On Stochastic Games with Stationary Optimal Strategies
Author(s): Truman Bewley and Elon Kohlberg
Source: *Mathematics of Operations Research*, Vol. 3, No. 2 (May, 1978), pp. 104-125
Published by: INFORMS
Stable URL: http://www.jstor.org/stable/3689337
Accessed: 13/02/2014 06:20

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

INFORMS is collaborating with JSTOR to digitize, preserve and extend access to *Mathematics of Operations Research.*

http://www.jstor.org

This content downloaded from 164.15.117.233 on Thu, 13 Feb 2014 06:20:14 AM
All use subject to JSTOR Terms and Conditions

MATHEMATICS OF OPERATIONS RESEARCH
Vol. 3, No. 2, May 1978
Printed in U.S.A.

ON STOCHASTIC GAMES WITH STATIONARY OPTIMAL STRATEGIES*

TRUMAN BEWLEY AND ELON KOHLBERG

Harvard University

We study two-person zero-sum stochastic games in which the state and action spaces are finite. We give both necessary and sufficient conditions for the players to have stationary optimal strategies in the infinite-stage game.

1. Introduction. In this paper, we apply results from previous work [3] to study two-person, zero-sum, undiscounted, infinite-stage, stochastic games in which the state and action spaces are finite and both players have stationary optimal strategies. The theory of such games is much simpler than that of stochastic games in general. The study of general, undiscounted, infinite-stage stochastic games remains quite incomplete. For instance, it is not known whether such games always have a value, even if the state and action spaces are finite. (We study the theory of stochastic games with finite state and action spaces in [2] and [3].)

There are many possible definitions of the value of an undiscounted, infinite-stage stochastic game. For instance, the definition of value depends on how the players evaluate a stream of payoffs. Two possible evaluations of the stream, $(\pi_1, \pi_2, \ldots)$ are

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \pi_n \quad \text{and} \quad \liminf_{r\to 0} \frac{r}{1+r} \sum_{n=1}^{\infty} \frac{1}{(1+r)^{n-1}} \pi_n.$$

In §4, we give a number of definitions of the value. In §5, we show that they are all equivalent for games in which both players have stationary optimal strategies.

In §6, we describe our main theorems. These theorems give necessary or sufficient conditions for the existence of stationary optimal strategies. The theorems are proved in §§10 through 16. §7 contains auxiliary results. §9 is devoted to examples.

In §8, we relate our work to previous work on Markov decision processes. In §17, we discuss previous work on stochastic games.

There have been many papers giving sufficient conditions for the existence of stationary optimal strategies in games for which the total payoff is unbounded. In §17, we show that all of these conditions follow from our theorems.

In the following two sections, we review some notation, definitions, and results from our previous paper [3]. Those familiar with that paper may skip these sections.

2. Definitions and notation. A *stochastic game* is played in stages. At each stage, the game is in one of the finitely many states, $s = 1, \ldots, S$. Each of the players observes the current state and then chooses one of finitely many actions. The actions of player 1 are indexed by $i = 1, \ldots, I$; those of player 2 by $j = 1, \ldots, J$. If player 1 chooses action i and player 2 chooses j, then a payoff of $\gamma_{ij;\,s}$ is made by player 2 to player 1 and transition to state s' occurs with probability $p_{ij;\,ss'}$, $s' = 1, \ldots, S$. If play stops after N stages, the game is said to be an *N-stage game*. Otherwise, the game is said to be an *infinite-stage game*.

* Received July 6, 1976; revised September 23, 1977.
AMS 1970 subject classification: Primary: 90D05, 90D15, 93E05. Secondary: 90C40.
IAOR 1973 subject classification: Main: Games. Cross References: Markov decision programming.
Key words: Game theory, stochastic games, minmax value, stationary optimal strategies, Markov decision processes.

Copyright © 1978, The Institute of Management Sciences

This content downloaded from 164.15.117.233 on Thu, 13 Feb 2014 06:20:14 AM
All use subject to JSTOR Terms and Conditions

A player's *strategy* is a specification of a probability distribution over his actions at each stage conditional on the current state and the history of the game up to that stage. If this distribution depends only on the current state, then the strategy is said to be *stationary*. A general strategy for player 1 will usually be denoted by q. If q is stationary, it will be understood that q also denotes the vector $(q_1, \ldots, q_S)$, where q_s is a probability vector with I components and is the probability distribution of player 1's actions in state s. Similarly, a general strategy for player 2 will usually be denoted by z and, if z is stationary, z will also denote $(z_1, \ldots, z_S)$, where z_s is the probability distribution of player 2's actions in state s.

R^n denotes n-dimensional Euclidean space.

If $x \in R^S$, $G_s(x)$ denotes the $I \times J$ matrix whose (i, j)th entry is

$$G_s(x)_{ij} = \gamma_{ij;\,s} + \sum_{s'=1}^{S} p_{ij;\,ss'} x_{s'}, \qquad s = 1, \ldots, S. \tag{2.1}$$

If A is a matrix, val A denotes the minmax value of A. val $G(x)$ denotes the vector $(\text{val } G_1(x), \ldots, \text{val } G_S(x)) \in R^S$.

$V_{N,s}$ denotes the minmax value of the N-stage game with initial state s, and V_N denotes $(V_{N,1}, \ldots, V_{N,S})$. The dynamic programming equation for V_N is

$$V_{N+1} = \text{val } G(V_N).$$

This equation will be referred to as the *recursion equation*.

If the payoffs are discounted at interest rate r, the infinite-stage game is called the *r-discount game*. Shapley [20] proved that this game has a value and that the players have optimal stationary strategies. $V_s(r)$ denotes the value of the r-discount game with initial state s and $V(r)$ denotes $(V_1(r), \ldots, V_S(r))$. The dynamic programming equation for $V(r)$ is

$$V(r) = \text{val } G\left(\frac{V(r)}{1 + r} \right).$$

The equation

$$x = \text{val } G\left(\frac{x}{1 + r} \right)$$

will be referred to as the *r-discount equation*.

F will denote the ordered field $\{\sum_{k=-\infty}^{K} a_k \theta^{k/M} \mid K$ is an integer, M is a positive integer, the a_k are real numbers, and the series $\sum_{k=-\infty}^{K} a_k t^{k/M}$ converges for all sufficiently large real numbers $t\}$. F is referred to as the field of real Puiseux series. $\sum_{k=-\infty}^{K} a_k \theta^{k/M} > 0$ if and only if $a_N > 0$, where N is the largest integer k such that $a_k \neq 0$.

F^n will denote the n-fold Cartesian product of F with itself.

If A is a matrix with entries in F, the minmax value of A exists and belongs to F (Weyl [22]). This value will be denoted by val A. If $x \in F^S$, $G_s(x)$ and val $G(x)$ are defined as when $x \in R^S$.

If $A = \sum_{k=-\infty}^{K} a_k \theta^{k/M} \in F$ and t is a real number, $\sigma_t A$ denotes the number $\sum_{k=-\infty}^{K} a_k t^{k/M}$. $\sigma_t A$ is well defined if t is sufficiently large. If $A = (a_{ij})$ is a matrix with entries in F, $\sigma_t A$ denotes $(\sigma_t a_{ij})$.

The equation

$$x = \text{val } G\left(\frac{x}{1 + \theta^{-1}} \right), \quad \text{where } x \in F^S,$$

will be called the *limit discount equation*.

This content downloaded from 164.15.117.233 on Thu, 13 Feb 2014 06:20:14 AM
All use subject to JSTOR Terms and Conditions

$\tau : F \to F$ is defined by

$$\tau\left(\sum_{k=-\infty}^{K} a_k \theta^{k/M}\right) = \sum_{k=-\infty}^{K} a_k (\theta + 1)^{k/M}, \tag{2.2}$$

where $(\theta + 1)^{k/M}$ is defined to be $\theta^{k/M} + k\theta^{(k/M)-1}/M + \cdots \in F$. It is not hard to show that after rearranging terms in (2.2), one obtains an element of F. Clearly, if $A \in F$, then

$$\sigma_n \tau A = \sigma_{n+1} A \tag{2.3}$$

for all n so large that $\sigma_n A$ and $\sigma_{n+1} A$ are defined.

If $x \in F^S$, τx denotes $(\tau x_1, \ldots, \tau x_S)$.

The equation

$$\tau x = \text{val } G(x), \quad \text{where } x \in F^S,$$

will be called the *limit recursion equation*.

If $A = \sum_{k=-\infty}^{K} a_k \theta^{k/M} \in F$, the *valuation* of A will be defined to be N/M, where N is the largest integer k such that $a_k \neq 0$.

The expression $o(\theta^\alpha)$ (resp., $O(\theta^\alpha)$) will be used to denote an element or vector of elements of F of valuation less than α (resp., at most α).

If $A \in F$, the absolute value of A will be defined to be $\max(A, -A)$, and will be denoted by $|A|$. If $A \in F^n$, $|A|$ will denote $\max(|A_1|, \ldots, |A_n|)$.

3. Previous results. The following is a list of results from a previous paper [3] that we use here.

The limit discount equation, $x = \text{val } G(x/(1 + \theta^{-1}))$, has a unique solution in F^S. The components of this solution are of valuation at most one. (3.1)

If $x \in F^S$ is the solution of the limit discount equation, then $V(r) = \sigma_{r^{-1}} x$, for $r > 0$ sufficiently small. (3.2)

$\lim_{r \to 0} rV(r)$ and $\lim_{N \to \infty} N^{-1} V_N$ exist and are equal to $a \in R^S$, where the solution to the limit discount equation is $x = a\theta + o(\theta)$. (3.3)

If A is a matrix with entries in F, $\sigma_t \text{val } A = \text{val } \sigma_t A$, if t is sufficiently large. If ξ is a strategy for A with entries in F, ξ is optimal in A if and only if $\sigma_t \xi$ is optimal in $\sigma_t A$ for t sufficiently large. (3.4)

The function val $G(\cdot)$ is Lipschitz with constant one with respect to the norm $|\cdot|$. (3.5)

If $x \in F^S$ solves the limit discount equation, then $y = x/(1 + \theta^{-1})$ satisfies $\tau y = \text{val } G(y) + o(1)$. (3.6)

4. Definitions of the value. The undiscounted infinite-stage game is a model for a game which is to be played for a large but unspecified number of stages and in which there is no reason to discount future payoffs. In such games, it is not at all clear what criterion the players should use to compare strategies.

One class of criteria, which we call *limit criteria*, are appropriate for situations in which the players believe that they are playing the game indefinitely and that they should not discount the payoffs. These criteria specify how the players evaluate the probability distributions over infinite streams of payoffs resulting from both their strategies.

This content downloaded from 164.15.117.233 on Thu, 13 Feb 2014 06:20:14 AM
All use subject to JSTOR Terms and Conditions

Another class of criteria, which we call the *asymptotic criteria*, is appropriate for situations in which the players believe they are playing an N-stage game or an r-discount game but they are not sure how large N is or how small r should be. An asymptotic criterion evaluates a strategy according to the expected payoff per stage which it can guarantee uniformly in N, for large N, or uniformly in r for small positive r.

We now describe the limit criteria in detail. A choice of strategy by each player and a specification of the initial state jointly determine a probability distribution over the sequences of states and actions, and hence over sequences of payoffs. Let $D_{q,z,s}$ denote the probability distribution over sequences of payoffs to player 1, when the strategies chosen are q and z and the initial state is s.

A limit criterion is specified by choice of an evaluation, $W(D)$, of probability distributions, D, over sequences of payoffs. Such an evaluation is a real valued function. Given W, a strategy q *guarantees* b_s to player 1 in the game with initial state s if $\inf_z W(D_{q,z,s}) \geqslant b_s$. $W(D_{q,z})$ will denote $(W(D_{q,z,1}), \ldots, (W(D_{q,z,S}))$. $\inf_z W(D_{q,z})$ will denote the vector whose sth component is $\inf_z W(D_{q,z,s})$. q will be said to guarantee $b \in R^S$ to player 1 if $\inf_z W(D_{q,z}) \geqslant b$. The symmetric definition applies to player 2.

We now give a list of the evaluations which we study. We also give the names of the criteria corresponding to these evaluations. In what follows, $E_{q,z,s}$ will denote expectation with respect to the distribution $D_{q,z,s}$. The vector $(E_{q,z,1}, \ldots, E_{q,z,S})$ will be denoted by $E_{q,z}$. If b_n is a sequence of vectors in R^S, $\liminf_{n\to\infty} b_n$ will denote the vector whose sth component is $\liminf_{n\to\infty} b_{ns}$. π_n will denote the payoff at stage n.

$$W(D_{q,z}) = E_{q,z} \liminf_{r\to 0} \frac{r}{1+r} \sum_{n=1}^{\infty} \frac{1}{(1+r)^{n-1}} \pi_n,$$

the *limit discount criterion*. (4.1)

$$W(D_{q,z}) = \liminf_{r\to 0} E_{q,z} \frac{r}{1+r} \sum_{n=1}^{\infty} \frac{1}{(1+r)^{n-1}} \pi_n,$$

the *limit expected discount criterion*. (4.2)

$$W(D_{q,z}) = E_{q,z} \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \pi_n,$$

the *limit average criterion*. (4.3)

$$W(D_{q,z}) = \liminf_{N\to\infty} E_{q,z} \frac{1}{N} \sum_{n=1}^{N} \pi_n,$$

the *limit expected average criterion*. (4.4)

The asymptotic criteria remain to be described. According to the *asymptotic discount criterion*, a strategy q guarantees the vector $b \in R^S$ to player 1 if

$$\liminf_{r\to 0} \inf_z E_{q,z} \frac{r}{1+r} \sum_{n=1}^{\infty} \frac{1}{(1+r)^{n-1}} \pi_n \geqslant b. \tag{4.5}$$

According to the *asymptotic average criterion*, q guarantees $b \in R^S$ to player 1 if

$$\liminf_{N\to\infty} \min_z E_{q,z} \frac{1}{N} \sum_{n=1}^{N} \pi_n \geqslant b. \tag{4.6}$$

This content downloaded from 164.15.117.233 on Thu, 13 Feb 2014 06:20:14 AM
All use subject to JSTOR Terms and Conditions

The symmetric definitions apply to player 2. For instance, z guarantees $b \in R^S$ to player 2 according to the asymptotic discount criterion if

$$\limsup_{r\to\infty} \sup_q E_{q,z} \frac{r}{1+r} \sum_{n=1}^{\infty} \frac{1}{(1+r)^{n-1}} \pi_n \leqslant b.$$

When both players use a certain criterion to choose their strategies, the game will be given the name of that criterion. In this way, we obtain the *limit discount game*, the *limit expected discount game*, and so forth. $v \in R^S$ is said to be the *value* of the game with a given criterion and q and z are said to be *optimal strategies* in this game if q guarantees v to player 1 according to the given criterion, and z guarantees v to player 2 according to the same criterion.

REMARK. The above definition of value is strong in the sense that both players adopt "pessimistic" points of view (lim inf for player 1, lim sup for player 2). In fact, if any of the criteria defined above are considered to be payoffs, then the sum of the payoffs to players 1 and 2 could be less than zero. (Consider, for example, the game with one state and with payoff matrix $\boxed{1}\boxed{0}$. Player 2 has a strategy that yields him a payoff of -1 and yields player 1 a payoff of 0.)

5. Equivalence of the notions of value. In this section, we show that $\lim_{N\to\infty} N^{-1}V_N$ is a natural bound for what either player can achieve with a stationary strategy in any of the versions of the undiscounted infinite-stage game. ($\lim_{N\to\infty} N^{-1}V_N$ exists by (3.3).) Also, if either player can guarantee this bound with a stationary strategy in one version of the game, this strategy guarantees it to him in every version.

v_∞ will denote $\lim_{N\to\infty} N^{-1}V_N$.

THEOREM 5.1. *Player* 1 *cannot guarantee more than* v_∞ *in either the asymptotic discount or the asymptotic average games, with any strategy, stationary or nonstationary.*

PROOF. We give the proof for the asymptotic average game. The proof for the asymptotic discount game is the same.

Let q be any strategy for player 1. Clearly,

$$\min_z E_{q,z} \sum_{n=1}^{N} \pi_n \leqslant V_N.$$

It follows that

$$\liminf_{N\to\infty} \min_z E_{q,z} \frac{1}{N} \sum_{n=1}^{N} \pi_n \leqslant \lim_{N\to\infty} \frac{1}{N} V_N = v_\infty.$$

THEOREM 5.2. *If a stationary strategy, q, guarantees $v \in R^S$ to player* 1 *in one of the games listed below, then q guarantees v to player* 1 *in all of the other games as well*:

(a) *limit discount*;
(b) *limit expected discount*;
(c) *limit average*;
(d) *limit expected average*;
(e) *asymptotic discount*;
(f) *asymptotic average*.

The proof of theorem 5.2 is given in the appendix.

COROLLARY 5.3. *If in one of the games* (*a*)–(*f*), *both players have stationary optimal strategies, then these strategies are optimal in all the games and the value of each of these games is* v_∞.

This content downloaded from 164.15.117.233 on Thu, 13 Feb 2014 06:20:14 AM
All use subject to JSTOR Terms and Conditions

Hereafter, we will call a pair of stationary strategies *optimal in the undiscounted infinite game* if they are optimal in any of the games (a)–(f).

6. Main theorems. A strategy is *uniformly discount optimal* if it is optimal in the r-discount game for all $r > 0$ sufficiently small. In what follows, a real vector is a vector with real components.

THEOREM 6.1. *Let $x \in F^S$ be the solution of the limit discount equation. A player has a uniformly discount optimal strategy in the stochastic game if and only if he has real strategies optimal in the matrices $G_s(x/(1 + \theta^{-1}))$, $s = 1, \dots, S$. Furthermore, a stationary strategy $q = (q_1, \dots, q_S)$ is uniformly discount optimal if and only if q_s is optimal in the matrix $G_s(x/(1 + \theta^{-1}))$, for all s.*

As before, let $V_{N,s}$ be the value of the N-stage stochastic game with initial state s. A strategy q for player 1 is said to be *uniformly N-stage optimal* if the sequence $V_{N,s} - \min_z E_{q,z,s}\sum_{n=1}^{N} \pi_n$ is bounded, for $s = 1, \dots, S$. It is easy to see that if both players have strategies which are uniformly N-stage optimal, then these strategies are optimal in the undiscounted infinite game.

In what follows, if $A = (a_{ij})$ is an $I \times J$ matrix with entries in F, we say that $\xi \in F^I$ guarantees a payoff $b \in F$ to player 1 in A if $\min_j \sum_{i=1}^{I} \xi_i a_{ij} \geqslant b$.

THEOREM 6.2. *Let $x \in F^S$ be the solution of the limit discount equation.* (i) *If players* 1 *and* 2 *both have real strategies, q_s and z_s respectively, which guarantee* val $G_s(x/(1 + \theta^{-1})) + o(1)$ *in the matrices $G_s(x/(1 + \theta^{-1}))$, for $s = 1, \dots, S$, then the stationary strategies $q = (q_1, \dots, q_S)$ and $z = (z_1, \dots, z_S)$ are uniformly N-stage optimal.* (ii) *If players* 1 *and* 2 *both have stationary strategies, $q = (q_1, \dots, q_S)$ and $z = (z_1, \dots, z_S)$ respectively, which are optimal in the undiscounted infinite-stage game, then q_s and z_s guarantee them* val $G_s(x/(1 + \theta^{-1})) + O(1)$ *in $G_s(x/(1 + \theta^{-1}))$, for all s.*

COROLLARY 6.3. *If both players have stationary uniformly discount optimal strategies, then these strategies are uniformly N-stage optimal and hence optimal according to the criteria* (a)–(f) *listed in Theorem* 5.2.

Let $R(\theta)$ be the field of rational functions of θ with real coefficients (a rational function of θ is a ratio of two polynomials in θ). Clearly, $R(\theta)$ may be viewed as a proper subfield of F. We will assume that $R(\theta)$ has the order inherited from F. The expansion of an element of $R(\theta)$ is of the form $\sum_{-\infty}^{K} a_k\theta^k$. That is, it contains no fractional powers of θ. Let $R(\theta)^S$ denote the S-fold Cartesian product of $R(\theta)$ with itself.

THEOREM 6.4. *Let $x \in F^S$ be the solution of the limit discount equation. If both players have uniformly discount optimal strategies, then $x \in R(\theta)^S$.*

Applying (3.2) and (3.3), we have:

COROLLARY 6.5. *If both players have uniformly discount optimal strategies, then $V(r)$ has an expansion of the form*

$$V(r) = v_\infty r^{-1} + \sum_{k=0}^{\infty} a_k r^k,$$

where $a_R \in R^S$, valid for all $r > 0$ sufficiently small.

THEOREM 6.6. *Let $x \in F^S$ be the solution of the limit discount equation. If both players have stationary optimal strategies in the undiscounted infinite game, then x must be of the form $v_\infty\theta + O(1)$.*

This content downloaded from 164.15.117.233 on Thu, 13 Feb 2014 06:20:14 AM
All use subject to JSTOR Terms and Conditions

7. The generalized Howard equation. The limit recursion equation, $\tau x = \text{val } G(x)$, may have no solution. In general, there is no finite algorithm either for finding a solution or for verifying that one does not exist. The reason for this is that the operator τ is not elementary (see [2]). If, however, one restricts attention to solutions of the form $v\theta + w$, the limit recursion equation reduces to the following elementary equation:

$$v\theta + v + w = \text{val } G(v\theta + w), \tag{7.1}$$

where $v, w \in R^S$. We will refer to (7.1) as the *generalized Howard equation*. Trying to solve the generalized Howard equation is a natural first step in a search for a solution of the limit recursion equation.

THEOREM 7.2. *Let $x \in F^S$ be the solution of the limit discount equation. If players* 1 *and* 2 *have real strategies, q_s and z_s, respectively, which guarantee* val $G_s(x/(1 + \theta^{-1})) + o(1)$ *in $G_s(x/(1 + \theta^{-1}))$, $s = 1, \ldots, S$, then the following are true.*

(i) *$v_\infty \theta + w$ solves the generalized Howard equation, where $x/(1 + \theta^{-1}) = v_\infty \theta + w + o(1)$.*

(ii) *q_s and z_s are optimal in $G_s(v_\infty \theta + w)$, $s = 1, \ldots, S$.*

REMARK. That $x/(1 + \theta^{-1})$ is of the form $v_\infty \theta + w + o(1)$ follows from theorems 6.2 and 6.6.

THEOREM 7.3. *Let $v\theta + w$ solve the generalized Howard equation. If the real strategies, q_s and z_s, are optimal for players* 1 *and* 2, *respectively, in $G_s(v\theta + w)$, $s = 1, \ldots, S$, then*

(i) *$q = (q_1, \ldots, q_S)$ and $z = (z_1, \ldots, z_S)$ are uniformly N-stage optimal.*

(ii) *$V_n = nv + O(1)$.*

8. Markov decision processes. In this section, we show how our results give some insight into the theory of Markov decision processes.

Recall that a Markov decision process is simply a one-person stochastic game. This one person is usually referred to as the *decision maker*, and his strategies are called *policies*. We will think of the decision maker as player 1, that is, as the maximizing player.

In the case of Markov decision processes, our conditions for optimality apply trivially. These conditions all require that the players have real strategies which are optimal or nearly optimal in the matrices $G_s(x)$ for particular values of x. In Markov decision processes, the matrices $G_s(x)$ are simply column vectors, so that the decision maker may restrict himself to pure strategies which are, of course, real. The corollary below summarizes the application of our optimality conditions. The results listed in this corollary are well known.

Payoffs and transition probabilities will now be denoted by γ_{is} and $p_{iss'}$, respectively. A policy is called *deterministic* if it does not involve randomization. A deterministic policy may be described by integers $(i_1, \ldots, i_S)$, where i_s is the action specified for state s. Such a policy will be said to be optimal in $G(x)$ if for all s, $G_s(x)_{i_s} = \max_{1 \leq i \leq I} G_s(x)_i$, where $G_s(x)_i$ is the ith component of the column vector $G_s(x)$.

The limit discount equation may be expressed as

$$x_s = \max_{1 \leq i \leq I} G_s\left(\frac{x}{1 + \theta^{-1}}\right)_i, \qquad s = 1, \ldots, S. \tag{8.1}$$

Similarly, the generalized Howard equation (7.1) may be written as follows.

8.2. *Howard's Equations*

$$v_s \theta + w_s + v_s = \max_{1 \leq i \leq I} G_s(v\theta + w)_i, \qquad s = 1, \ldots, S,$$

This content downloaded from 164.15.117.233 on Thu, 13 Feb 2014 06:20:14 AM
All use subject to JSTOR Terms and Conditions

where $v \in R^S$ and $w \in R^S$. Howard's equations may be rewritten in a form familiar in the theory of dynamic programming (Howard [11, p. 61]),

$$(v_s, v_s + w_s) = \operatorname*{lex\,max}_{1 \leqslant i \leqslant I} \left(\sum_{s'} p_{iss'} v_{s'}, \gamma_{is} + \sum_{s'} p_{iss'} w_{s'} \right),$$

where $\text{lex max}_{1 \leqslant i \leqslant I}\,(a_i, b_i)$ is the lexicographic maximum of $\{(a_i, b_i)\}_{i=1}^{I}$.

COROLLARY 8.3.

(i) *The limit discount equation* (8.1) *has a unique solution*, $x \in R(\theta)^S$.

(ii) *Any deterministic policy optimal in* $G_s(x/(1 + \theta^{-1}))$, $s = 1, \ldots, S$, *describes a stationary deterministic policy which is uniformly discount optimal and hence uniformly N-stage optimal and optimal according to the criteria* (a)–(f) *listed in theorem* 5.2.

(iii) $v_\infty \theta + w$ *solves Howard's equations, where* $x/(1 + \theta^{-1}) = v_\infty \theta + w + o(1)$.

(iv) *Let* $v\theta + w$ *be any solution to Howard's equations. Then, any deterministic policy optimal in* $G_s(v\theta + w)$, $s = 1, \ldots, S$ *describes a stationary deterministic policy which is uniformly N-stage optimal and hence optimal according to the criteria* (a)–(f) *listed in theorem* 5.2.

(v) $V(r)$ *has an expansion of the form* $V(r) = v_\infty r^{-1} + \sum_{k=0}^{\infty} a_k r^k$, *where* $a_k \in R^S$, *valid for all* $r > 0$ *sufficiently small.*

(vi) $V_n = nv_\infty + O(1)$.

Blackwell [4] proved that there exists a uniformly discount optimal policy. Blackwell [4] also proved that there exists a solution to a form of the limit discount equation. He proved that the solution of the r-discount equation is a rational function of r in an interval $(0, \epsilon)$ where $\epsilon > 0$. Howard [11] proved that there exists a solution to his equations (8.2). He also proved that any policy which maximizes the right-hand side of these equations is optimal in the limit expected average sense. Brown [6, theorem 4.2] proved that any policy which maximizes the right-hand side of Howard's equations is uniformly N-stage optimal.

REMARK. Miller and Veinott [18] present an algorithm which yields a uniformly discount optimal policy. Perhaps an alternative to their algorithm may be found by trying to solve the system of linear inequalities (8.1), using the theory of linear inequalities over ordered fields (see, for example, Kuhn [14] and Weyl [23]). Once the solution is found, (ii) of the above corollary may be applied.

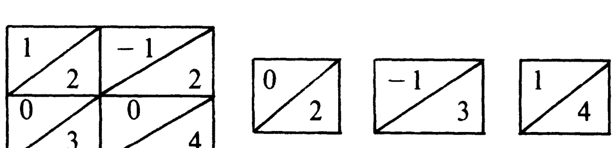
9. Examples. In what follows, we will use the following formula for the computation of the value.

$$\text{val}\, A = \frac{\det A}{\sum_{i,j} (-1)^{i+j} \det A_{ij}} \tag{9.1}$$

where A_{ij} is the minor obtained by deleting the ith row and jth column. This formula is valid for games which are completely mixed, that is, in which the players have optimal strategies, q and z, such that $q_i > 0$ for all i and $z_j > 0$ for all j. Formula (9.1) is proved in Karlin [12, p. 50].

The first example shows that the sufficient condition of theorem 6.2 is not necessary for the existence of stationary uniformly N-stage optimal strategies.

9.2



State 1 State 2 State 3 State 4

This content downloaded from 164.15.117.233 on Thu, 13 Feb 2014 06:20:14 AM
All use subject to JSTOR Terms and Conditions

The notation

γ / s

means that if the players choose the row and column corresponding to this box, then player 2 pays player 1 the amount γ and the next state is s.

The vector $x = (0, 0, -\theta(1+\theta^{-1}), \theta(1+\theta^{-1}))$ is the solution of the limit discount equation. Hence,

$$G_1\left(\frac{x}{1+\theta^{-1}}\right) = \begin{pmatrix} 1 & -1 \\ -\theta & \theta \end{pmatrix}.$$

Clearly, player 1 can guarantee only -1 in $G_1(x/(1+\theta^{-1}))$ by use of a real strategy. Since val $G_1(x/(1+\theta^{-1})) = 0$, the condition of theorem 6.2 is not satisfied. On the other hand, the players obviously have stationary strategies which are uniformly N-stage optimal.

The next example shows that the converse to both theorem 6.4 and to theorem 6.6 is not true.

9.3

State 1

1 / 1	0 / 1
0 / 2	0 / 3

State 2

0 / 2

State 3

1 / 3

Let x be the solution of the limit discount equation, and let $y = x/(1+\theta^{-1})$. y satisfies the equations $(1+\theta^{-1})y_s = \text{val } G_s(y)$, for $s = 1, 2, 3$. Clearly, $y_2 = 0$ and $y_3 = \theta$. Hence, the first equation is

$$(1+\theta^{-1})y_1 = \text{val}\begin{pmatrix} 1+y_1 & y_1 \\ 0 & \theta \end{pmatrix}.$$

Since $0 \leqslant y_1 \leqslant \theta$, the above matrix is completely mixed. We may therefore use formula (9.1) to obtain

$$(1+\theta^{-1})y_1 = \frac{\theta(1+y_1)}{\theta+1}$$

which implies that $y_1 = \theta/(2+\theta^{-1})$. Thus the components of y, and hence x, lie in $R(\theta)$. However, in this example, player 1 has no optimal strategy in the undiscounted infinite game. This fact was proved by Blackwell and Ferguson [5], who call this game the "Big Match."

The following example of Gillette [9] shows that a game with stationary uniformly N-stage optimal strategies may not have uniformly discount optimal strategies.

9.4

State 1

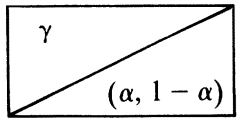
1 / $\left(\frac{1}{3}, \frac{2}{3}\right)$	0 / $\left(\frac{2}{3}, \frac{1}{3}\right)$
0 / $\left(\frac{1}{3}, \frac{2}{3}\right)$	1 / $\left(\frac{2}{3}, \frac{1}{3}\right)$

State 2

0 / $\left(\frac{1}{2}, \frac{1}{2}\right)$

This content downloaded from 164.15.117.233 on Thu, 13 Feb 2014 06:20:14 AM
All use subject to JSTOR Terms and Conditions

The notation is the same as above except that



means that the payoff is γ and the next state is 1 with probability α and 2 with probability $1-\alpha$.

In this example, the solution to the limit discount equation is $x = ((\theta+2)/4, \theta/4)$ and

$$G_1\left(\frac{x}{1+\theta^{-1}}\right) = \begin{bmatrix} 1-\epsilon+\dfrac{\theta}{4} & \epsilon+\dfrac{\theta}{4} \\ -\epsilon+\dfrac{\theta}{4} & 1+\epsilon+\dfrac{\theta}{4} \end{bmatrix},$$

where $\epsilon = 1/(12(1+\theta^{-1}))$. The unique optimal strategy for player 1 in $G_1(x/(1+\theta^{-1}))$ is

$$\left(\frac{1}{2}+\frac{1}{12(1+\theta^{-1})}, \frac{1}{2}-\frac{1}{12(1+\theta^{-1})}\right).$$

Therefore, player 1 does not have a uniformly discount optimal strategy (theorem 6.1). However, this game is irreducible. Irreducible games are defined in §17, where it is shown that in such games, both players have stationary uniformly N-stage optimal strategies.

10. Proof of Theorem 6.1. Shapley [20] proved that

$$q = (q_1, \ldots, q_S) \text{ is optimal in the } r\text{-discount game if and only if for all } s,\ q_s \text{ is optimal in } G_s(V(r)/(1+r)). \tag{10.1}$$

Hence,

$$q = (q_1, \ldots, q_S) \text{ is uniformly discount optimal} \tag{10.2}$$

if and only if for all s,

$$q_s \text{ is optimal in } G_s(V(r)/(1+r)), \text{ for all } r>0 \text{ sufficiently small.} \tag{10.3}$$

If $r>0$ is sufficiently small, $V(r) = \sigma_{r^{-1}}(x)$ (by 3.2). Since q_s is real, $\sigma_{r^{-1}}q_s = q_s$. Hence, (10.3) may be rewritten as follows:

$$\sigma_{r^{-1}}q_s \text{ is optimal in } \sigma_{r^{-1}}G_s\big(x/(1+\theta^{-1})\big), \text{ for } r>0 \text{ sufficiently small.}$$

This statement is equivalent to the following (by 3.4).

$$q_s \text{ is optimal in } G_s\big(x/(1+\theta^{-1})\big). \tag{10.4}$$

The equivalence between (10.2) and (10.4) proves the second half of theorem 6.1. To prove the first half, observe that if $\hat{q}$ is uniformly discount optimal and if q_s is the randomization specified by $\hat{q}$ in stage 1 and state s, then q_s must be optimal in $G_s(V(r)/(1+r))$, when $r>0$ is small. It follows from (10.1) that the stationary strategy $q \equiv (q_1, \ldots, q_S)$ is uniformly discount optimal. Q.E.D.

11. Proof of Theorem 6.4. By theorem 6.1, there are real strategies, q_s and z_s, optimal for players 1 and 2, respectively, in the matrices $G_s(x/(1+\theta^{-1}))$, $s = 1, \ldots, S$. Since x solves the limit discount equation, we have

$$x_s = q_s G_s\big(x/(1+\theta^{-1})\big) z_s, \qquad s = 1, \ldots, S. \tag{11.1}$$

This content downloaded from 164.15.117.233 on Thu, 13 Feb 2014 06:20:14 AM
All use subject to JSTOR Terms and Conditions

The solution of (11.1) is unique, for if y and $\bar{y}$ in F^S are solutions, then

$$|y - \bar{y}| = \max_s \left| q_s G_s\left(\frac{y}{1 + \theta^{-1}} \right) z_s - q_s G_s\left(\frac{\bar{y}}{1 + \theta^{-1}} \right) z_s \right| \leqslant \frac{y - \bar{y}}{1 + \theta^{-1}} .$$

Thus, x is the unique solution of a system of linear equations with coefficients in $R(\theta)$, and hence $x \in R(\theta)^S$. Q.E.D.

12. Proof of Theorem 6.6. By Corollary 5.3, the players have stationary strategies, q and z respectively, optimal in the asymptotic discount game. That is,

$$\liminf_{r \to 0} \frac{r}{1 + r} W_1(r) \geqslant \limsup_{r \to 0} \frac{r}{1 + r} W_2(r), \tag{12.1}$$

where $W_1(r)$ is the value of the r-discount Markov decision process that player 2 faces when player 1 chooses strategy q, and $W_2(r)$ is the value of the corresponding process for player 1. (The payoffs resulting from actions in this process are random variables. The theory of such Markov decision processes is the same as for processes with deterministic payoffs.)

Let W_1 and W_2 be the solutions to the limit discount equations of the above processes. Then, by (3.3), (12.1) may be rewritten as

$$W_1 \geqslant W_2 + o(\theta). \tag{12.2}$$

Obviously, $W_1(r) \leqslant V(r) \leqslant W_2(r)$. By (3.2), $\sigma_{r^{-1}} W_1 \leqslant \sigma_{r^{-1}} x \leqslant \sigma_{r^{-1}} W_2$ when $r > 0$ is small. Hence,

$$W_1 \leqslant x \leqslant W_2. \tag{12.3}$$

By Corollary 8.3 (i), W_1 and W_2 belong to $R(\theta)^S$. Hence, (12.2) and (12.3) imply

$$W_1 + O(1) = x = W_2 + O(1). \tag{12.4}$$

Since W_1 is of the form $v\theta + O(1)$, so is x. Q.E.D.

13. Proof of a lemma.

LEMMA 13.1. *Let $x \in F^S$ be the solution of the limit discount equation. If the real strategies, q_s and z_s, guarantee players* 1 *and* 2, *respectively*, val $G_s(x/(1 + \theta^{-1})) + o(1)$ *in* $G_s(x/(1 + \theta^{-1}))$, $s = 1, \ldots, S$, *then* $q = (q_1, \ldots, q_S)$ *and* $z = (z_1, \ldots, z_S)$ *are optimal in the undiscounted infinite game.*

PROOF. By theorem 5.2, it is sufficient to show that

$$\liminf_{N \to \infty} \min_z E_{q, z}\left(\frac{1}{N} \sum_{n=1}^{N} \pi_n \right) \geqslant v_\infty. \tag{13.2}$$

Let $y = x/(1 + \theta^{-1})$. Then the assumption of the lemma is that, for all $s = 1, \ldots, S$, val $q_s G_s(y) =$ val $G_s(y) + o(1)$. (If a is a row vector, then val a is simply the minimum of its components.) Also, by (3.6), $\tau y =$ val $G(y) + o(1)$. It follows that

$$\tau y = \text{val } qG(y) + o(1), \tag{13.3}$$

where val $qG(y) \equiv ($val $q_1 G_1(y), \ldots,$ val $q_S G_S(y))$. Applying σ_n to (13.3) and using (2.3) and (3.4), we obtain that, for some $\beta > 0$,

$$|\sigma_{n+1} y - \text{val } qG(\sigma_n y)| < n^{-\beta}, \tag{13.4}$$

for sufficiently large n.

This content downloaded from 164.15.117.233 on Thu, 13 Feb 2014 06:20:14 AM
All use subject to JSTOR Terms and Conditions

Let

$$m_N = \min_z E_{q,z}\left(\sum_{n=1}^{N} \pi_n \right).$$

Clearly,

$$m_{n+1} = \text{val } qG(m_n). \tag{13.5}$$

Combining (13.4) and (13.5), and using the fact that val $qG(\cdot)$ is Lipschitz with constant one, we obtain $|m_{n+1} - \sigma_{n+1} y| < |\text{val } qG(m_n) - \text{val } qG(\sigma_n y)| + n^{-\beta} \leqslant |m_n - \sigma_n y| + n^{-\beta}$. Adding over n, we find that for some $B > 0$,

$$|m_N - \sigma_N y| < B + \frac{N^{1-\beta}}{1-\beta}, \quad \text{for all } N.$$

Hence,

$$\liminf_{N\to\infty} \frac{m_N}{N} = \liminf_{N\to\infty} \frac{\sigma_N y}{N}.$$

That is,

$$\liminf_{N\to\infty} \min_z E_{q,z}\left(\frac{1}{N} \sum_{n=1}^{N} \pi_n \right) = \lim_{N\to\infty} \frac{\sigma_N y}{N} = \lim_{N\to\infty} \frac{\sigma_N x}{N} = v_\infty,$$

where the last equality follows from (3.3). This proves lemma 13.1. Q.E.D.

14. Proof of Theorem 7.2. By Lemma 13.1, the stationary strategies $q = (q_1, \ldots, q_S)$ and $z = (z_1, \ldots, z_S)$ are optimal in the undiscounted infinite game. Hence by Theorem 6.6, x is of the form $v_\infty \theta + O(1)$. It follows that $x/(1+\theta^{-1})$ is of the form $v_\infty \theta + w + o(1)$, where w is a vector with real components. Substituting $v_\infty \theta + w + o(1)$ for $x/(1+\theta^{-1})$ in the limit discount equation, we obtain

$$(1+\theta^{-1})(v_\infty \theta + w + o(1)) = \text{val } G(v_\infty \theta + w + o(1)).$$

Since val $G_s(\cdot)$ is Lipschitz with constant one, this equation becomes

$$v_\infty \theta + v_\infty + w = \text{val } G(v_\infty \theta + w) + o(1). \tag{14.1}$$

Since $v_\infty \theta + v_\infty + w = \tau(v_\infty \theta + w)$, (14.1) may be expressed as

$$\tau(v_\infty \theta + w) = \text{val } G(v_\infty \theta + w) + o(1).$$

Clearly, since q_s guarantees val $G_s(x/(1+\theta^{-1})) + o(1)$ in $G_s(x/(1+\theta^{-1}))$, q_s guarantees val $G_s(v_\infty \theta + w) + o(1)$ in $G_s\ (v_\infty \theta + w)$. Also, the payoff that q_s guarantees in $G_s\ (v_\infty \theta + w)$ must be of the form $a_s \theta + b_s$, where a_s and b_s are real numbers. Letting $a = (a_1, \ldots, a_S)$ and $b = (b_1, \ldots, b_S)$, we have

$$a\theta + b = \text{val } G(v_\infty \theta + w) + o(1). \tag{14.2}$$

Comparing (14.2) with (14.1), we obtain $a = v_\infty$ and $b = v_\infty + w$. A similar argument proves that z_s guarantees player 2 the payoff $v_{\infty s} \theta + v_{\infty s} + w_s$ in $G_s(v_\infty \theta + w)$. Hence, $v_{\infty s} \theta + v_{\infty s} + w_s = \text{val } G_s(v_\infty \theta + w)$ and q_s and z_s are optimal in $G_s(v_\infty \theta + w)$. The last equation may be expressed as

$$\tau(v_\infty \theta + w) = \text{val } G(v_\infty \theta + w). \quad \text{Q.E.D.}$$

15. Proof of Theorem 7.3. Let $x = v\theta + w$. Since $\tau x = \text{val } G(x)$, it follows from (2.3) and (3.4) that $\sigma_{n+1} x = \sigma_n \tau x = \sigma_n \text{val } G(x) = \text{val } G(\sigma_n x)$, for large n. Also, V_{n+1}

This content downloaded from 164.15.117.233 on Thu, 13 Feb 2014 06:20:14 AM
All use subject to JSTOR Terms and Conditions

$= \text{val } G(V_n)$ for all n. It follows that the tails of the sequences $\sigma_n x$ and V_n satisfy the same recursion equation. Since the function val $G(\cdot)$ is Lipschitz with constant one (see (3.5)), $|V_n - \sigma_n x|$ does not increase when n is large. Hence,

$$\text{the sequence } V_n - \sigma_n x \text{ is bounded.} \tag{15.1}$$

This proves Part (ii) of theorem 7.3. We now prove Part (i).

Let $m_N = \min_z E_{q,z} \sum_{n=1}^{N} \pi_n$. Then,

$$m_{n+1} = \text{val } qG(m_n), \tag{15.2}$$

where val $qG(\cdot)$ is as in (13.3). Also, $\tau x = \text{val } G(x) = \text{val } qG(x)$, where the second equality follows from the fact that q_s is optimal in $G_s(x)$, $s = 1, \ldots, S$. Thus, by (2.3) and (3.4),

$$\sigma_{n+1} x = \text{val } qG(\sigma_n x), \quad \text{when } n \text{ is sufficiently large.} \tag{15.3}$$

Since val $qG(\cdot)$ is Lipschitz with constant one, it follows from (15.2) and (15.3) that the sequence $m_n - \sigma_n x$ is bounded. Hence, by (15.1), the sequence $m_n - V_n$ is bounded, which proves that q is uniformly N-stage optimal. Q.E.D.

16. Proof of Theorem 6.2 By theorem 7.2,

$$\tau(v_\infty \theta + w) = \text{val } G(v_\infty \theta + w),$$

and q_s and z_s are optimal in $G_s(v_\infty \theta + w)$, for all s, where $x/(1 + \theta^{-1}) = v_\infty \theta + w + o(1)$. Hence, by theorem 7.3, q and z are uniformly N-stage optimal. This completes the proof of Part (i) of the theorem.

Let q be a stationary strategy optimal in the undiscounted infinite game and let W be the solution to the limit discount equation of the Markov decision process that player 2 faces when player 1 chooses q. Then

$$W_s = \text{val } q_s G_s\big(W/(1 + \theta^{-1})\big), \qquad s = 1, \ldots, S.$$

By (12.4), $W - x = O(1)$, hence

$$x_s = \text{val } q_s G_s\big(x/(1 + \theta^{-1})\big) + O(1).$$

Since x solves the limit discount equation, it follows that

$$\text{val } q_s G_s\Big(\frac{x}{1 + \theta^{-1}}\Big) = x_s + O(1) = \text{val } G_s\Big(\frac{x}{1 + \theta^{-1}}\Big) + O(1).$$

That is, q_s guarantees val $G_s(x/(1 + \theta^{-1})) + O(1)$ in $G_s(x/(1 + \theta^{-1}))$. Q.E.D.

17. Previous work on stochastic games. The criteria of optimality we use are all designed for games in which the total payoff is unbounded. For this reason, we deal with limits of averages. If the total payoff is bounded, there is no need to use such averages. Therefore, the theory of games with bounded total payoff is quite different from that which we give. This theory has been fully explored by Shapley [20] and Everett [8] for games with a finite state space. Shapley's results have been extended to games with an infinite state space by Maitra and Parthasarathy [16], [17].

The main contributions to the theory of stochastic games with unbounded total payoffs and a finite state space have been made by Gillette [9], Liggett and Lippman [15], Hoffman and Karp [10], Stern [21], and Kohlberg [13]. All of the above authors, except Kohlberg, prove that in special cases of stochastic games there exist stationary strategies which are optimal in the limit average sense. We now show that all the

This content downloaded from 164.15.117.233 on Thu, 13 Feb 2014 06:20:14 AM
All use subject to JSTOR Terms and Conditions

existence theorems for these special cases may be derived from results in this paper. We also discuss the work of Kohlberg.

Games with perfect information. Gillette [9] and Liggett and Lippman [15] prove that if a stochastic game has perfect information, then both players have stationary optimal strategies in the undiscounted infinite game. This result follows immediately from theorem 6.1. In a stochastic game, perfect information means that at each state, one of the players is restricted to one action. In this case, G_s has only one column or one row, for each s. As in the case of Markov decision processes, the players may restrict themselves to pure strategies in G_s and theorem 6.1 may be applied. In fact, corollary 8.3 holds with the words "Howard's equations" replaced by "generalized Howard's equations."

Irreducible games. Gillette [9] dealt with another class of stochastic games which he called cyclic. Hoffman and Karp [10] also studied these games, calling them irreducible. A game is said to be irreducible if when both players adopt stationary strategies, then the resulting transition probabilities on states determine an irreducible Markov process, no matter what the stationary strategies may be. (Recall that a Markov process is said to be irreducible if any state may be reached from any other state in a finite number of steps with positive probability.)

The irreducibility assumption removes the primary difficulty in the theory of stochastic games with unbounded total payoffs. The player in a stochastic game must, in choosing actions, weigh the current payoff against different growth rates of the payoff which could result from transition to other states. That is, the players must compare finite objects with infinite ones. This difficulty disappears if the rates of growth are all the same. It is quite easy to see that in irreducible games the rates of growth are, in fact, the same. To be more specific, one can easily prove that in an irreducible game $\lim_{N\to\infty} (V_{N,s}/N)$ exists and does not depend on the initial state. From this fact, it follows at once that the asymptotic average value exists. To see this, consider the following strategy for player 1 in the infinite-stage game. In stage 1, he plays optimally in the one-stage game. In the next two stages, he plays optimally in the two-stage game, given the state at stage 2. This brings him to stage 4. In the next three stages, he plays optimally in the three-stage game, given the state at stage 4. Continuing in this way, player 1 guarantees

$$C_N = \frac{\sum_{n=1}^{N} n \min_s (V_{n,s}/n)}{1 + 2 + \cdots + N}$$

in the first $1 + 2 + \cdots + N$ stages. Since all the sequences $V_{n,s}/n$, $s = 1, \ldots, S$, converge to a common limit, say v, it follows that C_N converges to v. A similar argument proves that player 2 can guarantee v in the asymptotic average sense. (The strategy described above is an adaptation of the one described in Aumann and Maschler [1, p. 76].)

The simple nature of irreducible stochastic games follows directly from the following fact.

$$\max_{1 \leq s \leq S} |V_s(r) - V_1(r)| \text{ is uniformly bounded in } r > 0. \tag{17.1}$$

In order to prove (17.1), we may assume (by adding a sufficiently large constant to all payoffs) that $\gamma_{ij;s} > 0$ for all i, j, and s. Since the game is irreducible, there is a number T with the following property. Given initial state s and any state s' and given any strategies for the players, the mean number of stages to the first occurrence of s'

This content downloaded from 164.15.117.233 on Thu, 13 Feb 2014 06:20:14 AM
All use subject to JSTOR Terms and Conditions

is not larger than T. Let q and z be optimal strategies for players 1 and 2, respectively, in the game with payoffs discounted at interest rate r. Let N be the random variable which is the stage at which state 1 first occurs. Let $\gamma = \max_{ij;\, s} |\gamma_{ij;\, s}|$. Then

$$\begin{aligned} V_s(r) &\leq E_{q,z,s}\left(\sum_{n=1}^{N} \pi_n + \left(\frac{1}{1+r}\right)^N V_1(r)\right) \\ &\leq E_{q,z,s}\left(N\gamma + \left(\frac{1}{1+r}\right)^N V_1(r)\right) \\ &\leq T\gamma + V_1(r). \end{aligned}$$

Hence, $V_s(r) - V_1(r) \leq T\gamma$. A symmetric argument proves $V_1(r) - V_s(r) \leq T\gamma$. This proves (17.1). This proof is a modification of an argument in Ross [19, p. 149].

Irreducible stochastic games satisfy the generalized Howard equations (7.1). In fact, they satisfy the following, more specialized equations obtained from (7.1) by adding the requirement $v_s = v_{s'}$, for all s, s'.

$$u + w_s = \text{val } G_s(w), \qquad s = 1, \ldots, S, \text{ where } u \in R \text{ and } w \in R^S. \tag{17.2}$$

Proposition 17.3 (Hoffman and Karp [10]). *If a game is irreducible, then there exists a solution to* (17.2).

Proposition 17.3 may be derived from (17.1) by elementary methods. (One need only adapt the arguments of Ross [19, pp. 146–149] to the context of games.) We present an alternative proof.

Proof. Let $x \in F^S$ be the solution of the limit discount equation and let $x/(1 + \theta^{-1}) = \sum_{k=-\infty}^{M} a_k \theta^{k/M}$. (By (3.1), M is the upper limit of summation.) Since, by (3.2), $V(r) = \sigma_{r^{-1}} x$ for $r > 0$ sufficiently small, it follows from (17.1) that if $k > 0$ then $a_{ks} = a_{ks'}$ for all s and s'. That is, $a_k = \alpha_k e$ where α_k is a real number and $e = (1, \ldots, 1) \in R^S$. Let $\alpha = \sum_{k=1}^{M} \alpha_k \theta^{k/M}$. Then the limit discount equation becomes

$$(1 + \theta^{-1})(\alpha e + a_0 + o(1)) = \text{val } G(\alpha e + a_0 + o(1)). \tag{17.4}$$

Observe that, for any $y \in F^S$ and $\alpha \in F$

$$\text{val } G(y + \alpha e) = \text{val } G(y) + \alpha e. \tag{17.5}$$

Applying (17.5) to equation (17.4), we obtain

$$\theta^{-1}\alpha e + (1 + \theta^{-1})(a_0 + o(1)) = \text{val } G(a_0 + o(1)).$$

Since $\theta^{-1}\alpha e = \alpha_M e + o(1)$, we obtain

$$\alpha_M e + a_0 = \text{val } G(a_0) + o(1).$$

Since both $\alpha_M e + a_0$ and val $G(a_0)$ are real, we must have

$$\alpha_M e + a_0 = \text{val } G(a_0).$$

Setting $u = \alpha_M$ and $w = a_0$, we obtain a solution to (17.2). Q.E.D.

Clearly, any strategy which is optimal in $G_s(w)$ is also optimal in $G_s(ue\theta + w)$. Hence, the players have real optimal strategies in $G_s(ue\theta + w)$ and Theorem 7.3 applies.

Corollary 17.6. *In an irreducible game, the players have stationary strategies which are uniformly N-stage optimal and hence optimal in the undiscounted infinite game.*

This content downloaded from 164.15.117.233 on Thu, 13 Feb 2014 06:20:14 AM
All use subject to JSTOR Terms and Conditions

This corollary is a slight extension of a result of Hoffman and Karp [10], which states that the players have stationary strategies which are optimal in the undiscounted infinite game in the limit expected average sense.

Games with transition probabilities controlled by one player. We say that player 1 controls the transition probabilities if $p_{ij;\,ss'}$ does not depend on j, for all i, s and s'. Stern [21, theorem 2.8] proved that in this case both players have optimal stationary strategies in the undiscounted infinite game. We show that this result follows from theorem 6.2.

Let $A = (a_{ij})$ be a matrix with entries in F. We claim that

$$\text{if } a_{ij} = \alpha_i + O(1) \text{ for all } i \text{ and for all } j, \text{ where } |\alpha_i| > O(1) \text{ or } \alpha_i = 0, \text{ then the players have real strategies which guarantee them val } A + o(1). \tag{17.7}$$

To see this, let $\hat{A}$ be the submatrix of A consisting of those rows, i, for which $\alpha_i = \max_j \alpha_j \equiv \alpha$. Clearly, $\operatorname{val} A = \operatorname{val} \hat{A} = \alpha + \operatorname{val}(B + o(1))$, where B is a real matrix. It is easy to see that a real strategy optimal in B guarantees val $A + o(1)$ in A.

If $p_{ij;\,ss'}$ does not depend on j, then $\sum_{s'=1}^{S} p_{ij;\,ss'} x_{s'}$ does not depend on j and hence

$$\gamma_{ij;\,s} + \frac{1}{1+\theta^{-1}} \sum_{s'=1}^{S} p_{ij;\,ss'} x_s{}' \equiv G_s\left(\frac{x}{1+\theta^{-1}} \right)_{ij}$$

varies by at most $O(1)$ as j varies. By (17.7), the players have real strategies which guarantee val $G_s(x/(1+\theta^{-1})) + o(1)$, and theorem 6.2 applies.

Repeated games with absorbing states. A *repeated game with absorbing states* is a stochastic game in which every state but one is absorbing. A state s is said to be absorbing if for all i and j, $p_{ij;\,ss'} = 1$ if $s' = s$ and equals zero otherwise.

Kohlberg [13] proved that in every repeated game with absorbing states, $N^{-1}V_N$ converges. This is, of course, a special case of (3.3).

The method of [13] can be interpreted in terms of F. Let state 1 be the nonabsorbing state, let v_s be the value per stage of the game in the absorbing state s, $s = 2, \ldots, S$. It is shown in [13] that $\lim_{N\to\infty} N^{-1}V_{N,1}$ equals the point at which the function $\Delta(u)$ crosses zero, where

$$\Delta(u) = \lim_{n\to\infty} \left(\operatorname{val} G_1(un, v_2 n, \ldots, v_s n) - (1 + n^{-1})un\right).$$

The main difficulty in the proof arises from the fact that $\Delta(u)$ may jump across zero. The use of the field F eliminates this problem. $\Delta(u)$ may be interpreted as the real part of val $G_1(u\theta, v_2\theta, \ldots, v_s\theta) - (1+\theta^{-1})u\theta$. This function may be discontinuous. However, if we replace $\Delta(u)$ by val $G_1(x, v_2\theta, \ldots, v_s\theta) - (1+\theta^{-1})x$ and allow x to vary over F, we, in effect, fill in the gaps in the graph of $\Delta(u)$. The solution of the limit discount equation is, of course, the value of x at which this function is equal to zero.

18. Open problems. There remain many open problems in the theory of undiscounted infinite-stage stochastic games. Four are listed below.

PROBLEM 1. Do both players have ϵ-optimal strategies in every stochastic game? That is, does every stochastic game have a value?

The players are said to have *ϵ-optimal strategies* and $v \in R^S$ is said to be the *value* of the game if for every $\epsilon > 0$ and for all s, player 1 (resp., 2) has a strategy which guarantees him $v_s - \epsilon$ (resp., $v_s + \epsilon$), if the initial state is s. The sense of "guarantee" could be any of those given in §4.

PROBLEM 2. Does any strategy guarantee player 1 more than v_∞ in some stochastic game, where $v_\infty = \lim_{N\to\infty} N^{-1}V_N$?

This content downloaded from 164.15.117.233 on Thu, 13 Feb 2014 06:20:14 AM
All use subject to JSTOR Terms and Conditions

Theorem 5.1 answers this question for the asymptotic average and asymptotic discount games.

PROBLEM 3. What characterizes games in which both players have optimal strategies?

PROBLEM 4. What characterizes games in which both players have stationary optimal strategies?

Theorem 6.2 gives an incomplete answer to this question.

Blackwell and Ferguson [5] proved that in the Big Match (9.3), player 1 has an ϵ-optimal strategy but no optimal strategy. The ϵ-optimal strategy is not stationary. Kohlberg [13] proves that every repeated game with absorbing states has a value. (See the previous section for a definition of repeated games with absorbing states.)

Stern [21] gives an example of a game in which both players have optimal strategies but one player has no stationary optimal strategy.

The most challenging problem seems to be the first. The following example illustrates some of the difficulties involved. We do not know if this game has a value.

0 / 2	−1 / 1
0 / 3	0 / 2

State 1

0 / 1	0 / 4
1 / 2	0 / 1

State 2

−1 / 3

State 3

1 / 4

State 4

The solution to the limit discount equation for this example is of the form $(-\theta^{1/2}/2, \theta^{1/2}/2, -\theta, \theta) + o(\theta^{1/2})$. It follows that $\lim_{N\to\infty}(1/N)V_N = (0, 0, -1, 1)$. The optimal strategies for players 1 and 2 in $G_1(x/(1+\theta^{-1}))$ are of the form $(1-\theta^{-1/2}, \theta^{-1/2}) + o(\theta^{-1/2})$ and $(\theta^{-1/2}, 1-\theta^{-1/2}) + o(\theta^{-1/2})$, respectively. In $G_2(x/(1+\theta^{-1}))$, the strategies are $(\theta^{-1/2}, 1-\theta^{-1/2}) + o(\theta^{-1/2})$ and $(1-\theta^{-1/2}, \theta^{-1/2}) + o(\theta^{-1/2})$, respectively. The real parts of these strategies guarantee at most -1 to player 1 and at most 1 to player 2 in the undiscounted infinite game with initial state 1 or 2. In order to do better, player 1 must try to force a shift to state 2 when the play is in state 1 and player 2 must try to force a shift to state 1 when play is in state 2.

Appendix. Much of Theorem 5.2 is scattered in Derman [7]. We state and prove it in order to avoid any possible confusion.

PROOF OF THEOREM 5.2. The pattern of the proof is (a) $\Rightarrow$ (b) $\Rightarrow$ (e) $\Rightarrow$ (f) $\Rightarrow$ (d) $\Rightarrow$ (c) $\Rightarrow$ (a), where "$(x) \Rightarrow (y)$" means "if q guarantees v in game (x) then q guarantees v in game (y)." The only hard step is the proof that (d) $\Rightarrow$ (c).

(a) $\Rightarrow$ (b). By Fatou's Lemma,

$$\inf_z \liminf_{r\to 0} E_{q,z} \frac{r}{1+r} \sum_{n=1}^{\infty} \frac{1}{(1+r)^{n-1}} \pi_n$$

$$\geq \inf_z E_{q,z} \liminf_{r\to 0} \frac{r}{1+r} \sum_{n=1}^{\infty} \frac{1}{(1+r)^{n-1}} \pi_n \geq v.$$

That is, q guarantees v to player 1 in the limit expected discount game.

This content downloaded from 164.15.117.233 on Thu, 13 Feb 2014 06:20:14 AM
All use subject to JSTOR Terms and Conditions

(b)$\Rightarrow$ (e). Let z^* be a uniformly discount optimal strategy for player 2 in the Markov decision process that he faces when player 1 chooses strategy q. We have

$$\begin{aligned}
&\liminf_{r\to 0} E_{q,z^*} \frac{r}{1+r} \sum_{n=1}^{\infty} \frac{1}{(1+r)^{n-1}} \pi_n \\
&\geqslant \inf_z \liminf_{r\to 0} E_{q,z} \frac{r}{1+r} \sum_{n=1}^{\infty} \frac{1}{(1+r)^{n-1}} \pi_n \\
&\geqslant \liminf_{r\to 0} \inf_z E_{q,z} \frac{r}{1+r} \sum_{n=1}^{\infty} \frac{1}{(1+r)^{n-1}} \pi_n \\
&\geqslant \liminf_{r\to 0} E_{q,z^*} \frac{r}{1+r} \sum_{n=1}^{\infty} \frac{1}{(1+r)^{n-1}} \pi_n,
\end{aligned}$$

where the last inequality follows from the fact that z^* is uniformly discount optimal. Since the first and the last expressions in the above string of inequalities are the same, we have

$$\begin{aligned}
&\liminf_{r\to 0} \inf_z E_{q,z} \frac{r}{1+r} \sum_{n=1}^{\infty} \frac{1}{(1+r)^{n-1}} \pi_n \\
&= \inf_z \liminf_{r\to 0} E_{q,z} \frac{r}{1+r} \sum_{n=1}^{\infty} \frac{1}{(1+r)^{n-1}} \pi_n \geqslant v.
\end{aligned}$$

That is, q guarantees v to player 1 in the asymptotic discount game.

(e)$\Rightarrow$ (f). As above, if player 1 uses strategy q, the game may be viewed by player 2 as a Markov decision process. Let W_N denote the value of the N-stage version of this process. Similarly, let $W(r)$ denote the value of the r-discount version. By assumption (e),

$$\liminf_{r\to 0} \frac{r}{1+r} W(r) = \liminf_{r\to 0} \inf_z E_{q,z} \frac{r}{1+r} \sum_{n=1}^{\infty} \frac{1}{(1+r)^{n-1}} \pi_n \geqslant v.$$

By (3.3),

$$\lim_{r\to 0} \frac{r}{1+r} W(r) \text{ exists and equals } \lim_{N\to\infty} \frac{1}{N} W_N.$$

Hence,

$$v \leqslant \liminf_{N\to\infty} \frac{1}{N} W_N = \liminf_{N\to\infty} \min_z E_{q,z} \frac{1}{N} \sum_{n=1}^{N} \pi_n,$$

that is, q guarantees v to player 1 in the asymptotic average game.

(f)$\Rightarrow$ (d). Clearly,

$$\inf_z \liminf_{N\to\infty} E_{q,z} \frac{1}{N} \sum_{n=1}^{N} \pi_n \geqslant \liminf_{N\to\infty} \min_z E_{q,z} \frac{1}{N} \sum_{n=1}^{N} \pi_n \geqslant v,$$

so that q guarantees v to player 1 in the limit expected average game.

(c)$\Rightarrow$ (a). By Abel's Theorem,

$$\begin{aligned}
&\inf_z E_{q,z} \liminf_{r\to 0} \frac{r}{1+r} \sum_{n=1}^{\infty} \frac{1}{(1+r)^{n-1}} \pi_n \\
&\geqslant \inf_z E_{q,z} \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \pi_n \geqslant v.
\end{aligned}$$

This content downloaded from 164.15.117.233 on Thu, 13 Feb 2014 06:20:14 AM
All use subject to JSTOR Terms and Conditions

That is, q guarantees v to player 1 in the limit discount game. Abel's Theorem is stated in Derman [7, p. 144] and proved in Widder [24, p. 18, Corollary 1c].

(d) $\Rightarrow$ (c). It is enough to prove that for every $\epsilon > 0$, for every state $\underline{s}$, and for each strategy z_1 for player 2, there exists a strategy z_2 such that

$$\liminf_{N\to\infty} E_{q, z_2, \underline{s}} N^{-1} \sum_{n=1}^{N} \pi_n \leqslant \epsilon + E_{q, z_1, \underline{s}} \liminf_{N\to\infty} N^{-1} \sum_{n=1}^{N} \pi_n. \tag{A.1}$$

Recall that if player 1 uses strategy q, then the game becomes a Markov decision process from the point of view of player 2. Call this process M_q. Define a state, s_1, to be *attainable* from s_2 if there exists a sequence of actions for player 2 which carries the play from s_1 to s_2 with positive probability. The relation "s_1 and s_2 are attainable from each other" is an equivalance relation on the set of states. Let $C_0(1), \ldots, C_0(k)$ be the equivalence classes of this relation.

It is not hard to see that with probability one, the play eventually remains forever in some one of the sets $C_0(i)$. Without loss of generality, it may be assumed that $C_0(1), \ldots, C_0(k_1)$ are the sets with the following property. If player 2 uses strategy z_1, then the game arrives in $C_0(i)$ with positive probability.

Suppose that the subset $C_j(i)$ of $C_0(i)$ has been defined, where $j \geqslant 0$. Let $C_{j+1}(i)$ be the set of states in $C_j(i)$ such that player 2 has at least one action which prevents transition to states outside of $C_j(i)$. Let $C(i) = \bigcap_{j=0}^{\infty} C_j(i)$. Clearly, if $1 \leqslant i \leqslant k_1$, $C(i)$ is not empty. Let M_i be the Markov decision process obtained by restricting M_q to states in $C(i)$ and by restricting player 2 to actions for which transition to states outside of $C(i)$ is impossible. Let $z_{C(i)}$ be a stationary optimal strategy for M_i.

It is not hard to see that there exists N_1 such that

$$P_{q, z_1, \underline{s}} \{\text{for some } i = 1, \ldots, k_1, \text{ the play is in } C(i) \text{ at stage } N_1 \text{ and remains in } C(i) \text{ thereafter}\} > 1 - \epsilon/6\gamma, \tag{A.2}$$

where $\gamma = \max_{ijs} \gamma_{ijs}$ and $P_{q, z_1, \underline{s}}$ is the probability when player 1 uses strategy q, player 2 uses z_1, and the initial state is $\underline{s}$.

z_2 instructs player 2 to play as follows. Play according to z_1 in stages $1, \ldots, N_1 - 1$. If at stage N_1, the game is in $C(i)$, $i = 1, \ldots, k_1$, use $z_{C(i)}$ at stage N_1 and thereafter. If at stage N_1 the game is not in any of the sets $C(i)$, $i = 1, \ldots, k_1$, use z_1 at that stage and thereafter.

With probability at least $1 - \epsilon/6\gamma$, player 2 eventually uses a stationary strategy (see (A.2)). If he uses a stationary strategy, the averages $N^{-1}\sum_{n=1}^{N} \pi_n$ converge, by the ergodic theorem for Markov processes, so that, by the Lebesgue dominated convergence theorem,

$$E \liminf_{N\to\infty} N^{-1} \sum_{n=1}^{N} \pi_n = \liminf_{N\to\infty} E N^{-1} \sum_{n=1}^{N} \pi_n.$$

Since $|N^{-1}\sum_{n=1}^{N} \pi_n|$ is bounded by γ, one obtains

$$\liminf_{N\to\infty} E_{q, z_2, \underline{s}} N^{-1} \sum_{n=1}^{N} \pi_n \leqslant E_{q, z_2, \underline{s}} \liminf_{N\to\infty} N^{-1} \sum_{n=1}^{N} \pi_n + \left(\frac{\epsilon}{6\gamma}\right) 2\gamma.$$

This content downloaded from 164.15.117.233 on Thu, 13 Feb 2014 06:20:14 AM
All use subject to JSTOR Terms and Conditions

Hence, in order to prove (A.1), it is enough to prove that

$$E_{q, z_2, \underline{s}} \liminf_{N\to\infty} N^{-1} \sum_{n=1}^{N} \pi_n \leqslant E_{q, z_1, \underline{s}} \liminf_{N\to\infty} N^{-1} \sum_{n=1}^{N} \pi_n + \frac{2\epsilon}{3}. \tag{A.3}$$

Let D_i be the event that the play is in $C(i)$ from stage N_1 onwards, $i = 1, \ldots, k_1$. It is now shown that it is sufficient to prove the following.

$$E_{q, z_2, \underline{s}}\left(\liminf_{N\to\infty} N^{-1} \sum_{n=1}^{N} \pi_n \mid D_i \right) \leqslant E_{q, z_1, \underline{s}}\left(\liminf_{N\to\infty} N^{-1} \sum_{n=1}^{N} \pi_n \mid D_i \right) \quad \text{for all } i = 1, \ldots, k_1. \tag{A.4}$$

Let D_0 be the complement of $\sum_{i=1}^{k_1} D_i$. (A.2) says that $P_{q, z_1, \underline{s}}(D_0) < \epsilon/6\gamma$. It is easy to see that $P_{q, z_1, s}(D_i) \leqslant P_{q, z_2, s}(D_i)$ for $i = 1, \ldots, k_1$. Hence,

$$\sum_{i=0}^{k_1} \left| P_{q, z_2, s}(D_i) - P_{q, z_1, s}(D_i) \right| < \frac{\epsilon}{3\gamma}.$$

If (A.4) is true, we obtain (A.3) as follows.

$$\begin{aligned}
E_{q, z_2, \underline{s}} \liminf_{N\to\infty} N^{-1} \sum_{n=1}^{N} \pi_n
&= \sum_{i=0}^{k_1} P_{q, z_2, \underline{s}}(D_i) E_{q, z_2, \underline{s}}\left(\liminf_{N\to\infty} N^{-1} \sum_{n=1}^{N} \pi_n \mid D_i \right) \\
&\leqslant \frac{\epsilon}{3} + \sum_{i=0}^{k_1} P_{q, z_1, \underline{s}}(D_i) E_{q, z_2, \underline{s}}\left(\liminf_{N\to\infty} N^{-1} \sum_{n=1}^{N} \pi_n \mid D_i \right) \\
&\leqslant \frac{2\epsilon}{3} + \sum_{i=0}^{k_1} P_{q, z_1, \underline{s}}(D_i) E_{q, z_1, \underline{s}}\left(\liminf_{N\to\infty} N^{-1} \sum_{n=1}^{N} \pi_n \mid D_i \right) \\
&= \frac{2\epsilon}{3} + E_{q, z_1, \underline{s}} \liminf_{N\to\infty} N^{-1} \sum_{n=1}^{N} \pi_n.
\end{aligned}$$

We now apply a theorem of Derman [7, p. 98] to prove (A.4). Fix i, $i = 1, \ldots, k_1$. For each $s \in C(i)$, let

$$h_s = \inf\Big\{ \liminf_{N\to\infty} E_{q, z, s} N^{-1} \textstyle\sum_{n=1}^{N} \pi_s \mid z \text{ is a stationary strategy for player 2 in } M_i \Big\}.$$

The following is a corollary of Derman's theorem.

A.5. *Suppose that player* 1 *uses* q, *that the initial state belongs to* $C(i)$, *and that player* 2 *uses a stationary strategy for* M_i. *Then, with probability one,*

$$\liminf_{N\to\infty} N^{-1} \sum_{n=1}^{N} \pi_n \geqslant \min_{s \in C(i)} h_s.$$

This content downloaded from 164.15.117.233 on Thu, 13 Feb 2014 06:20:14 AM
All use subject to JSTOR Terms and Conditions

(In the Markov decision processes that Derman studies, the current payoff resulting from an action is deterministic. In our case, it is a random variable. However, Derman's proof remains valid in this case.)

z_1 does not require player 2 to play in M_i, even if event D_i occurs. That is, when the play is in $C(i)$, z_1 may require player 2 to give positive weight to actions which could carry the play out of $C(i)$. However, it is easy to see that

$$\textstyle\sum_{n=N_1}^{\infty} P_{q, z_1, \underline{s}}(\{\text{at stage } n, \text{ player 2 chooses an action with a positive transition probability to some state outside } C(i)\} \mid D_i) < \infty.$$

This fact and (A.5) imply that

$$P_{q, z_1, \underline{s}}\left(\left\{ \liminf_{N\to\infty} N^{-1} \sum_{n=1}^{N} \pi_n \geq \min_{s\in C(i)} h_s \right\} \mid D_i \right) = 1. \tag{A.6}$$

It is easy to see that since in M_i every state is obtainable from every other state, $h_s = h_{s'}$, for all s and s' in $C(i)$ (see (17.1)). Let h be the common value of the numbers h_s.

Since $z_{C(i)}$ is a stationary optimal strategy for M_i, it follows by the ergodic theorem for Markov processes and the dominated convergence theorem that

$$\begin{aligned} &E_{q, z_2, \underline{s}}\left(\liminf_{N\to\infty} N^{-1} \sum_{n=1}^{N} \pi_n \mid D_i \right) \\ &\quad = \liminf_{N\to\infty} E_{q, z_2, \underline{s}}\left(N^{-1} \sum_{n=1}^{N} \pi_n \mid D_i \right) = h. \end{aligned} \tag{A.7}$$

(A.6) and (A.7) imply (A.4). Q.E.D.

REMARK. If one replaces $\liminf_{N\to\infty}$ by $\limsup_{N\to\infty}$ in formulas (4.1)–(4.6), one obtains new criteria, which we call the weak criteria. We thus obtain the weak limit discount criterion, the weak limit discount game, etc. It is easy to see that theorems 5.1 and 5.2 and corollary 5.3 apply to all these games, weak and nonweak. To see that theorem 5.2 applies, let (a_w) denote the weak limit discount game, (b_w) the weak limit expected discount game, etc. Obviously, (a) ⇒ (a_w), (b) ⇒ (b_w), . . . , and (f) ⇒ (f_w). We now prove the opposite implications.

Recall that when player 1 uses a stationary strategy, q, player 2 faces a Markov decision process, M_q. By Theorem 5.2, player 2 can do the most harm to player 1 according to any of the nonweak criteria by using a stationary uniformly discount optimal strategy for M_q. Let z^* be such a strategy. $N^{-1}\sum_{n=1}^{N} \pi_n$ converges almost everywhere according to the probability distribution induced by q and z^*. It follows that player 2 can do as much harm to player 1 according to the weak criteria as according to the nonweak criteria. Hence, (a_w) ⇒ (a), . . . , (f_w) ⇒ (f).

Acknowledgement. We would like to thank an anonymous referee for a very thorough report. He caught several errors in the paper and suggested many improvements.

References

[1] Aumann, R. J. and Maschler, M. (1968). Repeated Games with Incomplete Information: The Zero Sum Extensive Case. First Report on Contract ACDA/ST-143, prepared by Mathematica, Chapter II, pp. 25–108. Princeton, N.J.

[2] Bewley, T. and Kohlberg, E. (1976). The Asymptotic Solution of a Recursion Equation Occurring in Stochastic Games. *Math. Oper. Res.* **1** 321–336.

This content downloaded from 164.15.117.233 on Thu, 13 Feb 2014 06:20:14 AM
All use subject to JSTOR Terms and Conditions

[3] Bewley, T. and Kohlberg, E. (1976). The Asymptotic Theory of Stochastic Games. *Math. Oper. Res.* **1** 197–208.

[4] Blackwell, D. (1962). Discrete Dynamic Programming. *Ann. Math. Statist.* **33** 719–726.

[5] ——— and Ferguson, T. S. (1968). The Big Match. *Ann. Math. Statist.* **39** 159–168.

[6] Brown, B. (1965). On the Iterative Method of Dynamic Programming on a Finite Space Discrete Time Markov Process. *Ann. Math. Statist.* **36** 1279–1285.

[7] Derman, C. (1970). *Finite State Markovian Decision Processes*. Academic Press, New York.

[8] Everett, H. (1957). Recursive Games. *Contributions to the Theory of Games*, Vol. III *(Annals of Mathematics Studies*, No. 39). Princeton University Press, Princeton, N.J., pp. 47–78.

[9] Gillette, D. (1957). Stochastic Games with Zero Stop Probabilities. *Contributions to the Theory of Games*, Vol. III (*Annals of Mathematics Studies*, No. 39). Princeton University Press, Princeton, N.J., pp. 179–187.

[10] Hoffman, A. J. and Karp, R. M. (1966). On Non-Terminating Stochastic Games. *Management Sci.* **12** 359–370.

[11] Howard, Ronald (1960). *Dynamic Programming and Markov Processes*. Technology Press of MIT, Cambridge, Mass., and Wiley, New York.

[12] Karlin, S. (1959). *Mathematical Methods and Theory in Games, Programming, and Economics*, Vol. 1. Addison-Wesley, Reading, Mass.

[13] Kohlberg, E. (1974). Repeated Games with Absorbing States. *The Annals of Statistics* **2** 724–738.

[14] Kuhn, H. W. (1956). Solvability and Consistency for Linear Equations and Inequalities. *Amer. Math. Monthly* **63** 217–232.

[15] Liggett, T. and Lippman, S. (1969). Stochastic Games with Perfect Information and Time Average Payoff. *SIAM Review* **11** 604–607.

[16] Maitra, A. and Parthasarathy, T. (1970). On Stochastic Games. *J. Optimization Theory Appl.* **5** 289–300.

[17] ——— and ———. (1971). On Stochastic Games II. *J. Optimization Theory Appl.* **8** 154–160.

[18] Miller, B. and Veinott, A. (1969). Discrete Dynamic Programming with a Small Interest Rate. *Ann. Math. Statist.* **40** 366–370.

[19] Ross, S. (1970). *Applied Probability Models with Optimization Applications*. Holden-Day, San Francisco.

[20] Shapley, L. (1953). Stochastic Games. *Proc. Nat. Acad. Sci. U.S.A.* **39** 1095–1100.

[21] Stern, Martin A. (1975). *On Stochastic Games with Limiting Average Payoff*. Doctoral Dissertation in Mathematics, University of Illinois.

[22] Weyl, H. (1950). Elementary Proof of a Minmax Theorem Due to von Neumann. *Contributions to the Theory of Games*, Vol. I (Annals of Mathematical Studies, No. 24). Princeton University Press, Princeton, N.J., pp. 19–25.

[23] ———. (1950). The Elementary Theory of Convex Polyhedra. *Contributions to the Theory of Games*, Vol. I (Annals of Mathematical Study, No. 24). Princeton University Press, Princeton, N.J., pp. 3–18.

[24] Widder, David V. (1941). *The Laplace Transform*. Princeton University Press, Princeton, N.J.

DEPARTMENT OF ECONOMICS, HARVARD UNIVERSITY, CAMBRIDGE, MA 02138

GRADUATE SCHOOL OF BUSINESS ADMINISTRATION, HARVARD UNIVERSITY, BOSTON, MA 02163

This content downloaded from 164.15.117.233 on Thu, 13 Feb 2014 06:20:14 AM
All use subject to JSTOR Terms and Conditions