PURDUE UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING

[redacted]

Purdue University

Lafayette, Indiana 47907

ON BEHAVIOR STRATEGY SOLUTIONS IN FINITE EXTENDED DECISION PROCESSES

V. A. Utgoff
R. L. Kashyap

TR-EE 70-24
June, 1970

School of Electrical Engineering
Purdue University
Lafayette, Indiana 47907

This work was partially supported by NASA under Grant NGR 15-005-021, by NSF under Grant GK-1970, and by the AFOSR under Grant 69-1776.

PRECEDING PAGE BLANK NOT FILMED

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The problem of determination of best behavior strategies for a decision maker in a two-person, zero-sum, finite extended decision process is considered.

In this problem, best strategies are defined to be those strategies which maximize a player's expected return against all possible strategies that can be employed by the other player, and the decision process is thus considered as a two-person, zero-sum, finite extended game.

It is noted that most two-person, zero-sum, finite extended games arise in settings which:

(i) define the maximum amount of information on previous alternative choices that each player can gather from move to move, but

(ii) do not define how much of this information actually will be gathered, or how much of the actually gathered information will be remembered from move to move.

Defining any specification of (ii) for a player as an information collection scheme, and defining the complexity of a strategy as the

number of numbers that must be stored in a computer to implement it, it is clear that both the effectiveness and the complexity of a strategy depend upon the player's information collection scheme in a critical manner.

In the report then, techniques are developed for finding best behavior strategies on arbitrary information collection schemes. These techniques make it possible to search in any given game for strategies which are minimally complex but acceptably effective.

A doctor-patient medical game is "searched" using these techniques, and behavior strategies are found for the doctor which are more than 90% as effective but less than 10% as complex as the best strategy which results when the doctor gathers and retains all the information defined to be available to him.

A pursuit and evasion game is also studied and best behavior strategies are determined for the pursuer and the evader when both players gather and retain all the information defined to be available to each of them.

CHAPTER 1

INTRODUCTION

1.1 General Problem

We are interested in the determination of best strategies for a decision maker, in decision processes which are described as follows:

First, the alternative choices in the processes are made by three parties:

(i) a decision maker, our protagonist;

(ii) a second party; and

(iii) nature, operating under known statistics.

Second, the processes are extended, in that each party may make more than one decision as the process unfolds.

Third, the processes are finite, in that each party may make only a finite number of decisions or moves, and at each decision is presented with only a finite number of alternatives.

Finally, the processes will be considered to involve only two persons, in that nature will be removed from active consideration by considering the payoffs to the protagonist in terms of expected returns under nature's known statistics.

In our considerations, we will consider best strategies as those which maximize the protagonist's minimum expected return against all possible decision strategies that can be employed by the second party.

These strategies have the appealing property that they neither require nor imply an estimate of the second party's *intentions*; they only take into account his *capabilities*.

In order to find such max-min strategies for the protagonist, we shall pose the problem as a game in which the second party is considered to lose whatever return the protagonist wins. Specifically, we will label the second party as the antagonist, and we will treat the decision processes as finite, two-person, zero-sum extended games.

1.2 Information Gathering

We will assume that as play progresses, each player can become aware of *some* of the alternative choices that he and the other parties have made.[†] Specifically, we will assume that the settings in which our extended games arise:

(i) define the *maximum* amount of previous information about alternative choice that each player *can gather*, but

(ii) do *not* define how much of the theoretically available information each player *will gather*, and

(iii) do *not* define how much of the actually gathered information each player *will recall* from move to move.

If we call any specification of factors (ii) and (iii) above, a specification of that players *information collection scheme*, we can restate our interests as follows. We wish to consider the problem

---

[†]Games in which each player, at each move, knows all previous alternative choices by all parties are commonly known as games with *perfect information*. Such games can be solved by techniques which are far simpler than those which we will develop here.

of selection of a decision strategy for the protagonist in two parts:

(i) the selection of an information collection scheme for the protagonist, and

(ii) the determination of a max-min strategy for the protagonist given the selected information collection scheme.

1.3 Considerations in Selection of an Information Collection Scheme for the Protagonist

Suppose we adopt the following nomenclature:

(i) Any information collection scheme in which any piece of information that is collected at any move is recalled at each subsequent move, will be called a perfect-recall information collection scheme; and

(ii) any perfect-recall scheme which collects every piece of information defined as available by the settings in which the game arises, will be called a complete information collection scheme.

Using this nomenclature, we can describe the considerations involved in selection of an information collection scheme for the protagonist as follows.

By definition, any strategy for a player is a complete set of instructions telling the player how to choose alternatives in every situation that he can encounter in the course of a game. But what defines a situation for the player?--simply the values that can be taken on by the various collections of pieces of information on previous alternative choices that he is defined to know under his information collection scheme.

Taking the number of situations a strategy must allow for as a measure of its complexity, we can make the following statements.

(i) The more information on previous alternative choices that a player knows at each move, the more complex his strategies will be.

(ii) The more complex a player's strategies are, the more expensive they are to store and implement.

All of which says, with respect to storage and implementation costs of his strategy, the less information a player knows at each move, the better off he is going to be.

On the other hand, for any given information collection scheme for the opponent, the minimum expected return guaranteed by a player's max-min strategies is a maximum, when the player is using its complete information collection scheme. This information collection scheme, however, leads to the most complicated strategies.

In light of the above considerations, we will establish the following rationale for selection of the player's information collection schemes:

(i) Noting that we are taking the protagonist's part (hence the label), we will conservatively assign to the antagonist a complete information collection scheme, and

(ii) we will select for the protagonist an information collection scheme which leads to the simplest possible max-min or nearly max-min strategy guaranteeing a minimum expected return which is acceptably close† to that guaranteed by a complete information-collection scheme.

1.4 Types of Strategies

There are three basic types of strategies that can be employed by the players:

(i) the <u>pure strategy</u>, which is a list of instructions telling the player which alternative to choose in any situation he may encounter in a play of the game;

(ii) the <u>mixed strategy</u>, which is a probability distribution over the set of all possible pure strategies, which the player uses to randomly select a pure strategy when he begins a play of the game; and

(iii) the <u>behavior strategy</u>, which is a collection of probability distributions, one distribution for each situation the player can encounter, each distribution describing how the player should randomly choose an alternative in that situation.

The first two of these strategy types are relatively well known, and general techniques have been developed for finding a pure strategy which maximizes over a player's set of pure strategies its minimum return, and for finding a mixed strategy which maximizes over a player's set of mixed strategies its minimum return. The third strategy is not so well known, and no general techniques have been developed for finding a behavior strategy which maximizes over a player's set of behavior strategies its minimum return.

---

†To be defined.

We shall show below, however, that in all but the most trivial of extended games with less than perfect information, the determination and use of the best of a player's <u>pure</u> or <u>mixed</u> strategies is <u>not practical</u>.

We will, therefore, restrict our attention to the problem of determination of a best <u>behavior</u> strategy for the protagonist, and we will develop some computational techniques for this purpose.

1.5 Organization of the Argument

In approaching the general problem we have outlined above, we will organize our argument as follows.

(1) We will develop a formal technique of description for general extended games similar to that developed by Von Neumann[1] and later refined by Kuhn[2].

(2) We will formally define the concepts of a pure, a mixed, and a behavior strategy; and we will show why attempts to find and implement pure or mixed strategy solutions in extended games are usually impractical.

(3) Restricting ourselves to behavior strategies then, we will formally state the general problem that we wish to solve: find a minimally complex, acceptably good, behavior strategy for the protagonist.

(4) We will discuss the difficulties involved in generating a complete solution to the problem, and we will outline a heuristic technique for obtaining a partial solution.

(5) In order to obtain even a partial solution, however, we will need a method for finding max-min or nearly max-min behavior strategies for any given information collection scheme for the protagonist. We will split this problem into two parts, considering separately the special case where the protagonist employs a perfect-recall information collection scheme, and the more general case where the protagonist may employ any information collection scheme.

(6) We will develop an algorithm for determination of max-min behavior strategies for the perfect-recall information collection scheme case.

(7) For the arbitrary information collection scheme case, we will show that for every finite two-person, zero-sum, extended game in which one player has N moves and the other player has M moves, there is an associated (N + M)-person, non-cooperative game with solution strategies of a new type corresponding to behavior strategies in the original game:

(i) which meet a set of necessary conditions for max-min behavior strategies, and which

(ii) if the players' total sensitivity[†] to forgotten information is less than ϵ, will guarantee each player a minimum expected return within ϵ of the minimum expected return it can guarantee itself by playing a max-min mixed strategy.

[†]To be defined later; roughly, a measure of the influence of the forgotten information on the expected return.

(8) We will develop a heuristic algorithm for generating solutions in the associated game, or equivalently, for determining behavior strategies which meet a set of necessary conditions for max-min behavior strategies in games in which the players may employ arbitrary information schemes.

(9) We will demonstrate the use of the "perfect-recall" algorithm by solving several pursuit-evasion games.

(10) Finally, we will employ the general algorithm in a search for a minimally complex, acceptably good, behavior strategy for the doctor in an example doctor-patient medical decision process. In this example, we will first find a behavior strategy X on the doctor's complete information collection scheme. We will then carry out a search for a minimally complex, acceptably good, behavior strategy for the doctor; and the search will result in a behavior strategy X^* which is more than 90% as effective as X, but less than 10% as complex.

1.6 Contributions

The contributions made in this report are:

(i) the development of an algorithm for determination of max-min behavior strategies in two-person, zero-sum, finite extended games in which both players employ perfect-recall information collection schemes;

(ii) a demonstration of a relationship between solutions in an associated game and max-min behavior strategies in the original game; and

(iii) the development of a heuristic algorithm for determination of behavior strategies which meet a set of necessary conditions for max-min behavior strategies in games in which the players employ arbitrary information collection schemes.

These contributions are of value in that they provide, for relatively complicated, finite, two-person, extended decision processes, a means for determination of behavior strategies for the decision maker which are:

(i) relatively simple to store and implement, and

(ii) come acceptably close to maximizing the decision maker's minimum expected return against all possible strategies which can be employed by the other person.

CHAPTER 2

DESCRIPTIONS OF FINITE EXTENDED GAMES

2.1 Introduction

In this chapter we will carry out the first of several arguments which will lead us to a precise formulation of the problem that we wish to consider. Specifically,

(i) we will give a formal definition of a finite extended game which is slightly broader than that given by Von Neumann[1] and Kuhn[2];

(ii) we will develop a form of presentation for the rules specifying a finite extended game which will allow considerable insight into the strategic properties of such a game; and, in the process of this development,

(iii) we will define and examine the concept of an information collection scheme.

2.2 Notation and Conventions

In this and the following chapters, we will indicate that a function or variable is associated with the protagonist, the antagonist, or nature by superscripting it with p, a, or n, respectively.

Further, we will establish most of our definitions and notation with respect to the protagonist only, whenever the corresponding definition or notation for the antagonist is clear. When the corresponding definition or notation is not clear, we will give it in a separate statement, or we will indicate in square brackets[†] the changes that must be made to the protagonist's statement to make it true of the antagonist.

2.3 Word Usage[††]

In common usage the words game, play (as a noun), party, move, and choice, have more or less ambiguous meanings. From this point on, we will use these words only in the following formal ways.

(i) A game is the totality of rules that describe it.

(ii) A play is a particular instance of how a game is played from beginning to end.

(iii) Any player or nature is a party in the game.

(iv) A move is the occassion of a choice between various alternatives to be made by one of the parties under conditions which are precisely described by the rules of the game.

(v) A choice is an alternative chosen in a specific instance.

2.4 Definitions of Finite Extended Games

As mentioned above, a game is the totality of rules which describe it. The strategic interest of a game, however, lies in what is not directly described by the rules.

[†] In the verbal portions of the text, square brackets will be employed only for this purpose.

[††] cf. reference 1, page 49.

In the definition of a finite extended game that was given by Von Neumann[1] and later refined by Kuhn[2], the rules of the game include a rule precisely defining the information on previous choices that each player _will_ know at each of its moves. The only things _not_ described in the rules of the game are the strategies under which the players make their decisions at each of their moves.

We wish to leave to each player the additional problem of deciding, within the limits of what it _can_ know about previous choices at each of its moves, which pieces of information it actually _should_ know at each move in order to play acceptably well. Therefore, we shall state our definition of a finite extended game accordingly.

Definition 1: A _finite extended game_ is a collection of four rules:

(i) a rule describing _all possible plays_ in the game;

(ii) a rule describing the manner in which _nature_ chooses an alternative at each of her moves (a statistical description for nature);

(iii) a rule describing the _payoffs_ to each player for each possible play; and

(iv) a rule describing the _maximum amount of information_ on alternative choices by all parties, that each player can gather at each of the moves in the game.

2.5 An Example of a Finite Extended Game

An example of a finite extended game is described below. The four rules of definition 1 are completely (though perhaps not explicitly) specified by the given verbal description of the game.

Consider a simple card game between two players, α and β, the former consisting of two persons, $\alpha 1$ and $\alpha 2$, who are not allowed to communicate. The game is played as follows.

(i) $\alpha 1$ chooses a card from one of two well-shuffled decks of cards--a deck A which has twice as many red cards as black cards, or a deck B which has twice as many black cards as red cards.

(ii) The card chosen by $\alpha 1$ is given to β, and β is then asked to guess which deck the card came from.

(iii) $\alpha 2$ is then asked to guess which deck β's card came from. $\alpha 2$ is not allowed to know $\alpha 1$'s choice or to see β's card, but is allowed to know β's guess, provided that α had planned in advance to use this information in its strategy.

(iv) A payoff is then given to each player under the following rule. If both guess right, α wins \$1 from β; if neither one guesses right, α wins \$0 from β; if only $\alpha 2$ is right, α wins \$3 from β; and if only β is right, α wins -\$3 from β.

An interpretation of how this game is played is given by the flow diagram of figure 1.

In the next few sections, we will develop a form of presentation for the rules describing any finite extended game. This form of presentation, which is very similar to the one developed by Kuhn[2], allows considerable insight into the strategic properties of finite extended games. In the process of developing this form, we will

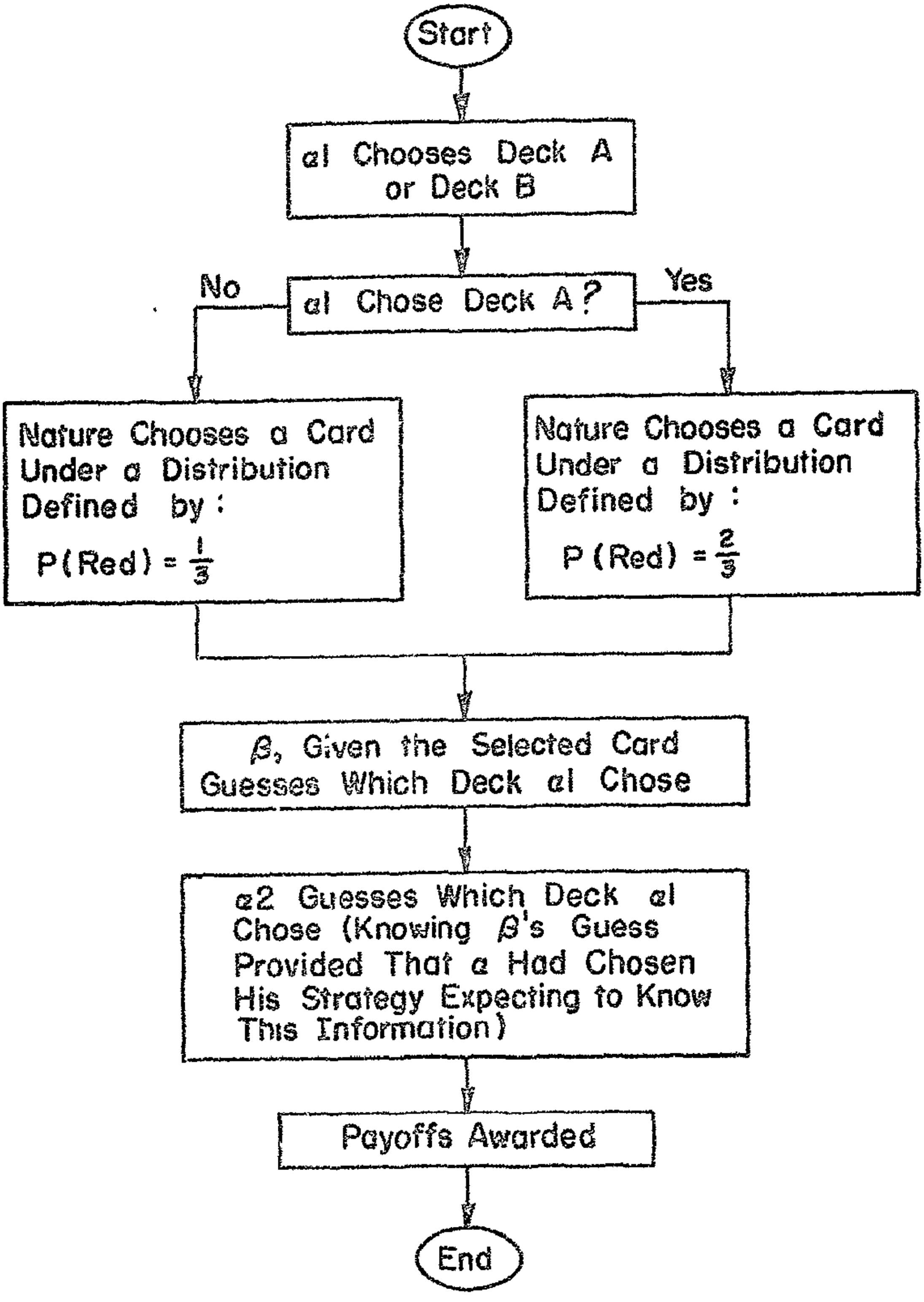


FIGURE 1. AN EXAMPLE OF A FINITE EXTENDED GAME.

define and examine the concept of an information collection scheme. As we proceed, we will use this form to describe the example finite extended game given above.

2.6 Rules Describing All Possible Plays: Game Trees

Let us consider any two moves as distinct if each one arises with a different string of alternative choices by the parties. With this convention, the rule describing all possible plays in a finite extended game can be easily expressed in the form of a finite tree,† usually known as a game tree.††

Definition 2: For a given finite extended game Γ, a game tree K is a finite tree such that:

(i) the nodes of K are in one-to-one correspondence with the moves of Γ;

(ii) one node, called the distinguished node, O, represents the first or starting move of the game;

(iii) the branches of K represent the alternatives at each move in Γ, with the conventions that:

(a) the i branches of O are indexed 1,...,i, in counterclockwise order, starting with any branch;

(b) for any node X with j branches, except for O, the branches are indexed 1,...,j in counterclockwise order starting with the branch following that branch which does not represent an alternative at the move corresponding to X, and

---

†A connected graph without loops.

††Reference 2: reference 3, page 41; and elsewhere.

(c) the ith branch at any node X represents the ith alternative of the move corresponding to X.

In figure 2 we establish an indexing of alternatives and show a corresponding game tree for the example game of section 2.5.

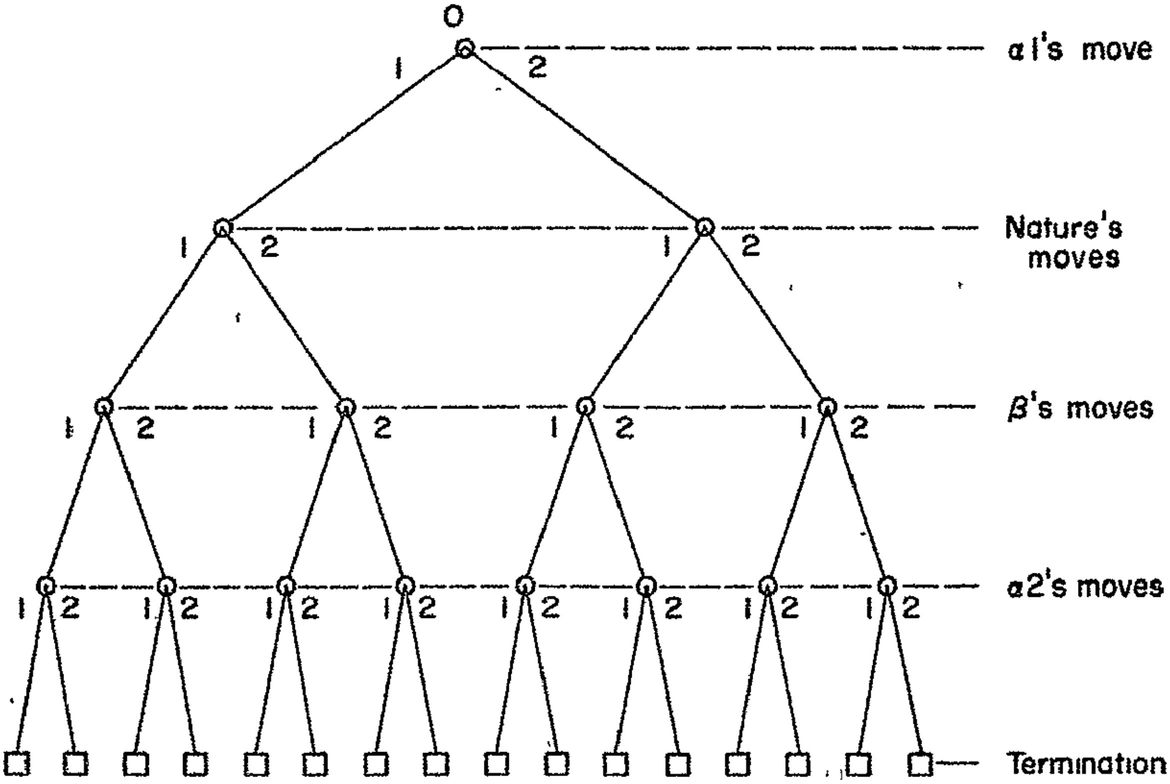
In figure 2 we have indicated the terminating branches, i.e., those which correspond to choices which do not lead to further moves, by ending them with squares.

Note that in a game tree K, drawn for a game Γ with a given indexing of alternatives, there is a one-to-one correspondence between the possible plays in Γ and the unicursal lines from O through terminating branches. In view of this one-to-one correspondence and those between moves and nodes, and alternatives and branches, we will use the same name for each element in Γ and its correspondent in K.

2.7 Rules Describing Nature

The second rule required in the description of a finite extended game is a rule describing the manner in which nature chooses an alternative at each of her moves. This is given by specifying a probability distribution over nature's alternatives at each of her moves. Such distributions are easily indicated in a game-tree diagram. We simply label each alternative at each of nature's moves with the probability that it will be selected if the given move arises.

Note that if all such distributions differ at each of nature's moves, then at every move, every previous alternative choice can be said to "condition" or influence nature's decision. On the other hand, if a particular previous choice does not influence nature's decision



<u>α's "1st" move</u>

Alternative	Name
1	Choose deck A
2	Choose deck B

<u>β's moves</u>

Alternative	Name
1	Guess deck A
2	Guess deck B

<u>Nature's moves</u>

Alternative	Name
1	Give β a red card
2	Give β a black card

<u>α's "2nd" moves</u>

Alternative	Name
1	Guess deck A
2	Guess deck B

FIGURE 2 AN ALTERNATIVE INDEXING AND A GAME TREE FOR THE EXAMPLE GAME

at some move, then at every move with the same alternatives and a choice history which is identical except for the non-influential choice, the probability distributions over nature's alternatives must be the same.

In figure 3, the game tree for the example given in section 2.5 is appropriately labeled to indicate how nature chooses an alternative at each of her moves. Note that nature's decision at each of her moves is conditioned by the choice made by $\alpha 1$.

2.8 Rules Describing Payoffs

The third rule required in the description of a finite extended game is a rule describing the payoffs to each player for each possible play. This rule will be denoted by the function $h_i(W)$, indicating the payoff to the ith player when play W has been realized.

In a game-tree diagram, we can indicate the payoffs to each of the N players when play W is realized by listing the vector $H(W) = \{h_i(W),\ i=1,\ldots,N\}$ at the terminating branch of W. In figure 3, the payoffs to the players in the example game of section 2.5 are indicated in this manner. However, in any two-person, zero-sum, finite extended game, we will hereafter list only the payoffs to the player who is considered to be the protagonist.

2.9 Rules Describing Maximum Gatherable Information

The fourth rule required in the description of a finite extended game is a rule describing the maximum amount of information on alternative choices that each player can gather at each move.

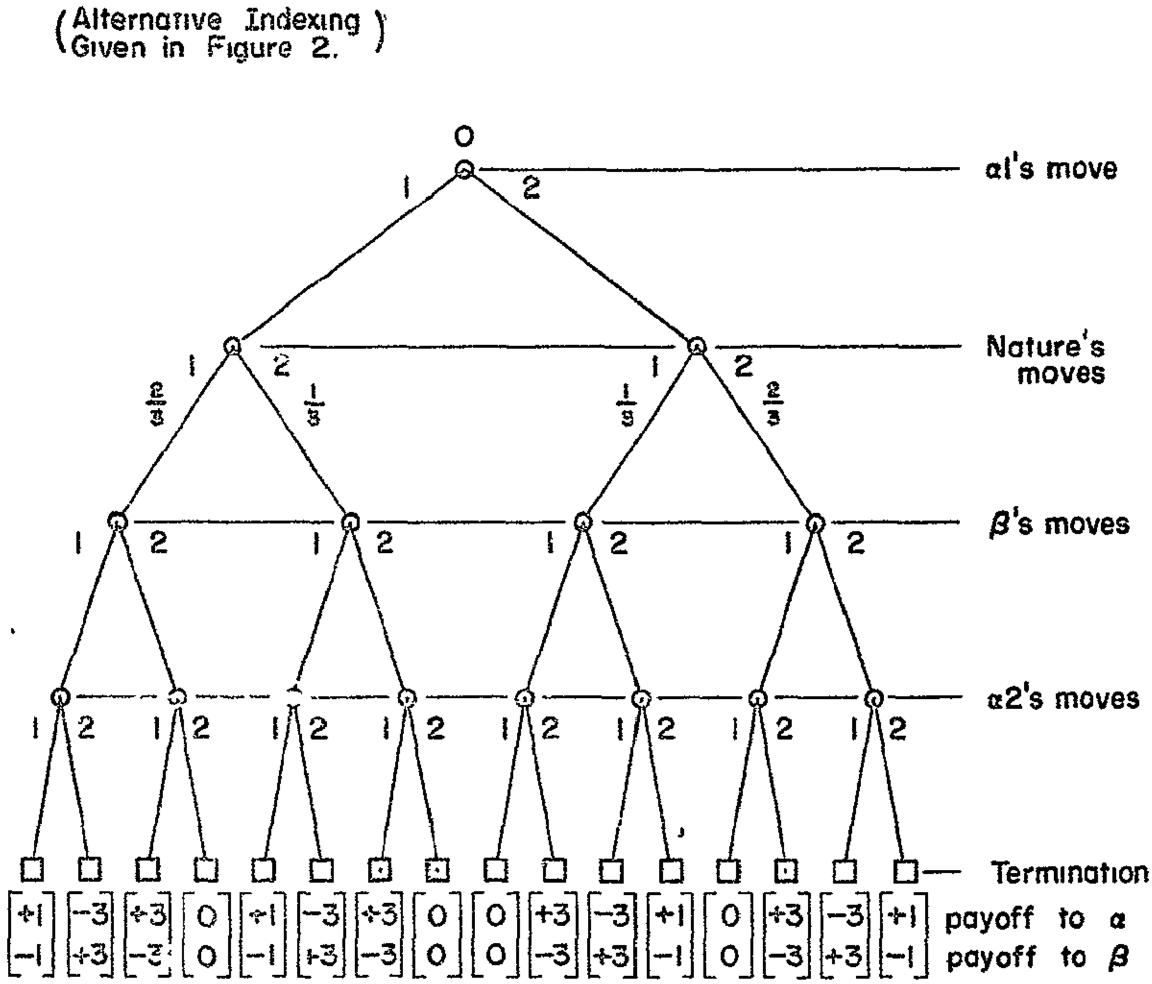


FIGURE 3. STATISTICAL DESCRIPTION FOR NATURE AND PAYOFF RULE FOR THE EXAMPLE GAME.

In order to explain the technique which we shall use to present this rule, we must first define the concept of an information collection scheme and develop some of the properties of this concept.

2.9.1 Information Collection Schemes and Induced Move Partitions

We begin by giving formal definitions for the concepts of an information collection scheme and an information set.

Definition 3: An information collection scheme for a player is a rule R, defining for each move X at which the player is required to choose an alternative, the portion of the alternative choice history establishing X which is known to that player.

Definition 4: Any set of pieces of information that can be known by a player at any of its moves and which consists of:

(i) a list of alternatives available at that move, and

(ii) the pieces of information on previous alternative choices which the player knows at that move under its given information collection scheme R,

will be called an information set.

Clearly, in any finite extended game, there are only a finite number of information sets possible for a player under any given information collection scheme R. Further, we can state the following property.

Property 1: An information collection scheme R for a given player induces a unique partitioning of that player's moves into a minimum number of sets, each of which contains only moves that the player cannot distinguish from one another.

Proof: Consider the partitioning of a player's moves which is defined by collecting into a single set of moves, every move corresponding to a given information set. Since, by definition 4, a single information set corresponds to each move, the property follows easily.

Definition 5: The partitioning of a player's moves into the minimum number of sets, each of which contains only moves which are indistinguishable under a given information collection scheme R, will be called the player's information partition under R and will be denoted by IP(R).

In view of the one-to-one correspondence between the information sets that are possible for a player under an information collection scheme R and the sets of indistinguishable moves of its information partition under R, we will use the same name to refer to a set of moves of the player's information partition under R or its corresponding information set.

Consider the following examples of information partitions.

(i) If a player employs an information collection scheme R such that it has perfect information on the previous alternative choices at each of its moves, then its information partition under R consists of sets, each containing a single move; and on the other extreme,

(ii) if a player employs an information collection scheme R such that it has no information on the previous alternative choices at each of its moves, then its information partition under R consists of the minimum possible number of sets of moves with identical lists of alternatives.

2.9.2 Equivalence of Information Collection Schemes

In section 1.3 we stated that a strategy is a complete set of instructions telling its player how to choose an alternative in every situation that he can encounter in the course of a game. From the arguments given above, it is clear that a situation is simply an information set. It follows then that the effectiveness of a strategy for a player depends upon its information collection rule only through the information partition defined by that rule.

Definition 6: Any two information collection schemes R and R′ will be considered *equivalent* if the information partitions under R and R′ are identical.

In view of this definition, we will henceforth consider any set of equivalent information collection schemes as a single information scheme. Note that by means of definition 5, definition 6, and property 1, we have established a one-to-one correspondence between each information collection scheme possible for a player and the information partition induced by that scheme.

2.9.3 The Complete Information Collection Scheme

By definition 3, the number of information collection schemes possible for a player is finite. Correspondingly, the number of possible information partitions is also finite. After stating a few required definitions, we will establish a property of a player's set of all possible information partitions.

Definition 7: The *rank* r(X) of a move X is the number of alternative choices made by all parties in establishing X.

Definition 8: Let X and Y be any two moves for a given player which fall on any common play W. Then, a perfect-recall information collection scheme for this player is any information collection scheme R such that: $r(X) > r(Y)$ implies that R defines as known to that player at X, any information on previous alternative choices known at Y and the given player's choice at Y.

Definition 9: For a given player, the complete information collection scheme in Γ is the perfect-recall information collection scheme R, under which the player collects at each move every piece of information on alternative choices that is defined to be gatherable at that move by the rules defining Γ.

Denoting a player's complete information collection scheme as C, we can show that the set of all information partitions possible for a player has the following property.

Property 2: A player's information partition under any information collection scheme R can be formed as appropriate unions of the sets of that player's information partition under C.

Proof: Property 2 is not true only if:

(i) some information set under information collection scheme R includes some but not all of the moves contained in a single information set under information collection scheme C;

but (i) is true only if:

(ii) information possessed under R allows the player to distinguish between moves which are not distinguishable under C.

But (ii) is false since the amount of information known to a player at each of his moves is a maximum under collection scheme C. |

We return now to our main line of argument.

2.9.4 Rules Describing Maximum Gatherable Information

In view of the one-to-one correspondence between each information collection scheme and the information partition under that scheme, it is obvious that we can represent any information collection scheme by indicating in a game-tree diagram the sets of the corresponding partition. Therefore, we will present the rule describing the maximum amount of information on alternative choices that a player can gather at each move by indicating in a game-tree diagram the sets of the player's information partition under its complete information collection scheme.

In figure 4A we employ this technique to present this "maximum information rule" for the example game of section 2.5. In figure 4B we show the move partitions which correspond to the information collection schemes in which β does not look at its card, and $\alpha 2$ does not examine β's guess as to which deck β's card came from. Recalling property 2 of the preceding subsection, note that the sets of the information partitions of figure 4B can be formed as unions of the sets of the information partitions of figure 4A.

2.10 Summary

In the above sections we have defined the concept of a finite extended game, and we have shown that the four rules describing

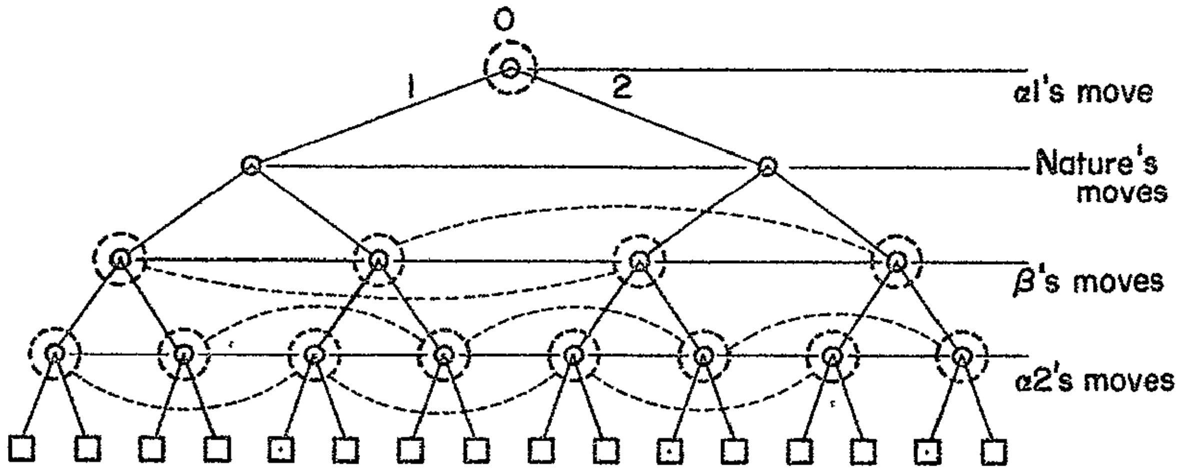


FIGURE 4(a) MOVE PARTITION DEFINING MAXIMUM GATHERABLE CHOICE INFORMATION FOR THE EXAMPLE GAME

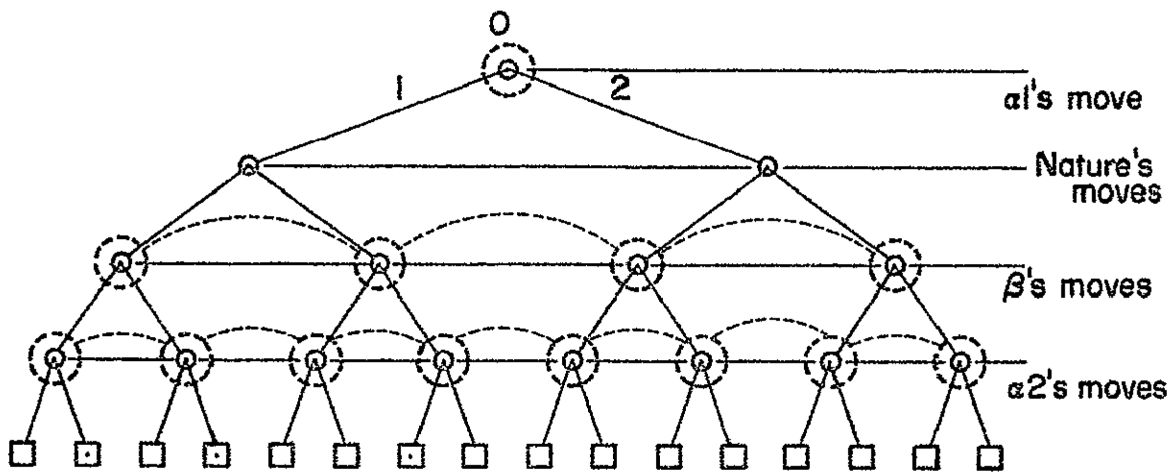


FIGURE 4(b) MOVE PARTITION FOR INFORMATION COLLECTION SCHEMES IN THE EXAMPLE GAME WHICH CORRESPOND TO · β DOES NOT LOOK AT ITS CARD AND α2 DOES NOT EXAMINE β's GUESS.

such a game can be presented by:

(i) assuming an indexing of the players' alternatives and constructing a corresponding finite tree to indicate which choice combinations or plays are possible (rule 1);

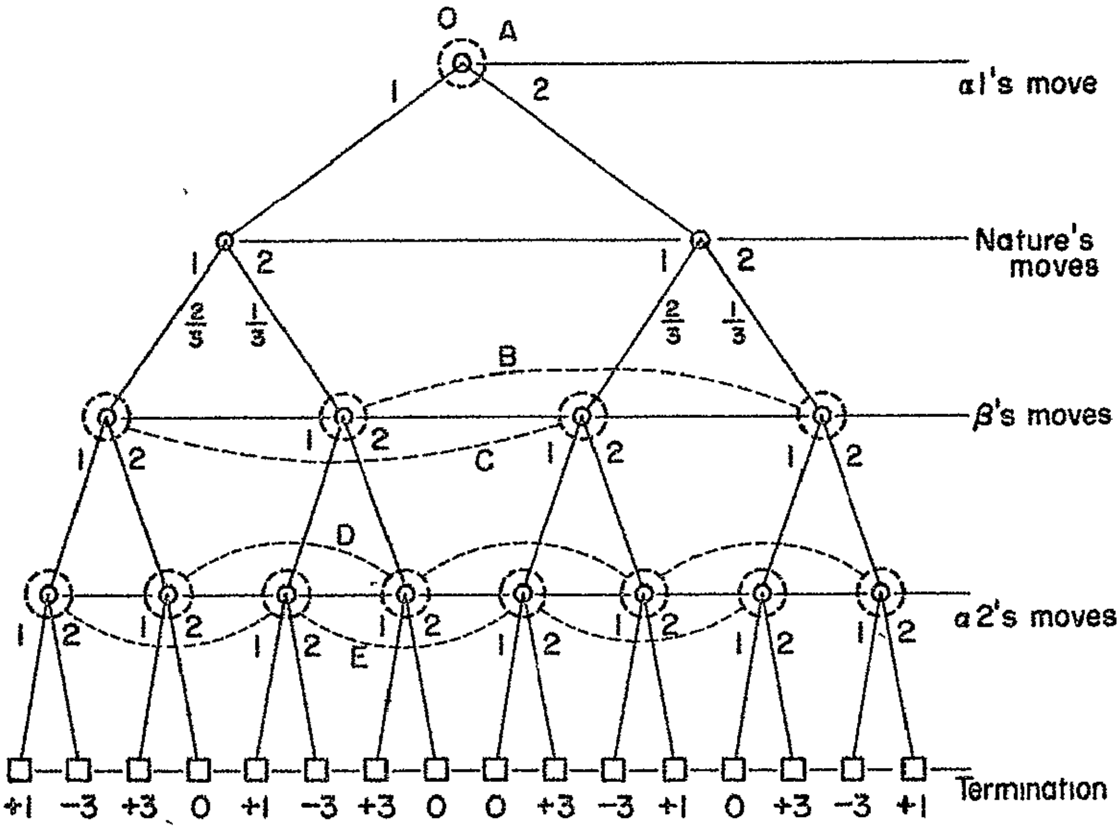
(ii) labeling the alternatives of nature's moves to indicate the probability with which each alternative is selected when the given move arises (rule 2);

(iii) labeling each possible play W with a vector[†] H(W) to indicate the payoffs to each player when that play arises (rule 3); and

(iv) partitioning each player's moves into sets, each of which contains only moves which are indistinguishable when the player is employing its complete information collection scheme, to indicate the maximum amount of "distinguishability" that a player can achieve, or equivalently, the maximum amount of alternative choice information that it can gather at each move.

In figure 5 we present a complete description in this form of the example game of section 2.5. In the example game we are taking the α player to be the protagonist.

[†]Or a scalar indicating the protagonist's payoff in the two-person, zero-sum case.



α's "1st" move

Alternative	Name
1	Choose deck A
2	Choose deck B

β's moves

Alternative	Name
1	Guess deck A
2	Guess deck B

Nature's moves

Alternative	Name
1	Give β a red card
2	Give β a black card

α's "2nd" moves

Alternative	Name
1	Guess deck A
2	Guess deck B

FIGURE 5. A COMPLETE DESCRIPTION OF THE EXAMPLE GAME.

CHAPTER 3

STRATEGIES IN TWO-PERSON, ZERO-SUM, FINITE EXTENDED GAMES

3.1 Introduction

In this chapter we will carry out the second of several arguments which will lead us to a precise formulation of the problem that we wish to consider. Specifically,

(i) we will examine the general concept of a strategy, and we will define a measure of strategy effectiveness;

(ii) we will give formal definitions for the concepts of a pure, a mixed, and a behavior strategy;

(iii) we will compare these three strategy types with respect to their effectiveness and the difficulties involved in their determination and implementation; and, in the process of this comparison,

(iv) we will show why behavior strategies provide the only practical means of solution to many two-person, zero-sum, finite extended games.

3.2 General Concept of a Strategy

A strategy is by definition a rule telling its player how to choose an alternative in any situation that he can encounter in the course of a game.

In chapter 2 we defined the concepts of an information set and an information collection scheme, and we noted that:

(i) the information collection scheme employed by a player defines the set of information sets which he can encounter in the course of a game, and

(ii) the defined information sets are the situations for which the player's strategy must give instructions.

In order to define a strategy then, we must first define the information collection scheme upon which it is based. In this chapter we will state our definitions of the various types of strategies assuming a given information collection scheme. We will defer until the next chapter the question of how to choose an information collection scheme upon which to base a strategy.

3.3 Types of Strategies

We will consider strategies of three types:

(i) pure strategies, which are deterministic rules for selection of an alternative in each situation that the player can encounter;

(ii) mixed strategies, which are rules for random selection of a deterministic rule (pure strategy), to be chosen when play begins and employed until a play has been completed; and

(iii) behavior strategies, which are rules for random selection of an alternative at each move encountered in any play of the game.

The use of randomized strategies may be justified by one or both of the following two arguments.

(i) In some cases, randomized strategies can be found which have greater effectiveness than any deterministic strategy†; and

(ii) even in cases where deterministic strategies exist which are as effective as the most effective of randomized strategies, randomized strategies of maximum effectiveness are usually easier to find than equally effective deterministic strategies.

3.4 Bases of Comparison of Strategy Types

We will compare the strategy types mentioned above by asking the following three questions about each one.

(i) What difficulties are encountered in the determination of a most effective strategy of the given type?

(ii) What difficulties are encountered in the implementation of a most effective strategy of the given type?

(iii) How does the effectiveness of the most effective strategy of the given type compare with the maximum effectiveness achievable over all strategy types?

3.4.1 Difficulty of Implementation

In implementation of a strategy, our concern is with the total amount of computer memory required to store the following three types of information:

---

†See Appendix A, Note 1 for comments on a common game with this property.

(i) information collected on previous alternative choices made by the parties as the game progresses;

(ii) information indicating how an alternative is to be selected in any information set that can arise (the strategy rule itself); and

(iii) information on how to combine the information of types (i) and (ii) to produce alternative choices in a given play of the game.

In making comparisons of strategy types with respect to computer storage required in their implementation, we will assume that the amount of type (iii) information is essentially the same for all three strategy types and/or small with respect to the storage requirement posed by information of type (ii). Further, as we proceed with the examination of the various strategy types, it will become obvious that the storage required for information of type (ii) is vastly greater than that required for information of type (i) in all but the most trivial of games.

Therefore, we will only be concerned with the relative amounts of computer storage required to store the strategy rules themselves.

3.4.2 Strategy Effectiveness

Our general objective is to find for the protagonist the simplest possible strategy which will maximize its minimum expected return against all possible strategies that can be employed by the antagonist. The appealing aspect of a strategy with this property is that it implies no assumptions concerning the antagonist's <u>intentions</u>; it is determined only by the players' <u>capabilities</u>.

If we wish to retain this feature in the protagonist's strategies while taking into account the antagonist's freedom to choose its information collection scheme, we must consider the strategies that can be employed by the antagonist <u>for any possible information collection scheme that he can use</u>. We therefore state our definition of strategy effectiveness, or <u>quality</u>, as follows.

<u>Definition 10</u>[†]: The <u>quality</u> $Q(X^p, R^p)$ of any strategy X^p for the <u>protagonist</u> which is based on an information collection scheme R^p, is defined by:

$$Q(X^p, R^p) = \min_{Y^a, R^a} H(X^p, R^p; Y^a, R^a)$$

where:

(i) the minimization indicated is over all antagonist strategies Y^a which can be constructed on an information collection scheme R^a, <u>and</u> over the finite set of all R^a which are possible for the antagonist under the rules of Γ; and

(ii) $H(X^p, R^p; Y^a, R^a)$ is the expected payoff to the protagonist when X^p on R^p and Y^a on R^a are employed.

The quality $Q(X^p, R^p)$ of any strategy X^p on R^p has the following property.

<u>Property 3</u>: If C^a is the complete information collection scheme for the antagonist, then:

[†]See Appendix A, Note 2 for comments on definitions 10, 11 and 12 as they would be stated for the antagonist.

$$Q(X^p, R^p) = \min_{Y^a} H(X^p, R^p; Y^a, C^a) \;.$$

Proof: By property 2, the sets of the antagonist's information partition under any information collection scheme R^a can be formed as unions of the sets of the antagonist's information partition under C^a. It follows that any antagonist strategy that can be constructed on R^a can also be constructed on C^a and, therefore, that:

$$\min_{Y^a} H(X^p, R^p; Y^a, R^a) \geq \min_{Y^a} H(X^p, R^p; Y^a, C^a) \text{ for any } R^a. \quad \blacksquare$$

Finally, note that by means of definition 10 above, we can state our general objective in the following simple form. We are interested in finding for the protagonist, a simplest possible strategy which has the maximum possible (or acceptably close to the maximum possible) quality.

3.5 Solution Strategies

In the course of our arguments, we will be interested in strategies for the protagonist which satisfy one or more of the following definitions.

Definition 11: Any strategy X^{p*} on an information collection scheme R^p is an ϵ-R solution if:

$$Q(X^{p*}, R^p) \geq \max_{X^p} Q(X^p, R^p) - \epsilon$$

where the maximization over X^p is over all possible strategies which can be constructed on information collection scheme R^p.

Definition 12: Any strategy X^{p*} on an information collection scheme R^{p*} is an ϵ-solution if:

$$Q(X^{p*}, R^{p*}) \geq \max_{X^p, R^p} Q(X^p, R^p) - \epsilon$$

where the maximization indicated is over all protagonist strategies X^p which can be constructed on an information collection scheme R^p and over the finite set of all R^p which are possible for the protagonist under the rules of Γ.

Finally, strategies which satisfy definitions 11 and 12 for $\epsilon = 0$ will simply be called R-solutions and solutions, respectively.

In the remaining sections of this chapter, we will formally define and examine each of the three strategy types mentioned in section 3.3 above. As we proceed, we will employ the following example to illustrate our arguments.

3.6 An Example Game

Consider the following simplified wheat-market game† in which a single wheat-futures trader is considered as the protagonist, and the remainder of the wheat market is considered to consist of two parts:

(i) a part for which the protagonist has a statistical description, which will be considered as nature; and

---

†We will describe this game using terms taken from the commodity market, but since the terms used are only alternative labels, an understanding of their technical meanings will not be required here. However, for interest's sake, the terms used are defined in Appendix A, Note 3.

(ii) a part for which the protagonist does not have a statistical description, which will be considered as the antagonist.[†]

The game is played as follows.

(i) To start the game, the antagonist chooses a market trend m from the set M = {rising market, falling market, stagnant market}.

(ii) The trader now begins a series of N days of trading, each of which proceeds as follows.

(a) At the beginning of the jth day, the trader chooses a "market position," q(j), from the set Q = {long-5000 bushels, short-5000 bushels, no position}.

(b) During the jth day, nature selects a "market change," c(j), from the set C = {up 1 cent per bushel, down 1 cent per bushel, no change}, under a probability distribution $= P^n(c(j)|m;\ c(k),\ k=1,\dots,j-1)$.

(iii) At the end of the Nth day, the game is terminated and the trader is paid an amount $h(c(i);\ q(i),\ i=1,\dots,N)$ which incorporates his total gains, losses, and the commissions he has had to pay to his broker.[††]

(iv) At the end of each day, the trader can record (gather) his own market position and nature's market change. (The

---

[†]See section 1.1; we are considering the "unknown" portion of the wheat market as the "second party."

[††]The broker, here considered as part of the market, wins a commision with every change in the trader's market position.

antagonist gathers no information since it chooses an alternative only at the starting move.)

The initial few moves of the game tree for any "wheat-market" game with $N \geq 2$ are shown in figure 6. Note that under its complete information collection scheme, each of the trader's information sets consist of three moves.

3.7 Pure Strategies

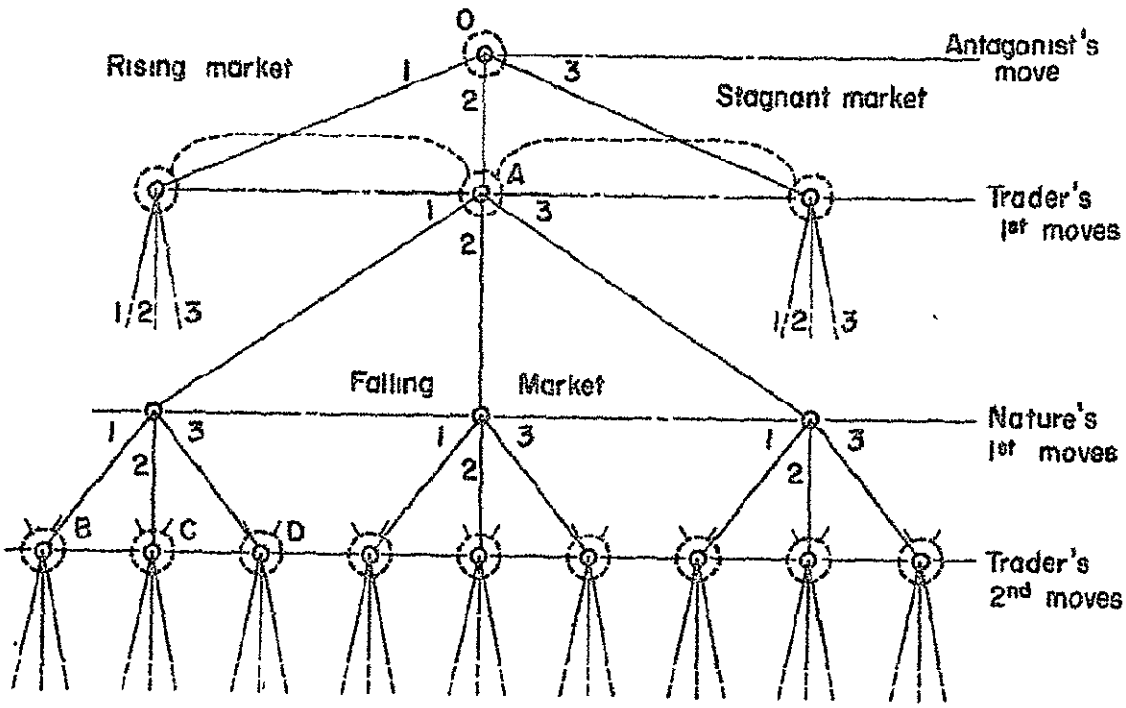
Definition 13: A pure strategy for the protagonist, based on an information collection scheme R^p, is any function σP^p which maps each information set $I^p \in IP(R^p)$ into an alternative of I^p.

It is possible for two pure strategies σP_1^p on R^p and σP_2^p on R^p to be equivalent in the sense that the probability distribution $P(W)$ for the plays $W \in K$ is identical under either pure strategy, for any given strategy X^a on R^a for the antagonist. We take this possibility into account in the definitions given below. In making these definitions, we assume that the probability of any alternative at any of nature's moves is positive.†

Definition 14: An information set $I^p \in IP(R^p)$ is realizable under a strategy X^p on R^p if there is some strategy X^a on R^a for the antagonist such that the probability that I^p occurs is positive.

For an illustration of the concept of realizability, consider the wheat-market game with $N = 2$ and a protagonist who is employing

†We can always construct a simplified equivalent game in which this assumption holds.



Antagonist's move

Alternative	Name
1	Rising market
2	Falling market
3	Stagnant market

Nature's moves

Alternative	Name
1	Up - 1 cent per bushel
2	Down- 1 cent per bushel
3	No change

Trader's moves

Alternative	Name
1	Long - 5000 bushels
2	Short - 5000 bushels
3	No position

FIGURE 6 GAME TREE FOR THE WHEAT-MARKET GAME

its complete information collection scheme C^p. Referring to the description of this game which is given in figure 6, note that if the protagonist employs a pure strategy σP^p on C^p which chooses alternative 1 at the information set labeled A, then only the information sets labeled A, B, C and D are realizable under σP^p on C^p.

Definition 15: Any two pure strategies σP_1^p on R^p and σP_2^p on R^p are equivalent if for every information set $I^p \in IP(R^p)$ which is realizable under σP_1^p on R^p or σP_2^p on R^p, the same alternative is chosen under either σP_1^p on R^p or σP_2^p on R^p.

In view of this definition, we will henceforth consider any set of equivalent pure strategies as if it were a single pure strategy.

In table 1 we show, for the wheat-market game with $N = 2$, a pure strategy for the protagonist, based on its complete information collection scheme. Note that as a consequence of definition 15, the pure strategy shown lists only alternative choices for realizable information sets.

3.7.1 Implementation of Pure Strategies

In order to implement a pure strategy, its player must store a complete list of pairs, each of which consists of a realizable information set and the alternative to be chosen when that information set occurs. This information can be conveniently stored as a vector with integer components, the first few digits of each component giving the alternative to be chosen when the information set represented by the remaining digits of that component is realized. We will refer to any pure strategy represented in this way as a <u>pure strategy in coded form</u>. In table 1, we show the example pure strategy mentioned above, in a coded form.

Clearly, the complexity of a pure strategy for a player depends upon the number of information sets which are realizable under it, and this number in turn depends upon the rules of the game and the information collection scheme he employs. In the wheat-market game with arbitrary finite N and a protagonist who is employing its complete information collection scheme, the number of information sets which are realizable for the protagonist under any pure strategy is given by:

$$\eta = \sum_{i=0}^{N-1} 3^i \; . \qquad (3.1)$$

Obviously, a pure strategy for the protagonist who is employing his complete information collection scheme in the wheat-market game, becomes quite complex for relatively small N. However, as we shall see below, for any given N and information collection scheme R, a pure strategy is the simplest of the strategy types to implement.

3.7.2 <u>Effectiveness of Pure Strategies</u>

Recall definitions 11 and 12 which established the various types of strategy solutions for a game. In terms of these definitions we can make the following statements on the effectiveness of pure strategies.

Table 1: An Example Pure Strategy for the Trader, Using a Complete Information Collection Scheme, in the Wheat-Market Game With N = 2

Example Pure Strategy

	Information Set	Alternative to Choose
A	no information	long on 1st day
B	long on 1st day, mkt. up 1st day	long on 2nd day
C	long on 1st day, mkt. down 1st day	short on 2nd day
D	long on 1st day, no change	long on 2nd day

Coded Form

$$\begin{matrix} A \\ B \\ C \\ D \end{matrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 2 & 1 & 2 \\ 1 & 1 & 3 \end{pmatrix}$$

Key to Coded Form

(i) alternatives indexed as in figure 6

(ii) left integer is alternative to choose

(iii) next two integers are observed choices in order of occurrence; i.e., trader's choice on 1st day, and nature's choice on 1st day

(iv) "0" means no information collected

(i) If a player can know at any move in the game, every previous alternative choice made and employs his complete information collection scheme†, then he has a pure strategy solution.††

(ii) In any other case, a pure strategy R-solution (or ϵ-R-solution for some given ϵ) may not exist.

(iii) The only way to tell whether or not a pure strategy R-solution (or ϵ-R-solution) exists for any given game and information collection scheme R (except for case (i)), is through the success or failure of an exhaustive attempt to find such a strategy.

3.7.3 Determination of Pure Strategy R-Solutions or ϵ-R-Solutions

Denote by $C(R^p)$, the finite set of all possible pure strategies for the protagonist when it is employing an information collection scheme R^p.

For any given game Γ and information collection scheme R^p for the protagonist, the only way to find a pure strategy ϵ-R-solution, if one exists, is to carry out a brute force search of $C(R^p)$. It is this fact which usually makes determination of a pure strategy ϵ-R-solution impractical even when such a strategy exists, for even in relatively simple games in which the protagonist employs an information collection scheme R^p under which he has relatively little information at each move on previous alternative choices, $C(R^p)$ can contain an enormous number of pure strategies.

---

†In such a situation, the protagonist is said to have "perfect information."

††Reference 1, page 123.

For example, in the wheat-market game of section 3.6 above, for general N and a complete information collection scheme R^{p*} for the protagonist, the number of pure strategies in $C(R^{p*})$ is given by:

$$\mu = \prod_{i=0}^{N-1} 3^{3^i} . \qquad (3.2)$$

In this game with N = 4, the protagonist has a maximum of only 6 pieces of alternative choice information on hand in any information set, yet, in this case, $\mu = 3^{40} = 10^{17.5}$.

3.8 Mixed Strategies

We will consider here and in the next section, two types of randomized strategies. With the introduction of randomized strategies, not just nature alone, but all of the parties in the game can be "sources of randomization." Note, however, that our definition of strategy quality has already been stated in terms of the expected return to the protagonist, and therefore need not be changed.

Definition 15: A mixed strategy for the protagonist based on an information collection scheme R^p is a probability distribution σM^p on the finite set $C(R^p)$ of all pure strategies for the protagonist which are based upon information collection scheme R^p.

The protagonist uses a mixed strategy to randomly select at the start of any play of the game, a pure strategy to be used until a play has been completed.

Note that any degenerate mixed strategy on an information collection scheme R, i.e., a mixed strategy $\sigma MD^{\dagger}$ assigning probability 1.0

† We will indicate degenerate mixed and behavior strategies with the extra letter D from now on.

to a single pure strategy σP on R, is equivalent to σP on R. Thus, we can say that the set of all possible mixed strategies on R "contains" the set of all possible pure strategies on R.

3.8.1 Implementation of Mixed Strategies

In order to implement a mixed strategy σM, a player must store two pieces of information:

(i) the positive components of σM, and

(ii) a list of all pure strategies assigned positive probability under σM.

This can be done conveniently by storing information (i) as a vector V, and information (ii) as an array M, in the following manner.

(i) Let the columns of M be the pure strategies assigned positive probability under σM, and

(ii) let the jth component of V be the probability that the pure strategy represented by the jth column in M will be selected under σM.

We will refer to any mixed strategy represented in this way as a <u>mixed strategy in coded form</u>. In table 2 we show for the wheat-market game with N = 2, an example mixed strategy in coded form, for the protagonist when he employs his complete information collection scheme.

Note that the complexity of a mixed strategy in coded form depends upon two factors:

(i) the number of pure strategies which are assigned positive probability, and

Table 2: An Example Mixed Strategy σM^{p*}, in Coded Form, for the Trader Using Its Complete Information Collection Scheme, in the Wheat-Market Game With N = 2

Coded Form

100	200	300
111	121	321
212	222	222
113	223	223
.300	.525	.175

(i) an array M, whose columns are the pure strategies given positive probability under σM^{p*}

(ii) a vector V, whose components are the positive components of σM^{p*}, with the jth component being the probability that the pure strategy represented by the jth column in M will be selected for use

(ii) the complexities of these individual pure strategies.

It is possible then, that in cases where the number of pure strategies in C(R) is enormous, any mixed strategy which has maximum quality among the mixed strategies may have a coded form which is exceedingly complex. For example, in the wheat-market game with $N = 4$, and a complete information collection scheme for the protagonist:

(i) the number of information sets which are realizable under any given pure strategy for the protagonist is given by equation (3.1) as 37, and

(ii) the number of possible pure strategies for the protagonist is given by equation (3.2) as $3^{40} = 10^{17.5}$.

In this case then, the coded form of any mixed strategy which has maximum quality among the mixed strategies may consist of a vector of real numbers of dimension $10^{17.5}$ and a rectangular array of dimension $37 \times 10^{17.5}$ with integer components. The storage requirement of such an amount of information is, of course, well beyond the capacity of any present-day computer.

3.8.2 Effectiveness of Mixed Strategies

Recalling again our definitions of the various types of strategy solutions for a game, we can make the following statement on the effectiveness of mixed strategies.

(i) For any given information collection scheme R^p for the protagonist, a mixed strategy R-solution always exists, and <u>thus</u>,

(ii) for any given game, a mixed strategy solution must exist. These facts follow directly from Von Neumann's minimax theorem.†

3.8.3 Determination of Mixed Strategy R-Solutions or ϵ-R-Solutions

Numerous methods have been developed for determination of mixed strategy R-solutions or ϵ-R-solutions in two-person, zero-sum, finite games.†† However, only one of these methods can be practically applied in situations which involve games and information collection schemes which give rise to more than perhaps 10^3 pure strategies for either player. This method, called the "method of fictitious play," is described in section 5.4.

This method, however, has an inherent drawback. Specifically, it involves the generation of a sequence of increasingly complicated mixed strategies which converges for infinitely many iterations to a mixed strategy R-solution. Since, however, only finitely many iterations can be carried out, the method, in practice, yields only mixed strategy ϵ-R-solutions, and for ϵ's of practical interest, often requires so many iterations that the mixed strategy ϵ-R-solutions it generates are too complicated to implement, even in cases where relatively simple mixed strategy ϵ-R-solutions may exist.

3.9 Behavior Strategies

The second type of randomized strategy that we will consider is defined as follows.

---

†Reference 1, section 17.

††Reference 2, Appendix 8; reference 4, pages 159-193; reference 6.

Definition 16: A behavior strategy for the protagonist, based on an information collection scheme R^p, is any function σB^p which maps each information set $I^p \in IP(R^p)$ into a probability distribution on the alternatives of I^p.

Note the conceptual difference between a mixed strategy and a behavior strategy: a mixed strategy involves a single randomization by its player at the beginning of any play, while a behavior strategy involves randomization at each of its player's moves on any play.†

Note also that any degenerate behavior strategy σBD on an information collection scheme R, i.e., a behavior strategy which chooses some alternative with probability 1.0 at every information set in IP(R), is equivalent to a pure strategy. In fact, for every pure strategy σP on R, there is an equivalent degenerate behavior strategy σBD on R, so the set of all possible behavior strategies on R can be said to "contain" the set of all possible pure strategies on R.

In table 3 we show for the wheat-market game with $N = 2$, an example behavior strategy for the protagonist, based on his complete information collection scheme.

3.9.1 Implementation of Behavior Strategies

In order to implement a behavior strategy, its player must store a complete list of pairs, each of which consist of a realizable information set and the probability distribution under which an alternative is chosen when that information set occurs. This

†Note that nature is considered to employ a behavior strategy.

Table 3: An Example Behavior Strategy for the Trader, Using a Complete Information Collection Scheme, in the Wheat-Market Game With N = 2

Example Behavior Strategy

Information Set	Probability Distribution on Alternatives					
no information	long--.50;		short--.50;		no position--.00	
long on day 1; mkt. up day 1	"	.90	"	.00	"	.10
" ; mkt. dwn. day 1	"	.50	"	.40	"	.10
" ; no change day 1	"	.90	"	.00	"	.10
short on day 1; mkt. up day 1	"	.40	"	.50	"	.10
" ; mkt. dwn. day 1	"	.00	"	.90	"	.10
" ; no change day 1	"	.00	"	.90	"	.10

Coded Form

00	11	12	13	21	22	23
.50	.90	.50	.90	.40	.00	.00
.50	.00	.40	.00	.50	.90	.90
.00	.10	.10	.10	.10	.10	.10

(i) a vector V, whose components each are the observed choices of a realizable information set in order of occurrence

(ii) an array M, whose components m_{jk} are each the probability of selecting alternative j when in the kth information set information can be conveniently stored in the form of a vector V and an array M as follows:

(i) let the components of V be descriptions of the information sets which can arise under σB^P, and

(ii) let the columns of M be the probability distributions for selection of alternatives in each information set with the jth column in M corresponding to the jth component of V.

We will refer to any behavior strategy represented in this way as a <u>behavior strategy in coded form.</u> In table 3 we show the example behavior strategy mentioned above in coded form.

Note that the complexity of a behavior strategy in coded form depends upon two factors:

(i) the number of information sets which can arise, and

(ii) the number of alternatives in each information set.

However, in most of the game and information collection scheme combinations that we wish to consider, the most complicated of behavior strategies in coded form is at most only slightly more complicated than the simplest of pure strategies in coded form, and may be vastly less complicated than the simplest of mixed strategy R-solutions in coded form.

For example, in the wheat-market game with N = 4 and a complete information collection scheme for the protagonist, the most complex of behavior strategies can be represented by a vector of dimension 820[†]

[†]Note that under some behavior strategies, every information set in $IP(R^P)$ will be realizable; and in the wheat-market game, the total number of information sets is given by: $\sum_{i=0}^{N-1} 3^{2i}$.

whose components are integers, and a rectangular array of dimension 820 x 3 whose components are real numbers. For this case:

(i) the most complicated of pure strategies can be represented as a vector of dimension 40, whose components are integers, and as we stated in section 3.8.1,

(ii) the simplest of mixed strategy R-solutions may have a coded form representation involving more than 10^{19} real numbers.

3.9.2 Effectiveness of Behavior Strategies

In reference 2, Kuhn has established several results which directly imply that:

(i) it is not necessarily true that a behavior strategy R-solution or ϵ-R-solution will exist in every game on every information collection scheme R^p for the protagonist,† but

(ii) a behavior strategy R-solution will exist in any game on every perfect-recall information collection scheme R^p for the protagonist.

From (ii) we can state the following property.

Property 4: In any game, at least one information collection scheme R^p exists upon which the protagonist can construct a behavior strategy solution.

Proof: By (ii) above, a behavior strategy R-solution σB^{p*} must exist on the protagonist's complete information collection scheme C^p, which by definition implies that:

†See Appendix A, Note 4 for an example.

$$Q(\sigma B^{p*}, C^p) \geq \max_{X^p} Q(X^p, C^p) \ .$$

But any strategy constructable on any information collection scheme R^p for the protagonist must also be constructable on C^p and, therefore:

$$Q(\sigma B^{p*}, C^p) \geq \max_{X^p} Q(X^p, C^p) \geq \max_{X^p, R^p} Q(X^p, R^p) \ .$$

By definition, then, σB^{p*} is a solution, so C^p is an information collection scheme upon which the protagonist can construct a behavior strategy solution. ▮

In Chapter 6 we will generalize on the above results by demonstrating that there are games in which behavior strategy R-solutions and ϵ-R-solutions will exist for other than perfect-recall information collection schemas for the protagonist.

3.9.3 Determination of Behavior Strategy R-Solutions or ϵ-R-Solutions

At present, no techniques† have been developed for determining behavior strategy R-solutions and ϵ-R-solutions, even where such solutions do exist. However,

(i) in Chapter 5 we will develop an algorithm for determination of behavior strategy ϵ-R-solutions in games in which the protagonist employs a perfect-recall information collection scheme; and

(ii) in Chapter 6 we will develop a heuristic algorithm for determination of behavior strategy ϵ-R-solutions (when

†We rule out games which are simple enough for algebraic techniques.

such solutions exist) in games in which the protagonist employs information collection schemes which are more general than perfect-recall information collection schemes.

Note that by property 4, the algorithm mentioned in statement (ii) above provides a method for determination of a behavior strategy solution for the protagonist in any game.

3.10 Summary

We have compared the three basic types of strategies which can be employed in games of the type that we wish to consider and have argued that in such games:

(i) pure strategy R-solutions and ϵ-R-solutions, while easy to implement, may not exist, or when they do exist, may be impossible to determine for practical reasons;

(ii) mixed strategy R-solutions and, therefore, mixed strategy ϵ-R-solutions always exist, but the only feasible technique for determining them is likely to lead to (or there may only exist), mixed strategy ϵ-R-solutions which are far too complex to implement;

(iii) behavior strategy R-solutions and ϵ-R-solutions are not much more difficult than pure strategies to implement, and in every game there are information collection schemes R^p for the protagonist, upon which behavior strategy solutions and ϵ-solutions will exist; and finally,

(iv) methods can be established for determining at least one behavior strategy ϵ-solution in any game.

From these facts we draw the conclusion that in many games of the type that we wish to consider, behavior strategies offer the only feasible means of solution.

CHAPTER 4

FORMAL PROBLEM STATEMENT AND CONTRIBUTIONS TO THE SOLUTION OF THE PROBLEM

4.0 Introduction

In previous chapters we have:

(i) defined the concept of an information collection scheme;

(ii) considered the questions of determination, implementation and effectiveness of various types of strategies that can be based on a given information collection scheme; and

(iii) concluded that in many problems of the sort that we wish to consider, only behavior strategies offer a feasible means of solution.

From this point on then, we will restrict our attention to the consideration of behavior strategies for the protagonist; further, we shall raise the question that we have so far suppressed: how do we select an information collection scheme for the protagonist?

4.1 Formal Problem Statement

When considering a given behavior strategy for possible use in his decision process, the protagonist has two questions.

(i) What is the *quality* of the given behavior strategy?

(ii) How difficult is the given behavior strategy to store and implement?

In comparing behavior strategies, some rationale for combining the answers to these questions must be assumed. We will assume here:

(i) that the protagonist has a hard lower limit on the quality of any behavior strategy that he will employ; specifically, that the behavior strategy must be an ϵ-solution for some prescribed ϵ; and

(ii) that within this limit, the behavior strategy employed should have the minimum possible complexity.

These assumptions imply the following formal problem statement: for a given $\epsilon \geq 0$, find for the protagonist, a behavior strategy ϵ-solution of minimum possible complexity.

4.2 An Approach to a Solution of the Problem

At the present point, any complete search for a minimally complex behavior strategy ϵ-solution in any non-trivial game is out of reach for the following reasons:

(i) for a given information collection scheme R^p, the techniques that we will develop, at best, find only a behavior strategy ϵ-R solution, not necessarily a minimally complex behavior strategy ϵ-R-solution; and

(ii) the number of information collection schemes R^p which conceivably must be examined is likely to be enormous even in a relatively simple problem.

However, note that any information collection scheme R^p establishes an upper bound on the complexity of any behavior strategy based upon it. For example, in the wheat-market game with $N = 2$, the most complex of behavior strategies on the protagonist's complete information collection scheme has a coded form consisting of a vector with a total of 10 components and a rectangular array with a total of 30 components.

This observation suggests a partial search organized in the following manner:

(i) rather than search for the simplest possible behavior strategy ϵ-solution on any given information collection scheme R^p, we will simply test R^p to see if a behavior strategy ϵ-solution can be constructed on it; and

(ii) rather than search over all possible information collection schemes R^p, we will examine an information collection scheme only if it looks "promising"; i.e., if an information collection scheme R^p is to be examined, then

(a) R^p should establish a bound on the complexity of any behavior strategy generated upon it which is substantially below the complexity of any behavior strategy ϵ-solution previously generated in the search; and yet

(b) R^p should seem likely to provide sufficient information so that a behavior strategy ϵ-solution on R^p will exist.

For example, in the wheat-market game with fairly high N, intuition suggests that "very old information" may be of little value to the trader. Thus, in looking for a simple behavior strategy ϵ-solution for the trader, we might try the information collection scheme R^{p*}, which on the last day of trading does not recall the market change and trader position realized on the first day. In going from the protagonist's complete information collection scheme C^p to information collection scheme R^{p*}, a considerable reduction in the maximum complexity of the protagonist's behavior strategy ϵ-solution is realized (if a behavior strategy ϵ-solution can be generated on R^{p*}). Specifically, for $N = 4$, the coded form of the most complex behavior strategy on C^p is represented by a total of 7380 real numbers while the coded form of the most complex behavior strategy on R^{p*} is represented by only 1204 real numbers.

4.3 Contributions to the Solution of the Problem

In the next two chapters we will develop the means to carry out a partial search of the kind mentioned above. Specifically,

(i) we will develop an algorithm for construction of behavior strategy ϵ-R-solutions in cases where the protagonist employs any perfect-recall information collection scheme; and

(ii) we will develop a heuristic algorithm for construction of behavior strategy ϵ-R-solutions in cases in which the protagonist employs information collection schemes which are more general than perfect-recall information collection schemes.

Then, in the final two chapters we will employ these algorithms:

(i) to find behavior strategy ϵ-solutions on the protagonist's complete information collection scheme in several pursuit-evasion decision processes; and

(ii) to carry out a search for a simple behavior strategy ϵ-solution for the doctor in a medical decision process.

CHAPTER 5

DETERMINATION OF BEHAVIOR STRATEGY ϵ-R-SOLUTIONS ON PERFECT-RECALL INFORMATION COLLECTION SCHEMES

5.0 Introduction

In this chapter we will restrict ourselves to the consideration of perfect-recall information collection schemes for the protagonist. For this case we will show that an algorithm for determination of behavior strategy ϵ-R-solutions can be obtained by combining:

(i) a transformation mapping any mixed strategy on a perfect-recall information collection scheme R, into an equivalent behavior strategy on R;

(ii) the technique for finding mixed strategy ϵ-R-solutions in finite two-person, zero-sum games, called "fictitious play"; and

(iii) the technique of dynamic programming.

5.1 Preliminaries

Before beginning our arguments, we will state formally several assumptions, and we will develop expressions for the expected payoff to the protagonist when:

(i) the players each employ a mixed strategy, and when

(ii) the players each employ a behavior strategy.

5.1.1 Assumptions

In order to facilitate the writing of simple payoff expressions, we make the following non-restrictive assumptions.

Assumption 1: Any two plays W and W′ in a game tree K for a game Γ, contain the same number of moves.

Assumption 2: Any two moves $X \in W$ and $X' \in W'$ belong to the same party if $r(X) = r(X')$.

These assumptions are not restrictive, since any game tree K can be altered to satisfy them by the insertion of insignificant† extra moves.

Assumption 3: Any two moves $X \in W$ and $X' \in W'$ contained in the same information set are of the same rank.

This assumption is not restrictive if we are considering only perfect-recall information collection schemes. In this case, each player knows at any move that has been realized, how many choices he has previously made, and thus, by Assumption 2, the rank of the realized move.

5.1.2 An Expression for the Protagonist's Return When Mixed Strategies Are Employed

Let us assume arbitrary orderings for the members of the sets $CP(R^p)$ and $CP(R^a)$, and let us denote as:

(i) σM_k^p and σM_k^a, the probabilities assugned to pure strategies $\sigma P_k^p \in CP(R^p)$ and $\sigma P_k^a \in CP(R^a)$ by mixed strategies σM^p on R^p and σM^a on R^a, respectively; and as

†A move is insignificant if identical play structures and payoffs result for any alternative chosen at that move.

(ii) $n(R^p)$ and $n(R^a)$, the total numbers of pure strategies in $CP(R^p)$ and $CP(R^a)$, respectively.

Using this notation, we can express the expected return to the protagonist when σM^p on R^p and σM^a on R^a are employed, as follows:

$$H(\sigma M^p, R^p; \sigma M^a, R^a) = \sum_{k=1}^{n(R^p)} \sum_{j=1}^{n(R^a)} H(\sigma P_k^p, R^p; \sigma P_j^a, R^a) \cdot \sigma M_k^p \cdot \sigma M_j^a \, . \qquad (5.2)$$

Note that this expression is linear in the components of the protagonist's [antagonist's] mixed strategy. Thus, for any given mixed strategy for the antagonist [protagonist], a degenerate† mixed strategy σMD^p on R^p [σMD^a on R^a] can always be found which maximizes [minimizes] the protagonist's expected return.

5.1.3 An Expression for the Protagonist's Return When Behavior Strategies Are Employed

For any play W, let us denote:

(i) by $E^p(k)$, $E^a(k)$, or $E^n(k)$, the alternative choice on W that is established by the protagonist, the antagonist, or nature, respectively, at that party's kth move on W; and

(ii) by $I^p(k)$, $I^a(k)$, or $I^n(k)$, the information set containing the protagonist's, the antagonist's, or nature's kth move on W; and

(iii) by N, M, and L, the total number of moves on W for the protagonist, the antagonist, or nature, respectively.

†Degenerate mixed and degenerate behavior strategies are introduced in sections 3.8 and 3.9, respectively.

Further, we shall assume for each $k=1,\ldots,N+M+L$, an arbitrary ordering of the information sets which contain moves of rank k, and we will denote by $I^p(k)_i$, $I^a(k)_i$, and $I^n(k)_i$, the ith possible information set for the protagonist, the antagonist, and nature, at that party's kth moves, under these assumed orderings.

Proceeding, the expected return to the protagonist when σB^p on R^p and σB^a on R^a are employed is first written as:

$$H(\sigma B^p, R^p; \sigma B^a, R^a) = \sum_{W \in K} h_1(W)\ p(W) \tag{5.3}$$

where p(W) is the probability distribution defining the probabilities of the plays $W \in K$, when σB^p on R^p and σB^a on R^a are employed, and $h_1(W)$ is the protagonist's payoff when W is realized.

By (iii) above, we can expand (5.3) to obtain:

$$H(\sigma B^p, R^p; \sigma B^a, R^a) = \sum_{C_W} p(E^p(i), i=1,N;\ E^a(j), j=1,M;\ E^n(k), k=1,L)$$

$$h_1(E^p(i), i=1,N;\ E^a(j), j=1,M;\ E^n(k), k=1,L) \tag{5.4}$$

where C_W is the set of all choice sequences which constitute plays in K. Noting that the use of behavior strategies means that the players make conditionally independent choices at every move, we can apply Bayes rule to (5.4) to obtain:

$$H(\sigma B^p, R^p; \sigma B^a, R^a) = \sum_{C_W} p(E^p(1)|I^p(1))p(E^p(2)|I^p(2)) \ldots$$

$$p(E^p(N)|I^p(N))\ p(E^a(1)|I^a(1))\ p(E^a(2)|I^a(2)) \ldots p(E^a(M)|I^a(M))$$

$$p(E^n(1)|I^n(1))\ p(E^n(2)|I^n(2))\ \ldots\ p(E^n(L)|I^n(L))\ h_1(E^p(i),$$
$$i=1,N;\ E^a(j),j=1,M;\ E^n(k),k=1,L)\ . \qquad (5.5)$$

Since there are finitely many choices and finitely many information sets, we can construct from $h_1(E^p(i),i=1,N;\ E^a(j),j=1,M;\ E^n(k),k=1,L)$ a function $g(E^p(i),\ I^p(i),i=1,N;\ E^a(j),\ I^a(j),j=1,M;\ E^n(k),\ I^n(k),$ $k=1,L)$ such that $H(\sigma B^p,R^p;\sigma B^a,R^a)$ can be written in the following form:[†]

$$H(\sigma B^p,R^p;\sigma B^a,R^a) = \sum_{I^p(1)}\ \sum_{E^p(1)|I^p(1)_i} p(E^p(1)|I^p(1)_i)\ \ldots$$

$$\sum_{I^p(N)}\ \sum_{E^p(N)|I^p(N)_i} p(E^p(N)|I^p(N)_i)\cdot \sum_{I^a(1)}\ \sum_{E^a(1)|I^a(1)_i}$$

$$p(E^a(1)|I^a(1)_i)\ \ldots\ \sum_{I^a(M)}\ \sum_{E^a(M)|I^a(M)_i} p(E^a(M)|I^a(M)_i)$$

$$\Big[\sum_{I^n(1)}\ \sum_{E^n(1)|I^n(1)_i} p(E^n(1)|I^n(1)_i)\ \ldots\ \sum_{I^n(L)}\ \sum_{E^n(L)|I^n(L)_i}$$

$$p(E^n(L)|I^n(L)_i)\cdot g(E^p(j),I^p(j),j=1,N;\ E^a(k),I^a(k),k=1,M;$$

$$E^n(\ell),I^n(\ell),\ell=1,L)\Big] \qquad (5.6)$$

where the notation $\sum_{X|Y}$ means summation over the set of alternatives X that are available in information set Y.

[†]See Appendix A, Note 5 for an example of this construction.

Carrying out the summations of the term enclosed by brackets in (5.6) (using the behavior strategies given for nature in the rules of the game), we can denote the result as $d(E^p(j), I^p(j), j=1,N;\ E^a(k), I^a(k), k=1,M)$, and rewrite (5.6) as:

$$H(\sigma B^p, R^p; \sigma B^a, R^a) = \sum_{I^p(1)} \sum_{E^p(1)|I^p(1)_i} \cdots \sum_{I^p(N)} \sum_{E^p(N)|I^p(N)_i}$$

$$\sum_{I^a(1)} \sum_{E^a(1)|I^a(1)_i} \cdots \sum_{I^a(M)} \sum_{E^a(M)|I^a(M)_i} p(E^p(1)|I^p(1)_i) \cdots$$

$$\cdot\ p(E^p(N)|I^p(N)_i)\, p(E^a(1)|I^a(1)_i) \cdots p(E^a(M)|I^a(M)_i)\, d(E^p(j),$$

$$I^p(j), j=1,N;\ E^a(k), I^a(k), k=1,M)\ . \qquad (5.7)$$

Having completed the preliminary arguments, we will now begin the development of our algorithm by stating the theorem and transformation upon which the algorithm depends.

5.2 Kuhn's Transformation and Theorem

In reference 2, Kuhn gives the following transformation mapping any mixed strategy σM on R into an associated behavior strategy $\ddot{\sigma}B$ on R.

Transformation (Kuhn): The behavior strategy $\ddot{\sigma}B$ on R which is associated with a mixed strategy σM on R, is defined by the following transformation:

$$P(\nu|I) = \Big[\sum_{D_\nu} \sigma M_i \Big] \div \Big[\sum_{D} \sigma M_i \Big]\ , \qquad (5.8)$$

where:

(i) $P(\nu|I)$ is the probability that the player of the associated behavior strategy selects alternative ν when information set $I \in IP(R)$ is realized,

(ii) D is the set of all pure strategies σP_i for the given player under which information set I is realizable, and

(iii) D_ν is the subset of D consisting of all pure strategies under which I is realizable and ν is chosen.

To interpret this transformation, suppose that the protagonist employs a mixed strategy σM^p on R^p in a game Γ, and imagine that an observer watches many occurrences of Γ and collects statistics on the relative frequencies of selection of each alternative in any given information set $I^p \in IP(R^p)$. If the observer watches "long enough," it will accumulate relative frequencies of selection for each alternative which are identical to the corresponding probabilities of selection defining the behavior strategy $\bar{\sigma}B^p$ on R^p which is associated with σM^p on R^p.

In reference 2, Kuhn employs the transformation given above to prove an interesting theorem which, in terms of our definitions and nomenclature, can be stated as follows:

Theorem (Kuhn): In any finite extended game in which the players each employ a perfect-recall information collection scheme, the payoff to each player is identical under any set of mixed strategies or the corresponding set of associated behavior strategies.

The significance of this theorem in the case of a two-person, zero-sum finite extended game is indicated by the following corollary.

Corollary 1:† If R^p is a perfect recall information collection scheme, and if $\bar{\sigma}B^{p*}$ on R^p is the behavior strategy associated with a mixed strategy ϵ-R-solution σM^{p*} on R^p, then $\bar{\sigma}B^{p*}$ on R^p is a behavior strategy ϵ-R-solution.

5.3 Use of the Transformation

In view of the above theorem and transformation, an obvious way to obtain behavior strategy ϵ-R-solutions when R is a perfect-recall information collection scheme, is to construct a mixed strategy ϵ-R-solution on R and transform it into an equivalent behavior strategy ϵ-R-solution.

However, in problems of the sort that we wish to consider, the method of fictitious play is the only feasible technique for generating mixed strategy ϵ-R-solutions, and this technique tends to generate mixed strategy ϵ-R-solutions which are so complex that the implied intermediate steps of storing these strategies will not be feasible.

In the following sections, however, we will show that there is a second way to employ Kuhn's transformation in conjunction with the method of fictitious play to generate behavior strategy ϵ-R-solutions while avoiding the above-mentioned storage problem.

†The proof is given in Appendix B.

5.4 The Method of Fictitious Play[†]

The method of fictitious play (or equivalently, the fictitious play algorithm) was first suggested by Brown.[7] The idea upon which the algorithm is based is both straightforward and intuitively appealing.[††] Specifically, if we wish to find a mixed strategy ϵ-R-solution for the protagonist on an information collection scheme R^p, we can assign to the antagonist its complete information collection scheme C^a, and have the players engage in a series of fictitious plays (or repetitions) of the game as follows:

(i) at the first repetition of the game, each player chooses an arbitrary pure strategy σP^p on R^p or σP^a on C^a, and

(ii) at subsequent repetitions, each player chooses a pure strategy on his given information collection scheme which maximizes his expected return, assuming a mixed strategy for the opponent, which weights each of the opponent's previously used pure strategies in proportion to the number of times it has been used, and then

(iii) as the number of repetitions becomes large, the expected return to the protagonist at each repetition should approach (and eventually surpass) the minimum expected return that it can guarantee itself by employing a mixed strategy ϵ-R-solution on R^p.

[†]The method and its properties will be described in terms of our definitions and notation.

[††]See Appendix A, Note 6, for observations on the intuitive appeal of the method.

In reference 8, Robinson proves that statement (iii) is indeed true. Note that since the initial pure strategies chosen by the players are arbitrary, and since there may be many pure strategies which maximize or minimize a player's return at any given repetition, the sequences of mixed strategies generated by the fictitious play algorithm are certainly not unique. The convergence of the method depends only on the fact that the mixed strategies of the sequences are related in the manner described above.

To describe the fictitious play algorithm formally then, we will state the following theorem which combines Robinson's result with a pair of recursive equations describing the manner in which the mixed strategies generated in a sequence of fictitious play repetitions are related.

Theorem (Brown-Robinson): Denote by:

(i) $\sigma M^p(j)$ on R^p [$\sigma M^a(j)$ on C^a] the mixed strategy that is generated for the protagonist [antagonist] at the jth repetition of the fictitious play algorithm, and by

(ii) $\underset{\sigma P^p}{\text{amax}}\, H(\sigma P^p, R^p; \sigma M^a(j), C^a)$ [$\underset{\sigma P^a}{\text{amin}}\, H(\sigma M^p(j), R^p; \sigma P^a, C^a)$] any pure strategy† which maximizes [minimizes] $H(\sigma P^p, R^p; \sigma M^a(j), C^a)$ [$H(\sigma M^p(j), R^p; \sigma P^a, C^a)$].

Then any pair of mixed strategy sequences ($\sigma M^p(1)$ on R^p, $\sigma M^p(2)$ on $R^p, \ldots$) and ($\sigma M^a(1)$ on C^a, $\sigma M^a(2)$ on $C^a, \ldots$) which satisfies the following pair of recursive equations:

---

†In section 5.1.2 we argued that pure strategies exist which maximize [minimize] this expression.

$$\text{(i)}\quad \sigma M^p(j+1) = \left(\frac{1}{j+1}\right)\left[(j)\sigma M^p(j) + \underset{\sigma P^p}{\text{amax}}\, H(\sigma P^p, R^p; \sigma M^a(j), C^a)\right] \qquad (5.9)$$

$$\text{(ii)}\quad \sigma M^a(j+1) = \left(\frac{1}{j+1}\right)\left[(j)\sigma M^a(j) + \underset{\sigma P^a}{\text{amin}}\, H(\sigma M^p(j), R^p; \sigma P^a, C^a)\right] \qquad (5.10)$$

(where the "plus" in these expressions has the obvious operational meaning),[††] must have the following properties: for any given $\epsilon > 0$, there exists an integer J such that for all $j \geq J$,

(i) $H(\sigma M^p(j), R^p; \sigma M^a, C^a) > \max_{\sigma M^p} H(\sigma M^p, R^p; \sigma M^a(j), C^a) - \epsilon$ for any mixed strategy σM^a on C^a, and

(ii) $H(\sigma M^p, R^p; \sigma M^a(j), C^a) < \min_{\sigma M^a} H(\sigma M^p(j), R^p; \sigma M^a, C^a) + \epsilon$ for any mixed strategy σM^p on R^p.

The significance of this theorem is indicated by the following corollary.

<u>Corollary 2:</u>[†] If ($\sigma M^p(1)$ on R^p, $\sigma M^p(2)$ on $R^p, \ldots$), ($\sigma M^a(1)$ on C^a, $\sigma M^a(2)$ on $C^a, \ldots$) is any pair of mixed strategy sequences generated by the fictitious play algorithm, then for any $\epsilon > 0$, there exists an integer J such that for all $j \geq J$, $\sigma M^p(j)$ on R^p is an ϵ-R-solution.

5.5 Use of the Transformation (Continued)

In practice, for games and ϵ's of the sort that we wish to consider, the mixed strategies generated by the fictitious play

---

[†] The proof is given in Appendix B.

[††] Represent the maximizing [minimizing] pure strategy as a degenerate mixed strategy and add.

algorithm become too complex to store long before sufficiently many repetitions have been carried out to generate a mixed strategy ϵ-R-solution for the protagonist. As mentioned in section 5.3, this means that the obvious route to a behavior strategy ϵ-R-solution, i.e., find a mixed strategy ϵ-R-solution and transform it, is not feasible.

This problem can be avoided, however, by transforming the fictitious play algorithm itself so that it generates a sequence of behavior strategies which converges to a behavior strategy ϵ-R-solution, rather than a sequence of mixed strategies which converges to a mixed strategy ϵ-R-solution. Denoting Kuhn's transformation as T, this transformation of the fictitious play algorithm is accomplished by replacing equations (5.9) and (5.10), respectively, by equations (5.11) and (5.12) shown below:[†]

$$\text{(i)}\quad \sigma S^p(j+1) = \left(\frac{1}{j+1}\right)\left[(j)\sigma B^p(j)+T\left[\underset{\sigma P^p}{\text{amax}}\ H(\sigma P^p, R^p; \sigma B^a(j), C^a)\right]\right]^{\dagger\dagger} \quad (5.11)$$

$$\text{(ii)}\quad \sigma B^a(j+1) = \left(\frac{1}{j+1}\right)\left[(j)\sigma B^a(j)+T\left[\underset{\sigma P^a}{\text{amin}}\ H(\sigma B^p(j), R^p, \sigma P^p, C^a)\right]\right] \quad (5.12)$$

where the indicated transformations of the pure strategies are

---

[†]The derivation of equations (5.11) and (5.12) from equations (5.9) and (5.10) is given in Appendix A, Note 7.

[††]In some cases, simpler expressions for $H(\sigma P^p, R^p; \sigma B^a(j), C^a)$ and $H(\sigma B^p(j), R^p; \sigma P^a, C^a)$ may be used which reduce the amount of computation required in each repetition; see Appendix A, Note 8.

accomplished by employing equation (5.8) and considering the pure strategies as degenerate mixed strategies, and where the "plus" in each of these expressions means addition of the probability distributions at each information set of $IP(R^p)$ or $IP(C^a)$.

With this replacement, however, we have not yet achieved a practical means for finding behavior strategy ϵ-R-solutions in games of the sort that we wish to consider. One more difficulty remains, and we consider it below.

5.6 Dynamic Programming and Fictitious Play

In implementing the fictitious play algorithm, note that at each repetition we must determine pure strategies σP^{p*} on R^p and σP^{a*} on C^a, respectively, satisfying expressions (5.13) and (5.14) below.

$$H(\sigma P^{p*}, R^p; \sigma B^a(j), C^a) = \max_{\sigma P^p} H(\sigma P^p, R^p; \sigma B^a(j), C^a) \tag{5.13}$$

$$H(\sigma B^p(j), R^p; \sigma P^{a*}, C^a) = \min_{\sigma P^a} H(\sigma B^p(j), R^p; \sigma P^a, C^a) \ . \tag{5.14}$$

Recall, however, that in section 3.7.3 we commented that in games of the sort that we wish to consider, the number of pure strategies for each player can be enormous.[†] Obviously then, the determination of pure strategies satisfying (5.13) and (5.14) by means of brute force searches of the players' pure strategy sets is not feasible.

[†]The protagonist in the wheat-market game with $N = 4$ has $10^{17.5}$ pure strategies when employing its complete information scheme.

Continuing our arguments with the maximization problem (the minimization problem follows easily), note that any degenerate behavior strategy σBD^{p*} on R^p which satisfies:

$$H(\sigma BD^{p*}, R^p; \sigma B^a(j), C^a) = \max_{\sigma B^p} H(\sigma B^p, R^p; \sigma B^a(j), C^a) \tag{5.15}$$

also satisfied expression (5.13).[†] Thus, we may consider the maximization problem as a search for a solution σBD^{p*} of (5.15).

Using expression (5.7), we can rewrite (5.15) as:

$$H(\sigma BD^{p*}, R^p; \sigma B^a(j), C^a) = \max_{\sigma B^p} \Bigg[\sum_{I^p(1)} \sum_{E^p(1)|I^p(1)} \cdots \sum_{I^p(N)}$$

$$\sum_{E^p(N)|I^p(N)_i} p(E^p(1)|I^p(1)_i) \cdots p(E^p(N)|I^p(N)_i) \Bigg[\sum_{I^a(1),}$$

$$\sum_{E^a(1)|I^a(1)_i} \cdots \sum_{I^a(M)} \sum_{E^a(M)|I^a(M)_i} p(E^a(1)|I^a(1)_i) \cdots$$

$$p(E^a(M)|I^a(M)_i) \cdot d(E^p(\ell), I^p(\ell), \ell=1,N; E^a(k), I^a(k), k=1,M) \Bigg] \Bigg] \tag{5.16}$$

where the antagonist's probability distributions are specified by $\sigma B^a(j)$ on C^a. Carrying out the summations of the inner bracketed terms in (5.16) and indicating the result as $e(E^p(i), I^p(i), i=1,N; \sigma B^a(j))$, and expanding σB^p to indicate the collections of distributions

---

[†]Recall that the set of all behavior strategies on any information collection scheme R^p contains a degenerate behavior strategy equivalent to any pure strategy on R^p.

specified for each rank at which the protagonist has moves, we can rewrite (5.16) as:

$$H(\sigma BD^{p*}, R^p; \sigma B^a(j), R^a) = \max_{p(E^p(1)|I^p(1))} \max_{p(E^p(2)|I^p(2))} \cdots$$

$$\max_{p(E^p(N)|I^p(N))} \Big[\sum_{I^p(1)} \sum_{E^p(1)|I^p(1)_i} p(E^p(1)|I^p(1)_i) \sum_{I^p(2)}$$

$$\sum_{E^p(2)|I^p(2)} p(E^p(2)|I^p(2)_i) \cdots \sum_{I^p(N)} \sum_{E^p(N)|I^p(N)_i} p(E^p(N)|I^p(N)_i)$$

$$\cdot\ e(E^p(k), I^p(k), k=1,N; \sigma B^a(j)) \Big]. \qquad (5.17)$$

Now, if R^p is a perfect-recall information collection scheme, then $I^p(j)$ completely specifies $E^p(k)$ and $I^p(k)$ for $k=1, j-1$. This implies that a different probability distribution is chosen for every realizable combination of values for $E^p(k), I^p(k), k=1, j-1$, which, in turn, allows us to rewrite equation (5.17) as:

$$H(\sigma BD^{p*}, R^p; \sigma B^a(j), C^a) = \sum_{I^p(1)} \max_{p(E^p(1)|I^p(1)_i)} \sum_{E^p(1)|I^p(1)_i}$$

$$p(E^p(1)|I^p(1)_i) \Big[\sum_{I^p(2)} \max_{p(E^p(2)|I^p(2)_i)} \sum_{E^p(2)|I^p(2)_i}$$

$$p(E^p(2)|I^p(2)_i \cdots \Big[\sum_{I^p(N)} \max_{p(E^p(N)|I^p(N)_i)} \sum_{E^p(N)|I^p(N)_i}$$

$$p(E^p(N)|I^p(N)_i) \cdot e(E^p(N), I^p(N); \sigma B^a(j))\Big] \ldots\Big] \qquad (5.18)$$

Examining the right side of equation (5.18), we see that it can be maximized sequentially, starting with the probability distributions $p(E^p(N)|I^p(N))$ and working backwards to the distributions $p(E^p(1)|I^p(1))$. This technique of backwards sequential maximization is commonly known as dynamic programming.[10,11]

Note that this sequential maximization can be accomplished with the choice of a degenerate probability distribution at each information set, thus producing a degenerate behavior strategy which maximizes $H(\sigma B^p, R^p; \sigma B^a(j), C^a)$ or $H(\sigma P^p, R^p; \sigma B^a(j), C^a)$.†

In light of these observations, and noting that $T(\sigma BD) = \sigma BD$ for any degenerate behavior strategy σBD, we will change the form of our algorithm once more. Specifically, in the final form of the algorithm, instead of equations (5.11) and (5.12), we will employ equations (5.19) and (5.20) given below:

$$\sigma B^p(j+1) = \left(\frac{1}{j+1}\right)\Big[(j)\sigma B^p(j) + \underset{\sigma BD^p}{\text{amax}}\ H(\sigma B^p, R^p; \sigma B^a(j), C^a)\Big] \qquad (5.19)$$

$$\sigma B^a(j+1) = \left(\frac{1}{j+1}\right)\Big[(j)\sigma B^a(j) + \underset{\sigma BD^a}{\text{amin}}\ H(\sigma B^p(j), R^p; \sigma B^a, C^a)\Big] \qquad (5.20)$$

where the max and min operations will be accomplished using dynamic programming.

In order to see the sort of computational advantage to be gained by using the dynamic programming technique to find a pure strategy

†Recall that this was our original aim; see equation (5.13).

maximizing $H(\sigma P^p, R^p; \sigma B^a(j), C^a)$, rather than a brute force search, consider again the wheat-market game with $N = 4$ and a protagonist who employs his complete information collection scheme. In this game:

(i) using a brute force search to find a maximizing pure strategy requires comparison of the values taken by $H(\sigma P^p, R^p; \sigma B^a(j), C^a)$ over $10^{17.5}$ possible pure strategies, or $10^{17.5}$ scalar comparisons, but

(ii) using the dynamic programming technique to find a maximizing pure strategy requires comparisons of values for 3 alternatives each, in 820 information sets, or $2460(10^{3.5})$ scalar comparisons.

Clearly then, the use of the dynamic programming technique to carry out the maximization and minimization operations required in fictitious play is essential in problems of the sort that we wish to consider.

5.7 <u>Summary</u>

Our objective in this chapter was to construct an algorithm for determination of behavior strategy ϵ-R-solutions when the protagonist employs a perfect-recall information collection scheme. In the preceding sections we have shown that such an algorithm can be constructed by combining Kuhn's transformation, the method of fictitious play, and the technique of dynamic programming.

We summarize the algorithm obtained, then, in the following theorem.

<u>Theorem 1</u>: Let R^p be any perfect recall information collection scheme for the protagonist, let C^a be the complete information collection scheme for the antagonist, and let the pair of behavior strategy sequences ($\sigma B^p(1)$ on R^p, $\sigma B^p(2)$ on $R^p, \ldots$), ($\sigma B^a(1)$ on C^a, $\sigma B^a(2)$ on $C^a, \ldots$) satisfy the recursive expressions:

$$\text{(i)} \quad \sigma B^p(j+1) = \left(\frac{1}{j+1}\right)\left[(j)\sigma B^p(j) + \underset{\sigma BD^p}{\text{amax}}\; H(\sigma B^p, R^p; \sigma B^a(j), R^a)\right]$$

$$\text{(ii)} \quad \sigma B^a(j+1) = \left(\frac{1}{j+1}\right)\left[(j)\sigma B^a(j) + \underset{\sigma BD^a}{\text{amin}}\; H(\sigma B^p(j), R^p; \sigma B^a, R^a)\right],$$

where $\sigma B^p(1)$ on R^p and $\sigma B^a(1)$ on C^a are both arbitrary. Then the sequence ($\sigma B^p(1)$ on R^p, $\sigma B^p(2)$ on $R^p, \ldots$) has the property that: for any $\epsilon > 0$, there exists an integer J such that for any $j \geq J$, $\sigma B^p(j)$ on R^p is a behavior strategy ϵ-R-solution.

In the flow diagram of figure 7, we indicate an implementation of this algorithm, which we will henceforth refer to as the <u>perfect-recall fictitious play algorithm</u>.

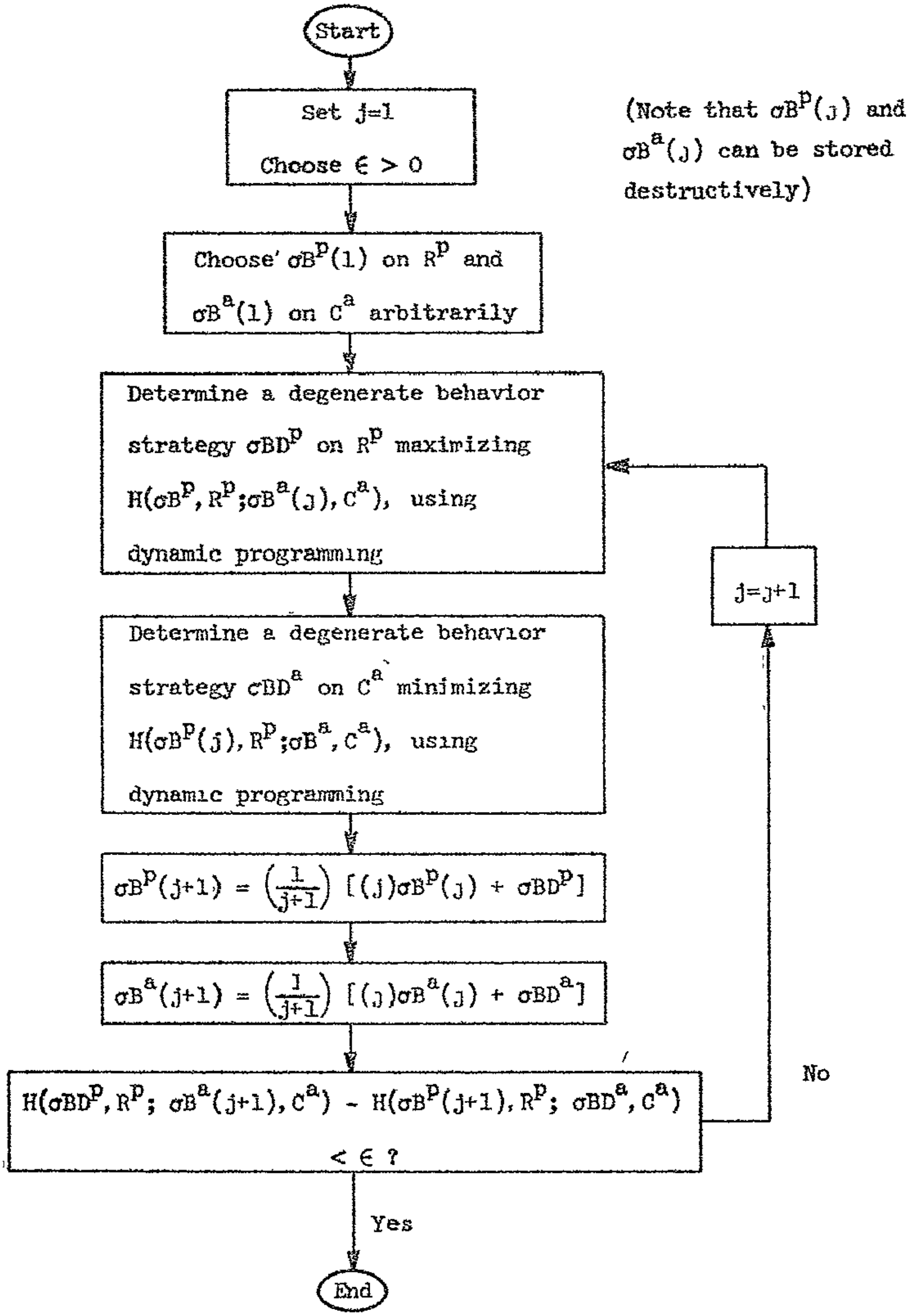


FIGURE 7. FLOW DIAGRAM FOR THE PERFECT-RECALL FICTITIOUS PLAY ALGORITHM.

CHAPTER 6

DETERMINATION OF BEHAVIOR STRATEGY ∈-R-SOLUTIONS ON RELATIVELY GENERAL INFORMATION COLLECTION SCHEMES

6.0 Introduction

The determination of a behavior strategy ∈-R-solution for the protagonist on a general information collection scheme is usually a much more difficult problem than the corresponding problem with a perfect-recall information collection scheme. This difference in difficulty is due to two factors:

(i) a behavior strategy ∈-R-solution does not necessarily exist on every information collection scheme that the protagonist can employ; and

(ii) the behavior strategy that is associated with a given mixed strategy ∈-R-solution on a general information collection scheme R^p, is not necessarily of maximum quality in the set of behavior strategies possible on R^p.

In this chapter we will develop an approach to the problem of determining a behavior strategy ∈-R-solution on a given information collection scheme which can be applied for relatively general information collection schemes R^p.

Our arguments will proceed as follows:

(i) We will describe an n-person, non-cooperative game Γ' which can be defined for any given two-person, zero-sum finite extended game with a given information collection scheme R^p for the protogonist.

(ii) We will define the concept of an entity-behavior equilibrium point in Γ', and we will show that a behavior strategy on R^p is of maximum quality in the set of behavior strategies possible on R^p only if it corresponds to an entity-behavior equilibrium point in Γ'.

(iii) We will show that in order for a behavior strategy on R^p, which corresponds to an entity-behavior equilibrium point in Γ', to be an R-solution, it is sufficient but not necessary for R^p to be a perfect-recall information collection scheme.

(iv) On the basis of these arguments, we will conclude that behavior strategies on R^p, which correspond to entity-behavior equilibrium points in Γ', are promising candidates for behavior strategy ϵ-R-solutions in Γ; and we will describe a form of the fictitious play algorithm which can be used to generate such behavior strategies.

6.1 Preliminaries

Before beginning our arguments, we will define the class of imperfect-recall information collection schemes that we intend to consider, and we will develop a convenient form of expression for the expected return to the protagonist when both players employ behavior strategies.

6.1.1 Imperfect-Recall Information Collection Schemes to be Considered

We will again make Assumptions 1, 2 and 3, which were stated in the previous chapter in section 5.1.1. Note, however, that if we are no longer limited to perfect-recall information collection schemes, then Assumption 3 is restrictive. Specifically, in making this assumption we are saying:

(i) that the protagonist has perfect-recall of the number of choices that it has previously made, or equivalently,

(ii) that the protagonist knows the rank of the move that has been realized whenever he is to choose an alternative.

In this chapter then, we will only consider information collection schemes which induce information partitions which satisfy Assumption 3.† We will refer to such information collection schemes as admissible information collection schemes.

6.1.2 An Expression for the Expected Return to the Protagonist

Employing the notation of section 5.1.3 and assuming for each $k=1,\overline{N+M+1}$ an arbitrary ordering for the information sets containing moves of rank k, we define partitioned vectors X^{pk} for $k=1,N$, and Y^{ak} for $k=1,M$ as follows.

(a) Let the ℓth component of X^{pk} be given by:

†A slightly more general assumption can be made; see Appendix A, Note 9.

$X_\ell^{pk} = p(E^p(k) = i \mid I^p(k)_j)$ where $\ell = j + \sum_{m=1}^{i-1} n_{mk}^p$ and

n_{mk}^p is the total number of alternatives of information set $I^p(k)_m$; and

(b) let the ℓth component of Y^{ak} be given by: $Y_\ell^{ak} = p(E^a(k) = i \mid I^a(k)_j)$ where $\ell = j + \sum_{m=1}^{i-1} n_{mk}^a$ and n_{mk}^a is the total number of alternatives of information set $I^a(k)_m$.

In figure 8 we show how such vectors can be defined in the example game of section 2.4[†] when each player is employing its complete information collection scheme.

Note that the set of vectors $(X^{p1}, X^{p2}, \ldots, X^{pN})$ constitutes a behavior strategy for the protagonist with the mth partition of X^{pk} being a probability distribution on the alternatives of information set $I^p(k)_m$.

Continuing, denote by ∇ the inner product operation defined by the following example of its use,

$$\sum_i \sum_j \sum_k \alpha_i \beta_j \gamma_k A_{ijk} \triangleq \alpha \nabla \beta \nabla \gamma \nabla A$$

and let $\prod_{i=1}^{N} X^{pi} \triangleq X^{p1} \nabla X^{p2} \nabla \ldots \nabla X^{pN}$.

Referring to expression (5.7) of section 5.1, we can construct an N + M-dimensional array F from the function $d(E^p(j), I^p(j), j=1,N; E^a(k), I^a(k), k=1,M)$ such that, by employing the notation given above,

---

[†]The "card" game.

<u>Key</u>

(i) Information sets A, B, C, D, E shown in figure 5

(ii) Assumed orderings on information sets at each rank are as follows: $I^p(1)_1 = A$; $I^a(1)_1 = B$; $I^a(1)_2 = C$; $I^p(2)_1 = D$; $I^p(2)_2 = E$

$$X^{p1} = \begin{pmatrix} P^p(1|A) \\ P^p(2|A) \end{pmatrix} \quad ; \quad Y^{a1} = \begin{pmatrix} P^a(1|B) \\ P^a(2|B) \\ \hdashline P^a(1|C) \\ P^a(2|C) \end{pmatrix} \quad ; \quad X^{p2} = \begin{pmatrix} P^p(1|D) \\ P^p(2|D) \\ \hdashline P^p(1|E) \\ P^p(2|E) \end{pmatrix}$$

FIGURE 8. BEHAVIOR STRATEGY VECTORS.

we can rewrite expression (5.7) in the following simple form:

$$H(\sigma B^p, R^p; \sigma B^a, R^a) = \left(\prod_{i=1}^{N} X^{pi} \right) \triangledown \left(\prod_{j=1}^{M} Y^{aj} \right) \triangledown F \, . \qquad (6.1)$$

Note that for a particular game Γ, the structure of this expression is determined by the information collection schemes employed by the players.

In the following sections we will define a particular (N + M)-person, non-cooperative game that is suggested by the structure of expression (6.1), and we will show that behavior strategies which correspond to equilibrium strategies in the defined non-cooperative game are promising candidates for ∈-R-solutions in the original game.

6.2 The Associated Non-Cooperative Game

Definition 18: The associated non-cooperative game Γ′ for a given two-person, zero-sum, finite extended game Γ with a given information collection scheme R^p for the protagonist, is the (N + M)-person non-cooperative† game defined by:

(i) assigning to the antagonist its complete information collection scheme C^a,

(ii) considering each rank of player moves in Γ to be under the control of a separate and independent entity, and

(iii) considering the payoff to each entity to be identical to that of the player whose moves the entity is controlling.

---

†The game we wish to consider is completely described in this definition; it is also a non-cooperative game as defined by Nash[12].

Denoting the entity controlling the kth moves of the protagonist [antagonist] as the k^p-entity [k^a-entity], we define the concept of an entity-behavior strategy as follows.

<u>Definition 19</u>: An <u>entity-behavior strategy for the k^p-entity</u> [k^a-entity] is any function which maps each information set $I^p(k)_i$ [$I^a(k)_i$] into a probability distribution on the alternatives of that information set.

Note that the set of all possible entity-behavior strategies for the k^p-entity [k^a-entity] corresponds to the set of all possible values for the vector X^{pk} [Y^{ak}]. In view of this correspondence, we can refer to the vector X^{pk} [Y^{ak}] as an entity-behavior strategy for the k^p-entity [k^a-entity], and we can consider any behavior strategy $(X^{pi}, i=1,N)$ [$(Y^{ai}, i=1,M)$] as a collection of entity-behavior strategies for the entities associated with the protagonist [antagonist].

We will consider the associated non-cooperative game to be played in the following manner:

(i) each entity k^p, for $k=1,N$, [k^a for $k=1,M$], employing its behavior strategy X^{pk} [Y^{ak}], simultaneously chooses an [illegible] $I^p(k)$ [illegible] $I^a(k)$] under its control, and then

(ii) for the complete set of choices thus determined, the expected return to the [illegible] is computed and awarded positively to each protagonist entity and negatively to each antagonist entity.

Finally, note that in any associated (N + M)-"person", non-cooperative game, the concept of information gathering has no explicit meaning since each entity acts only once.

6.3 A Necessary Condition for Behavior Strategies, σB^p on R^p, of Maximum Quality

We will begin this section by defining the concepts of an equilibrium point in behavior strategies and an equilibrium point in entity-behavior strategies. We will then show that in order for a behavior strategy σB^p on R^p to have maximum quality in the set of behavior strategies possible in R^p, it is necessary for σB^p on R^p to correspond to an entity-behavior equilibrium point in the associated non-cooperative game Γ'.

Definition 20: A pair of behavior strategies (X^{*pi}, i=1,N), (Y^{*ai}, i=1,M) in a game Γ, with information collection schemes R^p for the protagonist and R^a for the antagonist, is a player-behavior equilibrium point if:

$$\max_{(X^{pi},\, i=1,N)} \left(\prod_{i=1}^{N} X^{pi}\right) \triangledown \left(\prod_{j=1}^{M} Y^{*aj}\right) \triangledown F \leq \left(\prod_{i=1}^{N} X^{*pi}\right) \triangledown \left(\prod_{j=1}^{M} Y^{*aj}\right)$$

$$\triangledown F \leq \min_{(Y^{aj},\, j=1,M)} \left(\prod_{i=1}^{N} X^{*pi}\right) \triangledown \left(\prod_{j=1}^{M} Y^{aj}\right) \triangledown F .$$

Definition 21: A set of entity-behavior strategies (X^{*pi}, i=1,N; Y^{*aj}, j=1,M) in an associated (N + M)-person, non-cooperative game Γ', is an entity-behavior equilibrium point if:

$$\max_{X^{pi}} \left(\prod_{j=1}^{i-1} X^{*pj}\right) \nabla X^{pi} \nabla \left(\prod_{j=i+1}^{N} X^{*pj}\right) \nabla \left(\prod_{k=1}^{M} Y^{*ak}\right) \nabla F \leq$$

$$\left(\prod_{j=1}^{N} X^{*pj}\right) \nabla \left(\prod_{k=1}^{M} Y^{*ak}\right) \nabla F \leq$$

$$\min_{Y^{a\ell}} \left(\prod_{j=1}^{N} X^{*pj}\right) \nabla \left(\prod_{k=1}^{\ell-1} Y^{*ak}\right) \nabla Y^{a\ell} \nabla \left(\prod_{k=\ell-1}^{M} Y^{*ak}\right) \nabla F \qquad (6.2)$$

for $i=1,N$ and $\ell=1,M$.[†]

We complete this section by stating the following two theorems.

Theorem 2:[††] For any associated game Γ', there exists at least one entity-behavior equilibrium point.

Theorem 3: A behavior strategy σB^{p*} on R^p is of maximum quality in the set of all possible behavior strategies σB^p on R^p only if it corresponds to an entity behavior equilibrium point of the associated non-cooperative game Γ'.

Proof: A behavior strategy σB^p on R^p is of maximum possible quality on R^p only if it corresponds to some player-behavior equilibrium point based on R^p and the antagonist's complete information collection scheme C^a; but further, any player-behavior equilibrium point based on R^p and C^a must correspond to an entity-behavior equilibrium point of the associated non-cooperative game. |

Suppose, then, that we have found an entity-behavior equilibrium point in the associated non-cooperative game Γ' corresponding to a

---

[†]A total of NM expressions of the form of 6.2.

[††]The proof of this theorem is given in Appendix B.

given information collection scheme R^p and a given two-person, zero-sum, finite extended game Γ. By theorem 3, the corresponding behavior strategy σB^p on R^p must satisfy the set of inequalities described by expression (6.2), and these inequalities are necessary conditions for a behavior strategy of maximum quality in the set of behavior strategies possible in R^p. The obvious question is: are these necessary conditions also sufficient? The next section gives a partial answer to this question.

6.4 A Sufficient but Not Necessary Condition for Entity-Behavior Equilibrium Points to Correspond to Behavior Strategy R-Solutions

In this section we will define the concept of recall-sensitivity, and we will employ this concept to show that: in order for a behavior strategy σB^p on R^p to be an R-solution in a game Γ, given that it corresponds to an entity-behavior equilibrium point in the associated non-cooperative game Γ', it is sufficient but not necessary that R^p be a perfect-recall information collection scheme.

6.4.1 The Concept of Recall Sensitivity

Denoting by $M((X^{pi}, i=1,N;\ Y^{aj}, j=1,M)|I^p(k)_\ell)$, the expected return to the protagonist as a function of the behavior strategies employed by the players, given that information set $I^p(k)_\ell$ has been realized, we define the concept of recall-sensitivity as follows.

Definition 22: The recall-sensitivity of the protagonist about a given pair of behavior strategies, $(X^{*pi}, i=1,N)$ on R^p and $(Y^{*aj}, j=1,M)$ on R^a, is rs, where:

(a) $rs = \sum_{k=2}^{N} \delta^{pk}$, and

(b) $\{\delta^{p2}, \delta^{p3}, \ldots, \delta^{pN}\}$ is a set of real numbers defined as follows:

$$\delta^{pk}(X^{*pi}, i=1,N;\ Y^{*aj}, j=1,M) = \max_{X^{pk}, I^{p}(k)} 2 \Big[\max_{X^{p1}} \ldots \max_{X^{p(k-1)}} H((X^{p1}, X^{p2}, \ldots, X^{pk}, X^{*p(k+1)}, \ldots, X^{*pN};\ Y^{*aj}, j=1,M) | I^{p}(k)) - \min_{X^{p1}} \ldots \min_{X^{p(k-1)}} H((X^{p1}, X^{p2}, \ldots, X^{pk}, X^{*p(k+1)}, \ldots, X^{*pN};\ Y^{*aj}, j=1,M) | I^{p}(k))\Big], \text{ for } k=2,N.$$

The recall-sensitivity of the protagonist has the following property.

Property 5:† The recall-sensitivity of the protagonist about any pair of behavior strategies $(X^{pi}, i=1,N)$ on R^{p} and $(Y^{aj}, j=1,M)$ on R^{p}, is zero if R^{p} is a perfect-recall information collection scheme.

The concept of recall-sensitivity can be interpreted as follows. First, note that the protagonist's expected return, given that it has realized a particular information set $I^{p}(k)_{\ell}$, is dependent upon the functions $X^{pi}, i=1,k-1$ to the extent that these functions establish a probability distribution on the alternative choices he knew in the past but cannot recall. To see this, we will denote

†The proof is given in Appendix B.

the information that has been forgotten by the protagonist when $I^p(k)_\ell$ has been realized, as $IF^p(k)_\ell$, and we will express $H((X^{pi}, i=1,N; Y^{aj}, j=1,M)|I^p(k)_\ell)$ as follows:

$$H((X^{pi}, i=1,N; Y^{aj}, j=1,M)|I^p(k)_\ell) = \sum_{IF^p(k)_\ell} H((X^{pi}, i=1,N; Y^{aj}, j=1,M)|I^p(k)_\ell, IF^p(k)_\ell) \cdot P(IF^p(k)_\ell|I^p(k)_\ell) \quad (6.3)$$

where summation over $IF^p(k)_\ell$ means summation over all possible values that can be taken on by this forgotten information.

Now, for some perfect-recall information collection scheme $R^{p'}$, $H((X^{pi}, i=1,N; Y^{aj}, j=1,M)|I^p(k)_\ell, IF^p(k)_\ell)$ is the conditional expected return when the information set represented by $\{I^p(k)_\ell, IF^p(k)_\ell\}$ is realized. Therefore, by definition 22 and property 4, this conditional expected return cannot depend upon the functions $X^{pi}, i=1,k-1$. Further, by causality, $P(IF^p(k)_\ell|I^p(k)_\ell)$ is not dependent upon any of the functions $X^{pi}, i=k,N$.

With these observations, we can see that $\delta^{pk}(X^{*pi}, i=1,N; Y^{*aj}, j=1,M)$ is simply a bound on the effect of forgotten information on the protagonist's conditional expected return at its kth decision when $(X^{*pi}, i=1,N)$ on R^p and $(Y^{*aj}, j=1,M)$ on R^a are employed. Thus, the recall-sensitivity of the protagonist about $(X^{*pi}, i=1,N)$ on R^p and $(Y^{*aj}, j=1,M)$ on R^a is a measure of the total effect of forgotten information on the protagonist's expected return.

Examining expression (6.3), we can also see that under either of the following conditions, the loss of information $IF^p(k)_\ell$ will contribute little to $\delta^{pk}(X^{*pi}, i=1,N;\ Y^{*aj}, j=1,M)$.

(1) $H((X^{pk}, X^{*pi}, i=k+1,N;\ Y^{*aj}, j=1,M)\,|\,I^p(k)_\ell, IF^p(k)_\ell)$ is not strongly dependent upon $IF^p(k)_\ell$ for any given function X^{pk} and any given information set $I^p(k)_\ell$, or

(2) $P(IF^p(k)_\ell\,|\,I^p(k)_\ell)$ is not strongly dependent upon the functions $X^{pi}, i=1,k-1$.

6.4.2 A Sufficient but Not Necessary Condition for Behavior Strategy R-Solutions

By employing the concept of recall-sensitivity, we can obtain the theorems and corollaries given below.

Theorem 4:† If a pair of behavior strategies, σB^{*p} on R^p and σB^{*a} on C^a, in a two-person, zero-sum finite extended game Γ is such that:

(a) it corresponds to an entity-behavior equilibrium point in the associated non-cooperative game Γ', and

(b) the recall-sensitivity of the protagonist about this pair of behavior strategies is less than ϵ, then:

$$Q(\sigma B^{*p}, R^p) \triangleq \min_{\sigma B^a} H(\sigma B^{*p}, R^p; \sigma B^a, C^a) = H(\sigma B^{*p}, R^p; \sigma B^{*a}, C^a) \leq$$

$$\max_{\sigma B^p} H(\sigma B^p, R^p; \sigma B^{*a}, C^a) < H(\sigma B^{*p}, R^p; \sigma B^{*a}, C^a) + \epsilon .$$

†The proof of Theorem 4 is given in Appendix B.

From this theorem we can obtain the following corollary which states a sufficient condition for a behavior strategy which corresponds to an entity-behavior equilibrium point in Γ', to be an ϵ-R-solution in Γ.

<u>Corollary</u> [illegible][†] [illegible] strategy σB^{*p} on R^p, in a given pair of behavior strategies which correspond to an entity-behavior equilibrium point in Γ', is an ϵ-R-solution in Γ if [illegible] sensitivity of the protagonist about this pair of behavior strategies is less than ϵ.

Note that the sufficient condition of corollary 1 is given in terms of the recall-sensitivity of the protagonist about a given pair of behavior strategies, a quantity which is not easily evaluated in problems of the sort that we wish to consider.

However, corollary 2 given below, which establishes a more restrictive but simpler sufficient condition, follows directly from corollary 1 and property 4.

<u>Corollary 2</u>: A behavior strategy σB^{*p} on R^p which corresponds to an entity-behavior equilibrium point in Γ', is an R-solution in Γ <u>if</u> R^p is a perfect-recall information collection scheme.

The corollary given above raises a question. Specifically, in this corollary, is the condition "R^p is a perfect-recall information collection scheme" also a <u>necessary</u> condition? We can see that the answer to this question is "<u>no</u>" by considering the following example.

<u>Example</u>: In figure 9 we describe a two-person, zero-sum, finite extended game Γ. In this game, note that the moves of ranks 1 and 3

[†] The proof of Corollary 1 is given in Appendix B.

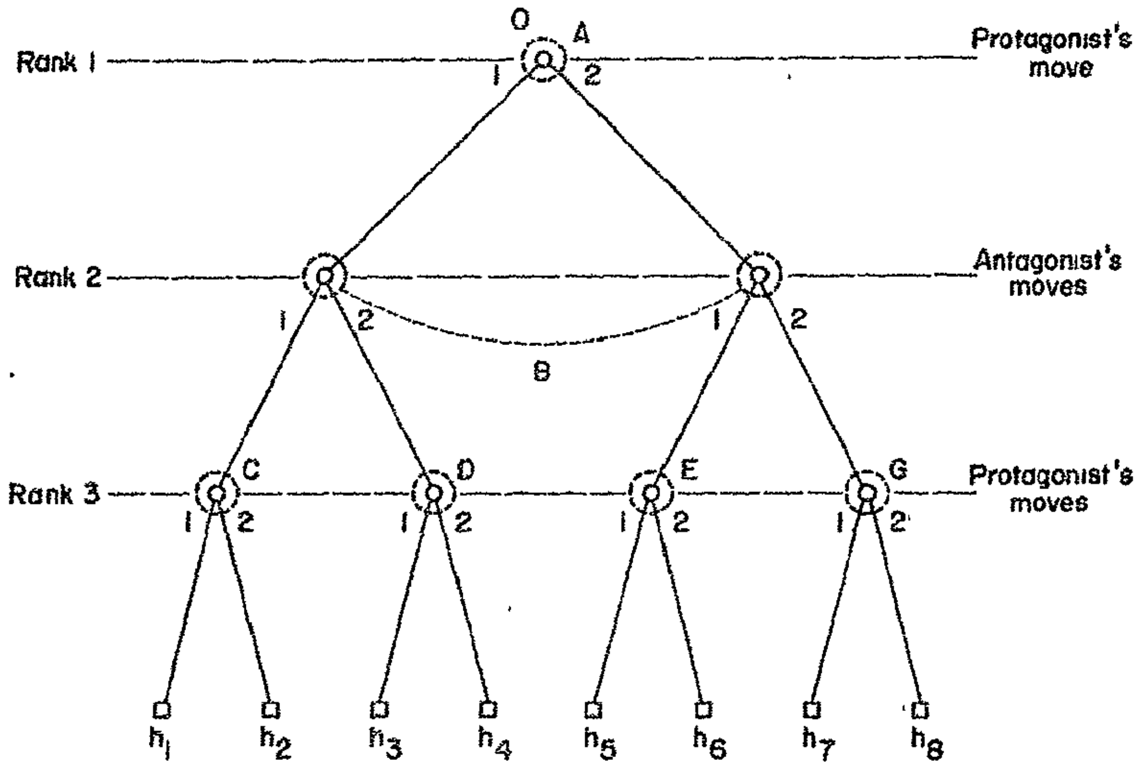


FIGURE 9 AN EXAMPLE GAME.

are protagonist moves while the moves of rank 2 are antagonist moves.

Consider then the imperfect-recall information collection scheme R^{p*} under which the protagonist forgets at any move of rank 3, the choice that he has made at his move of rank 1. The information partition induced by R^{p*} can be formed from the sets of the information partition induced by the protagonist's complete information collection scheme, and is given by $IP(R^{p*}) = \{A, C \cup E, D \cup G\}$ where A, C, D, E and G are the information sets labeled in figure 9.

Now, to investigate the recall-sensitivity of the protagonist, we must examine the protagonist's expected returns when information sets $C \cup E$ and $D \cup G$ are realized. The first of these returns can be expressed as:

$$H((X^{p1}, X^{p2}; Y^{a1}) | C \cup E) = \sum_{C_{W|C \cup E}} H(X^{p1}, X^{p2}; Y^{a1}) \div \sum_{C_{W\, C \cup E}} p(W) \quad (6.4)$$

where $C_{W|C \cup E}$ indicates summation over all plays W containing a move which is a member of information set $C \cup E$, and $C_{WC \cup E}$ indicates summation over the set of all partial plays leading from O to a move in $C \cup E$. Continuing, we can expand (6.4) to obtain:

$$H((X^{p1}, X^{p2}; Y^{a1}) | C \cup E) = p(1|A)\ p(1|B)\ [p(1|C \cup E)h_1 + p(2|C \cup E)h_2] + p(2|A)\ p(1|B)\ [p(1|C \cup E)h_5 + p(2|C \cup E)h_6] \div p(1|B)\ [p(1|A) + p(2|A)] = p(1|A)\ [p(1|C \cup E)h_1 + p(2|C \cup E)h_2] + p(2|A)\ [p(1|C \cup E)h_5 + p(2|C \cup E)h_6]\ . \quad (6.5)$$

Examining equation (6.5), we see that if $h_1 = \alpha h_5$ and $h_2 = \beta h_2$ where α and β are arbitrary constants, then $H((X^{p1},X^{p2};Y^{a1})|C \cup E)$ is not dependent upon X^{p1}.

The protagonist's expected return when information set $D \cup G$ is realized can be expressed as:

$$H((X^{p1},X^{p2};Y^{a1}|D \cup G) = p(1|A)\ p(2|B)\ [p(1|D \cup G)h_3 +$$

$$p(2|D \cup G)h_4] + p(2|A)\ p(2|B)\ [p(1|D \cup G)h_7 + p(2|D \cup G)h_8]$$

$$\div\ p(2|B)\ [p(1|A) + p(2|A)] = p(1|A)\ [p(1|D \cup G)h_3 + p(2|D \cup G)h_4]$$

$$+\ p(2|A)\ [p(1|D \cup G)h_7 + p(2|D \cup G)h_8] \tag{6.6}$$

Examining (6.6), we see that if $h_3 = \rho h_7$ and $h_4 = \mu h_8$ where ρ and μ are arbitrary constants, then $H((X^{p1},X^{p2};Y^{a1})|D \cup G)$ is also not dependent upon X^{p1}.

Therefore, by definition 22, if $h_1 = \alpha h_5$, $h_2 = \beta h_6$, $h_3 = \rho h_7$ and $h_4 = \mu h_8$ for arbitrary constants α, β, ρ, and μ, the protagonist has a recall-sensitivity of value zero about any pair of behavior strategies σB^p on R^{p*} and σB^a on C^a. By corollary 1, then, for the given game and the given information collection scheme R^{p*}, any behavior strategy σB^{p*} on R^{p*} which corresponds to an entity-behavior equilibrium point in the associated game Γ', is an R-solution.

It is also easy to show that for the given game and the given information collection scheme R^{p*}, if $h_1 = \alpha h_5 + \epsilon_1$, $h_2 = \beta h_6 + \epsilon_2$, $h_3 = \rho h_7 + \epsilon_3$ and $h_4 = \mu h_8 + \epsilon_4$, then the recall-sensitivity of the protagonist about any pair of behavior strategies is bounded by

$\epsilon^* = \max(\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4)$. In this case, by corollary 1, any behavior strategy σB^p on R^{p*} which corresponds to an entity-behavior equilibrium point in Γ' is an ϵ^*-R-solution.

Combining corollary 2 with the observations made in this example, we obtain the following theorem.

Theorem 5: In order for a behavior strategy σB^p on R^p to be an R-solution in a game Γ, given that σB^p on R^p corresponds to an entity-behavior equilibrium point in the associated non-cooperative game Γ', it is sufficient but not necessary that R^p be a perfect-recall information collection scheme.

6.5 Construction of Entity-Behavior Equilibrium Points

In the preceding two sections we have shown that if we can generate an entity-behavior equilibrium point $(X^{*pi}, i=1,N; Y^{*aj}, j=1,M)$ in the associated game Γ', then for a given game Γ and a given information collection scheme R^p, the behavior strategy $(X^{*pi}, i=1,N)$ on R^p:

(i) always meets a set of necessary conditions for a behavior strategy of maximum quality in the set of behavior strategies possible on R^p,

(ii) is an R-solution if R^p is a perfect-recall information collection scheme, and

(iii) may be an R-solution (or ϵ-R-solution) in cases where R^p is not a perfect-recall information collection scheme.

This raises an obvious question. "Given a particular associated non-cooperative game Γ', how can we find entity-behavior equilibrium points for Γ'?"

In the following subsection we will give an answer to this question by describing an algorithm that has been employed by the author to generate entity-behavior equilibrium points in a number of associated non-cooperative games. We will call this algorithm the "extended fictitious play algorithm" because of its similarity to the Brown-Robinson fictitious play algorithm described in section 5.4.

6.5.1 The Extended Fictitious Play Algorithm

Extended fictitious play in an associated non-cooperative game Γ' is analogous to fictitious play in a game Γ. Specifically, in extended fictitious play we carry out a series of fictitious plays (or repetitions) of Γ' in the following manner:†

(i) at the first repetition of Γ', each entity chooses an arbitrary degenerate behavior strategy, and

(ii) at subsequent repetitions of Γ', each entity chooses a degenerate behavior strategy which maximizes its return, assuming for each of the other entities the entity-behavior strategy that can be formed by summing (and appropriately normalizing) the degenerate behavior strategies previously employed by that other entity,

with the expectation that, as the number of repetitions becomes large, the set of entity-behavior strategies, which at any given repetition

†We described in section 6.2, the manner in which an associated game Γ' is considered to be played.

are assumed for each entity by the other entities, will approach an entity-behavior equilibrium point in Γ'.

In order to describe the algorithm formally, we establish the following notation. Denote:

(i) by $X^{pk}(j)$ [$Y^{ak}(j)$] the entity-behavior strategy that is produced at the jth iteration of the extended fictitious play algorithm, for the k^p-entity [k^a-entity],

(ii) by maxv X^{pi} [minv Y^{ai}], a vector partitioned in the same manner as X^{pi} [Y^{ai}] but having all 0.0 components except for a single 1.0 entry in each partition in a position corresponding to a maximum [minimum] element of X^{pi} [Y^{ai}] in that partition,† and

(iii) by $\prod_{\substack{\ell=1 \\ \ell\neq k}}^{N} X^{p\ell}$, the product: $X^{p1} \nabla X^{p2} \nabla \ldots \nabla X^{p(k-1)} \nabla X^{p(k+1)} \nabla \ldots \nabla X^{pN}$.††

Using this notation, we can formally describe the extended fictitious play algorithm by stating the following set of recursive equations which indicate the manner in which the entity-behavior strategies generated by the algorithm are related.

$$X^{pi}(j+1) = \left(\frac{1}{j+1}\right)\left[(j)X^{pi}(j) + \text{maxv}\left[\left(\prod_{\substack{\ell=1 \\ \ell\neq i}}^{N} X^{p\ell}\right) \nabla \left(\prod_{\ell=1}^{M} Y^{a\ell}\right) \nabla F\right]\right]$$

†In the case of several maximum [minimum] elements in a given partition, the choice of which one is to correspond to the 1.0 entry is arbitrary, for example, maxv [.5 .5 ⋮ .1 .2 .8] = [1.0 0.0 ⋮ 0.0 0.0 1.0] or [0.0 1.0 ⋮ 0.0 0.0 1.0].

††Note that this product yields a vector.

$$Y^{ak}(j+1) = \left(\frac{1}{j+1}\right)\left[(j)Y^{ak}(j) + \min\nabla\left[\left(\prod_{\ell=1}^{N} X^{p\ell}\right) \nabla \left(\prod_{\substack{\ell=1 \\ \ell\neq k}}^{M} Y^{a\ell}\right) \nabla F\right]\right]$$

for i=1,N and k=1,M, where $X^{pi}(0)$ for i=1,N and $Y^{ak}(0)$ for k=1,M are arbitrary degenerate entity-behavior strategies.

In the flow diagram of figure 10 we indicate an implementation of the extended fictitious play algorithm. The convergence of the algorithm is considered in the subsection below.

6.5.2 Convergence of the Extended Fictitious Play Algorithm

In any application of the extended fictitious play algorithm,

(i) we are interested in the quality of the behavior strategy σB^p on R^p which corresponds to any entity-behavior equilibrium point that may be generated, but

(ii) since only a finite number of iterations can be made, only an approximation to an entity-behavior equilibrium point will be generated.

Looking back to section 6.4.2, we see that theorem 4 establishes the quality of any behavior strategy σB^p on R^p which corresponds to an entity-behavior equilibrium point.

In light of these observations then, we will appropriately restate theorem 4 for the case of approximate entity-behavior equilibrium points, and we will then comment on the convergence of the algorithm in terms of this restatement.

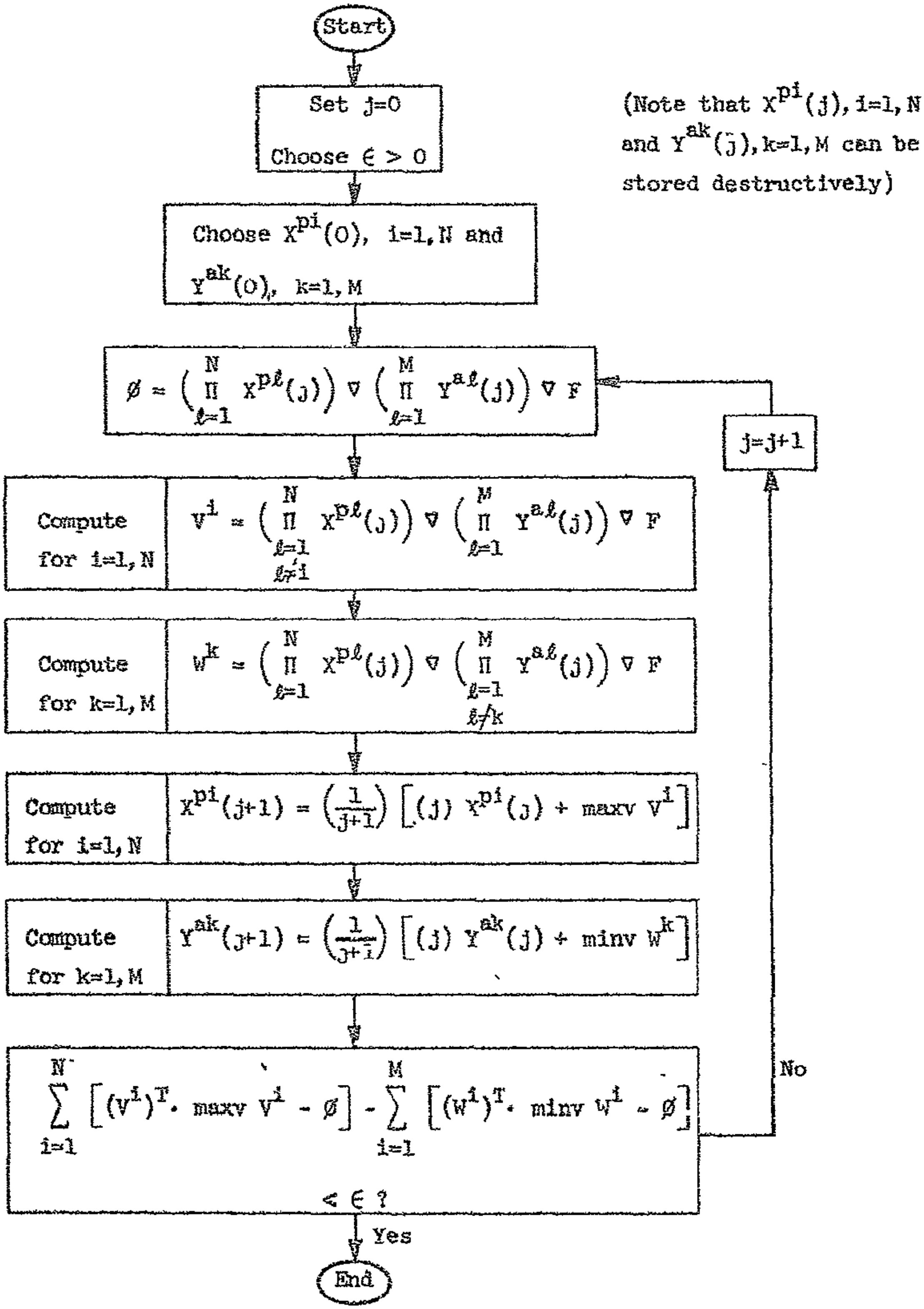


FIGURE 10. FLOW DIAGRAM FOR THE EXTENDED FICTITIOUS PLAY ALGORITHM.

Theorem 4[†] (restated): Consider a given two-person, zero-sum finite extended game Γ with a given information collection scheme R^p, and the associated non-cooperative game Γ'. Then, if a set of entity-behavior strategies $(X^{*pi}, i=1,N;\ Y^{*aj}, j=1,M)$ in Γ' is such that:

$$\max_{X^{pi}} \left(\prod_{\ell=1}^{i-1} X^{*p\ell}\right) \nabla X^{pi} \nabla \left(\prod_{\ell=i+1}^{N} X^{*p\ell}\right) \nabla \left(\prod_{\ell=1}^{M} Y^{*a\ell}\right) \nabla F - \epsilon^{*pi}$$

$$\leq \left(\prod_{\ell=1}^{N} X^{*p\ell}\right) \nabla \left(\prod_{\ell=1}^{M} Y^{*a\ell}\right) \nabla F \leq \min_{Y^{aj}} \left(\prod_{\ell=1}^{N} X^{*p\ell}\right) \nabla \left(\prod_{\ell=1}^{j-1} Y^{*a\ell}\right) \nabla$$

$$Y^{aj} \nabla \left(\prod_{\ell=j+1}^{M} Y^{*a\ell}\right) \nabla F + \epsilon^{*ai}$$

for $i=1,N$ and $j=1,M$, then the corresponding behavior strategies σB^{*p} on R^p and σB^{*a} on C^a in Γ are such that:

$$\left[H(\sigma B^{*p}, R^p; \sigma B^{*a}, C^a) - \sum_{i=1}^{M} \epsilon^{*ai}\right] \leq Q(\sigma B^{*p}, R^p) \leq$$

$$\left[H(\sigma B^{*p}, R^p; \sigma B^{*a}, C^a) + \sum_{i=1}^{N} \epsilon^{*pi}\right].$$

Considering convergence of the extended fictitious play algorithm then,[††] the author has employed this algorithm to obtain approximate

---

[†]Only minor modifications to the proof of theorem 4 are required to establish this restatement.

[††]Shapley (reference 13) has given a class of games for which the method of fictitious play will not converge, however, this class of games cannot be transformed into a class of games of the type that we are considering.

entity-behavior equilibrium points in a number of associated non-cooperative games and in each case convergence was observed in the following sense.

For $\delta = 0.05$, an integer $j \leq 200$ could always be found such that for $(X^{pi}(j), i=1,N; Y^{ai}(j), i=1,M)$:

$$\sum_{i=1}^{N} \epsilon^{pi}(j) \text{ and } \sum_{i=1}^{M} \epsilon^{ai}(j) \text{ are each } \leq \delta \left[\left(\prod_{i=1}^{N} X^{pi}(j) \right) \nabla \left(\prod_{i=1}^{M} Y^{ai}(j) \right) \nabla F \right].$$

In the following section we summarize our results and consider a practical question involved in their use.

6.6 Summary and Closing Arguments

In the preceding sections we have developed, for two-person, zero-sum finite-extended games, a heuristic algorithm for generating approximations to behavior strategies σB^p on R^p which:

(i) always meet a set of necessary conditions for a behavior strategy of maximum quality in the set of behavior strategies which are possible on R^p, and which

(ii) are R-solutions or ϵ-R-solutions if but not only if R^p is a perfect-recall information collection scheme.

We argue, then, that the algorithm can be said to produce promising candidates for behavior strategy ϵ-R-solutions on any information collection scheme.

However, if R^p is an imperfect-recall information collection scheme, and the algorithm is thus producing behavior strategies which are only known to be promising candidates, any candidate that is produced must be verified. But verification requires knowledge of the quality of an R-solution on R^p, and in games of the sort that we wish to consider, there are no practical techniques for determining the quality of an R-solution when R^p is an imperfect-recall information collection scheme.

All is not lost however, for recall that our primary interest is in ϵ-solutions on R^p, not in ϵ-R-solutions on R^p. In fact, we set out to develop an algorithm for generating ϵ-R-solutions only because we know that in order to be an ϵ-solution on R^p, a behavior strategy must first be an ϵ-R-solution on R^p.

In light of these comments then, we note that:

(i) the extended fictitious play algorithm generates promising candidates for ϵ-solutions on R^p, and, consequently,

(ii) what we are primarily interested in knowing is whether or not a behavior strategy produced by the algorithm is an ϵ-solution on R^p.

The verification implied by (ii) above, however, only requires knowledge of the quality of a solution in Γ, and this can be determined by appropriate employment of either the extended fictitious play algorithm or the perfect-recall fictitious play algorithm of the previous chapter.

CHAPTER 7

A PURSUIT AND EVASION GAME WITH MEASUREMENT UNCERTAINTY

7.0 Introduction

In this chapter we will consider a pursuit and evasion process in finite discrete time and in a finite discrete state space.

We will assume that the pursuer and the evader are each able to gather perfect measurements of their own states and uncertain measurements of their opponent's states, with the amount of uncertainty dependent upon the states of both pursuer and evader.

At each of the pursuer's decision times in the process, the pursuer either terminates the process or chooses a state to occupy at the next time instant. At each of the evader's decision times, the evader detects whether or not the pursuit has been previously terminated; and if it has not been, the evader chooses a state to occupy at the next time instant.

Finally, when the pursuer terminates the process, which he must do at or before some time t_f, the pursuer wins and the evader loses a payoff which is determined by the state histories that have been realized by the players at the time of termination.

7.0.1 Objectives

In considering this problem, our objectives will be

(i) to give a demonstration of the use of the perfect-recall fictitious play algorithm,

(ii) to give an example of a type of problem which may be much more easily solved in behavior strategies than in mixed strategies, and

(iii) to demonstrate (by showing that a most plausible simple method for generating "suboptimal" strategies leads to strategies of relatively poor quality) that the effort involved in determination of an ϵ-solution may be well justified.

In meeting these objectives, we will find behavior strategy ϵ-solutions for the pursuer and the evader when each is employing its complete information collection scheme. We will consider the problem of finding minimally complex behavior strategy ϵ-solutions in the example of the next chapter.

7.1 General Description of the Pursuit-Evasion Process

We give below a formal description of the pursuit and evasion process that we are going to consider. In this description and in the remainder of the chapter, we will consider the pursuer as the protagonist in the process and the evader as the antagonist.

To begin, we establish the following notation.

(i) Denote by S the state space of n points in which the process is considered to take place, and assign to the points of S an arbitrary ordering.

(ii) Denote by t_i, $i=0,1,2,\ldots$ the discrete time instants at which measurements are made, and decisions are made and executed.

(iii) Denote by $a(t_i)$, $b(t_i)$, $z(t_i)$, and $w(t_i)$, respectively, the pursuer's state at time t_i, the evader's state at t_i, the pursuer's measurement at t_i of the evader's state at t_i, and the evader's measurement at t_i of the pursuer's state at t_i.

(iv) Denote by $T(t_i)$, the set of termination alternatives open to the pursuer at t_i. This set will consist of the alternatives: terminate the pursuit, and/or continue the pursuit, which will be denoted, respectively, by the integers 1 and 0.

(v) Denote by $\tau(t_i) \in T(t_i)$, the pursuer's termination decision at time t_i.

The pursuit and evasion process that we will consider, then, is formally described by the following assumptions.

<u>Assumption A</u>: The <u>movement capability</u> of the pursuer in S is defined by a set function $S^p(a(t_o),a(t_1),\ldots,a(t_i))$ which maps the history at t_i, of past states occupied by the pursuer, into a subset of S consisting of the states that can be occupied by the pursuer at t_{i+1}. The <u>movement capability</u> of the evader in S is similarly defined by the set function $S^a(b(t_o),b(t_1),\ldots,b(t_i))$; and the pursuer and evader each know both functions.

Assumption B: The membership of the set T is dependent only upon time, and is known to both pursuer and evader. Further, the pursuer must terminate the pursuit at or before some time t_f; i.e., $T(t_f) = (1)$.

Assumption C: The pursuer and evader each have perfect measurements of their own states. Their measurements of their opponent's states $z(t_i)$ and $w(t_i)$ are random variables taking values in S under probability distributions, $P^{z}(z(t_i)|a(t_i), b(t_i))$ and $P^{w}(w(t_i)|a(t_i), b(t_i))$, which are dependent upon the pursuer and evader states at the time of measurement, and which are known to both pursuer and evader.

Assumption D: The initial states, $a(t_o)$ and $b(t_o)$, of the pursuer and evader are random variables taking values in S under probability distributions $P^{a}(a(t_o))$ and $P^{b}(b(t_o))$ which are known to both pursuer and evader.

Assumption E: The temporal order in which the pursuer and evader will be assumed to operate at each time instant t_i is as follows:

(i) the pursuer and evader each receive an uncertain measurement of its opponent's current state;

(ii) on the basis of the information it possesses, the pursuer decides whether or not to terminate the pursuit;

(iii) the evader receives the information that the pursuer has or has not terminated the pursuit;

(iv) on the basis of the information it possesses, the pursuer selects and occupies an accessible next state in S;

(v) on the basis of the information it possesses, but without knowledge of the pursuer's movement decision, the evader selects and occupies an accessible next state in S.

Assumption F: Upon termination of the process, a payoff $h(a(t_i), b(t_i), i=0, j)$ is awarded positively to the pursuer and negatively to the evader, where t_j is the time of termination.

In figure 11 we summarize the above assumptions in a flow diagram showing how this pursuit and evasion game is to be played when both pursuer and evader employ their complete information collection schemes.

7.2 Interpretation of the Pursuit and Evasion Game

In this section we will interpret the pursuit and evasion game in terms of the method of description that was developed for two-person, zero-sum, finite extended games in Chapter 2.

To begin, note that the four rules of our definition of a finite extended game are specified by the assumptions given in the previous section. In particular,

(i) the set of all possible plays is specified by Assumptions A and B,

(ii) the manner in which nature chooses an alternative at each of her moves is specified by Assumptions C and D,

(iii) the payoffs to each player for each possible play are specified by Assumption F, and

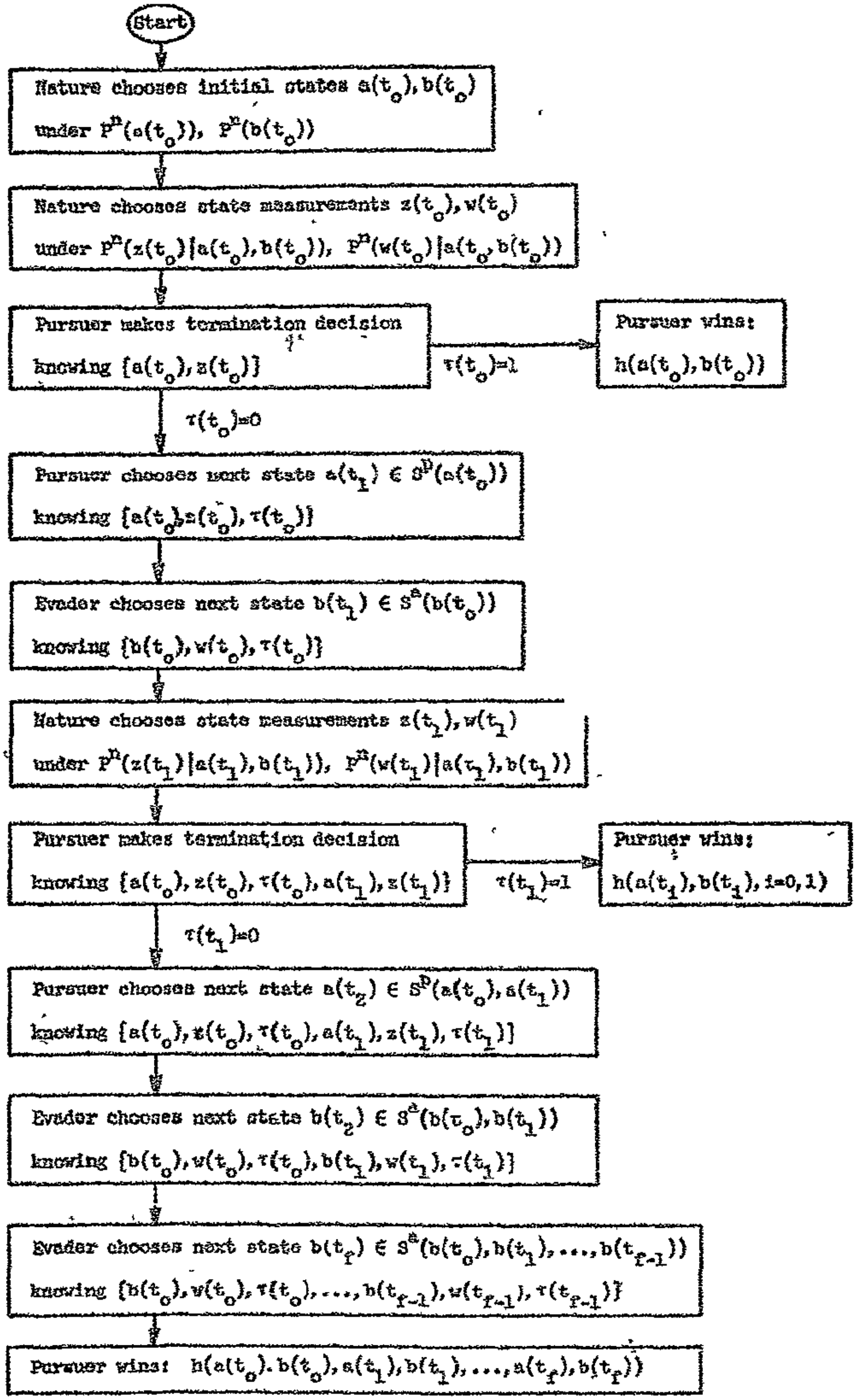


FIGURE 11. METHOD OF PLAY OF PURSUIT AND EVASION GAME WHEN BOTH PLAYERS EMPLOY THEIR COMPLETE INFORMATION COLLECTION SCHEMES.

(iv) the maximum amount of information that each player can gather is specified by assumptions C and E.

After establishing the following notation and conventions, we will describe a specific pursuit and evasion game in terms of a game-tree diagram.

7.2.1 Conventions and Notation

Let us assume the following indexing of alternatives.

First, at any move where the pursuer is making a termination decision, let the alternative "terminate the pursuit" be alternative 1 and the alternative "continue the pursuit" be alternative 2.

Next, at any move where the pursuer or the evader is choosing a next state, assume in accordance with the ordering of S, an ordering on the states of the relevant accessible state set, and let the ith state of the accessible state set under this ordering be the ith alternative at the given move.

For example, suppose that in a given pursuit and evasion game, the pursuer is making a "next-state" decision at a move where its accessible state set is $\{2,4\} \in S = \{1,2,3,4\}$. In this case we would let state 2 be alternative 1 and state 4 be alternative 2.

Finally, at any move where nature is choosing an alternative, we will assign, in a manner similar to that used to assign indexings at moves where the players are choosing next states, an indexing to the measurements with positive probability.

For example, suppose that in a given pursuit and evasion game, nature is selecting a measurement $z(t_i)$ at a move where $P^n(z(t_i)|a(t_i),b(t_i))$ gives positive probability only to states 2 and 4 in $S = \{1,2,3,4\}$. In this case we would let state 2 be alternative 1 and state 4 be alternative 2.

We will adopt the following convention in labeling information sets.

Recall that in Chapter 2 we denoted a player's information set at any move, as the string of alternative choices representing the portion which is known to that player of the alternative choice history defining that move. Noting then that, under the given indexing of alternatives, the string of past measurements and decisions known to a player at any given move corresponds to the player's information set at that move, we will adopt the convenient convention of describing information sets in terms of their corresponding measurement and decision strings.

Further, let us denote

(i) by $I^p(t_i^-)$ and $I^e(t_i^-)$, respectively, the pursuer's and evader's information sets at time t_i but prior to the termination decision of the pursuer; and

(ii) by $I^p(t_i)$ and $I^e(t_i)$, respectively, the pursuer's and the evader's information sets at time t_i but after the termination decision of the pursuer.

For example,[†]

$$I^p(t_1^-) = (a(t_o), z(t_o), \tau(t_o), a(t_1), z(t_1))$$

$$I^e(t_1^-) = (b(t_o), w(t_o), \tau(t_o), b(t_1), w(t_1))$$

and

$$I^p(t_1) = (a(t_o), z(t_o), \tau(t_o), a(t_1), z(t_1), \tau(t_1))$$

$$I^e(t_1) = (b(t_o), w(t_o), \tau(t_o), b(t_1), w(t_1), \tau(t_1)) \ .$$

Having established these conventions and notation, let us describe a simple pursuit and evasion game in terms of a game-tree diagram.

7.2.2 A Game-Tree Diagram for a Pursuit and Evasion Game

Consider a pursuit and evasion game described as follows:

(i) two stages: $t \in \{t_o, t_1\}$; $T(t_o) = \{0,1\}$; $T(t_1) = \{1\}$

(ii) state space of two states: $S = \{1,2\}$

(iii) pursuer movement capability: $S^p(1) = S^p(2) = \{1,2\}$

(iv) evader movement capability: $S^e(1) = \{1,2\}$; $S^e(2) = \{2\}$

(v) initial state distributions: $P^n(a(t_o)) = P^n(b(t_o)) = (.5\ .5)$

(vi) measurement statistics defined in table 4 below.

(vii) payoffs described as follows:

(a) if the pursuit terminates with the pursuer and the evader in the same state, the pursuer wins and the evader loses a payoff of 2.0 if the time of termination

[†]Recall that both the pursuer and the evader are employing their complete information collection schemes.

Table 4: Measurement Statistics

z(t) or w(t), a(t), b(t)	$P^n(z(t)\|a(t),b(t))$	$P^n(w(t)\|a(t),b(t))$
1,1,1	0.90	0.55
1,1,2	0.10	0.90
1,2,1	0.55	0.45
1,2,2	0.45	0.10
2,1,1	0.10	0.45
2,1,2	0.90	0.10
2,2,1	0.45	0.55
2,2,2	0.55	0.90

is t_o, and a payoff of 1.0 if the time of termination is t_1; and

(b) if the pursuit terminates with the pursuer and the evader in different states, the pursuer wins and the evader loses a payoff of 0.0 regardless of the time of termination.

In figure 12 we show a game-tree diagram for the pursuit and evasion game described directly above, and in figure 13 we give a detailed description of a single play.

In these figures, in order to clearly indicate the memberships of the players' information sets, we have indicated at each move the corresponding observed measurement and decision histories of the players. Specifically, at each move, LEGEND :

Numbers at each vertex are pursuer's observations/evader's observations, with an observation string alternating - own state, measurement of opponent's state, termination decision, own state, ...

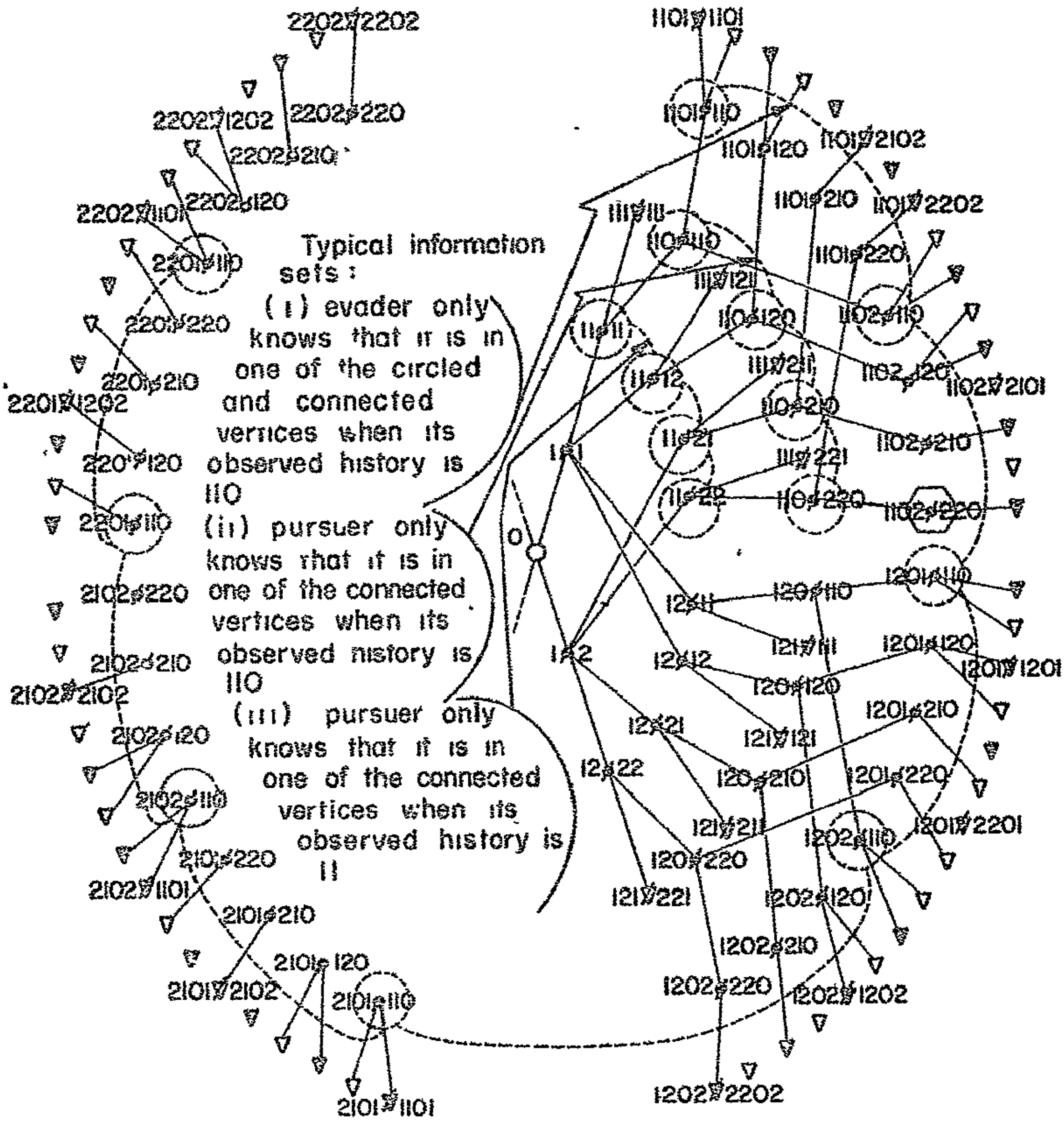


FIGURE 12. GAME TREE FOR 2-STAGE AND 2-STATE PURSUIT AND EVASION GAME.

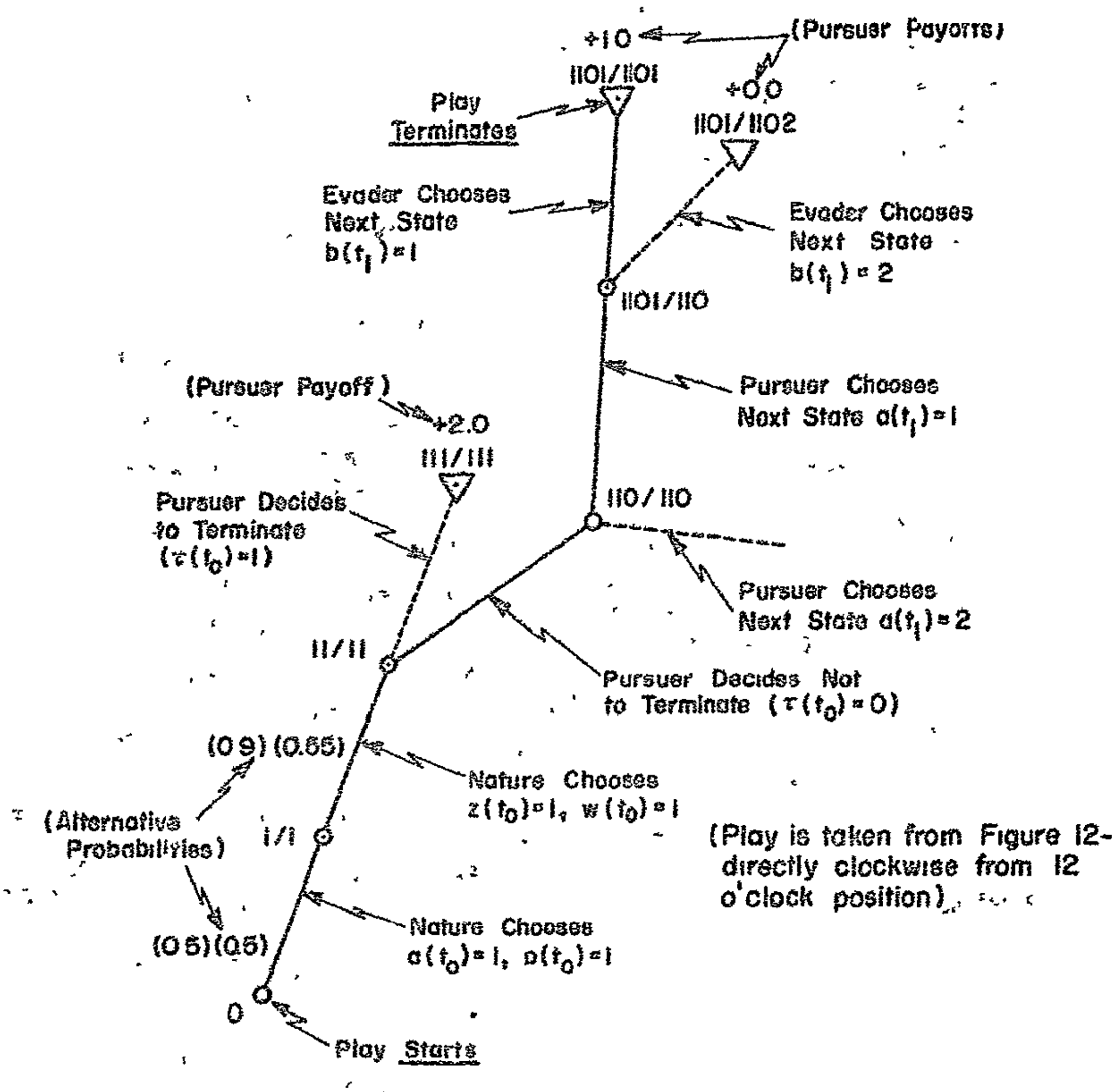


FIGURE 13. DETAIL OF A PLAY IN EXAMPLE PURSUIT AND EVASION GAME.

(i) the pursuer's measurement and decision history (information set) is indicated by the string of integers to the left of the slash, and the evader's measurement and decision history is indicated by the string of integers to the right of the slash, with

(ii) the numbers in each string indicating, from left to right, the player's measurements and decisions in the order in which they were made.

For example, the move which is circumscribed by a hexagon in figure 12 is labeled with the strings 1102/220 indicating that at this move:

(i) the pursuer has observed $\{a(t_o), z(t_o), \tau(t_o), a(t_1)\} = \{1, 1, 0, 2\}$ and

(ii) the evader has observed $\{b(t), w(t_o), \tau(t_o)\} = \{2, 2, 0\}$.

Note that together, the two strings at any move completely specify the alternative choice history defining that move.

Note also that in the figures:

(i) a single alternative is indicated at each vertex or move where the evader's admissible state set is $S^a(2) = \{2\}$,

(ii) nature's selection of initial states for the players has been combined into a single move, and

(iii) nature's selection of measurements for the players has been combined into a single move for each given pair of initial states.

Finally, note that the pursuit and evasion game that we have described meets Assumptions 1 and 3 of Chapter 5, but not Assumption 2. This failure is not significant, however, since Assumption 2 is not restrictive and was made in Chapter 5 only to allow the writing of simple general expressions for the protagonist's expected payoff.

7.3 Pursuit and Evasion Games Studied

The following example pursuit and evasion games were studied.

Example 1:

(i) three stages: $t \in \{t_o, t_1, t_2\}$; $T(t_o) = T(t_1) = \{0,1\}$,
$T(t_2) = \{1\}$

(ii) two states: $S = \{1,2\}$

(iii) pursuer movement capability: $S^p(a(t_o)) = S^p(a(t_o), a(t_1)) = S$ for any $a(t_o), a(t_1)$

(iv) evader movement capability: $S^e(b(t_o)) = S$ for any $b(t_o)$;
$S^e(b(t_o), b(t_1)) = S$ if $b(t_o) = b(t_1)$
$= b(t_1)$ otherwise

(v) initial state distributions: $P^a(a(t_o)) = (1.0 \;\; 0.0)$;
$P^a(b(t_o)) = (0.0 \;\; 1.0)$

(vi) measurement statistics as defined in table 4

(vii) payoff described as follows:

(a) if the pursuit terminates with the pursuer and the evader in the same state, the pursuer wins and the evader loses a payoff of 2.0 if the time is t_o, 1.5 if the time is t_1, and 1.0 if the time is t_2; and

(b) if the pursuit terminates with the pursuer and the evader in different states, the pursuer wins and the evader loses a payoff of 0.0 regardless of the time of termination.

Example 2:[†] The number of stages, movement capabilities of the players, and the payoffs in example 2 are the same as in example 1. We list the differences below.

(i) three states: $S = \{1,2,3\}$

(ii) initial state distributions: $P^n(a(t_o)) = (1.0 \quad 0.0 \quad 0.0)$;
$P^n(b(t_o)) = (0.0 \quad 0.0 \quad 1.0)$

(iii) measurement statistics defined as in table 5.

The use of the perfect-recall fictitious play algorithm in the generation of behavior strategy ϵ-solutions in these examples is discussed below.

7.4 Application of the Perfect-Recall Fictitious Play Algorithm

To apply the perfect-recall fictitious play algorithm, we must

(i) express the pursuer's (protagonist's) expected return as a function of the behavior strategies employed by the players; and employing this expression,

(ii) implement on a digital computer, the flow diagram shown in figure 7.

[†]In Appendix A, Note 10, we comment on how the numbers were chosen in this example.

Table 5: Measurement Statistics

z(t) or w(t),a(t),b(t)	p(z(t)|a(t),b(t))	p(w(t)|a(t),b(t))
1,1,1	0.60	0.60
1,1,2	0.20	0.90
1,1,3	0.20	0.50
1,2,1	0.90	0.20
1,2,2	0.05	0.05
1,2,3	0.05	0.25
1,3,1	0.50	0.20
1,3,2	0.25	0.05
1,3,3	0.25	0.25
2,1,1	0.20	0.20
2,1,2	0.60	0.05
2,1,3	0.20	0.25
2,2,1	0.05	0.60
2,2,2	0.90	0.90
2,2,3	0.05	0.50
2,3,1	0.25	0.20
2,3,2	0.50	0.05
2,3,3	0.25	0.25
3,1,1	0.20	0.20
3,1,2	0.20	0.05
3,1,3	0.60	0.25
3,2,1	0.05	0.20
3,2,2	0.05	0.05
3,2,3	0.90	0.25
3,3,1	0.25	0.60
3,3,2	0.25	0.90
3,3,3	0.50	0.50

7.4.1 An Expression for the Pursuer's Expected Return

With the common payoff expressions given above, we can write a single expression for the pursuer's expected payoff which is valid for both of the examples. In particular, denoting by C^p and C^a, respectively, the pursuer's and the evader's complete information collection schemes, we can take advantage of the simple structure of our examples to write:

$$H(\sigma B^p, C^p; \sigma B^a, C^a) \doteq 2.0 \times P(a(t_o) = b(t_o),\ \tau(t_o) = 1) +$$

$$1.5 \times P(a(t_1) = b(t_1),\ \tau(t_o) = 0,\ \tau(t_1) = 1) + 1.0 \times P(a(t_2) =$$

$$b(t_2),\ \tau(t_o) = \tau(t_1) = 0) \qquad (7.1)$$

where the probabilities of the three events with non-zero payoffs are defined by the players' and nature's behavior strategies.

If we denote by $u(\cdot,\cdot)$ an indicator function having value 1.0 where its arguments are identical and 0.0 otherwise, we can expand expression (7.1) to write:

$$H(\sigma B^p, C^p; \sigma B^a, C^a) = \sum_{a(t_o), z(t_o), b(t_o), w(t_o)} A \cdot \Big[2.0\ u(a(t_o), b(t_o)) \cdot$$

$$P^p(\tau(t_o)=1 | a(t_o), z(t_o)) + P^p(\tau(t_o)=0 | a(t_o), z(t_o)) \cdot$$

$$\sum_{a(t_1), z(t_1), b(t_1), w(t_1)} B \cdot \Big[1.5\ u(a(t_1), b(t_1)) \cdot P^p(\tau(t_1)=1 |$$

$$a(t_o), z(t_o), \tau(t_o), a(t_1), z(t_1)) + P^p(\tau(t_1)=0 | a(t_o), z(t_o), \tau(t_o),$$

$$a(t_1), z(t_1)) \cdot \sum_{a(t_2), b(t_2)} C \cdot \left[1.0\ u(a(t_2), b(t_2)) \right] \Big] \Big] \qquad (7.2)$$

where:

(i) $A = P^n(z(t_o)|a(t_o), b(t_o)) \cdot P^n(w(t_o)|a(t_o), b(t_o)) \cdot P^n(a(t_o)) \cdot P^n(b(t_o))$,

(ii) $B = P^n(z(t_1)|a(t_1), b(t_1)) \cdot P^n(w(t_1)|a(t_1), b(t_1)) \cdot P^p(a(t_1)|a(t_o), z(t_o), \tau(t_o)=0) \cdot P^a(b(t_1)|b(t_o), w(t_o), \xi(t_o)=0)$,

(iii) $C = P^p(a(t_2)|a(t_o), z(t_o), \tau(t_o)=0,\ a(t_1), z(t_1), \tau(t_1)=0) \cdot P^a(b(t_2)|b(t_o), w(t_o), \tau(t_o)=0,\ b(t_1), w(t_1), \tau(t_1)=0)$;

(iv) the indicated sums are over the states of S,

(v) $P^n(z(t_i)|a(t_i), b(t_i))$, $P^n(w(t_i)|a(t_i), b(t_i))$, $P^n(a(t_o))$ and $P^n(b(t_o))$ are nature's behavior strategies,

(vi) the probabilities superscripted by "p" and "a", respectively, are elements of the pursuer's and evader's behavior strategies with those elements restricted to taking values consistent with the movement and termination capabilities of the players (for example, in a state space of 3 states, if $S^p(1,2) = (2)$, then $P^p(3|1,\ z(t_o),\ 0,\ 2,\ z(t_1),\ 0) = P^p(1|1,\ z(t_o),\ 0,\ 2,\ z(t_1),\ 0) = 0$ for any pair of measurements $z(t_o)$ and $z(t_1)$); and

(vii) the sets of variables upon which the probabilities mentioned in (vi) are conditioned, correspond to the players' information sets in $IP(C^p)$ and $IP(C^e)$.

Note that, in spite of its complexity, equation 7.2 is much simpler than the expression that would have resulted if we had ignored the structure of the game and had attempted to express the expected return to the protagonist in the form of equation 5.18.

7.4.2 Generation of Behavior Strategy ϵ-Solutions for the Pursuer and the Evader

Using expression (7.2) for the pursuer's expected return, the perfect-recall fictitious play algorithm was programmed on a digital computer and behavior strategy ϵ-solutions were determined for both the pursuer and the evader.

Note that since we are considering the expected payoff when both players are employing their complete information collection schemes, a behavior strategy ϵ-solution is generated for both players at the same time.

In both examples, the value of ϵ that was selected was 0.02. The qualities of the behavior strategies found for the players are given in table 6 below.

Table 6: Qualities of Behavior Strategy ϵ-Solutions Found for Pursuer and Evader

	Example 1	Example 2
Quality of pursuer's ϵ-solution:	.845	.661
Quality of evader's ϵ-solution:	.864	.681

In table 7 we show the behavior strategy ϵ-solutions determined for the pursuer and the evader in example 1.[†] The "nr" indication in this table means that under the given strategies for the players and nature, the information set corresponding to the given measurement and decision history is not realizable.

7.4.3 Observed Convergence Rates

The computation of behavior strategy ϵ-solutions for the players by means of the perfect-recall fictitious play algorithm was carried out on an IBM 7094 digital computer.

For each example pursuit and evasion game, a general program was written to accomodate arbitrary initial state distributions, arbitrary measurement statistics, and arbitrary movement capabilities for the players. However, the programs were altered slightly to take advantage of the simple initial state distributions specified in examples 1 and 2 above. Specifically, alternative probability distributions were not determined on information sets that could not be realized under the given initial state distributions.

In table 8 we summarize the relevant information with respect to observed convergence rates of the algorithm for the examples considered. In this table the "number of variables to be determined" is the total number of probability distribution components in the pair of behavior strategies to be determined for the players. This number does not take into account the movement limitations of the players

[†]The ϵ-solutions for example 2 are too complex to include here.

Table 7: Behavior Strategy ∈-Solutions for Example 1

	Pursuer Observed History	Pursuer Termination Decision	Pursuer Movement Decision	Evader Observed History	Evader Movement Decision
	$I^p(t_i^-)$	$P^p(\tau(t_i)=0 \mid I^p(t_i^-))$	$P^p(a(t_{i-1})=1 \mid I^p(t_i^-), \tau(t_i)=0)$	$I^a(t_i^-)$	$P^a(b(t_{i+1})=1 \mid I^a(t_i^-), \tau(t_i)=0)$
	11	1.000	0.600	11	nr
t_o	12	1.000	0.605	12	nr
	21	nr	nr	21	0.430
	22	nr	nr	22	0.430
	11011	0.000	nr	11011	nr
	11012	1.000	0.483	11012	nr
	11021	0.000	nr	11021	nr
	11022	0.000	nr	11022	nr
	12011	0.000	nr	12011	nr
	12012	1.000	0.479	12012	nr
	12021	0.000	nr	12021	nr
t_1	12022	0.000	nr	12022	nr
	21011	nr	nr	21011	1.000
	21012	nr	nr	21012	1.000
	21021	nr	nr	21021	0.447
	21022	nr	nr	21022	0.447
	22011	nr	nr	22011	1.000
	22012	nr	nr	22012	1.000
	22021	nr	nr	22021	0.447
	22022	nr	nr	22022	0.447

or the fact that the components in any distribution must sum to 1.0 since the programs (as written) do not obtain any advantage from these facts.

Table 8: Observed Convergence Rates for Perfect-Recall Fictitious Play Algorithm

	Example 1	Example 2
Number of alternative probability distributions to be determined	28	90
Number of variables to be determined	56	240
Number of iterations	200	295
Total time (in seconds)	46	436

7.5 A Plausible Method for Generating "Good" Behavior Strategies for the Players

Knowing the qualities of ϵ-solutions for the pursuer and the evader, we are in a position to evaluate plausible heuristic techniques that might be proposed for generating "good" behavior strategies for the players. The following particularly plausible technique was examined for each of our examples.

First, behavior strategy ϵ-solutions were computed for each player in the pursuit and evasion game examples which are obtained from examples 1 and 2 by assuming that each player has perfect measurements of its opponent's states. Obviously, determination of these strategies is much easier than the determination of behavior strategy ϵ-solutions in the original games.

Then, the "decision operations" for the pursuer and the evader were separated into two parts:

(i) a filtering operation in which the pursuer or the evader computes from its measurements, the state history most likely to have been realized by its opponent (assuming that the opponent has perfect measurements and is employing its perfect measurement strategy computed above); and

(ii) a decision operation in which the pursuer or the evader employs its computed perfect measurement strategy, treating the opponent's most likely state history as if it were the opponent's actual state history.

Finally, the behavior strategies defined by the above separation were determined, and the qualities of these "separation strategies" were calculated.

In table 9 below, we give the qualities of these separation strategies and, for purposes of comparison, the qualities of the behavior strategy ϵ-solutions previously determined.

Table 9: Qualities of "Separation" Behavior Strategies and Behavior Strategy ϵ-Solutions

	Example 1	Example 2
Quality of evader separation strategy	1.12	.90
Quality of evader ϵ-solution	0.864	0.681
Quality of pursuer ϵ-solution	0.854	0.661
Quality of pursuer separation strategy	0.16	.56

From these results we can see that the losses risked by employment of separation strategies of this type are quite substantial for both pursuer and the evader.[†] In these examples, then, the effort involved in determination of behavior strategy ϵ-solutions for the players seems well justified.

[†]See Appendix A, Note 11, for comments on the sizes of these quality differences.

CHAPTER 8

A MEDICAL DECISION PROCESS

8.0 Introduction

In this chapter we will consider a finite medical decision process informally described as follows.

A doctor encounters a patient whose general condition is rapidly deteriorating with the spread of a disease but whose relative condition varies within a range from "relatively weak" to "relatively strong."

A surgical operation exists which can arrest the advance of the disease, but its effectiveness

(i) is a function of the patient's condition and the treatment being administered by the doctor at the time of the operation, and

(ii) is diminished as time goes on (and the disease spreads), reaching zero at some time t_f (when the spread of the disease can no longer be halted).

The doctor's control over the situation is through his choice of a time for the operation, and his choices of preoperative treatments for the patient. In particular:

(i) if the doctor administers treatment that is appropriate for the relative condition of the patient, the patient's relative condition changes in a random but statistically known fashion; and

(ii) if the doctor administers treatment that is inappropriate for the relative condition of the patient, the patient's relative condition changes in a random but unknown fashion.

The doctor's information on the patient's relative conditions as the process unfolds is uncertain in that the doctor does not know the actual relative conditions of the patient; he only has diagnoses which are statistically related to the patient's actual relative conditions.

Finally, the process is terminated when the doctor decides to operate, and at that time he is considered to receive a payoff which is:

(i) given in terms of an expected extension of the patient's lifetime, and

(ii) described by a known function of the total time elapsed in preoperative treatment and the patient's relative condition and the treatment he is receiving at the time of the operation.

8.0.1 Objectives

We will consider below, a formal version of the medical decision process described above. In considering this process, our specific objective will be to determine for the doctor a simplest possible treatment and operation behavior strategy which comes acceptably close to maximising the doctor's minimum expected return against all possible unknown probability distributions under which the patient's relative condition may be changing when the doctor administers inappropriate treatment.

In order to find such strategies, we will pose the problem as a two-person zero-sum game by considering that, in situations in which the doctor administers inappropriate treatment, the patient is under the control of a "devil" whose interests are diametrically opposed to those of the doctor.

Our specific objective, then, will be to determine for the doctor a behavior strategy ϵ-solution in this doctor (protagonist)-devil (antagonist) game.

Our general objectives in considering this problem will be

(i) to give a demonstration of the use of the extended fictitious play algorithm,

(ii) to give an example of a search (a very productive one, as it happens) for a simplest possible behavior strategy ϵ-solution, and

(iii) to give an example of a problem which can be solved practically in terms of behavior strategies but which is far too complex to solve in terms of mixed strategies.

8.1 Formal Description of the Doctor-Devil Game

In this section we will describe the formal version of the doctor-devil game that we are going to consider. In this formal version we will make the payoff function reflect the changes in the patient's general condition and we will concern ourselves only with the patient's relative conditions. For convenience then, we will henceforth drop the word relative.

The game that we will consider, then, is described by the following set of assumptions.

Assumption A: The set T of treatments τ that can be administered by the doctor, and the set C of conditions c that can be realized by the patient are both finite sets with their members in one-to-one correspondence, c_j paired with τ_j.†

Assumption B: The game terminates with the doctor's decision to operate, which must be made at or before some time t_f.

Assumption C: The treatment and operative decisions made by the doctor, and the condition changes made by the patient occur at discrete times $t_i \in \{t_1, t_2, \ldots, t_{f-1}\}$.

Assumption D: The patient is partially controlled by the doctor and partially controlled by the devil as follows:

(i) if $\tau(t_i) = \tau_j$, $c(t_i) = c_k$ and $j = k$, then nature selects the patient's condition at t_{i+1} under a known probability distribution $P^n(c(t_{i+1}) | \tau_k, c_k)$, but

---

† τ_j is considered "appropriate treatment" for c_j.

(ii) if $\tau(t_i) = \tau_j$, $c(t_i) = c_k$ and $j \neq k$, then the devil can select any condition $c \in C$ as the patient's condition at t_{i+1}.

<u>Assumption E</u>: If the game terminates at time t_k, the doctor wins and the devil loses a payoff $h(c(t_k), \tau(t_k))$ which is given in terms of an expected extension of the patient's lifetime.

<u>Assumption F</u>: At the end of each subinterval (t_o, t_1), (t_1, t_2), $\ldots, (t_{f-2}, t_{f-1})$, the doctor receives a diagnosis, $d(t_j) \in C$, of the patient's condition during that subinterval. This diagnosis is selected by nature under a known probability distribution $P^n(d(t_j) | \tau(t_j), c(t_j))$.

<u>Assumption G</u>: The information that the doctor can gather as the game progresses consists of the diagnoses he makes (receives), and the treatment and operation decisions that he makes.

<u>Assumption H</u>: The information that the devil can gather as the game progresses consists of the conditions assumed by the patient, and the treatment and operation decisions made by the doctor.

<u>Assumption I</u>: The temporal order in which the doctor and the patient (devil or nature) will be assumed to operate at each time instant t_i is as follows:

(i) the doctor "receives" an uncertain diagnosis of the patient's condition,

(ii) on the basis of the string of past diagnoses and treatments that he recalls, the doctor decides whether or not to operate,

(iii) the devil receives the information that the doctor has decided (decided not) to operate,

(iv) if the doctor has decided not to operate and has not applied appropriate treatment over the past time interval, then the devil, on the basis of the observation history it possesses, chooses a next condition for the patient,

(v) if the doctor has decided not to operate and has applied appropriate treatment over the past time interval, then nature chooses a next condition for the patient, and finally,

(vi) on the basis of the string of past diagnoses and treatments it recalls, but without knowledge of the patient's next condition, the doctor chooses a treatment to administer to the patient over the next time interval.

The manner in which the doctor-devil game described above is played is summarized in the flow diagram of figure 14. In this diagram we have assumed for the antagonist, a complete information collection scheme.

8.2 Interpretation of the Doctor-Devil Game

Note that the four rules of our definition of a finite extended game are specified by the assumptions given above. In particular,

(i) the set of all possible plays is specified by Assumptions A, B, D and I;

(ii) the manner in which nature chooses an alternative at each

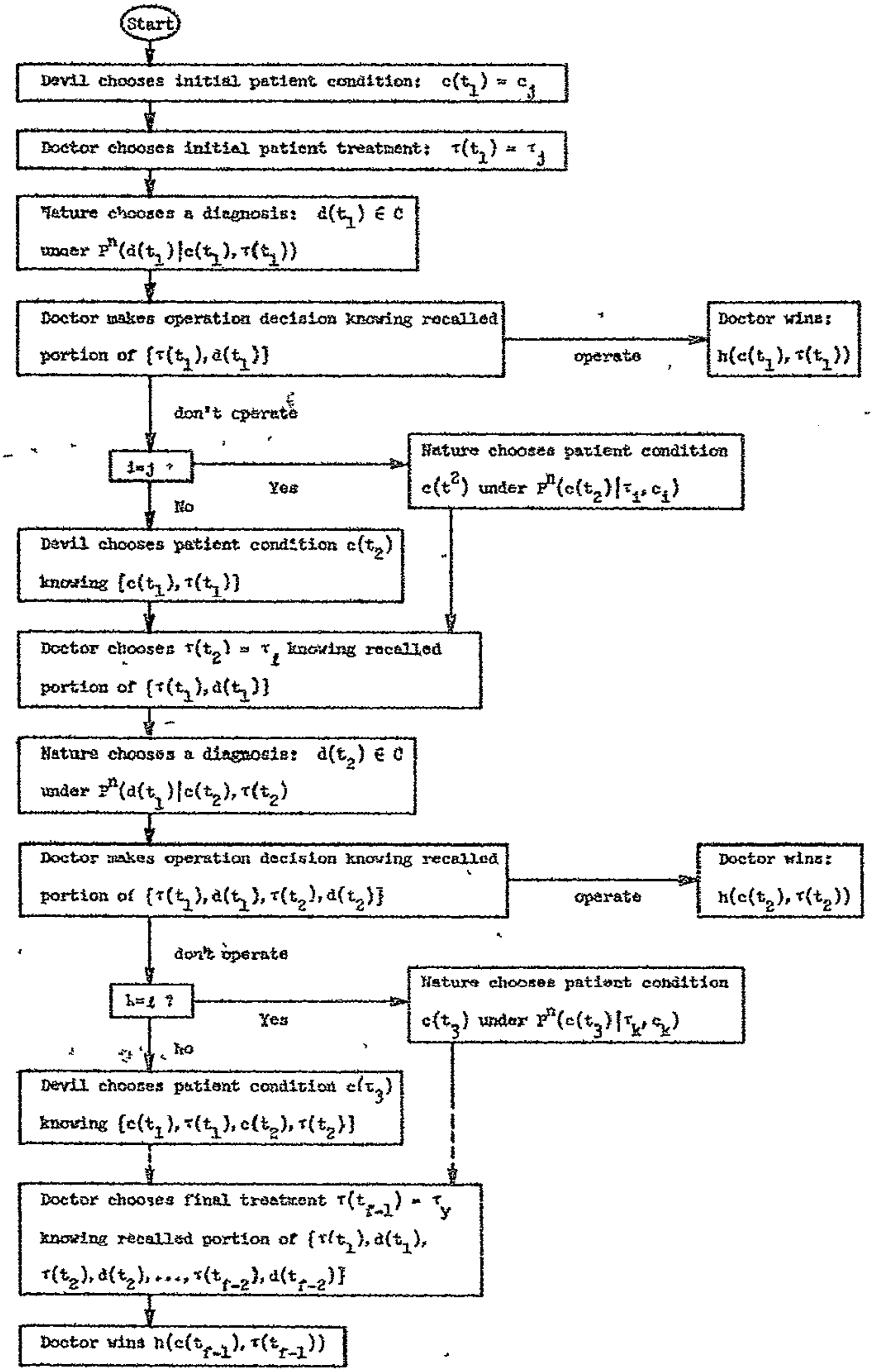


FIGURE 14 METHOD OF PLAY OF DOCTOR-DEVIL GAME WHEN THE DEVIL IS EMPLOYING ITS COMPLETE INFORMATION COLLECTION SCHEME

of her moves is specified by Assumption D and F;

(iii) the payoffs to each player for each possible play are specified by Assumption E, and

(iv) the maximum amount of information that each player can gather is specified by assumptions E, G, and H.

The interpretation of this game then, can be carried out by assigning an alternative indexing and constructing a tree, as was done for the pursuit and evasion game of Chapter 7.

Further, in interpreting this game we will again adopt the convention of referring to information sets by the corresponding strings of observations for the players, and we will denote by:

(i) $I^D(t_i^-)$ and $I^a(t_i^-)$, respectively, the doctor's and the devil's information sets at time t_i but prior to the operation decision of the doctor, and by

(ii) $I^D(t_i)$ and $I^a(t_i)$, respectively, the doctor's and the devil's information sets at time t_i but after the termination decision of the pursuer.

For example, when both the doctor and the devil employ their complete information collection schemes,

$$I^D(t_2^-) = \{r(t_i), d(t_i), i=1,2;\ \text{no operation thru } t_1\}$$

$$I^a(t_2^-) = \{c(t_i), r(t_i), i=1,2;\ \text{no operation thru } t_1\}$$

and

$$I^D(t_2^-) = \{r(t_i), d(t_i), i=1,2;\ \text{no operation thru } t_2\}$$

$$I^a(t_2^-) = \{c(t_i), r(t_i), i=1,2;\ \text{no operation thru } t_2\} .$$

8.3 The Doctor-Devil Game Studied

The following example doctor-devil game was studied.

(i) The sets of possible patient conditions and treatments are given by $C = \{c_1, c_2\}$ and $T = \{\tau_1, \tau_2\}$.

(ii) The game lasts for no more than 15 days, which is considered to consist of 5 three day periods.

(iii) The probability distributions under which the patient's conditions change when appropriate treatment is administered by the doctor (i.e.: $\tau(t_i) = \tau_j$ when $c(t_i) = c_j$ for $j = 1$ or 2) are constant with time. These distributions are given in table 10 below.

Table 10: Probability Distributions for Patient's Condition Changes Under Proper Treatment

	$c(t_{i+1}) = c_1$	$c(t_{i+1}) = c_2$
$P^n(c(t_{i+1}) \mid c(t_i) = c_1, \tau(t_i) = \tau_1)$	0.70	0.30
$P^n(c(t_{i+1}) \mid c(t_i) = c_2, \tau(t_i) = \tau_2)$	0.70	0.30

(iv) The payoffs to the doctor upon his decision to operate are given in terms of expected expressions of the patient's lifetime in days. These payoffs are described by the collection of matrices $A(i) = \{a(i)_{jk} = h(\tau(t_i) = \tau_j;\ c(t_i) = c_k)\}$ listed in table 11 below.

Table 11: Expected Lifetime Extension Payoff Matrices

$$A(1) = \begin{pmatrix} 1800 & 180 \\ 720 & 540 \end{pmatrix} \quad A(2) = \begin{pmatrix} 1620 & 160 \\ 650 & 485 \end{pmatrix} \quad A(3) = \begin{pmatrix} 1300 & 130 \\ 520 & 390 \end{pmatrix}$$

$$A(4) = \begin{pmatrix} 910 & 90 \\ 360 & 270 \end{pmatrix} \quad A(5) = \begin{pmatrix} 550 & 55 \\ 215 & 160 \end{pmatrix}$$

(v) The diagnoses received by the doctor at the end of each three day period are chosen by nature under probability distributions which are constant with time. These distributions are given in table 12 below. (Note that $d(t_i) = d_1$ and $d(t_i) = d_2$ are complementary events.)

Table 12: Probability Distributions for Diagnoses

j	1	1	2	2
k	1	2	1	2
$P^n(d(t_i) = d_1 \mid \tau(t_i) = \tau_j,\ c(t_i) = c_k)$	.80	.30	.70	.20

8.4 Application of the Extended Fictitious Play Algorithm

To apply the extended fictitious play algorithm, we carry out the following three steps.

Step 1: We assign to the antagonist, its complete information collection scheme C^a, and to the protagonist, the information collection scheme R^p that we wish to examine.

Step 2: We identify the entities of the associated game and we develop for these entities, the payoff expressions which are denoted in figure 10 by the vectors V^i for i=1,N and W^k for k=1,M.

In the associated game Γ', for the doctor-devil game and any admissible information collection scheme R^p, the entities are:

(i) 1^p, 3^p, 5^p, 7^p and 9^p, who choose treatments for the patient for the 1st, 2nd,..., and the 5th time periods, respectively,

(ii) 2^p, 4^p, 6^p and 8^p, who decide whether or not to operate at the end of the 1st, 2nd, 3rd and 4th time periods, respectively, and

(iii) 1^a, 2^a, 3^a, 4^a and 5^a, who choose (when nature does not) patient conditions for the 1st, 2nd,..., and 5th time periods, respectively.

In table 13 we list:

(i) the entities of an associated game and their functions, and

(ii) the general description and the number of information sets that each entity controls when the doctor employs its complete information collection scheme.

In writing the entity payoff expressions V^i for i=1,N and W^k for k=1,M, it should be noted that:

(i) we do not carry out the laborious task of determining the array F and forming these expressions as indicated in figure 10, but instead,

Table 13: Information Set Structure for the Complete Information Collection Scheme Case

Entity	Entity function	Information set description $I^p(k), I^a(k)$	No. of inf. sets "controlled"
1^a	choose $c(t_1)$	$\{0\}$	1
1^p	choose $\tau(t_1)$	$\{0\}$	1
2^p	operate at t_2^-?	$\{\tau(t_1), d(t_1)\}$	4
2^a	choose $c(t_2)$	$\{\tau(t_1), c(t_1)$; no operation thru $t_2\}$	2
3^p	choose $\tau(t_2)$	$\{\tau(t_1), d(t_1)$; no operation thru $t_2\}$	4
4^p	operate at t_3^-?	$\{\tau(t_i), d(t_i)$, $i=1,2$; no operation thru $t_2\}$	16
3^a	choose $c(t_3)$	$\{\tau(t_i), c(t_i)$, $i=1,2$; no operation thru $t_3\}$	8
5^p	choose $\tau(t_3)$	$\{\tau(t_i), d(t_i)$, $i=1,2$; no operation thru $t_3\}$	16
6^p	operate at t_4^-?	$\{\tau(t_i), d(t_i)$, $i=1,3$; no operation thru $t_3\}$	64
4^a	choose $c(t_4)$	$\{\tau(t_i), c(t_i)$, $i=1,3$; no operation thru $t_4\}$	32
7^p	choose $\tau(t_4)$	$\{\tau(t_i), d(t_i)$, $i=1,3$; no operation thru $t_4\}$	64
8^p	operate at t_5^-?	$\{\tau(t_i), d(t_i)$, $i=1,4$; no operation thru $t_5\}$	256
5^a	choose $c(t_5)$	$\{\tau(t_i), c(t_i)$, $i=1,r$; no operation thru $t_5\}$	256
9^p	choose $\tau(t_5)$	$\{\tau(t_i), d(t_i)$, $i=1,4$; no operation thru $t_5\}$	256

Total number of entities: 14; 9 for doctor, 5 for devil
Total number of controlled inf. sets: 852; 681 for doctor, 171 for devil

(ii) we take advantage of the structure of the game to write payoff expressions which have less elegant forms than the expressions of figure 10, but which lend themselves to more efficient computation.†

Step 3: We implement on a digital computer, the flow diagram of figure 10.

8.5 A Search for a Simplest Behavior Strategy ϵ-Solution for the Doctor

For the example doctor-devil game described above, we chose the following objective: find a simplest possible treatment and operation behavior strategy ϵ-solution for the doctor, for an $\epsilon = .10\,v$ where v is the quality of a solution.

Employing the extended fictitious play algorithm and evaluating the qualities of behavior strategies it produces by means of Theorem 4′ (section 6.5.2), we proceed as follows:

(i) First, we establish the quality of a behavior strategy solution for the doctor by generating an entity-behavior equilibrium point in the associated game corresponding to the doctor's complete information collection scheme.

(ii) Then we test other admissible†† information collection schemes for the doctor which might produce simple behavior

---

†As we did in the pursuit and evasion games of the last chapter.

††Defined in section 6.1.1.

strategy ϵ-solutions,[†] by generating entity-behavior equilibrium points in the corresponding associated games and comparing the qualities of behavior strategies thus obtained with the quality of a solution.

8.5.1 Information Collection Schemes Considered

In carrying out our search for a simplest behavior strategy ϵ-solution, we examined a total of 8 information collection schemes for the doctor (including his complete information collection scheme).

In table 14 we give the results of our search, specifically,

(i) we describe the information collection schemes examined by listing the information that is forgotten by the doctor under each scheme (in each scheme considered, the doctor gathers every available piece of information),

(ii) we list the qualities of the treatment and operation behavior strategies that were generated on each information collection scheme, and

(iii) for each strategy generated, we list the total number of information sets realizable (or an upper bound on this number), as an indication of the complexity of this strategy.

8.5.2 Comparison of the Behavior Strategies Generated

Comparing the strategies generated on each information collection scheme, we see that:

[†] See section 4.2.

Table 14: Summary of Results of Search for a Simplest Behavior Strategy ϵ-Solution

Information Collection Scheme	Information Forgotten	Number of Realizable Information Sets	Quality of Behavior Strategy (in days)
1	comp. inf. coll. scheme (none)	675	890^{+0}_{-45}
2	at t_5^- forget $\tau(t_1), d(t_1)$	291	891^{+0}_{-41}
3	at t_5^- forget $\tau(t_i), d(t_i), i=1,2$	195	885^{+0}_{-43}
4	at t_5^- forget $\tau(t_i), d(t_i), i=1,2$ at t_4^- forget $\tau(t_1), d(t_1)$	99	885^{+0}_{-10}
5	at t_5^- forget $\tau(t_i), d(t_i), i=1,3$ at t_4^- forget $\tau(t_1), d(t_1)$	77	891^{+0}_{-38}
6	at t_5^- forget $\tau(t_i), d(t_i), i=1,3$ at t_4^- forget $\tau(t_i), d(t_i), i=1,2$ at t_3^- forget $\tau(t_1), d(t_1)$	33 (max)	706^{+0}_{-13}
7	at t_5^- forget $\tau(t_i), d(t_i), i=1,3$ at t_4^- forget $\tau(t_i), d(t_i), i=1,2$	53	888^{+0}_{-34}
8	at t_5^- forget $\tau(t_i), d(t_i), i=1,4$ at t_4^- forget $\tau(t_i), d(t_i), i=1,2$	51 (max)	727^{+0}_{-17}

Quality of Solution: 890^{+29}_{-45} days

(i) the strategies generated for schemes 1, 2, 3, 4, 5 and 7 are all ϵ-solutions for $\epsilon = 0.10$ v, and that,

(ii) under the particular ϵ-solution generated for information collection scheme 7, only 53 information sets are realizable.

In finding "strategy 7" then, which is listed in table 15, we have found a behavior strategy ϵ-solution which is less than 8% as complex as the behavior strategy ϵ-solution generated on the doctor's complete information collection scheme.

In figure 15 we show for each information collection scheme examined, a plot of the amount of information recalled by the doctor at the beginning of each time interval. Comparing information collection schemes 6, 7 and 8 in this figure, it would seem that the minimum amount of "information recall" that the doctor requires for construction of a behavior strategy ϵ-solution is <u>not</u> non-decreasing.

This result seemed surprising, but upon examination of the strategy generated for information collection scheme 7, we found that the probability of an operation before time t_4 is quite high ($> .90$). This makes the occurrence of information sets corresponding to times t_4 and t_5 "rare events," which in turn makes the expected payoff to the doctor insensitive to crude play in these later situations.

$_1P$ $\quad p(\tau(t_1)=\tau_1) = .40$

Note that {operate, don't operate} are <u>complementary events</u>, as are $\{c_1, c_2\}$ in the game under study.

	$\tau(t_1)$	1	1	2	2
	$d(t_1)$	1	2	1	2
$_2P$	$p(\text{operate}\mid\tau(t_1), d(t_1))$	.94	.01	.19	.04
$_3P$	$p(\tau(t_2)=\tau_1\mid\tau(t_1), d(t_1))$	.93	.26	.26	.97

	$\tau(t_1)$	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2
	$\tau(t_2)$	1	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2
	$d(t_1)$	1	1	2	2	1	1	2	2	1	1	2	2	1	1	2	2
	$d(t_2)$	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
$_4P$	$p(\text{operate}\mid\tau(t_i), d(t_i), i=1,2)$	.99	.53	.99	.05	.97	.01	.93	.01	.99	.11	.99	.99	.99	.06	.99	.11
$_5P$	$p(\tau(t_2)=\tau_1\mid\tau(t_i), d(t_i), i=1,2)$	1.0	.35	.93	.09	.11	.99	.41	.99	.97	.10	.97	.10	.53	.97	.01	.97

	$\tau(t_3)$	1	1	2	2
	$d(t_3)$	1	2	1	2
$_6P$	$p(\text{operate}\mid\tau(t_3), d(t_3))$	.99	.99	.99	.08
$_7P$	$p(\tau(t_4)=\tau_1\mid\tau(t_2), d(t_3))$	.99	.02	.73	.99

<u>Quality:</u> 888^{+0}_{-34} days of expected patient lifetime extension (compared with 890^{+0}_{-45} for behavior strategy generated for complete information collection scheme).

	$\tau(t_4)$	1	1	2	2
	$d(t_4)$	1	2	1	2
$_8P$	$p(\text{operate}\mid\tau(t_4), d(t_4))$	.94	.01	.99	.99
$_9P$	$p(\tau(t_5)=\tau_1\mid\tau(t_4), d(t_4))$	.93	.25	1.0	1.0

Table 15: Treatment and Operation Behavior Strategy for Information Collection Scheme 7

Information Collection Scheme Number	1	2	3	4	5	6	7	8
Quality of Generated Treatment and Operation Behavior Strategy	890^{+0}_{-45}	891^{+0}_{-41}	885^{+0}_{-43}	886^{+0}_{-10}	891^{+0}_{-38}	706^{+0}_{-13}	888^{+0}_{-32}	727^{+0}_{-17}

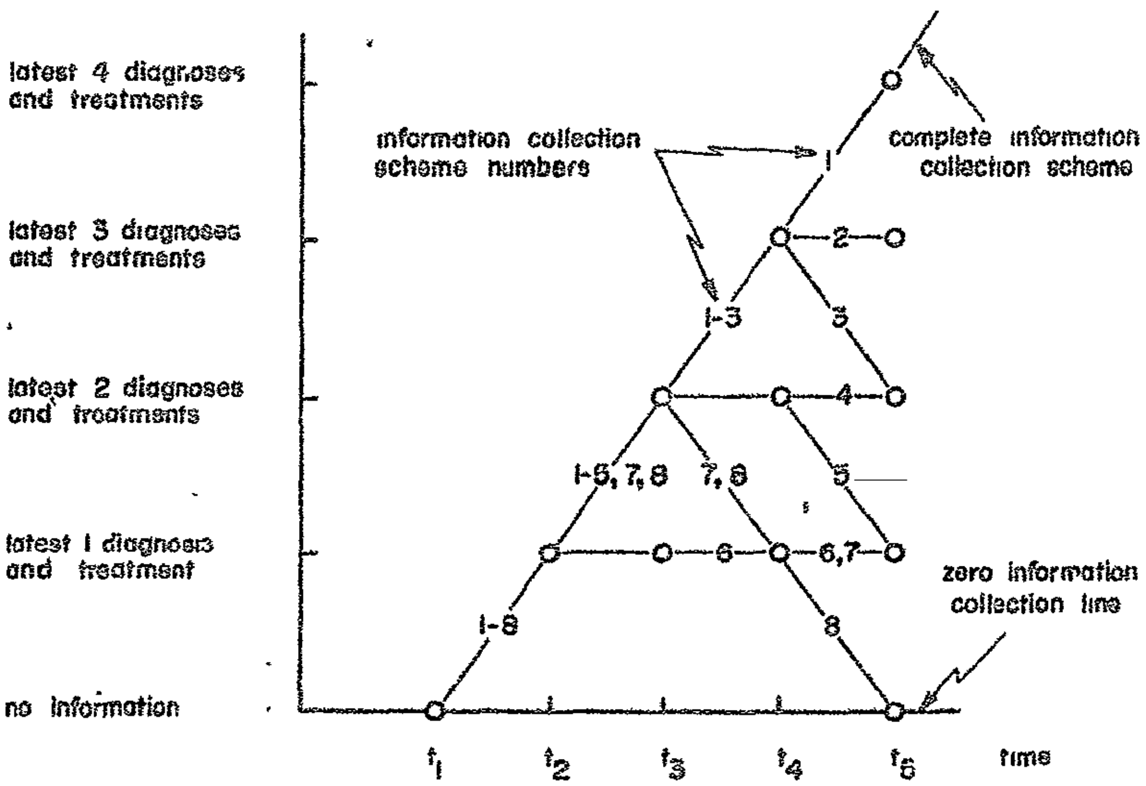


FIGURE 15. INFORMATION KNOWN BY DOCTOR FOR INFORMATION COLLECTION SCHEMES 1 THRU 8.

8.6 Observed Convergence of the Extended Fictitious Play Algorithm

In our studies of the doctor-devil problem, the extended fictitious play algorithm was implemented on a CDC 6500 digital computer. The particular program that was written was constructed in a sufficiently general form to allow study of all possible admissable information collection schemes with a minimum of modification required in changing from scheme to scheme. As a consequence, very little change in the amount of calculation per iteration of the algorithm results when we change information collection schemes. It is not surprising, therefore, that convergence of the algorithm, in the sense indicated in section 6.5.2 (for $\delta = 0.05$), took approximately the same amount of time for every information collection scheme considered.

In table 16, we summarize the relevant information with respect to the observed convergence of the extended fictitious play algorithm in the doctor-devil game with the doctor employing its complete information collection scheme. In this table, the number of variables to be determined is the total number of probability distribution components in the pair of behavior strategies determined for both players. This number does not take into account the fact that these components must sum to 1.0 over the alternatives of each information set, since the program (as written) does not take advantage of this fact.

Table 16: Observed Convergence for the Complete Information Collection Scheme Case ($\delta = 0.05$)

Number of alternative probability distributions to be determined	852
Number of variables to be determined	1704
Number of iterations	200
Total time (in seconds)	2000

8.7 Summary and Conclusions

In this chapter,

(i) we have demonstrated the use of the extended fictitious play algorithm, and

(ii) we have carried out a search for a simplest behavior strategy ϵ-solution that has resulted in a behavior strategy ϵ-solution which is less than 8% as complex as the behavior strategy solution that we generated on the doctor's complete information collection scheme.

Finally, then, is the trouble of generating a behavior strategy ϵ-solution justified in this problem? We argue, as follows, that it is.

Most of the doctors that we know seem fairly conservative, and if faced with a problem such as posed by our example, we suspect that they might be tempted to apply treatment 2 over period 1 and operate at the first opportunity, thus achieving a minimum expected return of 540 days extension and avoiding the possibility of getting an expected return of only 180 days extension.[†]

By applying the simple treatment and operation behavior strategy given in figure 15, however, the doctor can raise his minimum expected return to 856 days extension. This represents a 50% improvement over the "conservative policy" mentioned above.

[†]Refer to A(1) in table 11.

LIST OF REFERENCES

Direct References

1. Von Neumann, J., and Morgenstern, O., Theory of Games and Economic Behavior, Princeton University Press, 1944.

2. Kuhn, H. W., "Extensive Games and the Problem of Information," Contributions to the Theory of Games, Annals of Math Studies, No. 28, Princeton University Press, 1965.

3. Luce, R. D., and Raiffa, H., Games and Decisions, John Wiley and Sons, Inc., 1957.

4. Karlin, S., Mathematical Methods and Theory in Games, Programming and Economics, Addison-Wesley, 1959, Vol. 1.

5. Rapoport, A., Two-Person Game Theory, The Essential Ideas, Ann Arbor, The University of Michigan Press, 1960.

6. Von Neumann, J., "A Numerical Method to Determine an Optimal Strategy," Naval Research Logistics Quarterly, No. 1, 1954, pp. 109-115.

7. Brown, G. W., "Iterative Solutions of Games by Fictitious Play," Activity Analysis of Production and Allocation, Cowles Commission Monograph 13, Koopmans, T. C., editor, John Wiley and Sons, New York, 1951, pp. 374-376.

8. Robinson, J., "An Iterative Method of Solving a Game," Annals of Mathematics, Vol. 54, No. 2, Sept. 1956.

9. Danskin, J. M., "Fictitious Play for Continuous Games," Naval Research Logistics Quarterly, No. 1, 1954, pp. 313-320.

10. Bellman, R., Dynamic Programming, Princeton University Press, 1957.

11. Bellman, R., and Dreyfus, S., Applied Dynamic Programming, Princeton University Press, 1962.

12. Nash, J., "Non-Cooperative Games," Annals of Mathematics, Vol. 54, No. 2, Sept. 1951.

13. Shapley, L. S., "Some Topics in Two-Person Games," Annals of Math Studies, Vol. 52, 1964.

14. Kuhn, H. W., "Extensive Games," Proceedings of the National Academy of Sciences, U.S.A., No. 36, 1950, pp. 570-576.

15. Utgoff, V., and Kashyap, R., "A Pursuit and Evasion Problem With Measurement Uncertainty," Tech. Rept. 69-39, School of Electrical Engineering, Purdue University, Lafayette, Ind., Oct. 1969.

16. Utgoff, V., and Kashyap, R., "On Behavior Strategy Solutions of Finite Two-Person Constant-Sum Extended Games," Tech. Rept. 70-9, School of Electrical Engineering, Purdue University, Lafayette, Ind., Feb. 1970.

General References

17. Ferguson, T., Mathematical Statistics, A Decision Theoretic Approach, Academic Press, 1967.

18. Game Theory and Related Approaches to Social Behavior, Shubik, M., editor, John Wiley and Sons, Inc., 1964.

19. Danskin, J., The Theory of Max-Min, Econometrics and Operations Research V, Springer-Verlag, New York, Inc., 1967.

20. Blackwell, D., and Gershick, M., Theory of Games and Statistical Decisions, John Wiley and Sons, Inc., 1954.

21. Aumann, R., "Mixed and Behavior Strategies in Infinite Extensive Games," Annals of Math Studies, No. 52, Princeton University Press, 1964, pp. 627-650.

22. Dalkey, N., "Equivalence of Information Patterns and Essentially Determinate Games," Annals of Math Studies, No. 28, Princeton University Press, 1953, pp. 217-244.

23. Thompson, G., "Signaling Strategies in n-Person Games," Annals of Math Studies, No. 28, Princeton University Press, 1953, pp. 267-278.

24. Thompson, G., "Bridge and Signaling," Annals of Math Studies, No. 28, Princeton University Press, 1953, pp. 279-290.

25. Kuhn, H. W., "A Simplified Two-Person Poker," Annals of Math Studies, No. 24, Princeton University Press, 1950, pp. 97-104.

26. Aström, K., "Optimal Control of Markov Processes With Incomplete State Information," J. Math. Anal. and Appl., 10, 1965, pp. 174-205.

27. Deyst, J. J., "Optimal Control in the Presence of Measurement Uncertainties," M.I.T. Expr. Astronom. Lab. Rept. TE-17, Jan. 1967.

28. Isaacs, R., "Differential Games: Their Scope, Nature, and Future," Journal of Optimization Theory and Appl., Vol. 3, No. 5, 1969.

29. Dresher, M., Games of Strategy: Theory and Applications, Prentiss Hall, 1961.

30. Danskin, J., "An Algorithm for the Solution of Concave-Convex Games," Computing Methods in Optimization Problems - 2, Academic Press, Inc., 1969, pp. 67-76.

31. Galton, F., "Statistical Inquiries into the Efficacy of Prayer," Fortnightly Review, No. LXVIII, New Series, Aug. 1872.

32. Eaton, J. H., and Zadeh, L. A., "Optimal Pursuit Strategies in Discrete-State Probabilistic Systems," Journal of Basic Engineering, Vol. 23, March 1962.

33. Krylov, N. V., "The Construction of an Optimal Strategy for a Finite Controlled Chain," (translated by A. R. Kraiman), Theory of Probability and Its Applications, Vol. X, No. 1, 1965.

34. Zachrisson, L. E., "Markov Games," Contributions to the Theory of Games, Annals of Math Studies, No. 39, Princeton University Press, 1957.

35. Raviv, J., "Decision Making in Incompletely Known Stochastic Systems," Int. J. Eng. Sci., Vol. 3, 1965.

36. Howard, R. A., "Value of Information Lotteries," IEEE Trans. on Systems Science and Cybernetics, Vol. SSC-3, No. 1, June 1967.

37. Howard, R. A., "Information Value Theory," IEEE Trans. on Systems Science and Cybernetics, Vol. SSC-2, No. 1, Aug. 1966.

38. Dynkin, E. B., "Controlled Random Sequences," (translated by A. R. Kraiman), Theory of Probability and Its Applications, Vol. X, No. 1, 1965.

APPENDIX A

NOTES

1. Advantages of Randomization

Games can easily be constructed such that for some information collection scheme for the protagonist, randomized strategies exist which guarantee the protagonist a greater minimum expected return than any pure strategy on the same information collection scheme.

For an example of a "real life" game with this property, see reference 1, pages 186-219, where Von Neumann shows that poker is such a game, and specifically, that maximally effective strategies in poker call for a randomization in betting which is commonly called bluffing.

2. Quality and Solution Strategy Definitions for the Antagonist

If in a given game we wish to retain a given labeling for the protagonist and the antagonist, we can define strategy quality for the antagonist as follows.

Definition: The quality $Q(Y^a, R^a)$ of any strategy Y^a for the antagonist which is based on an information collection scheme R^a is defined by:

$$Q(Y^a, R^a) = \min_{X^p, R^p} [-H(X^p, R^p; Y^a, R^a)] .$$

Using this definition, the definitions of an ϵ-R-solution and an ϵ-solution for the antagonist are analogous to the corresponding definitions for the protagonist.

3. Definition of Terms Used in the Wheat-Market Game

On the Chicago Mercantile Exchange, it is possible to buy and sell contracts for future delivery of wheat.

The size of a contract is fixed at 5000 bushels; the date of delivery of any contract is fixed; and until the date of delivery, only a small deposit is required to carry out a purchase or sale.

A trader in wheat-futures contracts can realize profits or losses by carrying out transactions of the following sorts.

(i) He can buy a wheat contract, and "cancel" his purchase at or before the delivery date by selling a wheat contract with the same delivery date. This is called taking a "long position" and the trader realizes a profit (loss) if the value of a contract for delivery rises (falls) while he is long.

(ii) He can sell a wheat contract, and "cancel" his purchase at or before the delivery date by buying a wheat contract with the same delivery date. This is called taking a "short position" and the trader realizes a profit (loss) if the value of a contract for delivery falls (rises) while he is short.

4. A Game With an Information Collection Scheme Upon Which a Behavior Strategy Solution Does Not Exist

Consider the game shown in figure 16 and the information collection scheme R^p in which the protagonist fails to recall at any of his 2nd moves, the choice he has made at his first move.

The information partition for this information collection scheme is given by $IP(R^p) = \{A, C \cup D\}$. In table 17 below, we list the protagonist's pure strategies on R^p and the antagonist's pure strategies on its complete information collection scheme. In the listed strategies, the first integer in each entry indicates the alternative to be chosen when the information set indicated by the letters in that entry is realized.

Table 17: Pure Strategies in Coded Form

$$\sigma P_1^p = \begin{pmatrix} 1 & A \\ 1 & C \cup D \end{pmatrix} \quad \sigma P_2^p = \begin{pmatrix} 1 & A \\ 2 & C \cup D \end{pmatrix} \quad \sigma P_3^p = \begin{pmatrix} 2 & A \\ 1 & C \cup D \end{pmatrix}$$

$$\sigma P_4^p = \begin{pmatrix} 2 & A \\ 2 & C \cup D \end{pmatrix} \quad \sigma P_1^a = (1\ B) \quad \sigma P_2^a = (2\ B)$$

Using the notation established in section 5.1.2, we can write the expected return to the protagonist when mixed strategies σM^p on R^p and σM^a on C^a are employed, as:

$$H(\sigma M^p, R^p; \sigma M^a, C^a) = \sigma M_1^a\,(\sigma M_1^p \cdot 0 + \sigma M_2^p \cdot 10 + \sigma M_3^p \cdot 5 + \sigma M_4^p \cdot 0)$$

$$+ \sigma M_2^a\,(\sigma M_1^p \cdot 0 + \sigma M_2^p \cdot 5 + \sigma M_3^p \cdot 10 + \sigma M_4^p \cdot 0)\ . \qquad (A4.1)$$

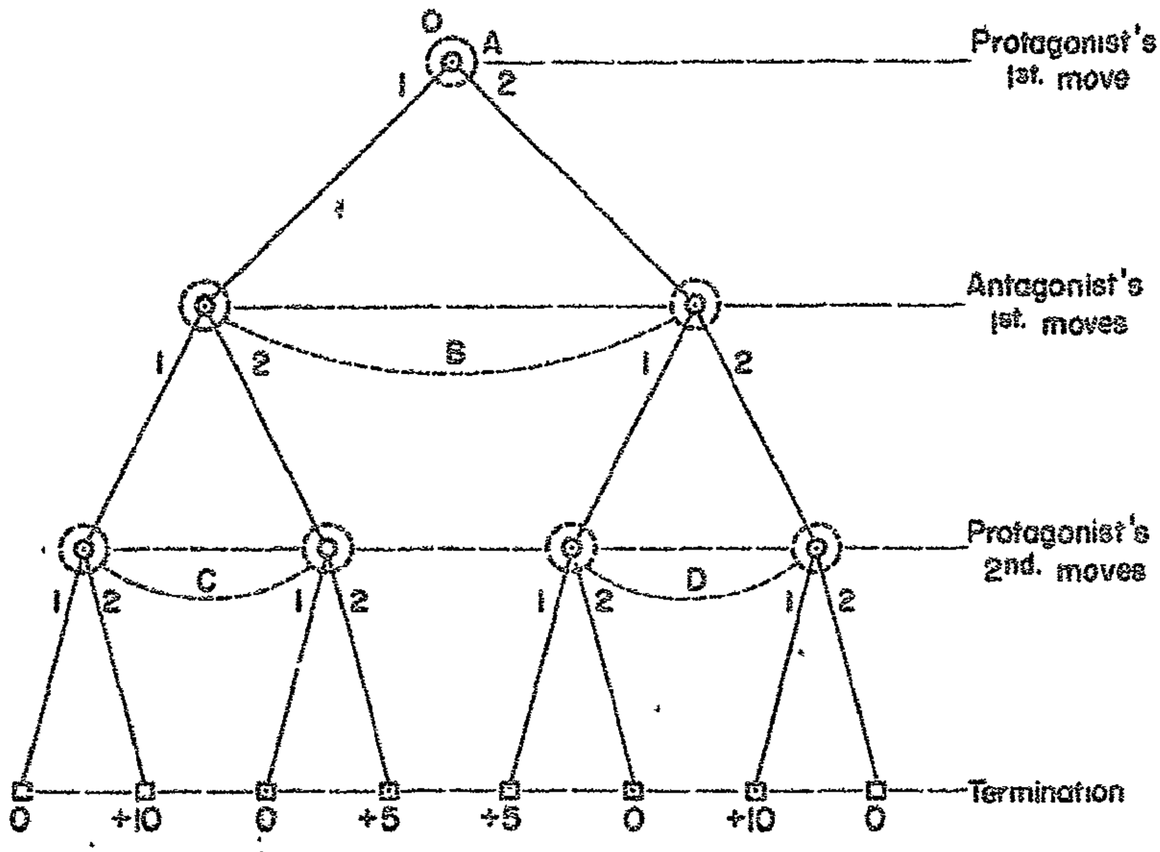


FIGURE 16. A GAME WITH AN INFORMATION COLLECTION SCHEME UPON WHICH A BEHAVIOR STRATEGY R-SOLUTION DOES NOT EXIST.

If we examine expression (A4.1), we find that the mixed strategy σM^p on R^p defined by $\sigma M_2^p = \sigma M_3^p = 0.5$ has maximum quality in the set of all possible mixed strategies on R^p. The quality of this strategy (and therefore the quality of an R-solution on R^p) is 7.5.

If the players employ behavior strategies σB^p on R^p and σB^a on C^a, respectively, the protagonist's expected return is given by:

$$H(\sigma B^p, R^p; \sigma B^a, C^a) = p(1|A) \left[p(1|B) \left[p(1|C \cup D) \cdot 0 + p(2|C \cup D) \cdot 10\right] + p(2|B) \left[p(1|C \cup D) \cdot 0 + p(2|C \cup D) \cdot 5\right] \right] + p(2|A) \left[p(1|B) \left[p(1|C \cup D) \cdot 5 + p(2|C \cup D) \cdot 0\right] + p(2|B) \left[p(1|C) \cdot 10 + p(2|C) \cdot 0\right] \right] \qquad (A4.2)$$

where the probabilities indicated are defined by the given behavior strategies.

If we examine expression (A4.2), we find that the behavior strategy σB^p on R^p defined by $p(1|A) = p(1|C \cup D) = 0.5$ has maximum quality in the set of all possible behavior strategies on R^p.

The quality of this behavior strategy, however, is only 3.75, and therefore a behavior strategy R-solution on R^p does not exist.

5. Construction of $g(E^p(i), I^p(i),\ i=1,N;\ E^a(j), I^a(j),\ j=1,M;\ E^n(k), I^n(k),\ k=1,L)$

This function can be constructed by listing for each play $W \in K$, the corresponding string of alternative choices and information sets under the players' given information collection schemes.

Then, for each listed string, the function g takes the value of the protagonist's payoff when the corresponding play W is realized. For example, in the game shown in figure 16, when the players are both employing their complete information collection schemes, g is defined as shown in table 18 below.

Table 18: Definition of g for Game of Figure 16 When Both Players Employ Their Complete Information Collection Schemes

g(1,A,1,C U D; 1,B) = 0	g(2,A,1,C U D; 1,B) = 5
g(1,A,2,C U D; 1,B) = 10	g(2,A,2,C U D; 1,B) = 0
g(1,A,1,C U D; 2,B) = 0	g(2,A,1,C U D; 2,B) = 10
g(1,A,2,C U D; 2,B) = 5	g(2,A,2,C U D; 2,B) = 0

6. Intuitive Appeal of the Method of Fictitious Play

We quote Brown in reference 7:

"A very simple iterative method for approximating to solutions (R-solutions in mixed strategies) of discrete zero-sum games. This method is related to some particular systems of differential equations whose steady state solutions correspond to solutions of a game.... The iterative method in question can be loosely characterized by the fact that it rests on the traditional statisticians philosophy of basing future decisions on the relevant past history. Visualize two statisticians, perhaps ignorant of minimax theory, playing many plays (engaging in many repetitions) of the same discrete zero-sum game. One might naturally expect a statistician to keep track of the opponent's past plays (chosen pure strategies) and, in the absence of a more sophisticated calculation, perhaps to choose at each play (repetition) the optimal pure strategy against the mixture represented by all the opponent's past plays (chosen pure strategies).

7. Derivation of Behavior Strategy Fictitious Play Equations

Starting with equation (5.9) repeated below,

$$\sigma M^{p}(j+1) = \left(\frac{1}{j+1}\right)\left[j\ \sigma M^{p}(j) + \underset{\sigma P^{p}}{\text{amax}}\ H(\sigma P^{p}, R^{p}; \sigma M^{a}(j), C^{a})\right], \quad (A7.1)$$

we can apply Kuhn's transformation (T) to both sides of (A7.1) to obtain:

$$\sigma B^{p}(j+1) = \left(\frac{1}{j+1}\right)\left[j\ \sigma B^{p}(j) + T\left[\underset{\sigma P^{p}}{\text{amax}}\ H(\sigma P^{p}, R^{p}; \sigma M^{a}(j), C^{a})\right]\right]. \quad (A7.2)$$

Further, since the protagonist's expected return is identical for all pure strategies σP^{p} on R^{p} under any mixed strategy σM^{a} on C^{a} or its associated behavior strategy, $T(\sigma M^{a}$ on $C^{a}) = \sigma B^{a}$ on C^{a}, we can write (A7.2) as:

$$\sigma B^{p}(j+1) = \left(\frac{1}{j+1}\right)\left[j\ \sigma B^{p}(j) + T\left[\underset{\sigma P^{p}}{\text{amax}}\ H(\sigma P^{p}, R^{p}; \sigma B^{a}(j), C^{a})\right]\right]. \quad (A7.3)$$

The derivation of equation (5.12) from (5.10), can be carried out in the same manner.

8. Simpler Expressions For $H(\sigma P^p, R^p; \sigma B^a(j), C^a)$ and $H(\sigma B^p(j), R^p; \sigma P^a, C^a)$

In carrying out repetitions of the fictitious play algorithm on a digital computer using equations (5.11) and (5.12) directly, many multiplication operations may be required to generate $H(\sigma P^p, R^p; \sigma B^a(j), C^a)$ and $H(\sigma B^p(j), R^p; \sigma P^a, C^a)$.

Note, however, that:

$$\min_{\sigma P^p} H(\sigma B^p(j), R^p; \sigma P^a, C^a) = \min_{\sigma P^a} \sum_{i=1}^{j} H(\sigma P^p(i), R^p; \sigma P^a, C^a) \qquad (A8.1)$$

and

$$\max_{\sigma P^p} H(\sigma P^p, R^p; \sigma B^a(j), C^a) = \max_{\sigma P^p} \sum_{i=1}^{j} H(\sigma P^p, R^p; \sigma P^a(i), C^a) \qquad (A8.2)$$

where $\sigma P^p(i)$ [$\sigma P^a(i)$] is the pure strategy selected for the protagonist [antagonist] at the ith iteration of the algorithm. Using these expressions, we can substitute addition operations for many of the multiplication operations required if (5.11) and (5.12) are used directly. This leads to a faster implementation of the algorithm since a digital computer performs addition operations much more rapidly than multiplication operations.

9. A Less Restrictive Assumption

Assumption 3 can be replaced by any other assumption which allows the arguments following definition 18 to be made. A minimally restrictive assumption allowing these arguments is stated below.

Assumption: No information set contains more than one move on the same play.

This assumption is identical to that made by Kuhn in reference 2. With this assumption, definition 18 must be stated as follows:

Definition 18 (restated): The associated non-cooperative game Γ^{*} for a given two-person, zero-sum, finite extended game Γ with a given information collection scheme R^{p} for the protagonist, is the (N + M)-person non-cooperative game defined by:

(i) assigning to the antagonist its complete information collection scheme C^{a};

(ii) considering every information set in $IP(R^{p})$ $\{I^{p}(C^{a})\}$ to be under the control of a separate and independent entity such that

(a) the entity controls information sets corresponding to only one player, and

(b) the probability of occurrence of exactly one of the entity's information sets is 1.0 in any repetition of the game for any strategies employed by the entities; and

(iii) considering the payoff to each entity to be identical to that of the player whose moves the entity is controlling.

The arguments subsequent to definition 18 in the text can be carried to identical conclusions using definition 18 (restated) given above.

We employed assumption 3 in the text because it is only slightly more restrictive than the assumption given above, but it makes the arguments following definition 18 relatively simple to explain.

10. The Statistical Description for Nature in the Second Pursuit and Evasion Game Example

If we examine table 5, we can see that in state 2 the players can obtain measurements of their opponent's states which are much better than the measurements obtainable in states 1 and 3 (the pursuer's and the evader's initial states, respectively).

This forces the pursuer to make a compromise between:

(i) selection of a state $a(t_1)$ to effect immediate capture, or

(ii) selection of a next state $a(t_1)$ to gain a good measurement which will be useful in achieving a later capture.

In examining the solutions generated for this problem, the effect of choosing these measurement statistics was clearly visible. Specifically, the pursuer's solution takes him to state 2 at t_1 with relatively high probability (0.53), even though the evader's solution takes him to state 2 at t_1 with relatively low probability (0.15).

11. Differences in Qualities of Solutions and Separation Strategies

We conjecture that the large differences between the qualities of the generated solutions and the qualities of the generated separation strategies are due to the strong dependence of the measurement statistics upon the players' states.

In cases where such dependence is present, use of a separation strategy of this type may be particularly dangerous, for such a strategy does not place a value upon a state on the basis of the information that may be gained (or lost) by entering it.

APPENDIX B

PROOFS

1. Proof of Corollary 1 (Section 5.2)

Let σM^{p*} on R^p be an ϵ-R-solution of Γ, and let ϕ be the quality of an R-solution on R^p. Then since

(i) a degenerate mixed strategy σMD^a on C^a (a pure strategy) can always be found which minimizes $H(\sigma M^{p*}, R^p; \sigma M^a, C^a)$[†],

and

(ii) σM^{p*} on R^p is an ϵ-R-solution on R^p,

it follows that: there does not exist a degenerate mixed strategy σMD^a on C^a such that

$$H(\sigma M^{p*}, R^p; \sigma MD^a, C^a) < \phi - \epsilon \text{ .} \qquad \text{(B1.1)}$$

Now, assume that the behavior strategy $\bar{\sigma} B^{p*}$ on R^p which is associated with σM^{p*} on R^p is not an ϵ-R-solution. Then since

(i) a degenerate behavior strategy σBD^a on C^a (a pure strategy) can always be found which minimizes $H(\bar{\sigma} B^{p*}, R^p; \sigma B^a, C^a)$[††],

and

(ii) $\bar{\sigma} B^{p*}$ on R^p is not an ϵ-R-solution,

it follows that: there exists a degenerate behavior strategy σBD^{a*}

[†]See argument in section 5.1.2.

[††]See argument in section 5.6.

on C^a such that

$$H(\tilde{\sigma}B^{p*}, R^p; \sigma BD^{a*}, C^a) < \emptyset - \epsilon \ .$$

But the pure strategy corresponding to σBD^{a*} on C^a is equivalent to some degenerate mixed strategy σMD^{a*} on C^a, and further, $T(\sigma MD^{a*}) = \sigma BD^{a*}$ where T is Kuhn's transformation (expression 2.8 in section 5.2).

By Kuhn's Theorem, then, (section 5.2):

$$H(\sigma M^{p*}, R^p; \sigma MD^{a*}, C^a) = H(\tilde{\sigma}B^{p*}, R^p; \sigma BD^{a*}, C^a) < \emptyset - \epsilon \ .$$

But this <u>contradicts</u> our assumption that $\tilde{\sigma}B^{p*}$ on R^p is not a behavior strategy ξ-R-solution on R^p. ∎

2. <u>Proof of Corollary 2 (Section 5.4)</u>

By Von Neumann's minimax theorem[†] and definition 11, for any game Γ and any information collection scheme R^p, there exists a pair of mixed strategies σM^{p*} on R^p and σM^{a*} on C^a such that:

$$\min_{\sigma M^a} H(\sigma M^p, R^p; \sigma M^a, C^a) \leq \min_{\sigma M^a} H(\sigma M^{p*}, R^p; \sigma M^a, C^a) =$$

$$\max_{\sigma M^p} H(\sigma M^p, R^p; \sigma M^{a*}, C^a) \leq \max_{\sigma M^p} H(\sigma M^p, R^p; \sigma M^a, C^a) \qquad (B2.1)$$

for any other pair of mixed strategies σM^p on R^p and σM^a on C^a.

But since an R-solution on R^p always exists in the set of all possible mixed strategies σM^p on R^p, σM^{p*} on R^p in (B2.1) must be an R-solution. Denoting the quality of an R-solution on R^p as $\emptyset$, we can therefore write from (B2.1):

[†]Reference 1, section 17.

$$\emptyset \leq \max_{\sigma M^p} H(\sigma M^p, R^p; \sigma M^a, C^a), \text{ for any } \sigma M^a \text{ on } C^a. \quad (B2.2)$$

Now from the Brown-Robinson Theorem (section 5.4), for any $\epsilon > 0$ there exists a J such that for all $j \geq J$:

$$H(\sigma M^p(j), R^p; \sigma M^a, C^a) > \max_{\sigma M^p} H(\sigma M^p, R^p; \sigma M^a(j), C^a) - \epsilon/2 \quad (B2.3)$$

for any mixed strategy σM^a on C^a, and

$$H(\sigma M^p, R^p; \sigma M^a(j), C^a) < \min_{\sigma M^a} H(\sigma M^p(j), R^p; \sigma M^a, C^a) + \epsilon/2 \quad (B2.4)$$

for any mixed strategy σM^p on R^p.

From (B2.3) and (B2.4), then,

$$\max_{\sigma M^p} H(\sigma M^p, R^p; \sigma M^a(j), C^a) - \epsilon/2 \leq \min_{\sigma M^a} H(\sigma M^p(j), R^p; \sigma M^a, C^a) + \epsilon/2$$

for all $j \geq J$. (B2.5)

Combining (B2.2) and B2.5), we obtain:

$$\emptyset \leq \min_{\sigma M^a} H(\sigma M^p(j), R^p; \sigma M^a, C^a) + \epsilon = Q(\sigma M^p(j), R^p) + \epsilon$$

for all $j \geq J$.

By definition, then, $\sigma M^p(j)$ on R^p is an ϵ-R-solution for any $j \geq J$. ▮

3.0 <u>Proof of Theorem 2 (Section 6.3)</u>

In order to prove Theorem 2, we establish the following definitions and notation.

Definition B1: Any entity-behavior strategy X^{pi} [Y^{ai}] which consists of only zeros and a single 1.0 in each partition will be called an entity-pure strategy and will be denoted as XP^{pi} [YP^{ai}].

Since there are finitely many information sets and alternatives, there are finitely many entity-pure strategies. We shall assign an arbitrary ordering to each entity's set of entity-pure strategies and denote for the k^p-entity $k=1,N$ [k^a-entity, $k=1,M$]:

(a) the ith entity-pure strategy as XP_i^{pk} [YP_i^{ak}],

(b) the total number of entity-pure strategies as ℓ^{pk} [ℓ^{ak}], and

(c) the set of all entity-pure strategies as S^{pk} [S^{ak}].

Definition B2: For any entity k^p, $k=1,N$ [k^a, $k=1,M$], any probability distribution over S^{pk} [S^{ak}] will be called an entity-mixed strategy and will be denoted as XM^{pk} [XM^{ak}].

We shall denote the probability assigned under XM^{pk} [YM^{ak}] to the entity-pure strategy XP_i^{pk} [YP_i^{ak}] as XM_i^{pk} [YM_i^{ak}].

We shall state definitions B3, B4 and B5 below for the k^p-entities only. The corresponding definitions for the k^a-entities are analogous.

Definition B3: For any entity mixed strategy XM^{pi}, $i=1,N$, the associated entity-behavior strategy is the entity-behavior strategy defined by:

$$X^{pi} = \sum_{k=1}^{\ell^{pi}} XM_k^{pi} \cdot XP_k^{pi}. \tag{B3.1}$$

<u>Definition B4</u>: For any entity-behavior strategy X^{pi}, i=1,N, the <u>associated entity-mixed strategy</u> is the entity-mixed strategy defined by

$$\bar{XM}_k^{pi} = \prod_{\substack{\text{all components } \ell|j \\ \text{such that } (XP_k^{pi})_{\ell|j}=1.0}} (X^{pi})_{\ell|j} \qquad \text{(B3.2)}$$

where the $\ell|j$ component of X^{pi} (or XP_k^{pi}) is the component corresponding to the ℓth alternative in the partition corresponding to the jth information set.

As a direct consequence of definitions B1 - B4, an entity-behavior strategy generates under equation (B3.2) an entity-mixed strategy which in turn generates under equation (B3.1) the original entity-behavior strategy.

<u>Definition B5</u>: A set of entity-mixed strategies (XM^{*pi}, i=1,N; YM^{*ak}, k=1,K) is called an <u>entity-mixed strategy equilibrium point</u> if:

$$\max_{XM^{pj}} H(XM^{*p1},\ldots,XM^{*p(j-1)},XM^{pj},XM^{*p(j+1)},\ldots,XM^{*pN};YM^{*ak},k=1,M)$$

$$\leq H(XM^{*pi},i=1,N;\ YM^{*ak},k=1,M) \leq \min_{YM^{a\ell}} H(XM^{*pi},i=1,N;\ YM^{*a1},$$

$$YM^{*a2},\ldots,YM^{*a(\ell-1)},YM^{a\ell},YM^{*a(\ell+1)},\ldots,YM^{*aM})$$

for j=1,N and ℓ=1,M.

We now proceed with the proof of Theorem 2 by stating and proving the following two lemmas.

Lemma 1: The expected return under any given set of entity-mixed strategies (XM^{pi}, $i=1,N$: YM^{ak}, $k=1,M$) is equal to the expected return under the corresponding set of associated entity-behavior strategies ($\bar{X}^{pi}$, $i=1,N$; $\bar{Y}^{ak}$, $k=1,M$).

Proof: Under (XM^{pi}, $i=1,N$; YM^{ak}, $k=1,M$), the expected return is given by

$$H(XM^{pi},i=1,N;YM^{aj},j=1,M) = \sum_{C}\left[\left(\prod_{i=1}^{N}\left(XM^{pi}_{k_i}\cdot XP^{pi}_{k_i}\right)^{\dagger}\right)\nabla\right.$$

$$\left.\left(\prod_{j=1}^{M}\left(YM^{aj}_{\ell_j}\cdot YP^{aj}_{\ell_j}\right)\right)\nabla F\right] \qquad \text{(B3.3)}$$

where C is the set of all possible integer sequences $(k_1,k_2,\ldots,k_N,$ $\ell_1,\ell_2,\ldots,\ell_M)$ such that $1 \le k_i \le \ell^{pi}$ for $i=1,N$ and $1 \le \ell_j \le \ell^{aj}$ for $j=1,M$. Expression (B3.3), however, can be written as:[††]

$$H(XM^{pi},i=1,N;YM^{aj},j=1,M) = \left(\prod_{i=1}^{N}\left(\sum_{k_i=1}^{\ell^{pi}} XM^{pi}_{k_i}\cdot XP^{pi}_{k_i}\right)\right)\nabla$$

$$\left(\prod_{j=1}^{M}\left(\sum_{\ell_j=1}^{\ell^{aj}} YM^{aj}_{\ell_j}\cdot YP^{aj}_{\ell_j}\right)\right)\nabla F. \qquad \text{(B3.4)}$$

By definition B3, however, each of the strategy terms in the innermost parentheses in expression (B3.4) is the associated entity-behavior strategy for the original entity-mixed strategy which

[†]Note that $XM^{pi}_{k_i}$ is a scalar while $XP^{pi}_{k_i}$ is a vector.

[††]This can be verified by expanding B3.4.

corresponds to it. Thus, we can rewrite (B3.4) as:

$$H(XM^{pi},i=1,N;YM^{aj},j=1,M) = \left(\prod_{i=1}^{N} \bar{X}^{pi}\right) \nabla \left(\prod_{j=1}^{M} \bar{Y}^{aj}\right) \nabla F$$

$$= H(\bar{X}^{pi},i=1,N;\bar{Y}^{aj},j=1,M). \quad \text{(B3.5)}$$

<u>Lemma 2</u>: If any set of entity-mixed strategies is an entity-mixed strategy equilibrium point, then the corresponding set of associated entity-behavior strategies is an entity-behavior equilibrium point.

<u>Proof</u>: Suppose $(XM^{*pi},i=1,N;\ YM^{*aj},j=1,M)$ is an entity-mixed strategy equilibrium. By definition B5, then, for any k=1,N and ℓ=1,M:

$$\max_{XM^{pk}} H(XM^{*p1},XM^{*p2},\ldots,XM^{*p(k-1)},XM^{pk},XM^{*p(k+1)},\ldots,XM^{*pN};$$

$$YM^{*aj},j=1,M) \leq H(XM^{*pi},i=1,N;YM^{*aj},j=1,M) \leq \min_{YM^{a\ell}} \quad \text{(B3.6)}$$

$$H(XM^{*pi},i=1,N;YM^{*a1},YM^{*a2},\ldots,YM^{*a(\ell-1)},YM^{a\ell},YM^{*a(\ell+1)},\ldots,YM^{*aM}).$$

Now, suppose there exists an X^{pk} such that:

$$H(\bar{X}^{*p1},\bar{X}^{*p2},\ldots,\bar{X}^{*p(k-1)},X^{pk},\bar{X}^{*p(k+1)},\ldots,\bar{X}^{*pN};\bar{Y}^{*a\ell},\ell=1,M) >$$

$$H(\bar{X}^{*pi},i=1,N;\bar{Y}^{*aj},j=1,M) \quad \text{(B3.7)}$$

where $(\bar{X}^{*pi},i=1,N;\bar{Y}^{*aj},j=1,M)$ is the set of associated entity-behavior strategies corresponding to the given entity-mixed strategy equilibrium point.

Applying Lemma 1 and definition 14, we can write from equation (B3.7) above, that:

$$H(XM^{*p1}, XM^{*p2}, \ldots, XM^{*p(k-1)}, XM^{pk}, XM^{*p(k+1)}, \ldots, XM^{*pN}; YM^{*aj},$$

$$j=1,M) = H(\bar{X}^{*p1}, \bar{X}^{*p2}, \ldots, \bar{X}^{*p(k-1)}, X^{pk}, \bar{X}^{*p(k+1)}, \ldots, \bar{X}^{*pN}; \bar{Y}^{*aj},$$

$$j=1,M) > H(\bar{X}^{*pi}, i=1,N; \bar{Y}^{*aj}, j=1,M) = H(XM^{*pi}, i=1,N; YM^{*aj}, j=1,M) \tag{B3.8}$$

Expression (B3.8), however, contradicts expression (B3.6) so no such entity-behavior strategy can exist, and we can write:

$$\max_{X^{pk}} H(\bar{X}^{*p1}, \bar{X}^{*p2}, \ldots, \bar{X}^{*p(k-1)}, X^{pk}, \bar{X}^{*p(k+1)}, \ldots, \bar{X}^{*pN}; \bar{Y}^{*pj}, j=1,M)$$

$$\leq H(\bar{X}^{*pi}, i=1,N; \bar{Y}^{*aj}, j=1,M). \tag{B3.9}$$

Similar arguments will establish this same result for any $k=1,N$ and for any $\ell=1,M$ that:

$$\min_{Y^{a\ell}} H(\bar{X}^{*pi}, i=1,N; \bar{Y}^{*a1}, \bar{Y}^{*a2}, \ldots, \bar{Y}^{*a(\ell-1)}, Y^{a\ell}, \bar{Y}^{*a(\ell+1)}, \ldots, \bar{Y}^{*aM})$$

$$\geq H(\bar{X}^{*pi}, i=1,N; \bar{Y}^{*aj}, j=1,M). \tag{B3.10}$$

But by definition, this implies that $(\bar{X}^{*pi}, i=1,N; \bar{Y}^{*aj}, j=1,M)$ is an entity-behavior equilibrium point. ‖

To complete the proof of Theorem 2, we now note that the (N+M)-entity non-cooperative game Γ^{θ} with the entities employing entity-mixed strategies is an "n-person non-cooperative finite game" as defined by Nash.[12] For such games, Nash gives the following theorem.

<u>Theorem(Nash)</u>: Every finite (n-person non-cooperative) game has an equilibrium point (in mixed strategies).

The proof of Theorem 2 of section 6.3 follows directly from Nash's theorem and Lemma 2 above. ▮

4.0 Proof of Property 5 (Section 6.4.1)

Suppose that the protagonist is employing a perfect-recall information collection scheme R^p.

From expression (5.5) we can write for any information set $I^p(k)_\ell$,

$$H((X^{pi}, i=1,N; Y^{aj}, j=1,M) | I^p(k)_\ell) = \Big[\sum_{C_W | I^p(k)_\ell} p(E^p(1)|I^p(1)) \cdot p(E^p(2)|I^p(2)) \ldots p(E^p(N)|I^p(N))\, p(E^a(1)|I^a(1)) p(E^a(2)|I^a(2)) \ldots p(E^a(M)|I^a(M))\, p(E^n(1)|I^n(1)) p(E^n(2)|I^n(2)) \ldots p(E^n(L)|I^n(L)) \cdot h(E^p(i), i=1,N; E^a(j), j=1,M; E^n(m), m=1,L) \Big] \div \Big[\sum_{C_{WI^p(k)_\ell}} p(E^p(1)| I^p(1)) p(E^p(2)|I^p(2)) \ldots p(E^p(k)|I^p(k))\, p(E^a(1)|I^a(1)) p(E^a(2)| I^a(2)) \ldots p(E^a(M_k)|I^a(M_k)) p(E^n(1)|I^n(1)) p(E^n(2)|I^n(2)) \ldots p(E^n(L_k)|I^n(L_k)) \Big] \quad \text{(B4.1)}$$

where:

(i) $C_{W|I^p(k)_\ell}$ indicates summation over all plays W which contain a move in $I^p(k)_\ell$,

(ii) $C_{WI^p(k)_\ell}$ indicates summation over all partial plays leading to a move in $I^p(k)_\ell$,

(iii) M_k is the number of moves for the antagonist on any partial play leading to a move in $I^p(k)_\ell$, and

(iv) L_k is the number of moves for nature on any partial play leading to a move in $I^p(k)_\ell$.

Now, if R^p is a perfect-recall information collection scheme, then $I^p(k)_\ell$ specifies $E^p(i)$, $I^p(i)$ for $i=1,k-1$,[†] so we can rewrite (B4.1) as follows:

$$H((X^{pi},i=1,N;Y^{aj},j=1,M)|I^p(k)_\ell) = \left[\left[\prod_{n=1}^{k-1} p(E^p(n)|I^p(n))\right] \cdot\right.$$

$$\left[\sum_{C_{W|I^p(k)_\ell}} p(E^p(k)|I^p(k))\ldots p(E^p(N)|I^p(N))p(E^a(1)|I^a(1))\ldots\right.$$

$$p(E^a(M)|I^a(M))p(E^n(1)|I^n(1))\ldots p(E^n(L)|I^n(L)) \cdot h(E^p(i),$$

$$\left.i=1,N;E^a(j),j=1,M;E^n(m),m=1,L)\right] \div \left[\left[\prod_{n=1}^{k-1} p(E^p(n)|I^p(n))\right] \cdot\right.$$

$$\left[\sum_{C_{W|I^p(k)_\ell}} p(E^p(k)|I^p(k))p(E^a(1)|I^a(1))\ldots p(E^a(M_k)|I^a(M_k)) \cdot\right.$$

[†] For every information set containing any move on any partial play leading to a move in $I^p(k)_\ell$.

$$p(E^n(1)|I^n(1))\ldots p(E^n(L_k)|I^n(L_k))\Big] \qquad (B4.2)$$

Now, noting that all terms defined by $(X^{p1},\ldots,X^{p(k-1)})$ cancel out, we can write:

$$H((X^{pi}, i=1,N;Y^{aj}, j=1,M)|I^p(k)_\ell) - H((X^{*pi}, i=1,k-1;X^{pm}, m=k,N;$$

$$Y^{aj}, j=1,M)|I^p(k)_\ell) = 0 \quad \text{for any } (X^{pi}, i=1,N;X^{*pi}, i=1,k-1;$$

$$Y^{aj}, j=1,M).$$

But similar arguments can be carried out for any information set $I^p(k)_\ell$ for k=2,N. By definition then, $\delta^{pk}=0$ for k=2,N and thus the recall sensitivity of the protagonist about any pair of behavior strategies $(X^{pi}, i=1,N)$ on R^p and $(Y^{aj}, j=1,M)$ on R^a is zero if R^p is a perfect-recall information collection scheme. ▮

5.0 Proof of Theorem 4 (Section 6.4.2)

Suppose that the pair of behavior strategies $(X^{*pi}, i=1,N)$ on R^p and $(Y^{*aj}, j=1,M)$ on C^a corresponds to an entity-behavior equilibrium point. Taking the protagonist's part, we can then write:

$$\max_{X^{pN}} H(X^{*pi}, i=1,N-1;X^{pN};Y^{*aj}, j=1,M) = \max_{X^{pN}} \Big[\sum_{I^p(N)} H((X^{*pi}, i=1,N-1;$$

$$X^{pN};Y^{*aj}, j=1,M)|I^p(N)_\ell)p(I^p(N)_\ell;(X^{*pi}, i=1,N-1;X^{pN};Y^{*aj}, j=1,M))\Big]$$

$$\leq H(X^{*pi}, i=1,N;Y^{*aj}, j=1,M) \qquad (B5.1)$$

where $p(I^p(N)_\ell;\ (X^{*pi}, i=1,N-1; X^{pN}; Y^{*aj}, j=1,M))$ is the probability that information set $I^p(N)_\ell$ occurs, as a function of the behavior strategies employed by the players.

By the definition of recall-sensitivity, we can write:

$$\max_{X^{p1}} \cdots \max_{X^{p(N-1)}} H((X^{pi}, i=1,N; Y^{*aj}, j=1,M) | I^p(N)_\ell) - H((X^{*pi}, i=1,N-1; X^{pN}; Y^{*aj}, j=1,M) | I^p(N)_\ell) \leq \delta^{pN}/2 \quad \text{for all } I^p(N)_\ell \ \underline{\text{and all } X^{pN}}\ . \qquad (B5.2)$$

Combining (B5.1) and (B5.2), and noting that since X^{pN} is a function on $I^p(N)$, the summation and maximization operations can be interchanged, we obtain:

$$\sum_{I^p(N)} \Big[\max_{X^{p1}} \cdots \max_{X^{pN}} H((X^{pi}, i=1,N; Y^{*aj}, j=1,M) | I^p(N)_\ell)\ p(I^p(N)_\ell; (X^{*pi}, i=1,N-1; X^{pN}; Y^{*aj}, j=1,M))\Big] \leq H(X^{*pi}, i=1,N; Y^{*aj}, j=1,M) + \delta^{pN}/2\ . \qquad (B5.3)$$

Now, $H(X^{*pi}, i=1,N;\ Y^{*aj}, j=1,M)$ can also be expressed as a sum over $I^p(N)$, and by causality $p(I^p(k);\ (X^{pi}, i=1,N;\ Y^{aj}, j=1,M))$ is not dependent upon $(X^{pi}, i=k,N)$ or $(Y^{aj}, j=m_k,M)$ where m_k is the number of moves for the antagonist on any partial play leading to a move in $I^p(k)$, so (B5.3) can be rewritten as:

$$\sum_{I^p(N)} \Big[\Big[\max_{X^{p1}} \cdots \max_{X^{pN}} H((X^{pi}, i=1,N; Y^{*aj}, j=1,M) | I^p(N)_\ell) -$$

$$H((X^{*pi}, i=1,N; Y^{*aj}, j=1,M) | I^p(N)_\ell) \Big] \, p(I^p(N)_\ell; \; (X^{*pi}, i=1,N-1;$$

$$Y^{*aj}, j=1,m_N)) \Big] \leq \delta^{pN}/2 \; . \qquad (B5.4)$$

Now, $p(I^p(N)_\ell; \; (X^{*pi}, i=1,N-1; \; Y^{*aj}, j=1,M_N))$ takes its values in $[0,1]$, and the inner bracketed term in (B5.4) is positive or zero so we can write from (B5.4):

$$\Big[\max_{X^{p1}} \cdots \max_{X^{pN}} H((X^{pi}, i=1,N; Y^{*aj}, j=1,M) | I^p(N)_\ell) - H((X^{*pi},$$

$$i=1,N; Y^{*aj}, j=1,M) | I^p(N)_\ell \Big] \leq \delta^{pN}/2 \; . \qquad (B5.5)$$

Multiplying (B5.5) by $p(I^p(N)_\ell; \; (X^{pi}, i=1,N-1; \; Y^{*aj}, j=1,m_N))$ and summing over $I^p(N)$, we obtain:

$$\sum_{I^p(N)} \Big[\max_{X^{p1}} \cdots \max_{X^{pN}} \; H((X^{pi}, i=1,N; Y^{*aj}, j=1,M) | I^p(N)_\ell) \cdot p(I^p(N)_\ell;$$

$$(X^{pi}, i=1,N-1; Y^{*aj}, j=1,m_N)) \Big] \leq \; \sum_{I^p(N)} \Big[H((X^{*pi}, i=1,N; Y^{*aj}, j=1,M) |$$

$$I^p(N)_\ell) \; p(I^p(N)_\ell; \; (X^{pi}, i=1,N-1; Y^{*aj}, j=1,m_N)) \Big] + \delta^{pN}/2. \qquad (B5.6)$$

Now, reversing the maximization and summation operations in (B5.6) either has no effect on this inequality or "strengthens" it. Further, by the definition of recall-sensitivity, we have:

$$|H((X^{*pi}, i=1,N; Y^{*aj}, j=1,M) | I^p(N)_\ell) - H((X^{pi}, i=1,N-1; X^{*pN};$$

$$Y^{*aj}, j=1,M) | I^p(N)_\ell)| \leq \delta^{pN}/2 \ . \qquad (B5.7)$$

Combining (B5.6) and (B5.7), we obtain:

$$\max_{X^{p1}} \dots \max_{X^{pN}} H(X^{pi}, i=1,N; Y^{*aj}, j=1,M) \leq H(X^{pi}, i=1,N-1; X^{*pN};$$

$$Y^{*aj}, j=1,M) \leq \delta^{pN} \quad \text{for any } (X^{pi}, i=1,N-1) \ . \qquad (B5.8)$$

Starting with the equation corresponding to (B5.1), but for maximization over $X^{p(N-1)}$, we can repeat the above arguments to obtain:

$$\max_{X^{p1}} \dots \max_{X^{pN}} \ H(X^{pi}, i=1,N-1; X^{*pN}; Y^{*aj}, j=1,M) \leq H(X^{pi}, i=1,N-2;$$

$$X^{*pk}, k=N-1,N; Y^{*aj}, j=1,M) + \delta^{p(N-1)} \text{ for any } (X^{pi}, i=1,N-2). \quad (B5.9)$$

By similar arguments, we can obtain for $\ell=2,N$ that:

$$\max_{X^{p1}} \dots \max_{X^{p\ell}} \ H(X^{pi}, i=1,\ell; X^{*pk}, k=\ell+1,N; Y^{*aj}, j=1,M) \leq H(X^{pi},$$

$$i=1,\ell-1; X^{*pk}, k=\ell,N; Y^{*aj}, j=1,M) + \delta^{p\ell} \text{ for any } (X^{pi}, i=1,\ell-1). (B5.10)$$

Now, combining the expression corresponding to expression (B5.1) but for maximization over X^{p1}, with expression (B5.10) for $\ell=2$, we obtain:

$$\max_{X^{p1}} \max_{X^{p2}} H(X^{p1}, X^{p2}; X^{*pi}, i=3,N; Y^{*aj}, j=1,M) \leq H(X^{p1}; X^{*pi}, i=2,N;$$

$$Y^{*aj}, j=1,M) + \delta^{p2} \leq \max_{X^{p1}} H(X^{p1};X^{*pi}, i=2,N;Y^{*aj}, j=1,M) + \delta^{p2}$$

$$\leq H(X^{*pi}, i=1,N;Y^{*aj}, j=1,M) + \delta^{p2} \qquad (B5.11)$$

Now, combining (B5.11) and (B5.10) for $\ell=3$, we obtain:

$$\max_{X^{p1}} \cdots \max_{X^{p3}} H(X^{p1},X^{p2},X^{p3};X^{*pi}, i=4,N;Y^{*aj}, j=1,M) \leq H(X^{*pi},$$

$$i=1,N;Y^{*aj}, j=1,M) + \delta^{p2} + \delta^{p3}. \qquad (B5.12)$$

Combining for $\ell=4,\ldots,N$, we finally obtain:

$$\max_{X^{p1}} \cdots \max_{X^{pN}} H(X^{pi}, i=1,N;Y^{*aj}, j=1,M) \leq H(X^{*pi}, i=1,N;$$

$$Y^{*aj}, j=1,M) + \sum_{i=2}^{N} \delta^{pi}. \qquad (B5.13)$$

But the summation term is the protagonist's recall-sensitivity about $(X^{*pi}, i=1,N)$ on R^p and $(Y^{*aj}, j=1,M)$ on C^a. Denoting this term as $\in$, we can write from (B5.13):

$$\max_{\sigma B^p} H(\sigma B^p,R^p;\sigma B^{*a},C^a) \leq H(\sigma B^{*p},R^p;\sigma B^{*a},C^a) + \in . \qquad (B5.14)$$

Now, establishing an analogous definition for the recall-sensitivity of the antagonist, we can carry out for the antagonist, arguments of the same type as those given above for the protagonist, obtaining:

$$\min_{\sigma B^a} H(\sigma B^{*p},R^p;\sigma B^a,C^a) \geq H(\sigma B^{*p},R^p;\sigma B^{*a},C^a) - \rho \qquad (B5.15)$$

where ρ is the recall-sensitivity of the antagonist about (X^{*pi}, $i=1,N$) on R^p and (X^{*aj}, $j=1,M$) on C^a.

Finally then, noting that the antagonist is employing a perfect-recall information collection scheme and thus that ρ is zero by property 5, and noting that the left hand side of (B5.15) is the quality of σB^{*p} on R^p, we can combine (B5.14) and (B5.15) to obtain:

$$Q(\sigma B^{*p},R^p) = \min_{\sigma B^a} H(\sigma B^{*p},R^p;\sigma B^a,C^a) = H(\sigma B^{*p},R^p;\sigma B^{*a},C^a)$$

$$\leq \max_{\sigma B^p} H(\sigma B^p,R^p;\sigma B^{*a},C^a) \leq H(\sigma B^{*p},R^p;\sigma B^{*a},C^a) + \epsilon \quad \| \qquad (B5.16)$$

6.0 Proof of Corollary 1 (Section 6.4.2)

Let $\emptyset$ be the quality of an R-solution. Let σB^{*p} on R^p and σB^{*a} on C^a be a pair of behavior strategies

(i) which corresponds to an entity-behavior equilibrium point in the associated non-cooperative game Γ', and

(ii) about which the protagonist has a recall-sensitivity which is less than ϵ.

From Theorem 4, then, we have:

$$\max_{\sigma B^p} H(\sigma B^p,R^p;\sigma B^{*a},C^a) \leq Q(\sigma B^{*p},R^p;\sigma B^{*a},C^a) + \epsilon \,. \qquad (B6.1)$$

From Von Neumann's minmax theorem (expression B2.1) we have:

$$\emptyset \leq \max_{\sigma B^p} H(\sigma B^p,R^p;\sigma B^{*a},C^a) \,. \qquad (B6.2)$$

Combining (B6.1) and (B6.2) and noting the definition of an R-solution, we obtain:

$$\phi - \epsilon \leq G(\sigma B^{*p}, R^{p}; \sigma B^{*a}, C^{a}) \leq \phi \ . \qquad (B6.3)$$

By definition, then, σB^{*p} on R^{p} is an ϵ-R-solution. ▮

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) School of Electrical Engineering Purdue University Lafayette, Indiana 47907	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

ON BEHAVIOR STRATEGY SOLUTIONS IN FINITE EXTENDED DECISION PROCESSES

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Scientific Report

5. AUTHOR(S) (Last name, first name, initial)

Utgoff, Victor A.
Kashyap, Rangasami L.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
June 1970	178 + 12	38

8a. CONTRACT OR GRANT NO. AFOSR 69-1776 b. PROJECT NO. c. d	9a. ORIGINATOR'S REPORT NUMBER(S) TR-EE 70-24 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. AVAILABILITY/LIMITATION NOTICES

Unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Air Force Office of Scientific Research

13. ABSTRACT

The problem of determination of best behavior strategies for a decision maker in a two-person, zero-sum, finite extended decision process is considered.

In this problem, best strategies are defined to be those strategies which maximize a player's expected return against all possible strategies that can be employed by the other player, and the decision process is thus considered as a two-person, zero-sum, finite extended game.

In the report then, techniques are developed for finding best behavior strategies on arbitrary information collection schemes. These techniques make it possible to search in any given game for strategies which are minimally complex but acceptably effective.

A doctor-patient medical game is "searched" using these techniques, and behavior strategies are found for the doctor which are more than 90% as effective but less than 10% as complex as the best strategy which results when the doctor gathers and retains all the information defined to be available to him.

A pursuit and evasion game is also studied and best behavior strategies are determined for the pursuer and the evader when both players gather and retain all the information defined to be available to each of them.

DD FORM 1 JAN 64 1473

UNCLASSIFIED

Security Classification

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT

DD FORM 1 NOV 65 1473 (BACK)

S/N 0101-807-6821

Security Classification

A-31409