stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE BESLISKUNDE BW 57/75 DECEMBER

A. FEDERGRUEN, IN COOPERATION WITH
O.J. VRIEZE & G.L WANROOY

ON THE EXISTENCE OF DISCOUNTED AND AVERAGE RETURN EQUILIBRIUM POLICIES IN N-PERSON STOCHASTIC GAMES

Prepublication

2e boerhaavestraat 49 amsterdam

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O), by the Municipality of Amsterdam, by the University of Amsterdam, by the Free University at Amsterdam, and by industries.

AMS(MOS) subject classification scheme (1970): 93E05

On the existence of discounted and average return equilibrium policies in N-person stochastic games. [*)]

by

A. Federgruen, in cooperation with
O.J. Vrieze & G.L. Wanrooy

ABSTRACT AND SUMMARY

This paper considers noncooperative stochastic games with N players. For the case where the state space is *arbitrary* and the action space compact, and under some continuity assumptions with respect to the immediate return and the transition probability function, the existence of a stationary equilibrium policy under the criterion of the total discounted return, is proven.

Next, for the case where the state space is denumerable, we give a number of recurrency conditions with respect to the transition probability matrices associated with the stationary policies that guarantee the existence of an equilibrium policy under the criterion of the average return per unit time.

Finally, in Section 4, we review and extend the results that are known for the case where both the state space and the action space are finite.

KEY WORDS & PHRASES: *Stochastic games, non-cooperative games, discounted return model, average return model, equilibrium points, optimal stationary policies.*

[*)] This paper is not for review; it is meant for publication elsewhere.

1. INTRODUCTION

This paper treats an N-person noncooperative stochastic game, specified by a five-tuple (ψ,S,A,q,r).

$\psi = \{1,\dots,N\}$ is a finite set of players, S is a locally compact Borel subset of some complete, separable metric space, and A is a compact metric space, where the set S denotes the state space of some system, and A denotes the set of possible actions. With C defined as

$$(1.1) \qquad C = X_{i=1}^{N} A,$$

q associates with each pair $(s,\underline{a}) \in S \times C$ a probability distribution $q(.|s;\underline{a})$ on the class $\mathcal{B}_S$ of Borel subsets of S, and r^i is a bounded real-valued and Borel-measurable function on $S \times C$, for all $i \in \psi$.

Behaviorally, a stochastic game is a sequence $\gamma_1,\gamma_2,\dots$ of noncooperative nonzero sum games played by the members of ψ, where $s \in S$ indexes the set $\{\Gamma_s \mid s \in S\}$ from which γ_t $(t=1,2,\dots)$ is drawn.

Without loss of generality, we assume that the set of actions A available to player i in the state s is identical for all $s \in S$ and $i \in \psi$ such that all the players' actions in γ_t $(t=1,2,\dots)$ are a vector $\underline{a} = [a^1,\dots,a^N] \in C$.

When $\gamma_t = s$, i.e. when the system is in state s and the vector $\underline{a} \in C$ denotes all the players' actions in γ_t, then the one-step expected reward to player i is given by $r^i(s;\underline{a})$, and the state next visited (i.e. the game next drawn) is distributed according to $q(.|s;\underline{a})$.

Let F(A) denote the set of all finite measures on $\mathcal{B}_A$, the Borel subsets of A, endowed with the weak topology (cf. [13], p.40). Observe, using Th. 5.9 of PARTHASARATHY [13] that F(A) is a linear Hausdorff topological space, and let M(A) be the subspace of all *probability* measures on $\mathcal{B}_A$, with the induced topology.

It then follows from the fact that A is a compact metric space, using Th. 6.4 of [13], that M(A) can be metrized as a compact metric subspace of F(A).

Next we define $F(C) = X_{i\in\psi} F(A)$ as the set of all finite *product* measures on $\mathcal{B}_C$, the product σ-field in C, endowed with the product topology.

Let $M(C) = X_{i\in\psi} M(A)$ be the subspace of all product *probability* measures on $\mathcal{B}_C$.

Observe that F(C) is again a linear Hausdorff topological space and that, as a result of Tychonoff's Theorem, M(C) can be metrized as a compact metric subspace of F(C) as well.

We use the (abbreviated) notation $[\underline{\mu}^{-i},\nu]$ for the N-person randomized action $[\mu^1,\dots,\mu^{i-1},\nu,\mu^{i+1},\dots,\mu^N]$ that results from $\underline{\mu} = [\mu^1,\dots,\mu^i,\dots,\mu^N]$, when the i-th player changes from μ^i to ν, the other players continuing to use their respective actions in $\underline{\mu}$.

Defining $r^i(s;\underline{\mu}) = E_{\underline{\mu}} r^i(s;\underline{a})$ and $q(.|s;\underline{\mu}) = E_{\underline{\mu}} q(.|s;\underline{a})$ for all $\underline{\mu} = [\mu^1,\dots,\mu^N] \in M(C)$, $s \in S$ and $i \in \psi$, we obtain:

$$r^i(s;\underline{\mu}) = \int_C r^i(s;\underline{a})d\underline{\mu}(\underline{a}) = \int_A\dots\int_A r^i(s;a^1,\dots,a^N)d\mu^1(a^1)\dots d\mu^N(a^N), \tag{1.2}$$

$$q(.|s;\underline{\mu}) = \int_C q(.|s;\underline{a})d\underline{\mu}(\underline{a}) = \int_A\dots\int_A q(.|s;a^1\dots a^N)d\mu^1(a^1)\dots d\mu^N(a^N), \tag{1.3}$$

where the second equality in (1.2) and (1.3) results from Fubini's Theorem. Observe that $r^i(s;\underline{\mu})$ and $q(.|s;\underline{\mu})$ are both multilinear in $\underline{\mu}$:

$$r^i(s;\mu^1,\dots,\lambda\mu^j+(1-\lambda)\nu^j,\dots,\mu^N) = \lambda r^i(s;\mu^1,\dots,\mu^j,\dots,\mu^N) + (1-\lambda)r^i(s;\mu^1,\dots,\nu^j,\dots,\mu^N), \tag{1.4}$$

$$q(.|s;\mu^1,\dots,\lambda\mu^j+(1-\lambda)\nu^j,\dots,\mu^N) = \lambda q(.|s;\mu^1,\dots,\mu^j,\dots,\mu^N) + (1-\lambda)q(.|s;\mu^1,\dots,\nu^j,\dots,\mu^N). \tag{1.5}$$

A policy π^i for player i is a rule that prescribes to player i, for each time t, which randomized action $\mu \in M(A)$ to choose at time t, as a Borel measurable function of the state s_t and the history

$$H_t = (s_0,\underline{a}_0;s_1,\underline{a}_1;\dots;s_{t-1},\underline{a}_{t-1})$$

of the system up to t.

Let Π denote the class of all one-player policies, and denote by Δ the set of all Borel measurable mappings $\delta\colon S \to M(A)$. For any $\delta \in \Delta$, let $\delta^{(\infty)}$ indicate the *stationary* (one-player) policy that prescribes the randomized action $\delta(s) \in M(A)$ whenever the system is in state s.

As a consequence we let Δ represent the class of all stationary one-player policies as well.

A stationary policy $\delta^{(\infty)}$ is said to be *pure* if in each state of the system it prescribes a specific action in A with probability one.

Finally, let $\underline{\Pi} = X_{i\in\psi}\ \Pi$ denote the class of all N players' policies, and let $\underline{\Delta} = X_{i\in\psi}\ \Delta$ denote the subset of the *stationary* N players' policies. For any policy $\underline{\pi} = [\pi^1,\dots,\pi^N] \in \underline{\Pi}$, we define $V^i_\alpha(\underline{\pi},s)$ and $g^i(\underline{\pi},s)$ as the total expected α-discounted return, and the long-run expected average return per unit time to player i, when the initial state is s:

$$(1.6)\qquad V^i_\alpha(\underline{\pi};s) = E_{\underline{\pi}}\{\sum_{k=0}^{\infty} \alpha^k r^i(s_k;\underline{a}_k) \mid s_0 = s\},\ i \in \psi,\ s \in S,\ 0 \le \alpha < 1,$$

$$(1.7)\qquad g^i(\underline{\pi};s) = \limsup_{t\to\infty} \frac{1}{t+1} E_{\underline{\pi}}\{\sum_{k=0}^{t} r^i(s_k;a_k) \mid s_0 = s\},\ i \in \psi,\ s \in S,$$

where $E_{\underline{\pi}}$ indicates the expectation given the players' common policy $\underline{\pi} \in \underline{\Pi}$ is used, and where $\{s_k;\ k = 0,1,2,\dots\}$ and $\{a_k;\ k = 0,1,\dots\}$ denote the stochastic processes of the states and actions that result from policy $\underline{\pi}$.

An N-tuple of policies $\underline{\pi}^* = [\pi^{*1},\dots,\pi^{*N}] \in \underline{\Pi}$ is said to be an *α-discounted equilibrium point of policies (α-DEP)* if

$$(1.8)\qquad V^i_\alpha(\underline{\pi}^*;s) \ge V^i_\alpha(\underline{\pi};s) \text{ for all } i \in \psi,\ s \in S, \text{ and } \underline{\pi} \in \Pi^{-i}(\underline{\pi}^*),$$

where

$$(1.9)\qquad \Pi^{-i}(\underline{\pi}^*) = \{\underline{\pi} = [\pi^1,\dots,\pi^N] \in \underline{\Pi} \mid \pi^j = \pi^{*j},\ j \ne i\}.$$

Similarly, we define $\underline{\pi}^*$ as an *average return equilibrium point of policies (AEP)*, if

$$(1.10)\qquad g^i(\underline{\pi}^*;s) \ge g^i(\underline{\pi};s) \text{ for all } i \in \psi,\ s \in S \text{ and } \underline{\pi} \in \Pi^{-i}(\underline{\pi}^*).$$

Hence, whenever the players choose an α-DEP (AEP) $\underline{\pi}^*$, none of them, whatever the initial state of the system, can increase his own total expected α-discounted return (expected average return per unit time) by changing to some other policy $\pi^i \neq \pi^{*i} \in \Pi$, the other players continuing to use their respective policies in $\underline{\pi}^*$.

In the following sections conditions will be given under which the existence of stationary α-DEP and AEPs will be proved.

We conclude this section by observing that if A is a subset of some linear metric space itself, such that for all $i \in \psi$, $r^i(s;\underline{a})$ is linear, or even concave in the i-th component of $\underline{a}$, and $q(.|s;\underline{a})$ is multilinear in $\underline{a}$ (cf. (1.4) and (1.5)), then the existence of a *pure*, instead of a randomized, stationary α-DEP and AEP is guaranteed, under the same conditions, as follows from an examination of the analysis below.

2. EXISTENCE OF STATIONARY α-DEPs

Hereafter we assume:

A1. $r^i(s;\underline{a})$ is continuous on $S \times C$, for all $i \in \psi$.

A2. $q(.|s_n;\underline{a}_n)$ converges weakly to $q(.|s;a)$ as $s_n \to s$, and $\underline{a}_n \to \underline{a}$, whereas $q(.|s;\underline{a}_n)$ converges setwise to $q(.|s;\underline{a})$ as $\underline{a}_n \to \underline{a}$, for all $s \in S$.

LEMMA 2.1. *Suppose* A1,A2 *hold. Then*

(a) $r^i(s;\underline{\mu})$ *is continuous on* $S \times M(C)$ *for all* $i \in \psi$;

(b) $q(.|s_n;\underline{\mu}_n)$ *converges weakly to* $q(.|s;\mu)$ *as* $s_n \to s$, *and* $\underline{\mu}_n \to \underline{\mu}$, *whereas* $q(.|s;\underline{\mu}_n)$ *converges setwise to* $q(.|s;\underline{\mu})$ *as* $\underline{\mu}_n \to \underline{\mu}$, *for all* $s \in S$.

PROOF. This proof proceeds along similar lines to the one in MAITRA & PARTHASARATHY ([12], Lemma 2.1).

(a) Let $s_n \to s_0$, and $\underline{\mu}_n \to \underline{\mu}_0$, fix $i \in \psi$, and pick $\varepsilon > 0$. We have (cf. (1.2)):

$$(2.1)\qquad |r^i(s_n;\underline{\mu}_n)-r^i(s_0,\underline{\mu})| \le |\int_A\cdots\int_A r^i(s_n;a^1,\dots,a^N)d\mu_n^1(a^1)\dots d\mu_n^N(a^N) +$$

$$- \int_A\cdots\int_A r^i(s_0;a^1,\dots,a^N)d\mu_n^1(a^1)\dots d\mu_n^N(a^N)| +$$

$$+ |\int_A\cdots\int_A r^i(s_0;a^1,\dots,a^N)d\mu_n^1(a^1)\dots d\mu_n^N(a^N) +$$

$$- \int_A\cdots\int_A r^i(s_0;a^1,\dots,a^N)d\mu_0^1(a^1)\dots d\mu_0^N(a^N)|.$$

Since S is locally compact, there is an open set O containing s, such that its closure $\bar{O}$ is compact. As $\bar{O} \times C$ is a compact metric space, $r^i(s;\underline{a})$ is uniformly continuous on $\bar{O} \times C$ and hence there exists an integer N_1 such that for all $n \ge N_1$, $|r^i(s_n;\underline{a})-r^i(s_0;\underline{a})| < \varepsilon/2$ for all $\underline{a} \in C$. This implies that the first term on the right-hand side of (3.1) is at most $\varepsilon/2$, for all $n \ge N_1$.

Next, observe that since for all $j \in \psi$, $\mu_n^j(.)$ converges to $\mu_0^j(.)$ in the weak topology, we have, as a consequence of Fubini's theorem, for every N-tuple $(f^1,\dots,f^N)$ of real-valued continuous functions on A:

$$(2.2)\qquad \int_A\cdots\int_A f^1(a^1)f^2(a^2)\dots f^N(a^N)d\mu_n^1(a^1)\dots d\mu_n^N(a^N) \to$$

$$\int_A\cdots\int_A f^1(a^1)\dots f^N(a^N)d\mu_0^1(a^1)\dots d\mu_0^N(a^N).$$

Since, due to the Stone-Weierstrass Theorem (cf. ROYDEN [16], p.174), $r^i(s_0;a^1,\dots,a^N)$, as a continuous function on C, can be approximated uniformly by a sequence of functions of the form $\sum_{\ell=1}^k f_\ell^1(a^1)\dots f_\ell^N(a^N)$, where for each $i \in \psi$, and $\ell = 1,2,\dots$, $f_\ell^i(.)$ is continuous on A, we obtain, using (2.2) that there exists an integer N_2, such that the second term on the right-hand side of (2.1) is at most $\varepsilon/2$, for all $n \ge N_2$, as well. Hence, $r^i(s_n;\underline{\mu}_n)$ converges to $r^i(s_0;\underline{\mu}_0)$ as $s_n \to s_0$, $\underline{\mu}_n \to \underline{\mu}_0$, which proves part (a).

(b) Show that A2 implies that

$$\int_S u(s')q(ds'|s_n;\underline{\mu}_n) \to \int_S u(s')q(ds'|s_0;\underline{\mu}_0), \text{ as } s_n \to s_0, \ \underline{\mu}_n \to \underline{\mu}_0$$

and that

$$q(B|s_0;\underline{\mu}_n) \to q(B|s_0;\underline{\mu}_0) \text{ as } \underline{\mu}_n \to \underline{\mu}_0,$$

for every real valued, continuous and bounded function u on S, and for every Borel set $B \in \mathcal{B}_S$, by repeating the proof of part (a) with $r^i(.;.)$ replaced by $\int_S u(s')q(ds'|.;.)$ and $q(B|.;.)$ respectively. □

Observe that $X_{s\in S}\, F(C)$, the space of all mappings $f\colon S \to F(C)$ endowed with the product topology, is a linear Hausdorff topological space.

Likewise, $X_{s\in S}\, M(C)$, the space of all mappings $f\colon S \to M(C)$ (with the induced topology), is a compact subspace, as a consequence of Tychonoff's Theorem.

Let $\{f_n\}_{n=1}^{\infty}$ be a sequence in $X_{s\in S}\, M(C)$. Then, since convergence of $f_n \to f$, in the product topology, implies $f_n(s) \to f(s)$, for all $s \in S$, it follows from KURATOWSKI ([10], part I, p.386) that $\underline{\Delta}$ is a closed, and hence compact subspace of $X_{s\in S}\, M(C) \subseteq X_{s\in S}\, F(C)$. We thus obtain:

(2.3) $\underline{\Delta}$ is a compact convex subspace of the linear Hausdorff topological space $X_{s\in S}\, F(C)$.

For any stationary policy $\underline{\delta}^{(\infty)} \in \underline{\Delta}$, the total expected α-discounted return to player i, $V^i_\alpha(\underline{\delta}^{(\infty)};s)$ is given by:

(2.4) $$V^i_\alpha(\underline{\delta}^{(\infty)};s) = \sum_{n=0}^{\infty} \alpha^n . \int_S r^i(y;\underline{\delta}(y))q^n_{\underline{\delta}}(dy|s)$$

where $q^n_{\underline{\delta}}(B|s)$, with $s \in S$ and $B \in \mathcal{B}_S$ denotes the n-step transition probability function of the Markov Chain $\{s_t\}$ associated with the stationary policy $\underline{\delta}^{(\infty)}$.

The following lemma proves that $V^i_\alpha(\underline{\delta}^{(\infty)};s)$ is a continuous function on $\underline{\Delta}$ for all $i \in \psi$, $s \in S$, $\alpha \in [0,1)$.

LEMMA 2.2. *Assume* A1-A2 *hold. Fix* $s_0 \in S$, $i \in \psi$, *and* $\alpha \in [0,1)$. *Then* $\lim_{n\to\infty} V^i_\alpha(\underline{\delta}_n^{(\infty)};s_0) = V^i_\alpha(\underline{\delta}^{(\infty)},s_0)$ *whenever* $\{\underline{\delta}_n\}_{n=1}^{\infty} \to \underline{\delta}$, *with* $\underline{\delta}_n \in \underline{\Delta}$.

PROOF. We first observe that $\underline{\delta} \in \underline{\Delta}$ (cf. (2.3)) and that $V^i_\alpha(\underline{\eta}^{(\infty)};s)$ is uniformly bounded in $\underline{\eta} \in \underline{\Delta}$. For, let M be such that $|r(s;\underline{a})| \leq M$ for all $s \in S$, and $\underline{a} \in C$. Then, it follows from (1.2) that $|r^i(s;\underline{\mu})| \leq M$ for all $s \in S$, and $\underline{\mu} \in M(C)$, and next, using (2.4), that

$$(2.5) \qquad |V^i_\alpha(\underline{\eta}^{(\infty)};s)| \leq \frac{M}{1-\alpha}, \text{ for all } \underline{\eta}^{(\infty)} \in \underline{\Delta}, \text{ and } s \in S.$$

Let M(S) denote the class of all bounded measurable and real-valued functions on S, and define for each $\underline{\eta} \in \underline{\Delta}$ the operator $H_{\underline{\eta}}: M(S) \to M(S)$ as follows: $H_{\underline{\eta}}(u)(.) = r^i(.;\underline{\eta}(.)) + \alpha \int_S u(s')q(ds'|.;\underline{\eta}(.))$ for all $u \in M(S)$. We next show that

$$(2.6) \qquad \lim_{n\to\infty} \underline{\delta}_n = \underline{\delta}^* \Rightarrow \lim_{n\to\infty} H^k_{\underline{\delta}_n}(u) = H^k_{\underline{\delta}^*}(u), \text{ for all } k = 0,1,\ldots \text{ and } u \in M(S),$$

where the convergence of $\{H^k_{\underline{\delta}_n}(u)\}_{n=1}^{\infty}$ is pointwise, and where $H^k_{\underline{\eta}}(u)$ is recursively defined by:

$$(2.7) \qquad H^{k+1}_{\underline{\eta}}(u)(.) = r^i(.;\underline{\eta}(.)) + \alpha \int H^k_{\underline{\eta}}(u)(s')q(ds'|.;\underline{\eta}(.)), \text{ for } k \geq 1;$$

$$H^0_{\underline{\eta}}(u) = u.$$

Proceeding by complete induction, we first observe that (2.6) trivially holds for k = 0. Suppose now it holds for $k = k_0$. Then, as a consequence of (2.7), it follows from $\underline{\delta}_n(s) \to \underline{\delta}(s)$, assumptions A1-A2, the boundedness of $H^{k_0}_{\underline{\delta}_n}(u)$ and Proposition 18 on p.232 in ROYDEN [16] that (2.6) holds for $k = k_0 + 1$ as well.

We next observe from (2.4) that

$$V^i_\alpha(\underline{\eta}^{(\infty)};s) = H^k_{\underline{\eta}}(0)(s) + \alpha^k \int_S V^i_\alpha(\underline{\eta}^{(\infty)};s')q^k_{\underline{\eta}}(ds'|s), \text{ for } k = 1,2,\ldots \text{ and } \underline{\eta} \in \underline{\Delta}.$$

Finally, let $\{\underline{\delta}_n\}_{n=1}^{\infty} \to \underline{\delta}^*$, where $\underline{\delta}_n \in \underline{\Delta}$, and pick $\varepsilon > 0$. Choose k such that $\alpha^k \le \frac{\varepsilon(1-\alpha)}{4M}$ and, in view of (2.6), an integer N_1 such that $|H^k_{\underline{\delta}_n}(0)(s_0)-H^k_{\underline{\delta}^*}(0)(s_0)| < \varepsilon/2$, for all $n \ge N_0$. Then for all $n \ge N_0$ we obtain, using (2.5), that:

$$\begin{aligned}|V^i_\alpha(\underline{\delta}_n^{(\infty)};s_0)-V^i_\alpha(\underline{\delta}^{*(\infty)};s_0)| &< |H^k_{\underline{\delta}_n}(0)(s_0)-H^k_{\underline{\delta}^*}(0)(s_0)| + \\ &+ \alpha^k \left|\int_S V^i_\alpha(\underline{\delta}_n^{(\infty)};s')q^k_{\underline{\delta}_n}(ds'|s_0) - \int_S V^i_\alpha(\underline{\delta}^{*(\infty)};s')q^k_{\underline{\delta}^*}(ds'|s_0)\right| < \\ &< \varepsilon/2 + \frac{\varepsilon(1-\alpha)}{4M} \cdot \frac{2M}{(1-\alpha)} = \varepsilon,\end{aligned}$$

which proves the lemma. □

We now turn to the existence of an α-DEP.

For a compact state space, and under the assumptions A1 and the first part of A2, the existence of an α-DEP was first proved in SOBEL [22], where, however, the considered class of policies had to be restricted to $X_{s\in S}\, M(C)$, the stationary, though not necessarily measurable ones. In addition, it appears that the proof of the Theorem in [22] is either incorrect or incomplete (cf. VRIEZE [25]).

Theorem 1 below proves the existence of an α-DEP, under the assumptions A1 and A2, within Π, the class of all measurable stationary and nonstationary policies, using an extension of the Kakutani fixed-point theorem by GLICKSBERG [8].

Moreover, we need the following lemma, the proof of which is given in the appendix.

<u>LEMMA 2.3</u>. *Fix* $0 \le \alpha < 1$. *A stationary policy* $\underline{\delta}^{(\infty)} = [\delta^{1(\infty)},\ldots,\delta^{N(\infty)}]$ *is an α-DEP, iff* $V^i_\alpha(\underline{\delta}^{(\infty)};s)$ *satisfies the optimality equation:*

$$(2.8)\qquad V^i_\alpha(\underline{\delta}^{(\infty)};s) = \max_{\mu\in M(A)}\{r^i(s;[\delta^{-i}(s),\mu]) + \alpha\int_S V^i_\alpha(\underline{\delta}^{(\infty)};s')q(ds'|s;[\delta^{-i}(s),\mu])\}$$

for all $s \in S$, $i \in \psi$.

THEOREM 1. *Assume* A1-A2 *to hold, and let* $0 < \alpha < 1$. *Then there exists a stationary* α*-DEP.*

PROOF. We construct a mapping $\Phi: \underline{\Delta} \to 2^{\underline{\Delta}}$, where $2^{\underline{\Delta}}$ denotes the class of all closed subsets of $\underline{\Delta}$. We first show that for each $\underline{\delta} \in \underline{\Delta}$ and $i \in \psi$, there exists an $\eta \in \Delta$ such that

$$r^i(s;[\delta^{-i}(s),\eta(s)]) + \alpha \int_S V_\alpha^i(\underline{\delta}^{(\infty)};y)q(dy|s;[\delta^{-i}(s),\eta(s)]) = \tag{2.9}$$
$$= \max_{\mu\in M(A)}\{r^i(s;[\delta^{-i}(s),\mu]) + \alpha \int_S V_\alpha^i(\underline{\delta}^{(\infty)};y)q(dy|s;[\delta^{-i}(s),\mu])\},$$
$$\forall s \in S.$$

Observe, as a consequence of the assumptions A1-A2, the boundedness of $V_\alpha^i(\underline{\delta}^{(\infty)};s')$ (cf. (2.5)), and Proposition 18 on p.232 of ROYDEN [16], that the expression within { } in (2.9) is a Borel-measurable function on $S \times M(A)$ that is continuous in μ. The existence of an $\eta \in \Delta$ satisfying (2.9) then follows from Th. 12.1 in SCHÄL [17].

For any $i \in \psi$, and $\underline{\delta} \in \underline{\Delta}$, let $\Phi^i(\underline{\delta})$ denote the set of all $\eta \in \Delta$ that satisfy (2.9), and define the point-to-convex-set mapping

$$\Phi: \underline{\Delta} \to 2^{\underline{\Delta}}: \underline{\delta} \to \Phi(\underline{\delta}) = X_{i\in\psi}\, \Phi^i(\underline{\delta}).$$

We next show the upper semi-continuity (in the sense of Kuratowski) of this point-to-set mapping:

$$\{\underline{\delta}_n\}_{n=1}^{\infty} \to \underline{\delta},\ \underline{\eta}_n \in \Phi(\underline{\delta}_n),\ \{\underline{\eta}_n\}_{n=1}^{\infty} \to \underline{\eta} \Rightarrow \underline{\eta} \in \Phi(\underline{\delta}). \tag{2.10}$$

Fix $\{\underline{\delta}_n\}_{n=1}^{\infty}$, $\{\underline{\eta}_n\}_{n=1}^{\infty}$ satisfying the conditions in (2.10) and fix $s \in S$. Substitute $\underline{\delta}_n$ for $\underline{\delta}$ and η_n^i for η in (2.9) and let n tend to infinity. It then follows that η^i satisfies (2.9) for $\underline{\delta}$, and this for all $i \in \psi$ and $s \in S$, as a consequence of the assumptions A1-A2, Lemma 2.2, the boundedness of $V_\alpha^i(\underline{\delta}_n^{(\infty)};s)$ (cf. (2.5)), and Proposition 18 on p.232 in ROYDEN [16].

As a consequence of (2.10) and the fact that Φ is a point-to-convex-set mapping of a convex compact subset $\underline{\Delta}$ of the linear Hausdorff topological space $X_{s\in S}$ F(C) into itself (cf. (2.3)), it follows from GLICKSBERG's [8] extension of the Kakutani fixed-point theorem that there exists a $\underline{\delta}^* \in \underline{\Delta}$ such that $\underline{\delta}^* \in \Phi(\underline{\delta}^*)$, which implies (2.8) and hence proves the theorem (cf. Lemma 2.3). □

REMARK. We sideways observe that the argument in the proof of Lemma 2.3 can be used in order to derive e.g. Theorem 3.1 in MAITRA & PARTHASARATHY [12] in a straightforward manner from the analogous result in Markov Decision Theory.

3. THE EXISTENCE OF AN AEP IN STOCHASTIC GAMES WITH A DENUMERABLE STATE SPACE

In the remainder of this paper we will restrict ourselves to stochastic games with a denumerable state space S.

As a consequence, we henceforth need the following notations: Let $q_{st}(\underline{a})$ denote the transition probability to state t, when the N players' actions in s are given by $\underline{a} \in C$.

We associate with each $\underline{\delta} \in \underline{\Delta}$, the transition probability matrix (tpm) $P(\underline{\delta})$, where $P(\underline{\delta})_{st} = q_{st}(\underline{\delta}(s))$, for all $s,t \in S$.

For any $\underline{\delta} \in \underline{\Delta}$, we define the matrix $P^*(\underline{\delta})$ as the Cesaro limit of the sequence $\{P^n(\underline{\delta})\}_{n=1}^{\infty}$. Let $R(\underline{\delta}) = \{t \mid P^*(\underline{\delta})_{tt} > 0\}$, i.e. $R(\underline{\delta})$ is the set of positive recurrent states. If $P(\underline{\delta})$ has exactly one positive recurrent class of states, then there exists a (unique) stationary probability distribution $\pi(\underline{\delta})(.)$, such that $\pi(\underline{\delta})_t = P^*(\underline{\delta})_{st}$, for all $s \in S$.

We henceforth assume the assumptions A1, A2 to hold, and introduce a number of conditions, each of which will be shown to guarantee the existence of an AEP:

A3.1. There is an integer $\nu \geq 1$, and a number $\rho > 0$ such that for each pair of states (s_1,s_2) and for each $\underline{\delta} \in \underline{\Delta}$:

$$(3.1) \qquad \sum_{t=1}^{\infty} \min\{P^{\nu}(\underline{\delta})_{s_1 t}, P^{\nu}(\underline{\delta})_{s_2 t}\} \geq \rho.$$

A3.2. For each policy $\underline{\delta}^{(\infty)} \in \underline{\Delta}$ there exists a state $s_{\underline{\delta}}$, such that the mean first passage time $m_{\underline{\delta}}(s,s_{\underline{\delta}})$, i.e. the expected number of transitions needed to get from state s to state s_δ under policy $\underline{\delta}^{(\infty)} \in \underline{\Delta}$ is finite and uniformly bounded in $s \in S$, and $\underline{\delta}^{(\infty)} \in \underline{\Delta}$.

A3.3. There exists a number R such that for every player $i \in \psi$, and for any combination of stationary policies $\{\delta^1,\dots,\delta^{i-1},\delta^{i+1},\dots,\delta^N\}$ of the other players, there is a policy $\delta^i \in \Delta$ for player i for which the mean first passage time $m_{\underline{\delta}}(s,t)$ from any state s, to any state t under policy $\underline{\delta} = [\delta^1\dots\delta^N]$ is bounded by R; i.e. for each $\{\delta^1,\dots,\delta^{i-1},\delta^{i+1},\dots,\delta^N\}$ with $\delta^j \in \Delta$ for all $j \neq i$, there exists a $\delta^i \in \Delta$, such that

$$(3.2) \qquad m_{\underline{\delta}}(s,t) \le R \text{ for all } s,t \in S, \text{ where } \underline{\delta} = [\delta^1,\dots,\delta^N].$$

The assumption A3.1 is an adaptation of a condition introduced in TIJMS [24] as an extension of the Doeblin condition (cf. e.g. DOOB [6], p.197) to a collection of Markov Chains. We note that A3.1 with $\nu = 1$, is equivalent to the condition that there is a number $\rho > 0$, such that for each four elements $(s_1,s_2,\underline{a}_1,\underline{a}_2)$ with $s_1 \neq s_2$ and $\underline{a}_1,\underline{a}_2 \in C$:

$$(3.3) \qquad \sum_{t=1}^{\infty} \min\{q_{s_1 t}(\underline{a}_1), q_{s_2 t}(\underline{a}_2)\} \ge \rho.$$

For, fix $s_1,s_2 \in S$ and $\underline{\mu}_1,\underline{\mu}_2 \in M(C)$, and observe that as a consequence of (3.3):

$$(3.4) \qquad \rho \le E_{\underline{\mu}_1,\underline{\mu}_2}\Big[\sum_{t=1}^{\infty} \min\{q_{s_1 t}(\underline{a}_1), q_{s_2 t}(\underline{a}_2)\}\Big]$$

$$= \sum_{t=1}^{\infty} E_{\underline{\mu}_1,\underline{\mu}_2} \min\{q_{s_1 t}(\underline{a}_1), q_{s_2 t}(\underline{a}_2)\} \le \sum_{t=1}^{\infty} \min\{q_{s_1 t}(\underline{\mu}_1), q_{s_2 t}(\underline{\mu}_2)\},$$

where the interchange of expectation and summation is justified by the nonnegativity of $\min\{q_{s_1 t}(\underline{a}_1), q_{s_2 t}(\underline{a}_2)\}$, and where the inequality part follows from Jensen's inequality and the concaveness of min(.,.) on R^2. Note finally that (3.4) coincides with the special case of (3.1) where $\nu = 1$. In Markov Chain terminology, the condition (3.3) is known as the assumption that for each stationary and *pure* policy $\underline{\delta}^{(\infty)}$, the associated tpm $P(\underline{\delta})$ is scrambling (cf. [1]) and has an ergodic coefficient of at least ρ.

The following lemma shows that under A3.1 the sequence $\{P^n(\underline{\delta})\}_{n=1}^{\infty}$ converges exponentially fast to a constant stochastic matrix for any $\underline{\delta} \in \underline{\Delta}$, where in addition the convergence rate is uniformly bounded in $\underline{\delta} \in \underline{\Delta}$. The proof proceeds along the same lines as the proof of Theorem 1 in ANTHONISSE & TIJMS [1] (cf. also pp.173-174 in DOOB [6]). For the sake of completeness, we include the proof in the appendix.

LEMMA 3.1. *Under* A3.1, *we have for any* $\underline{\delta} \in \underline{\Delta}$, *and* s,t ∈ S:

$$(3.5) \qquad |P^n(\underline{\delta})_{st} - \pi(\underline{\delta})_t| \le (1-\rho)^{[n/\nu]}, \qquad \textit{for all } n \ge 1,$$

where [x] *is the largest integer less than or equal to* x.

A3.2 and A3.3 are immediate adaptations of assumptions introduced in ROSS [15] and HORDIJK [9] (Th. 12.8) respectively.

For each α (0≤α<1), we choose a specific α-DEP $\underline{\delta}_\alpha \in \underline{\Delta}$. Next, we fix any state s^*, and define:

$$(3.8) \qquad v^i_\alpha(s) = V^i_\alpha(\underline{\delta}_\alpha^{(\infty)};s) - V^i_\alpha(\underline{\delta}_\alpha^{(\infty)};s^*), \qquad \text{for all } s \in S, \text{ and } i \in \psi.$$

Following the proofs of Lemma 1 in TIJMS [24] *), in ROSS [15], and Th. 12.8 in HORDIJK [9], we obtain the following lemma:

LEMMA 3.2. *Under any one of the conditions* A3, *the family of functions* $\{v^i_\alpha(.),\ 0 \le \alpha < 1\}$ *is uniformly bounded.* □

We now prove the existence of an AEP, making use of a technique introduced by TAYLOR [23], and a.o. used in ROSS [15]:

THEOREM 2. *Suppose that* A1-A3 *hold. Then there exists a stationary AEP* $\underline{\delta}^{(\infty)} \in \underline{\Delta}$, *and for each* $i \in \psi$ *a constant* g^i *and a bounded function* $v^i(.)$, *such that*

$$(3.9) \qquad g^i + v^i(s) = \max_{\mu\in M(A)}\{r^i(s;[\underline{\delta}^{-i}(s),\mu]) + \\ + \sum_{t=1}^{\infty} q_{st}([\underline{\delta}^{-i}(s),\mu])v^i(t)\}, \quad \forall s \in S,$$

where $\delta^{i}(s)$ *attains the maximum in the right-hand side of* (3.9) *for all* $s \in S$. *Moreover,* $g^{i}(\underline{\delta}^{(\infty)};s) = g^{i}$, *for all* $s \in S$, $i \in \psi$.

PROOF. We first observe using (2.5) that $|(1-\alpha)V^{i}_{\alpha}(\underline{\delta}^{(\infty)}_{\alpha};s^{*})| \leq M$, for all $\alpha \in [0,1)$ and $i \in \psi$. This, together with Lemma 3.1 and the fact that any sequence of points in the compact metric space M(C) has a cluster point imply, using the diagonalization procedure, the existence of N constants g^{i}, N bounded functions $v^{i}(.)$, a policy $\underline{\delta}^{(\infty)} \in \underline{\Delta}$, and a sequence $\{\alpha_{k}\}^{\infty}_{k=1}$, with $\alpha_{k} \in [0,1)$ and $\lim_{k\to\infty} \alpha_{k} = 1$, such that:

(a) $\lim_{k\to\infty} \underline{\delta}_{\alpha_k} = \underline{\delta}$;

(b) $\lim_{k\to\infty} (1-\alpha_{k})V^{i}_{\alpha_k}(\underline{\delta}^{(\infty)}_{\alpha_k};s^{*}) = g^{i}$, $\quad i \in \psi$;

(c) $\lim_{k\to\infty} v^{i}_{\alpha_k}(s) = v^{i}(s)$, $\quad$ for all $s \in S$.

Now, fix $i \in \psi$, and $s = s_{0} \in S$, and subtract $V^{i}_{\alpha_k}(s^{*})$ from both sides of (2.8) with $\alpha = \alpha_{k}$, and $s = s_{0}$, in order to obtain (cf. (3.8)):

$$(3.10) \qquad v^{i}_{\alpha_k}(s_{0}) = \max_{\mu\in M(A)}\{r^{i}(s_{0};[\underline{\delta}^{-i}_{\alpha_k}(s_{0}),\mu]) - (1-\alpha_{k})V^{i}_{\alpha_k}(s^{*}) + \sum_{t=1}^{\infty} q_{s_0 t}([\underline{\delta}^{-i}_{\alpha_k}(s_{0}),\mu]) \cdot v^{i}_{\alpha_k}(t)\},$$

where $\delta^{i}_{\alpha_k}(s_{0})$ attains the maximum in the right-hand side of (3.10). Letting k tend to infinity in (3.10) we obtain (3.9), with $\delta^{i}(s_{0})$ attaining the maximum in the right-hand side of (3.9), as a consequence of (a), (b), (c), the assumptions A1-A2, and Proposition 18 on p.232 in ROYDEN [16].

Next, it follows from the proof of Theorem 6.17 in ROSS [15] that policy $\underline{\delta}^{(\infty)}$ is an AEP and that $g^{i}(\underline{\delta}^{(\infty)};s) = g^{i}$ for all $s \in S$ and $i \in \psi$. □

The proof of Theorem 2 also shows the following corollary:

COROLLARY 3.3. *If the assumptions* A1-A3 *are satisfied, then each limit policy obtained from a sequence of* α*-DEPs with discount factor tending to one, is an AEP.* □

We conclude this section by considering the two person - zero sum case, where $N = 2$ and $r^{1}(s,\underline{a}) = -r^{2}(s,\underline{a})$ for all $s \in S$, and $\underline{a} \in A$.

Unlike the general N-person stochastic game, we have in this case that the total expected α-discounted return (the expected average return per unit time to player 1) is identical for *any* α-DEP (AEP), whatever the initial state of the system.

Henceforth, dropping the superindices 1 and 2, let $r(s;\underline{a}) = r^1(s;\underline{a})$ and let $g_{eq}(s)$ denote the average gain (to player 1) associated with an AEP, when the initial state of the system is s.

For any bounded function $w\colon S \to \mathbb{R}$ we define $K_w(s;[\mu,\nu]) = r(s;[\mu,\nu]) + \sum_{t=1}^{\infty} q_{st}([\mu,\nu])w(t)$, for all $s \in S$, $\mu,\nu \in M(A)$, and the dummy game $\Gamma_s(w)$ as the one-step game, with M(A) as the action space for both players and with $K_w(s;[\mu,\nu])$ as pay-off function. It then easily follows from (3.9), using standard arguments, that the functional equation:

$$(3.11) \qquad v(s) + g = \max_{\mu\in M(A)} \min_{\nu\in M(A)} K_V(s;[\mu,\nu]) + \text{val}\ \Gamma_s(V), \qquad s \in S$$

has a bounded solution $<v^*(.),g^*>$ (val $\Gamma_s(v)$ is the value of the game $\Gamma_s(v)$).

The following proposition shows that the correspondence between the solutions to the functional equation (3.11) and the AEPs is similar to the one in Markov Decision Theory:

PROPOSITION 3.4.

(a) *Under any of the assumptions* A3, *the functional equation* (3.11) *has a bounded solution* $<v(.),g>$. *Moreover, every solution* $<v^*(.),g>$ *to* (3.11) *has* $g = g_{eq} = g_{eq}(s)$, *for all* $s \in S$, *and any policy* $\underline{\delta} = (\delta^1,\delta^2)$ *such that* $[\delta^1(s),\delta^2(s)]$ *is an equilibrium pair of actions with respect to the dummy games* $\Gamma_s(v^*)$, *is an AEP.*

(b) *Assume* A3.1 *or* A3.2 *to hold, and let* $<v(.),g_{eq}>$ *be a particular solution to* (3.11). *Then the set of all solution pairs* V *is given by:*
$V = \{<v(.)+c,g_{eq}> \mid c \in \mathbb{R}\}$.

PROOF.

(a) The existence of a bounded solution $<v(.),g>$ to (3.9) was shown above.

Next, fix a solution $<v^*(.),g>$ and let $\underline{\delta} = (\delta^1,\delta^2)$ be such that $[\delta^1(s),\delta^2(s)]$ is an equilibrium pair of actions with respect to the dummy games $\Gamma_s(v^*)$. It then, once again, follows from the proof of Th. 6.17 in ROSS [15] that $\underline{\delta}^{(\infty)}$ is an AEP, with $g(\underline{\delta}^{(\infty)};s) = g = g_{eq}$.

(b) We first observe that under A3.1 and A3.2, for every policy $\underline{\delta} \in \underline{\Delta}$, the associated tpm $P(\underline{\delta})$ has a single (positive) recurrent class of states. Moreover, for any pair of policies $\underline{\delta},\underline{\eta} \in \underline{\Delta}$ we have $R(\underline{\delta}) \cap R(\underline{\eta}) \neq \emptyset$, since otherwise it would be possible to construct a policy ζ with two (positive) recurrent classes (let e.g. $\zeta(s) = \delta(s)$ for $s \in R(\underline{\delta})$, and $\zeta(s) = \eta(s)$ otherwise).

We now show that $V \subseteq \{<v(.)+c,g_{eq}> \mid c \in \mathbb{R}\}$, the other inclusion being trivial.

Let $<w(.),g_{eq}>$ be a second solution pair to (3.11), and fix $\underline{\delta},\underline{\eta} \in \underline{\Delta}$ such that $\underline{\delta}(s),\underline{\eta}(s)$ are equilibrium pairs of actions with respect to the dummy games $\Gamma_s(v)$ and $\Gamma_s(w)$ resp., for all $s \in S$. We then obtain from (3.11) that

$$v(s) + g \leq r(s;[\delta^1(s),\eta^2(s)]) + [P(\delta^1,\eta^2)v]_s,$$

and

$$w(s) + g \geq r(s;[\delta^2(s),\eta^1(s)]) + [P(\delta^2,\eta^1)w]_s.$$

Subtracting the second inequality from the first one, and iterating the resulting inequality k times, we get:

$$(3.12) \qquad v(s) - w(s) \leq [P^k(\delta^1,\eta^2)(v-w)]_s, \qquad s \in S \text{ and } k = 1,2,\ldots,$$

and by taking the Cesaro limit as $k \to \infty$ on both sides of the inequality (3.12),

$$(3.13) \qquad v(s) - w(s) \leq c_1 = \textstyle\sum_t \pi(\delta^1,\eta^2)_t(v-w)_t, \qquad s \in S.$$

Similarly, we obtain

$$(3.14) \qquad \textstyle\sum_t \pi(\delta^2,\eta^1)_t(v-w)_t = c_2 \leq v(s) - w(s), \qquad s \in S.$$

We finally prove $c_1 = c_2 = c$, which by the combination of (3.13) and (3.14) implies $V \subseteq \{<v(.)+c,g_{eq}> \mid c \in \mathbb{R}\}$, and hence part (b) of the proposition.

Multiplying both sides of each inequality in (3.13) by $\pi(\delta^1,\eta^2)_s$, summing over $s \in S$, we find $v(s) - w(s) = c_1$ for all $s \in R(\delta^1,\eta^2)$. Similarly, we obtain $v(s) - w(s) = c_2$ for all $s \in R(\delta^2,\eta^1)$, which implies $c_1 = c_2 = c$, as a consequence of $R(\delta^1,\eta^2) \cap R(\delta^2,\eta^1) \neq \emptyset$. □

REMARK. Although any policy $\underline{\delta}$ is an AEP if for all $s \in S$ $[\delta^1(s),\delta^2(s)]$ is an equilibrium pair of actions with respect to the dummy game $\Gamma_s(v)$, $<v(.),g_{eq}>$ being a solution pair to (3.11), we know from Markov Decision Theory that this condition is at all events not necessary when for some of the policies $\underline{\delta}$, $P(\underline{\delta})$ has more than one subchain (cf. SCHWEITZER & FEDERGRUEN [19], p.6d, Th. 3.1(e) and Th. 4.1(b)).

As a consequence, we observe that the conditions mentioned in Th. 4 of SOBEL [21] need not be necessary for an AEP.

4. STOCHASTIC GAMES WITH A FINITE STATE AND ACTION SPACE

In this section, we finally consider the N-person stochastic games with finite state and action space, as studied in ROGERS [14] and SOBEL [21].

We first need the following supplementary notations:
Let $A = \{1,\dots,K\}$ and let δ^i_{sk}, for any policy $\underline{\delta} \in \underline{\Delta}$, denote the probability with which the kth alternative $(1\le k\le K)$ is chosen by player i when entering state $s \in S$.

For any policy $\underline{\delta} \in \underline{\Delta}$, we define the fundamental matrix $Z(\underline{\delta}) = [I-P(\underline{\delta})+P^*(\underline{\delta})]^{-1}$ and for each $i \in \psi$ the bias-vector $w^i(\underline{\delta})$ by (cf. BLACKWELL [3]):

$$w^i(\underline{\delta})_s = \sum_t Z(\underline{\delta})_{st}[r^i(t;\underline{\delta}(t))-g^i(\underline{\delta}^{(\infty)};t)].$$

Observe that for each $\underline{\delta} \in \underline{\Delta}$, $g^i(\underline{\delta}^{(\infty)};s) = \sum_t P^*(\underline{\delta})_{st}\, r^i(t;\underline{\delta}(t))$ for all $i \in \psi$, $s \in S$, and that: (cf. [3])

$$(4.1) \qquad V^i_\alpha(\underline{\delta}^{(\infty)};s) = \frac{g^i(\underline{\delta}^{(\infty)};s)}{1-\alpha} + w^i(\underline{\delta})_s + o^i(\alpha;\underline{\delta})_s, \text{ for all } i \in \psi,\ s \in S,\ \alpha \in [0,1),$$

where $|o^i(\alpha;\underline{\delta})_s|$ decreases monotonically to zero as $\alpha \uparrow 1$.

Denote by $n(\underline{\delta})$ the number of subchains (closed, irreducible sets of states) for $P(\underline{\delta})$ and let $C^m(\underline{\delta})$ indicate the mth subchain $(1\le m\le n(\underline{\delta}))$. Finally, let $\underline{\Delta}_p \subseteq \underline{\Delta}$ denote the *finite* set of pure and stationary policies and define (cf. SCHWEITZER & FEDERGRUEN [19]):

(4.2) $$R^* = \{s \mid s \in R(\underline{\delta}) \text{ for some policy } \underline{\delta} \in \underline{\Delta}_p\},$$

the set of states that are recurrent under some pure policy.

Although the existence of an α-DEP is always guaranteed, it is known from a well-known counterexample by GILLETTE [7] that even in the two person - zero sum case an AEP does not need to exist when for some of the policies $\underline{\delta}^{(\infty)} \in \underline{\Delta}$, $P(\underline{\delta})$ is multichained (i.e. $n(\underline{\delta}) \geq 2$). This seeming contrast with the Markov Decision Processes (MDPs) with finite state and action space is explained by the fact that in stochastic games, as distinct from the former, an essential use is made of the set of all randomized actions, whereas in addition the above result perfectly corresponds with what is known to be the case in MDPs with a finite state space, but arbitrary compact action space (cf. BATHER [2]). Under the assumption that for each $\underline{\delta}^{(\infty)} \in \underline{\Delta}_p$, $P(\underline{\delta})$ is unichained, the existence of an AEP was first proved in ROGERS [14] and SOBEL [21]. Moreover, in SOBEL [21], as a still stronger property, the existence of a (g,w)- or bias-equilibrium policy $\underline{\delta}^* \in \underline{\Delta}$ was treated, which we believe should be defined as an AEP $\underline{\delta}^*$, for which:

(4.3) $$w^i(\underline{\delta}^*)_s \geq w^i(\underline{\eta})_s \text{ for all } i \in \psi,\ s \in S \text{ and } \underline{\eta} \in \Pi^{-i}(\underline{\delta}^*) \cap \Pi_{AEP}(\underline{\delta}^*),$$

where

$$\Pi_{AEP}(\delta^*) = \{\underline{\eta} \in \underline{\Pi} \mid g^i(\underline{\eta})_s = g^i(\underline{\delta})_s \text{ for all } s \in S\}$$

(the Definition 3 in [21] does not extend the (g,w)-optimality notion in Markov Decision Theory; moreover, with the definition in [21], a (g,w)-optimal policy does not even need to exist in the case N = 1, i.e. in the case of an MDP).

In SOBEL [21], the question of the existence of a (g,w)-equilibrium policy was treated using the Brouwer fixed-point theorem with respect to the point-to-point mapping $\Phi\colon \underline{\Delta} \to \underline{\Delta}$, with for all $i \in \psi$, $s \in S$ and $k \in A$:

$$\Phi(\underline{\delta})^i_{sk} = (\delta^i_{sk} + \phi^i_{sk}(\underline{\delta}))/(1 + \textstyle\sum_{\ell \in A} \phi^i_{s\ell}(\underline{\delta})),$$

where $\phi^i_{sk}(\underline{\delta}) = a^i_{sk} + b^i_{sk} + c^i_{sk}$, and

$$(1) \qquad a^i_{sk} = \max\{0, \textstyle\sum_{t\in S} q_{st}([\delta^{-i}(s),k])g^i(\underline{\delta}^{(\infty)};t) - g^i(\underline{\delta}^{(\infty)};s)\},$$

$$(2) \qquad b^i_{sk} = \begin{cases} 0, & \text{if } \sum_s \sum_k a^i_{sk} > 0, \\ \max\{0, r^i(s;[\delta^{-i}(s),k]) + \sum_t q_{st}([\delta^{-i}(s),k])w^i(\underline{\delta})_t \\ \qquad - g^i(\underline{\delta}^{(\infty)};s) - w^i(\underline{\delta})_s\}, & \text{otherwise,} \end{cases}$$

$$(3) \qquad c^i_{sk} = \begin{cases} 0, & \text{if } \sum_s \sum_k b^i_{sk} > 0, \\ \max\{0, \sum_t q_{st}([\delta^{-i}(s),k])z^i(\underline{\delta})_t - w^i(\underline{\delta})_s - z^i(\underline{\delta})_s, & \text{otherwise,} \end{cases}$$

where $z^i(\underline{\delta}) = -Z(\underline{\delta})w^i(\underline{\delta})$.

Unfortunately, the mapping Φ may be discontinuous in $\underline{\delta}$, since the $\phi^i_{sk}(\underline{\delta})$ can be discontinuous in those $\underline{\delta}$ that satisfy, for all $i \in \psi$, $s \in S$ the functional equation:

$$(4.4) \qquad g^i(\underline{\delta}^{(\infty)};s) = \max_{k\in A} \textstyle\sum_t q_{st}([\delta^{-i}(s),k])g^i(\underline{\delta}^{(\infty)};t),$$

or the functional equation (4.5)

$$(4.5) \qquad w^i(\underline{\delta})_s + g^i(\underline{\delta}^{(\infty)};s) = \max_{k\in A}\{r^i(s;[\delta^{-i}(s),k]) + $$

$$+ \textstyle\sum_t q_{st}([\delta^{-i}(s),k])w^i(\underline{\delta})_t\},$$

but for which, in any sphere in $\underline{\Delta}$ containing $\underline{\delta}$, policies $\underline{\eta}$ can be found that do *not* satisfy (4.4) (or (4.5) respectively). (As an example consider the MDP with $S = \{1,2,3\}$, $A = \{1,2,3\}$, $q_{11}(.) = q_{22}(.) = 1$; $q_{31}(1) = q_{31}(2) = 1$; $q_{32}(3) = 1$; $r(1,.) = 1$; $r(2,.) = 0$; $r(3,1) = -M$; $r(3,2) = r(3,3) = 0$; where $M >> 0$. Let δ_x denote the policy that chooses action 1 in state 1 and 2 with probability one, and in state 3 with probability x, whereas in state 3 action 3 is chosen with probability $1 - x$. Observe that $\phi^1_{32}(\delta)$ is discontinuous in δ_1.)

Although under the assumption that for every policy $\underline{\delta} \in \underline{\Delta}_p$, $P(\underline{\delta})$ is *unichained* the proof in SOBEL [21] can be rectified in order to show the existence of an AEP (merely by redefining $\phi^i_{sk}(\delta) = b^i_{sk}$, since in this case only criterion (2) is needed), we give a different proof, which shows the existence of an AEP within a wider class of stochastic games, including certain cases with *multichained* policies.

In addition, our approach has the advantage of showing that AEPs can be obtained as limit policies from a sequence of α-DEPs with discount factor tending to one (cf. Corollary 3.3).

Observe that in both the counterexamples (to the existence of an AEP) by BATHER [2], example 2.3 and GILLETTE [7], the matrix $P^*(\underline{\delta})$ is discontinuous in $\underline{\delta} \in \underline{\Delta}$.

In fact, Theorem 3 below shows that the existence of an AEP is guaranteed, as soon as $P^*(\underline{\delta})$ is continuous in $\underline{\delta} \in \underline{\Delta}$, and that this property is met under condition B.1 below, which is an assumption upon the chain structure of the policies belonging to $\underline{\Delta}_p$.

Let $\underline{\delta}_1,\dots,\underline{\delta}_\ell$ be an enumeration of $\underline{\Delta}_p$, and consider the following equivalence relation on (cf. SCHWEITZER & FEDERGRUEN [19], proof of Th. 3.2):

$$\mathcal{C} = \{C^m(\underline{\delta}^r) \mid 1 \le r \le L;\quad 1 \le m \le n(\underline{\delta}^r)\}.$$

Let $C \simeq C'$ if there exists $\{C^{(1)} = C, C^{(2)},\dots,C^{(n)} = C'\}$ with $C^{(i)} \in \mathcal{C}$, and $C^{(i)} \cap C^{(i+1)} \neq \emptyset$, for $i = 1,\dots,n-1$.

Let $\mathcal{C}^{(1)},\dots,\mathcal{C}^{(n^*)}$ be the corresponding equivalence classes on $\mathcal{C}$, and let $R^{*(1)},\dots,R^{*n^*}$ be the corresponding partition of R^* (cf. (4.2)):

$$R^{*(\ell)} = \cup_{\{(m,r)\mid C^m(\underline{\delta}^r)\in\mathcal{C}^{(\ell)}\}}\, C^m(\underline{\delta}^r).$$

The following lemma shows that under assumption B.1, all policies in $\underline{\Delta}$ have the same number of subchains, i.e. $n(\underline{\delta})$ is constant on $\underline{\Delta}$:

B.1. Every (pure) policy $\underline{\delta} \in \underline{\Delta}_p$ has exactly one subchain within each $R^{*(\ell)}$, $\ell = 1,\dots,n^*$.

<u>LEMMA 4.1</u>. *If* B.1 *holds, then all the policies in* $\underline{\Delta}$ *have the same number of subchains.*

<u>PROOF</u>. Fix $\underline{\delta}^0 \in \underline{\Delta}$. We prove that $P(\underline{\delta}^0)$ has exactly one subchain within each $R^{*(\ell)}$ ($\ell=1,\dots,n^*$) by showing subsequently:

(1) $R(\underline{\delta}^0) \subseteq R^*$; (2) any subchain of $P(\underline{\delta}^0)$ is contained within one of the sets $R^{*(\ell)}$; (3) in every one of the sets $R^{*(\ell)}$ there is exactly one subchain of $P(\underline{\delta}^0)$.

(1) and (2) follow immediately from parts (a) and (c) of Th. 3.2 in [19], so that (3) remains to be shown.

Fix ℓ $(1 \le \ell \le n^*)$ and assume first that $R(\underline{\delta}^0) \cap R^{*(\ell)} = \emptyset$. It then follows from Lemma 2.2 in [19] that there exists a pure policy $\underline{\eta} \in \underline{\Delta}_p$, with $R(\underline{\eta}) \subseteq R(\underline{\delta}^0)$, such that $R(\underline{\eta}) \cap R^{*(\ell)} = \emptyset$, contradicting B.1. Finally, observe that for any pair $\delta_1, \delta_2 \in \underline{\Delta}_p$, the subchains of $\underline{\delta}_1$ and $\underline{\delta}_2$ that are contained within $R^{*(\ell)}$ must intersect, since it would otherwise be possible to construct a $\underline{\delta}_3 \in \underline{\Delta}_p$ with two subchains within $R^{*(\ell)}$, contradicting B.1, and verify that this property implies that $P(\underline{\delta})$ cannot have two or more subchains within $R^{*(\ell)}$.

REMARK. Assume that every policy in $\underline{\Delta}_p$ is unichained (cf. SOBEL [21], ROGERS [14]) and observe that this assumption implies for any pair $(\underline{\delta}_1, \underline{\delta}_2) \in \underline{\Delta}_p$ that their subchains must intersect, so that all the subchains in C belong to the same equivalence class, i.e. $n^* = 1$.

It hence follows that the assumption in SOBEL [21] and ROGERS [14] is identical with the special case of B.1 where $n^* = 1$.

We next introduce assumption B.2:

B.2. For every $i \in \psi$ for every pair of states $s,t \in S$, and for every combination $\{\delta^j \in \Delta \mid j \neq i\}$ of the other players, there is a policy $\delta^i \in \Delta$ for player i and an integer ℓ such that $P(\underline{\delta})^{\ell}_{st} > 0$,

which is the relaxation of assumption A3.3 to the finite space model, and which can be seen as an extension of the *communicatingness*-property (cf. BATHER [2], HORDIJK [9]). Alternatively, one might say that B.2 expresses that for every $i \in \psi$, and whatever stationary policy the other players choose, each state is accessible from every other state for player i. Finally, Theorem 3 proves, under B.1 as well as under B.2, the existence of an AEP.

THEOREM 3. *There exists a stationary* AEP, *if either* B1 *or* B2 *holds.*

PROOF. Assume first that B.1 holds. Fix $i \in \psi$, $s \in S$. It follows from Lemma 4.1 that $n(\underline{\delta})$ is constant on $\underline{\Delta}$, and hence from Th. 5 in SCHWEITZER [18] that $P^*(\underline{\delta})$ is continuous in $\underline{\delta} \in \underline{\Delta}$, which in its turn invokes, by their very definition, the continuity of $g^i(\underline{\delta}^{(\infty)};s)$ and $w^i(\underline{\delta})_s$ in $\underline{\delta} \in \underline{\Delta}$.

We first fix an α-DEP $\underline{\delta}_\alpha \in \underline{\Delta}$, for each $\alpha \in [0,1)$.

Inserting (4.1) into both sides of (1.8) and multiplying both sides of the resulting inequality by $(1-\alpha)$ we obtain for all $\underline{\eta} \in \underline{\Delta}$

(4.6)
$$g^i(\underline{\delta}_\alpha^{(\infty)};s) + (1-\alpha)w^i(\underline{\delta}_\alpha)_s + (1-\alpha)o^i(\alpha;\underline{\delta}_\alpha)_s \geq$$
$$\geq g^i([\delta_\alpha^{-i},\eta]^{(\infty)};s) + (1-\alpha)w^i([\delta_\alpha^{-i},\eta])_s + (1-\alpha)o^i(\alpha;[\delta_\alpha^{-i},\eta])_s.$$

It next follows from the fact that $\underline{\Delta}$ is a compact metric space that one can find a policy $\underline{\delta}^{*(\infty)} \in \underline{\Delta}$, and a sequence $\{\alpha_k\}_{k=1}^\infty$, with $\alpha_k \in [0,1)$ and $\lim_{k\to\infty} \alpha_k = 1$, such that $\lim_{k\to\infty} \underline{\delta}_{\alpha_k} = \underline{\delta}^*$. We further show:

(4.7)
$$\lim_{k\to\infty} (1-\alpha_k)o^i(\alpha_k;\underline{\delta}_{\alpha_k})_s = 0 = \lim_{k\to\infty} (1-\alpha_k)o^i(\alpha_k;[\delta_{\alpha_k}^{-i},\eta])_s.$$

Merely proving the first equality in (4.7) (the proof of the second one being analogous), we observe that for each $\alpha \in [0,1)$, $o^i(\alpha,\delta)_s$ is continuous in $\underline{\delta} \in \underline{\Delta}$, as a result of Lemma (2.2), relation (4.1) and the continuity of $g^i(\underline{\delta}^{(\infty)};s)$ and $w^i(\underline{\delta})_s$ in $\underline{\delta} \in \underline{\Delta}$.

(4.7) then follows from the fact that for any $\underline{\eta} \in \underline{\Delta}$, $|(1-\alpha)o^i(\alpha;\eta)_s|$ decreases monotonically to zero, as $\alpha \to 1$, using e.g. Dini's Theorem (cf. ROYDEN [16], p.162).

Finally, let k tend to infinity in both sides of (4.6) with $\alpha = \alpha_k$, and use (4.7) as well as the continuity of $g^i(\underline{\delta}^{(\infty)};s)$ and $w^i(\underline{\delta})_s$ in $\underline{\delta} \in \underline{\Delta}$, in order to obtain:

(4.8)
$$g^i(\underline{\delta}^{*(\infty)};s) \geq g^i([\delta^{*-i},\eta]^{(\infty)};s), \quad \text{for all } i \in \psi,\ s \in S \text{ and } \eta \in \Delta.$$

Consider next the "decision problem" that arises when all players but player i tie themselves down to their respective policies in $\underline{\delta}^*$, and observe from (4.8) that in this decision problem, δ^{*i} is a maximal gain policy to player i within Δ. It then follows from the proof of Theorem 2 in DERMAN [5] that δ^{*i} is also optimal within $\tilde{\Pi}$ (cf. appendix), and hence using the argument(s) in the proof of Lemma 2.3, even within Π. This proves the theorem under B.1, whereas the existence of an AEP under B.2 follows immediately from Theorem 2, B.2 being the relaxation of A3.3 to the finite space model. □

We finally turn to the question under which condition(s) a pure instead of a randomized AEP exists, for every choice of the one-step expected rewards $r^i(s;\underline{a})$.

So far the only stochastic games known to have this property are the so-called two person-zero sum games with perfect information, in which in each state of the system, one of the two players has not more than one alternative.

The existence of a pure AEP for this class of stochastic games was first treated by GILLETTE [7]. Unfortunately an incorrect extension of the Hardy-Littlewood Theorem was used, as has been pointed out by LIGGETT & LIPPMAN [11].

The existence of a pure AEP, and, as an even stronger result, the existence of a pure bias-equilibrium policy, may, however, be derived from the fact that a pure stationary α-DEP exists for each $\alpha \in [0,1)$, where the latter has already been proved by SHAPLEY [20].

Since $\underline{\Delta}_p$ is a finite set, we can therefore find a policy $\underline{\delta}^* = (\delta^{*1},\delta^{*2}) \in \underline{\Delta}_p$ and a sequence $\{\alpha_n\}_{n=1}^{\infty}$, with $\alpha_n \uparrow 1$, such that $\underline{\delta}^*$ is an α_n-DEP for $n = 1,2,\ldots$. Let $r(s;\underline{a}) = r^1(s;\underline{a}) = -r^2(s;\underline{a})$ and $V_\alpha(\underline{\eta};s) = V_\alpha^1(\underline{\eta};s) = -V_\alpha^2(\underline{\eta};s)$, and observe that $V_\alpha(\underline{\eta};s) = \sum_t [I-\alpha P(\underline{\eta})]^{-1}_{st} r(t;\underline{\eta}(t))$ is a rational function in α for all $\underline{\eta} \in \underline{\Delta}_p$ and $s \in S$.

Since $V_\alpha([\eta^1,\delta^{*2}];s) - V_\alpha(\underline{\delta}^*;s)$ and $V_\alpha([\delta^{*1},\eta^2];s) - V_\alpha(\underline{\delta}^*;s)$ are also rational functions in α, for all $\eta^1,\eta^2 \in \Delta$ and $s \in S$, and hence are either identically zero or have a finite number of zeros, there exists an $\tilde{\alpha}(\eta^1,\eta^2,s)$ such that, for all $\alpha > \tilde{\alpha}(\eta^1,\eta^2,s)$:

$$(4.9) \qquad V_\alpha([\eta^1,\delta^{*2}];s) \le V_\alpha(\delta^*;s) \le V_\alpha([\delta^{*1},\eta^2];s).$$

Since S and $\underline{\Delta}_p$ are finite, we thus obtain an α^* such that $\underline{\delta}^*$ is an α-DEP for all $\alpha > \alpha^*$. It then follows by comparing (4.1) for $\underline{\delta}^*$ and $[\eta^1,\delta^{*2}]$, as well as $\underline{\delta}^*$ and $[\delta^{*1},\eta^2]$, that $\underline{\delta}^*$ is a bias-equilibrium policy as well.

REMARK. The proof in LIGGETT & LIPPMAN [11] for the existence of a pure AEP is more complicated than the one above; moreover, it requires an additional argument. More specifically, instead of Th. 5 in BLACKWELL [3] we need the stronger result that in each Markov Decision Model there exists a discount factor α^* such that any policy that is α-optimal for some $\alpha > \alpha^*$ is α-optimal for all $\alpha > \alpha^*$, which is immediate from the proof of Th. 5. Relation (5) in [11] should be adapted in this sense.

One might wonder whether the existence of a pure AEP is also guaranteed in the case of two-person, *nonzero*-sum, or even more generally in the case of N-person games with perfect information. The following two-person game is, however, a counterexample:
Let $S = \{1,2\}$ and assume player 2 has one alternative in state 1 and player 1 has one alternative in state 2. Let

$$r^2(1;(1,1)) = r^1(2;(1,1)) = 1$$

and

$$r^2(1;(2,1)) = r^1(2;(1,2)) = -1,$$

the other rewards being zero, and let

$$q_{11}(1,1) = q_{21}(1,1) = 2/3$$

and

$$q_{11}(2,1) = q_{21}(1,2) = 1/3.$$

We finally observe that the question of whether, and under which conditions, a (g,w)-equilibrium pair of policies exists still remains to be solved.

ACKNOWLEDGEMENT. We would like to thank Henk Tijms for many helpful comments and suggestions, and Arie Hordijk for stimulating discussions.

APPENDIX

PROOF OF LEMMA 2.3. Fix $i \in \psi$, and consider the Markov Decision Problem (MDP) with the ith player as decision-maker that results from our stochastic game, when each of the other players tie themselves down to a specific policy $\pi^j \in \Pi$. Define $\tilde{\Pi}$ as the class of all policies in this MDP, i.e. the class of all rules that prescribe for each time t which randomized action $\mu \in M(A)$ to choose as a Borel-measurable function of the state s_t and the history $\tilde{H}_t = (s_0,a_0,\ldots,s_{t-1},a_{t-1})$ of the system up to t. Observe that $\tilde{\Pi}$ is a *strict* subset of Π, as a consequence of H_t, the history of the system in the *stochastic game*, embodying the realizations of the random variables (since randomized actions) $\{\pi^j(s_\tau) \mid j \neq i,\ 0 \le \tau < t\}$ next to the actions $\{a_\tau \mid 0 \le \tau < t\}$ of the ith player.

Note that

(a) $\tilde{P}(s_t \in B \mid s_{t-1}=s, a_{t-1}=a, \tilde{H}_t) = \tilde{P}(s_t \in B \mid s_{t-1}=s, a_{t-1}=a)$ for all $B \in \mathcal{B}_S$

(b) $\tilde{P}(s_t \in B \mid s_{t-1}=s, a_{t-1}=a, H_t) = \tilde{P}(s_t \in B \mid s_{t-1}=s, a_{t-1}=a)$ for all $B \in \mathcal{B}_S$,

where the tilde ~ indicates the transition probability in the above-described MDP.

Now, if $\underline{\delta}^{(\infty)}$ is an α-DEP, then $\delta^{i(\infty)}$ is an optimal strategy in the above-mentioned MDP, within the class Π, and, a fortiori, within $\tilde{\Pi}$ (cf. (1.9)). It then follows from Th. 6 part (f) of BLACKWELL [4] that $V^i_\alpha(\underline{\delta}^{(\infty)};s)$ satisfies (2.8) for all $i \in \psi$, $s \in S$. Conversely, since the transition probability distribution in the above-described MDP is independent of the "extended" history H_t, just as it is independent of $\tilde{H}_t$ (cf. (a) and (b)), it follows from a straightforward extension of the same theorem in [4] that $\delta^{i(\infty)}$ is an optimal strategy in the MDP, not only within $\tilde{\Pi}$ but also within the class Π.

PROOF OF LEMMA 3.1. Fix $\underline{\delta} \in \underline{\Delta}$, and define for any $t \in S$, and $n = 1,2,\dots$:

$$M_t^n = \sup_{s\in S} P^n(\underline{\delta})_{st}, \quad \text{and} \quad m_t^n = \inf_{s\in S} P^n(\underline{\delta})_{st}.$$

From $P(\underline{\delta})_{st}^{n+1} = \sum_{u=1}^{\infty} P(\underline{\delta})_{su} . P(\underline{\delta})_{ut}^{n}$, it follows that for all $t \in S$:

$$(3.6) \qquad m_t^n \le m_t^{n+1} \le M_t^{n+1} \le M_t^n, \qquad \text{for all } n \ge 1.$$

Now, fix s,u and $n > \nu$. For any number a, let $a^+ = \max(a,0)$, and $a^- = -\min(a,0)$. Then using the facts that $(a-b)^+ = a - \min(a,b)$ and $\sum_k a_k^+ = \sum_k a_k^-$ when $\sum a_k = 0$, we get, for any $t \in S$:

$$P^n(\underline{\delta})_{st} - P^n(\underline{\delta})_{ut} = \sum_{k=1}^{\infty} [P^\nu(\underline{\delta})_{sk} - P^\nu(\underline{\delta})_{uk}]\, P^{n-\nu}(\underline{\delta})_{kt} =$$

$$= \sum_{k=1}^{\infty} [P^\nu(\underline{\delta})_{sk} - P^\nu(\underline{\delta})_{uk}]^+ P^{n-\nu}(\underline{\delta})_{kt} - \sum_{k=1}^{\infty} [P^\nu(\underline{\delta})_{sk} - P^\nu(\underline{\delta})_{uk}]^- P^{n-\nu}(\underline{\delta})_{kt}$$

$$\le \sum_{k=1}^{\infty} [P^\nu(\underline{\delta})_{sk} - P^\nu(\underline{\delta})_{uk}]^+ \{M_t^{n-\nu} - m_t^{n-\nu}\}$$

$$= [1 - \sum_{k=1}^{\infty} \min\{P^\nu(\underline{\delta})_{sk}, P^\nu(\underline{\delta})_{uk}\}]\ \{M_t^{n-\nu} - m_t^{n-\nu}\} \le (1-\rho)(M_t^{n-\nu} - m_t^{n-\nu}).$$

Since s and w were arbitrarily chosen, it now follows that for all $t \in S$:

$$M_t^n - m_t^n \le (1-\rho)(M_t^{n-\nu} - m_t^{n-\nu}).$$

Iterating this inequality and using the fact that $M_t^k - m_t^k \le 1$ for all $t \in S$ and $k = 1,2,\dots$ we obtain:

$$(3.7) \qquad M_t^n - m_t^n \le (1-\rho)^{[n/\nu]}.$$

Together, (3.6) and (3.7) prove that for any $t \in S$ there is a finite number $\pi_t \ge 0$ such that $\{M_t^n\}_{n=1}^{\infty} \downarrow \pi_t$, and $\{m_t^n\}_{n=1}^{\infty} \uparrow \pi_t$. It then follows from $m_t^n \le P(\underline{\delta})_{st}^n \le M_t^n$ that $\lim_{n\to\infty} P(\underline{\delta})_{st}^n$ exists and hence, $\pi_t = \lim_{n\to\infty} P(\underline{\delta})_{st}^n =$

$P^*(\underline{\delta})_{st} = \pi(\underline{\delta})_t$. Finally, this equality, together with $m_t^n \leq \pi_t$, $P^n(\underline{\delta})_{st} \leq M_t^n$ and (3.7) imply (3.5).

REFERENCES

[1] ANTHONISSE, J. & H. TIJMS, *On the stability of products of stochastic matrices*, to appear in J. Math. Anal. Appl.

[2] BATHER, J., *Optimal decision procedures for finite Markov chains, Parts I, II, III*, Adv. in Applied Prob. 5 (1973) 328-340, 521-540, 541-554.

[3] BLACKWELL, D., *Discrete dynamic programming*, Ann. Math. Stat. 36 (1962) 719-726.

[4] BLACKWELL, D., *Discounted dynamic programming*, Ann. Math. Stat. 39 (1968) 226-235.

[5] DERMAN, C., *Finite State Markovian Decision Processes*, Academic Press, New York, 1970.

[6] DOOB, J., *Stochastic Processes*, Wiley, New York, 1953.

[7] GILLETTE, D., *Stochastic games with zero stop probabilities*, in: M. Dresher et al. (eds), Contributions to the Theory of Games, Vol. III (Princeton Univ. Press, Princeton, N.J., 1957) 179-188.

[8] GLICKSBERG, I., *A further generalization of the Kakutani fixed point theorem with application to Nash equilibrium points*, Proc Amer. Math. Soc. 3 (1952) 170-174.

[9] HORDIJK, A., *Dynamic Programming and Markov Potential Theory*, MC Tract 51, Mathematisch Centrum, Amsterdam, 1974.

[10] KURATOWSKI, *Topology I*, 2nd rev. ed., Academic Press, New York, 1952.

[11] LIGGETT, T. & S. LIPPMAN, *Stochastic games with perfect information and time average payoff*, SIAM Review 11 (1969) 604-607.

[12] MAITRA, A. & T. PARTHASARATHY, *On stochastic games*, JOTA 5 (1970) 289-300.

[13] PARTHASARATHY, K., *Probability Measures on Metric Spaces*, Academic Press, New York, 1967.

[14] ROGERS, P., *Nonzero-sum stochastic games*, Report ORC69-8, Operations Res. Center, Univ. of California, Berkeley, Calif., 1969.

[15] ROSS, S.M., *Applied Probability Models with Optimization Applications*, Holden-Day, San Francisco, Calif., 1970.

[16] ROYDEN, H., *Real Analysis*, 2^{nd} ed., MacMillan, New York, 1968.

[17] SCHÄL, M., *Conditions for optimality in dynamic programming and for the limit of n-stage optimal policies to be optimal*, Z. Wahrscheinlichkeitstheorie Verw. Gebiete 32 (1975) 179-196.

[18] SCHWEITZER, P., *Perturbation theory and finite Markov chains*, J. Appl. Prob. (1968) 401-413.

[19] SCHWEITZER, P. & A. FEDERGRUEN, *The functional equations of undiscounted Markov renewal programming*, to appear (1975).

[20] SHAPLEY, L., *Stochastic games*, Proc. Nat. Acad. Sci. U.S.A. 39 (1953) 1095-1100.

[21] SOBEL, M., *Noncooperative stochastic games*, Ann. of Math. Stat. 42 (1971) 1930-1935.

[22] SOBEL, M., *Continuous stochastic games*, J. Appl. Prob. 10 (1973) 597-604.

[23] TAYLOR, H., *Markovian sequential replacement processes*, Ann. Math. Statist. 36 (1965) 1677-1694.

[24] TIJMS, H., *On dynamic programming with arbitrary state space compact action space and the average return as criterion*, Report BW 55/75, Mathematisch Centrum, Amsterdam, 1975.

[25] VRIEZE, O., to appear.