pr(a_{11} = -2) = 1/2 \\ \Pr(a_{11} = +1) = 1/2 \\ \\ \Pr(a_{22} = -1) = 1/2 \\ \Pr(a_{22} = 0) = 1/2 \end{array}$$

with a_{11} and a_{22} statistically independent. The following quantities are easily calculated:

a_{11}, a_{22}	$v(a_{11}, a_{22})$
-2 , -1	$-2/5$
-2 , 0	$-2/5$
$+1$, -1	$-2/5$
$+1$, 0	$-1/4$.

Since $v_{max} = -1/4$, the constraint space is defined by

$$\begin{array}{ll} \displaystyle\sum_{i=1}^{3} p_i a_{i3} \geqslant v_{max} & \\ \displaystyle\sum_{i=1}^{3} p_i = 1, & \\ p_i \geqslant 0 \qquad \text{for } i = 1,\ 2,\ 3 & \end{array} \qquad \text{or} \left\{ \begin{array}{l} p_2 \leqslant 1/4 \\ \\ p_1 + p_2 + p_3 = 1, \\ \\ p_1 \geqslant 0,\ p_2 \geqslant 0,\ p_3 \geqslant 0. \end{array} \right.$$

The four vertices of the space are

$$\begin{array}{ll} p_a = (0, 0, 1), & p_c = (3/4, 1/4, 0), \\ p_b = (1, 0, 0), & p_d = (0, 1/4, 3/4). \end{array} \tag{A-19}$$

The following results have been obtained for $L_{min}|a_{11}$. When $a_{11} = -2$, $p_{opt} = (0, 2/5, 3/5)$ for the problem with the single unknown payoff a_{22}. It is true that $L_{min}|a_{11} = 0$ $(a_{11} = -2)$ for both possible values of a_{22}. When $a_{11} = +1$, $p_{opt} = (1/4, 1/4, 1/2)$, and

$$L_{\min}|a_{11} = \begin{cases} 2/5 & \text{if } a_{22} = -1 \\ 0 & \text{if } a_{22} = 0 \end{cases} \qquad (a_{11}=+1).$$

The calculations for $L_{\min}|a_{22}$ have produced the following results. When $a_{22} = -1$, $p_{opt} = (0, 2/5, 3/5)$, and

$$L_{\min}|a_{22} = 0 \qquad (a_{22}=-1)$$

for both possible values of a_{11}. When $a_{22} = 0$, $p_{opt} = (1/4, 1/4, 1/2)$, and

$$L_{\min}|a_{22} = \begin{cases} 12/5 & \text{if } a_{11} = -2 \\ 0 & \text{if } a_{11} = +1 \end{cases} \qquad (a_{22}=0).$$

These results can be checked by using the techniques learned from section 4.2.

If p lies inside the constraint space, $L_{\max}$ can be written as Eq. 16. The reader can easily confirm the fact that Eq. 16 is valid when $v = v_{\max}$ at all four vertices (A-19).

$$L_{\max} = \begin{cases} \max\left[\dfrac{-2/5 - (-2p_1+2p_2-2p_3) + 0}{p_1}, \dfrac{-2/5 - (-p_1-p_2) + 0}{p_2}\right] & \text{if } (a_{11}, a_{22}) = (-2, -1) \\ \max\left[\dfrac{-2/5 - (-2p_1+2p_2-2p_3) + 0}{p_1}, \dfrac{-2/5 - (-p_1) + 12/5p_2}{p_2}\right] & \text{if } (a_{11}, a_{22}) = (-2, 0) \\ \max\left[\dfrac{-2/5 - (p_1+2p_2-2p_3) + 2/5p_1}{p_1}, \dfrac{-2/5 - (-p_1-p_2) + 0}{p_2}\right] & \text{if } (a_{11}, a_{22}) = (+1, -1) \\ \max\left[\dfrac{-1/4 - (p_1+2p_2-2p_3) + 0}{p_1}, \dfrac{-1/4 - (-p_1) + 0}{p_2}\right] & \text{if } (a_{11}, a_{22}) = (+1, 0) \end{cases}$$

$\bar{L}(p)$ is calculated for the four vertices of the constraint polyhedron:

$$\bar{L}(p_a) = 1/4[(+\infty)+(+\infty)+(+\infty)+(+\infty)] = +\infty$$

$$\bar{L}(p_b) = 1/4[(+\infty)+(+\infty)+(+\infty)+(+\infty)] = +\infty$$

$$\bar{L}(p_c) = 1/4[(12/5)+(19/5)+(12/5)+(2)] = 53/20 = 2.65$$

$$\bar{L}(p_d) = 1/4[(+\infty)+(+\infty)+(+\infty)+(+\infty)] = +\infty.$$

Now consider some other point not at a vertex of the polyhedron:

$$p_e \equiv (6/20, 5/20, 9/20).$$

$$\bar{L}(p_e) = 1/4[(2)+(2)+(3/5)+(1/5)] = 1.2 < 2.65 = \bar{L}(p_c).$$

(Fairly thorough trial-and-error calculations indicate that probably $p_{opt} = p_e$.) Although p_e lies on one of the surfaces of the constraint polyhedron, it is not a vertex.

APPENDIX VIII

PROOF THAT THERE EXISTS A MINIMAX SOLUTION FOR THE PROBLEM OF ADAPTIVE COMPETITIVE DECISION WITH EQUAL INFORMATION – SINGLE UNKNOWN PAYOFF

First consider an auxiliary decision process with a loss matrix that is related to the payoff matrix of the original problem of adaptive competitive decision in the following manner. The (1, 1) entry is $\bar{v} - \bar{a}_{11}$, and all other (i, j) entries are $\bar{v} - a_{ij} + x$. Then the expected loss for a single play of the auxiliary process is

$$\begin{aligned}\bar{L} &= p_1 q_1(\bar{v}-\bar{a}_{11}) + \sum_{(i,j)\neq(1,1)} p_i q_j(\bar{v}-a_{ij}+x)\\ &= \bar{v} - \bar{r} + (1-p_1 q_1)x. \qquad\qquad (A\text{-}20)\end{aligned}$$

If player A is the minimizing player and B is the maximizing player, this single-step process has a minimax solution with value y(x) and optimum strategies p_x and q_x. The value and minimax strategies are functions of the variable x.

Two sets of points are defined on the real axis:

$$\begin{aligned}&x \text{ is in Set 1 if} \begin{cases} y(x) < x & \text{and } x < 0, \text{ or}\\ y(x) \leqslant x & \text{and } x \geqslant 0;\end{cases}\\ &x \text{ is in Set 2 if} \begin{cases} y(x) > x & \text{and } x > 0, \text{ or}\\ y(x) \geqslant x & \text{and } x \leqslant 0.\end{cases}\end{aligned} \qquad (A\text{-}21)$$

Note that every real number is either in one or the other of the two sets; however, the point x = 0 may be in both sets. If x is in Set 1 and player A uses strategy p_x, then the following inequality is true for any mixed strategy[9] of player B:

$$\bar{v} - \bar{r} + (1-p_{1x}q_1)x \leqslant y(x) \begin{cases} < x \text{ if } x < 0\\ \leqslant x \text{ if } x \geqslant 0\end{cases}$$

or

$$\bar{v} - \bar{r} \leqslant \begin{cases} p_{1x}q_1 x - c & \text{if } x < 0\\ p_{1x}q_1 x & \text{if } x \geqslant 0,\end{cases} \qquad (A\text{-}22)$$

where c is some positive constant, and p_{1x} is the first component of p_x.

It will be shown that if x is in Set 1 and if player A uses the piecewise-stationary strategy that begins with the repeated use of probability distribution p_x for the original competitive decision process, then the mean loss is no greater than x. Likewise, if x is in Set 2 and if player B uses the piecewise-stationary strategy that begins with the repeated use of distribution q_x, then the mean loss is no less than x. Finally, it will be shown that there is a unique value $x = \bar{L}_{opt}$ which is such that points exist in both Sets 1 and 2 which are within an arbitrarily small distance of $\bar{L}_{opt}$. Thus, $\bar{L}_{opt}$ is a minimax (more precisely, a "supinf") value for the adaptive competitive decision process.

The total loss for N steps of the competitive decision process is

$$L_N = \sum_{k=1}^{N} (v-r^k) \prod_{t=0}^{k-1} \left(1-p_1^t q_1^t\right),$$

and the mean value of the total loss for N steps is

$$\bar{L}_N = \sum_{k=1}^{N} \left(\bar{v}-\bar{r}^k\right) \prod_{t=0}^{k-1} \left(1-p_1^t q_1^t\right), \qquad \text{(A-23)}$$

where $p_1^0 q_1^0$ is equal to zero. The derivation of these expressions is based upon the assumption that once a_{11} is received the loss will equal zero from that point on. Although Eq. A-23 is valid only for memoryless sequences of distributions, $p^1, \ldots, p^N$ and $q^1, \ldots, q^N$, the use of (A-23) will be restricted to cases in which one may assume, with complete generality, that the sequences are memoryless (see Appendix II).

If x is in Set 1, if N steps of the adaptive competitive process are played, and if A uses distribution p_x repeatedly until a_{11} is received or until the number of steps reaches N, then an upper bound to $\bar{L}_N$ can be found by substituting inequalities (A-22) in (A-23).

$$\bar{L}_N \leqslant \begin{cases} \displaystyle\sum_{k=1}^{N} \left(p_{1x} q_1^k x - c\right) \prod_{t=0}^{k-1} \left(1-p_{1x} q_1^t\right) & \text{if } x < 0 \\[2em] \displaystyle\sum_{k=1}^{N} \left(p_{1x} q_1^k x\right) \prod_{t=0}^{k-1} \left(1-p_{1x} q_1^t\right) & \text{if } x \geqslant 0. \end{cases} \qquad \text{(A-24)}$$

This is true no matter what sequence of distributions, $q^1, \ldots, q^N$, player B uses. It is easily shown that

$$\sum_{k=1}^{N} \left(p_{1x} q_1^k x\right) \prod_{t=0}^{k-1} \left(1-p_{1x} q_1^t\right) = x - x \prod_{k=0}^{N} \left(1-p_{1x} q_1^k\right).$$

When this identity is substituted in inequality (A-24), the following inequality results:

$$\bar{L}_N \leqslant \begin{cases} x - x\prod_{k=0}^{N}\left(1-p_{1x}q_1^k\right) - \sum_{k=1}^{N} c \prod_{t=0}^{k-1}\left(1-p_{1x}q_1^t\right) & \text{if } x < 0 \\ \\ x - x\prod_{k=0}^{N}\left(1-p_{1x}q_1^k\right) & \text{if } x \geqslant 0. \end{cases} \tag{A-25}$$

Clearly, when x is non-negative, $\bar{L}_N \leqslant x$ for all N. It is now desired to prove $\bar{L}_N \leqslant x$ if x is negative. The following expression is taken from the right-hand side of inequality (A-25):

$$-\left[\sum_{k=1}^{N} c \prod_{t=0}^{k-1}\left(1-p_{1x}q_1^t\right) - |x|\prod_{k=0}^{N}\left(1-p_{1x}q_1^k\right)\right], \tag{A-26}$$

where c is positive. All elements of the first term are non-negative. The sum converges as N approaches infinity only if the individual elements of the sum approach zero. But the individual elements approach zero only if the second term in expression (A-26) approaches zero. The conclusion is that expression (A-26) either converges to a non-positive quantity or approaches $-\infty$, as N grows large. In the former case

$$\lim_{N\to\infty} \bar{L}_N \leqslant x;$$

in the latter case there exists an integer N_o that is such that if N is greater than N_o, $\bar{L}_N \leqslant x$. In other words, by using distribution p_x repeatedly until a_{11} is received and then playing optimally, player A can guarantee either that his mean loss converges to a quantity no greater than x, or after a sufficient number of plays his mean loss is no greater than x.

An analogous derivation leads to the conclusion that if x is in Set 2 and if player B uses a piecewise-stationary strategy beginning with distribution q_x, then either

$$\lim_{N\to\infty} \bar{L}_N \geqslant x,$$

or there is an integer N_o that is such that $\bar{L}_N \geqslant x$ for all $N > N_o$, no matter what strategy player A uses.

The point $x = +\infty$ is in Set 1, and the point $x = -\infty$ is in Set 2; therefore, there is at least one point $x = \bar{L}_{opt}$, in any epsilon neighborhood of which there is a point in Set 1 and a point in Set 2. It can be shown that the point $\bar{L}_{opt}$ is unique. The assumption that $\bar{L}_{opt}$ is not unique, implies that there must exist two points, x_1 and x_2, in Sets 1 and 2, respectively, which are such that $x_2 > x_1$. The fact that x_1 and x_2 are in Sets 1 and 2, respectively, implies that $y(x_1) \leqslant x_1$ and $y(x_2) \geqslant x_2$, or that

$$y(x_2) - y(x_1) \geqslant x_2 - x_1. \tag{A-27}$$

Because y(x) is the minimax value of a particular matrix in which x is added to all but one of the entries of the matrix, if x changes by an amount d, y(x) cannot change by more than d. (If A uses strategy p_x in the auxiliary game, the mean loss is no greater than y(x); therefore, if A uses p_x when d is added to all but one of the entries of the matrix, the loss is no greater than y(x) + d.) As a result,

$$y(x_2) - y(x_1) \leq x_2 - x_1.$$

This inequality contradicts inequality (A-27), unless $y(x_1) = x_1$ and $y(x_2) = x_2$. But if $y(x_1) = x_1$, the definition of Set 1 implies that $x_1 \geq 0$. Also, $y(x_2) = x_2$ implies that $x_2 \leq 0$. Thus $x_1 \geq x_2$. But because $\bar{L}_{opt}$ was assumed to be non-unique, we were allowed initially to choose $x_2 > x_1$. The contradiction leads to the conclusion that $\bar{L}_{opt}$ is unique. This concludes the proof of the theorem.

The continuity of y(x), which has just been proved, and the fact that in any epsilon neighborhood of $\bar{L}_{opt}$ there exist points in both Sets 1 and 2 lead to the conclusion that $y(\bar{L}_{opt}) = \bar{L}_{opt}$.

APPENDIX IX

PROOF THAT AN EXAMPLE OF ADAPTIVE COMPETITIVE DECISION WITH UNEQUAL INFORMATION HAS A MINIMAX SOLUTION AND THAT PLAYER B CANNOT ATTAIN THE MINIMAX VALUE BY USING A PIECEWISE-STATIONARY STRATEGY

$$\begin{array}{ccc} & & B \\ & & \begin{array}{cc} 1 & \quad 2 \end{array} \\ A & \begin{array}{c} 1 \\ \\ 2 \end{array} & \begin{bmatrix} a_{11} & 0 \\ & \\ 0 & -1 \end{bmatrix} \end{array} \qquad \begin{array}{l} \Pr(a_{11} = -1) = 1/2 \\ \Pr(a_{11} = +1) = 1/2 \end{array}$$

Assume that A uses alternative 1 repeatedly until a_{11} is received. What strategy can B use that will maximize the loss to player A? If $a_{11} = +1$, then $v = 0$, so B's optimum strategy is to use alternative 2 repeatedly. If $a_{11} = -1$, then $v = -1/2$. In this case the single-step loss is $-1/2$ for each step at which player B uses alternative 2, and the single step loss is $+1/2$ for the first step at which B uses alternative 1 and 0 thereafter. Therefore, B's optimum strategy is to use alternative 1 at the first step and anything thereafter, since A uses $p = (1/2, 1/2)$ repeatedly after he learns that $a_{11} = -1$. If B uses these strategies,the mean loss is

$$\bar{L} = \lim_{N \to \infty} \bar{L}_N = 1/4.$$

Assume that B uses the appropriate minimax distribution repeatedly from the second step on ($q = (0, 1)$ if $a_{11} = +1$; $q = (1/2, 1/2)$ if $a_{11} = -1$), but at the first step uses alternative 2 if $a_{11} = +1$ and alternative 1 if $a_{11} = -1$. If player A plays in an optimum fashion, he sustains no loss after the first step because B's choice of alternative at the first step "spills the beans" about the true value of a_{11}, and A can use this information to play in an optimum fashion from the second step on. If A uses alternative 1 at step 1, his single-step loss is zero if $a_{11} = +1$ and $+1/2$ if $a_{11} = -1$; if A uses alternative 2 at step 1, his single-step loss is 1 if $a_{11} = +1$ and $-1/2$ if $a_{11} = -1$. Therefore, no matter what A does at the first step, $L = 1/4$. This proves that the given strategies are minimax strategies.

Does a piecewise-stationary strategy exist for player B which is such that $\bar{L} \geqslant 1/4$ for all strategies of A? Assume that B uses distribution $(q_+, 1-q_+)$ if $a_{11} = +1$ and distribution $(q_-, 1-q_-)$ if $a_{11} = -1$. These represent probability distributions used repeatedly until a_{11} is received. This is as general a piecewise-stationary strategy as B can use. Suppose that A uses alternative 1 repeatedly until a_{11} is received. Then if $a_{11} = +1$,

$$\overline{L} = \begin{cases} \dfrac{0 - q_+}{q_+} & \text{if } q_+ \neq 0 \\ \\ 0 & \text{if } q_+ = 0 \end{cases} = \begin{cases} -1 & \text{if } q_+ \neq 0 \\ \\ 0 & \text{if } q_+ = 0, \end{cases}$$

and if $a_{11} = -1$,

$$\overline{L} = \begin{cases} \dfrac{-1/2 + q_-}{q_-} & \text{if } q_- \neq 1/2 \\ \\ 0 & \text{if } q_- = 1/2 \end{cases} = 1 - \frac{1}{2q_-}.$$

It follows that

$$\overline{L} = \begin{cases} -\frac{1}{2} + \left(1 - \frac{1}{2q_-}\right)\Big/ 2 & \text{if } q_+ \neq 0 \\ \\ \frac{1}{2} - \frac{1}{4q_-} & \text{if } q_+ = 0 \end{cases} = \begin{cases} -\frac{1}{4q_-} \\ \\ \frac{1}{2} - \frac{1}{4q_-}. \end{cases}$$

In order that $\overline{L}$ be no less than 1/4, q_+ must equal zero, and q_- must equal 1. So, in order to ensure that $\overline{L} \geqslant 1/4$ when A uses alternative 1 repeatedly, B must use alternative 2 repeatedly when $a_{11} = +1$ and alternative 1 repeatedly when $a_{11} = -1$, until a_{11} is received. Using this strategy, B divulges his knowledge of the true value of a_{11} at the first step. Player A can use this information to ruin B. Suppose that A uses alternative 2 at the first step. Then if $a_{11} = +1$ he sustains a loss of 1 at the first step but no further loss because he will use alternative 1 repeatedly from the second step on. But if $a_{11} = -1$, A sustains a loss of $-1/2$ at the first step and should select alternative 2 repeatedly thereafter. Thus he never receives a_{11}, but sustains a loss of $-1/2$ at each step. As a result $\overline{L} = -\infty$. Therefore, there is no piecewise-stationary strategy that B can use to guarantee that $\overline{L} \geqslant 1/4$, irrespective of A's strategy. Player B's minimax strategy must be nonstationary.

APPENDIX X

SOLUTION TO AN N-TRUNCATED PROBLEM OF ADAPTIVE BAYES DECISION WITH A SINGLE UNKNOWN PAYOFF

Assume that payoff a_{11} is unknown in an $m \times n$ matrix. Notation used in Section III is employed here. It can be demonstrated that only a $2 \times n$ payoff matrix need be considered.

Assume that E(row 2) $\geq$ E(row i) for $i = 3, \ldots, m$. (The rows can be relabeled so that this inequality is satisfied.) In this case player A should never use any alternatives except 1 and/or 2. Therefore, all rows of the matrix but the first two may be eliminated from consideration without affecting the generality of the solution. Furthermore, if $\bar{v} = \bar{E}$(row 1) or $\bar{v} =$ E(row 2), then A can make the total loss zero by using, respectively, alternative 1 or 2 at all N steps. This is because $\bar{v} = \bar{E}$(row 1) implies

$$v = \sum_{j=1}^{n} q_j a_{1j} \geq \sum_{j=1}^{n} q_j a_{2j}$$

for all possible values of a_{11}, and $v =$ E(row 2) implies

$$v = \sum_{j=1}^{n} q_j a_{2j} \geq \sum_{j=1}^{n} q_j a_{1j}$$

for all possible values of a_{11}. In the remaining cases the following definitions can be used:

$$d_1 \equiv \bar{v} - \bar{E}(\text{row } 1) > 0$$

$$d_2 \equiv \bar{v} - \bar{E}(\text{row } 2) > 0.$$

The mean loss for a 1-truncated process when A uses distribution (p_1, p_2) is $\bar{L}_1 = p_1 d_1 + p_2 d_2$. Subscripts of $\bar{L}$ represent the number of steps in the truncated process, and the minimum value of the mean loss can be written

$$\bar{L}_{1\ \text{opt}} = \min\,[d_1, d_2].$$

The minimum mean loss for a 2-truncated process is written

$$\bar{L}_{2\ \text{opt}} = \min\,[d_1 + (1-q_1)L_{1\ \text{opt}}, d_2 + L_{1\ \text{opt}}]. \qquad \text{(A-28)}$$

An expression that equals zero has been omitted from the first term in the brackets of Eq. A-28. The complete expression is

$d_1 + (1-q_1)$(minimum mean loss for 1-truncated process when a_{11} is unknown)

$+\, q_1$(minimum mean loss for 1-truncated process when a_{11} is known).

The last term equals zero. The general expression for the minimum mean loss of an N-truncated process is

$$\bar{L}_{N\ opt} = \min\,[d_1+(1-q_1)\bar{L}_{N-1\ opt},\, d_2+\bar{L}_{N-1\ opt}].$$

Assume that

$$\bar{L}_{k\ opt} = d_1 + (1-q_1)\bar{L}_{k-1\ opt} < d_2 + \bar{L}_{k-1\ opt}. \qquad \text{(A-29)}$$

This assumption means that the mean loss for a k-truncated process can be minimized only if alternative 1 is used at the first step.

$$\bar{L}_{k+1\ opt} = \min\,[d_1+(1-q_1)(d_1+(1-q_1)\bar{L}_{k-1\ opt}),\, d_2+(d_1+(1-q_1)\bar{L}_{k-1\ opt})]$$
$$= \min\,[d_1+(1-q_1)(d_1+(1-q_1)\bar{L}_{k-1\ opt}),\, d_1+q_1d_2+(1-q_1)(d_2+\bar{L}_{k-1\ opt})]. \qquad \text{(A-30)}$$

If relationships (A-31) and (A-29) are applied to Eq. A-30, then Eq. A-32 results.

$$d_1 < d_1 + q_1d_2 \qquad \text{(A-31)}$$

$$\bar{L}_{k+1\ opt} = d_1 + (1-q_1)\bar{L}_{k\ opt} < d_2 + \bar{L}_{k\ opt}\,, \qquad \text{(A-32)}$$

Thus, if the mean loss for a k-truncated process can be minimized only if alternative 1 is used at the first step, then the mean loss for an n-truncated process can be minimized only if alternative 1 is used at the first step when $n \geqslant k$.

Next, assume that $\bar{L}_{1\ opt} = d_2 \leqslant d_1$. This assumption means that the mean loss for a 1-truncated process can be minimized by using alternative 2 at the single step.

$$\bar{L}_{2\ opt} = \min\,[d_1+(1-q_2)d_2,\, 2d_2]. \qquad \text{(A-33)}$$

If the second term in the brackets of (A-33) is no greater than the first, then $\bar{L}_{3\ opt} = \min\,[d_1+(1-q_1)2d_2,\, 3d_2]$.

In general, if $\bar{L}_k$ can be optimized by using alternative 2 at the first step, then

$$\bar{L}_{k+1\ opt} = \min\,[d_1+(1-q_1)kd_2,\, (k+1)d_2]. \qquad \text{(A-34)}$$

The second expression in the brackets of (A-34) is no larger than the first if $d_2 \leqslant d_1/(1+kq_1)$, or

$$k \leqslant (d_1-d_2)/q_1d_2 \equiv N_o - 1. \qquad \text{(A-35)}$$

In other words, if $d_2 \leqslant d_1$, the optimum strategy for an N-truncated process is to use alternative 2 at all N steps if $N \leqslant N_o$. But if $N_o < N \leqslant N_o + 1$, then the optimum strategy is to use alternative 1 at the first step and alternative 2 at the remaining N−1 steps.

The results just stated lead to the conclusion that the optimum alternative for A to use at the first step of an N-truncated process is 1 if $N > N_o$ and 2 if $N \leqslant N_o$. This statement also includes the conditions $\bar{v} = \bar{E}(\text{row } 1)$ or $\bar{v} = E(\text{row } 2)$. In other words, the optimum strategy is to use alternative 1 repeatedly until the number of steps remaining is no more than N_o (defined in Eq. A-35), and then use alternative 2 repeatedly until the process terminates.

Acknowledgment

I am deeply in debt to Professor Edward Arthurs for his enthusiasm, advice, suggestions, and critical appraisal. I should also like to thank Dr. P. M. Lewis II, who launched me upon this research topic and who repeatedly offered encouragement.

Many thanks are due to Professor D. A. Huffman, Dr. M. P. Schützenberger, Professor C. E. Shannon, and Dr. L. S. Shapley for invaluable suggestions and criticism.

References

1. W. A. Clark and B. G. Farley, Generalization of Pattern Recognition in a Self-Organizing System, Proc. Western Joint Computer Conference, March 1955, pp. 86-91.

2. B. G. Farley and W. A. Clark, Simulation of self-organizing systems by digital computer, Trans. IRE, Vol. PGIT-4, pp. 76-84, September 1954.

3. F. Rosenblatt, Perception simulation experiments, Proc. IRE 48, 301-309 (1960).

4. A. G. Oettinger, Programming a digital computer to learn, Phil. Mag., Ser. 7, Vol. 43, pp. 1243-1263 (1952).

5. R. Bellman and R. Kalaba, On Communication Processes Involving Learning and Random Duration, IRE National Convention Record, Vol. 6, Part 4, 1958, pp. 16-21.

6. B. Widrow, Adaptive Sampled-Data Systems — A Statistical Theory of Adaption, IRE WESCON Convention Record, 1959, Part 4, pp. 74-85.

7. G. M. White, Penny matching machines, Information and Control 2, 349-363 (1959).

8. R. L. Mattson, A Self-Organizing Binary System, Proc. of the Eastern Joint Computer Conference, December 1959, pp. 212-217.

9. B. Widrow and M. E. Hoff, Adaptive Switching Circuits, Technical Report No. 1553-1, Solid State Electronics Laboratory, Stanford Electronics Laboratory, Stanford University, Stanford, California, June 1960.

10. J. A. Aseltine, A. R. Mancini, and C. W. Sarture, A survey of adaptive control systems, Trans. IRE, Vol. PGAC-6, pp. 102-108, December 1958.

11. R. R. Bush and F. Mosteller, Stochastic Models for Learning (John Wiley and Sons, Inc., New York, 1955).

12. H. Robbins, A sequential decision problem with a finite memory, Proc. Nat. Acad. Sci. 42, 920-923 (1956).

13. J. R. Isbell, On a problem of Robbins, Ann. Math. Statist. 30, 606-610 (1958).

14. H. Robbins, Some aspects of the sequential design of experiments, Bull. Am. Math. Soc. 58, 527-535 (1952).

15. M. M. Flood, On Game-Learning Theory and Some Decision-Making Experiments, Decision Processes, edited by R. M. Thrall, C. H. Coombs, and R. L. Davis (John Wiley and Sons, Inc., New York, 1954), pp. 139-158.

16. R. N. Bradt, S. M. Johnson, and S. Karlin, On sequential designs for maximizing the sum of n observations, Ann. Math. Statist. 27, 1060-1074 (1956).

17. M. Kochen and E. H. Galanter, The acquisition and utilization of information in problem solving and thinking, Information and Control 1, 267-288 (1958).

18. R. M. Friedberg, A Learning Machine: Part I, IBM J. Res. Develop. 2, 2-13 (January 1958).

19. R. M. Friedberg, B. Dunham, and J. H. North, A Learning Machine: Part II, IBM J. Res. Develop. 3, 282-287 (July 1959).

20. H. Everett, Recursive Games, Contributions to the Theory of Games, Vol. III, edited by M. Dresher, A. W. Tucker, and P. Wolfe, Annals of Mathematics Studies No. 39 (Princeton University Press, Princeton, N. J., 1957), pp. 47-78.

21. R. Duncan Luce and H. Raiffa, Games and Decisions (John Wiley and Sons, Inc., New York, 1957).

22. S. Vajda, The Theory of Games and Linear Programming (Methuen and Company, Ltd., London, 1957).

23. J. D. Williams, The Compleat Strategist (McGraw-Hill Book Company, New York, 1954).

24. J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior* (Princeton University Press, Princeton, N. J., 1953).

25. D. Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions* (John Wiley and Sons, Inc., New York, 1954).

26. K. J. Arrow, T. Harris, and J. Marschak, Optimal inventory policy, Econometrica *19*, 250-272 (1951).

27. D. Gillette, Stochastic Games with Zero Stop Probabilities, *Contributions to the Theory of Games*, Vol. III, *op. cit.*, pp. 179-187.

28. H. W. Kuhn, Extensive Games and the Problem of Information, *Contributions to the Theory of Games*, Vol. II, edited by H. W. Kuhn and A. W. Tucker, Annals of Mathematics Studies No. 28 (Princeton University Press, Princeton, N. J., 1953), pp. 193-216.

29. L. J. Savage, *The Foundations of Statistics* (John Wiley and Sons, Inc., New York, 1954); see especially pp. 163-164.

30. J. Hannan, Approximation to Bayes Risk in Repeated Play, *Contributions to the Theory of Games*, Vol. III, *op. cit.*, pp. 97-139.

31. A. Charnes, W. W. Cooper, and A. Henderson, *An Introduction to Linear Programming* (John Wiley and Sons, Inc., New York, 1953).

32. L. S. Shapley, Stochastic games, Proc. Nat. Acad. Sci. *39*, 1095-1100 (1953).

33. H. Everett, *op. cit.*; see Theorem 8.

34. R. Duncan Luce and H. Raiffa, *op. cit.*, pp. 408-412.