stichting
mathematisch
centrum

AFDELING MATHEMATISCHE BESLISKUNDE BW 66/76 NOVEMBER
(DEPARTMENT OF OPERATIONS RESEARCH)

O.J. VRIEZE

THE STOCHASTIC NON-COOPERATIVE COUNTABLE-PERSON GAME WITH COUNTABLE STATE SPACE AND COMPACT ACTION SPACES UNDER THE DISCOUNTED PAYOFF CRITERIUM

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

The stochastic non-cooperative countable-person game with countable state space and compact action spaces under the discounted payoff criterium.

by

O.J. Vrieze

ABSTRACT

This paper considers stochastic non-cooperative countable-person games with countable state space and compact action spaces. Under some continuity assumptions on the payoff functions and the transition probabilities it is shown in section 2 that under the discounted criterium there exists an equilibrium point within the class of stationary strategies.

Next in section 3 we show for the most general class of stochastic games, under the only condition that the payoffs should be uniform bounded, that, if there exists an equilibrium point within a subclass of the stationary strategies, which has the property that all the players pure stationary strategies belong to these subclass, then this point is also equilibrium point within the most general class of strategies.

KEY WORDS & PHRASES: *Stochastic game, non-cooperative-game, countable-person game, discounted payoff model, equilibrium point, optimal stationary strategies, most general stochastic game with most general strategy spaces.*

1. INTRODUCTION

This paper treats a countable-person non-cooperative stochastic game, specified by a five-tuple $\Gamma = (I,S,A,g,P)$

I : the set of players

S : the state space

$A = \underset{s\in S}{X}\ \underset{i\in I}{X}\ A_i(s)$, where $A_i(s)$ is the set from which player i in state s will take his actions.

$g = \{g_i \mid i \in I\}$. Let $A(s) = \underset{i\in I}{X}\ A_i(s)$ and $SA = \{(s,a) \mid a \in A(s)\}$ then $g_i : SA \to \mathbb{R}^1$, $\forall i \in I$.
$g_i(s,a(s))$ is the immediate payoff to player i if in state s the joint players action is $a(s) \in A(s)$.

$P = \{p(H|s, a(s)) \mid \forall H \in \mathcal{B}_S, \forall (s,a(s)) \in SA\}$, where $\mathcal{B}_S$ is the σ-algebra of Borel subsets of S. P is the set of transition probabilities. For each $(s,a(s)) \in SA$ is $p(\cdot|s,a(s))$ a probability measure on S defined for each $H \in \mathcal{B}_S$. $p(H|s,a(s))$ denotes the probability that the next state is an element of H, if in state s the players joint action is a(s).

We make the following assumptions on these game parameters:

A_1 : I is a countable set

A_2 : S is a countable set

A_3 : $A_i(s)$ is a compact subset of some metric space, $\forall s \in S$, $\forall i \in I$.

A_4 : $g_i(.,.) \in V(SA)$ (V(SA) will be defined below)
and $\sup_{SA} |g_i(s,a(s))| \le M$, $\forall i \in I$.

A_5 : $p(s'|.,.) \in V(SA)$, $\forall s' \in S$.

Let M(S) be the class of all bounded real valued functions on S. Let C(A(s)) be the class of continuous real-valued functions on $A_i(s)$, $\forall s \in S$.
Let V(SA) be the class of functions such that $f \in V(SA)$ iff $f : SA \to \mathbb{R}^1$ and $f(\cdot,a(\cdot)) \in M(S)$, $\forall a(\cdot) \in A$ and $f(s,\cdot) \in C(A(s))$, $\forall s \in S$.

Behaviorally the game runs as follows:
An initial state $s_0 \in S$ is specified; simultaneously each player $i \in I$ chooses an action $a_i(s_0) \in A_i(s_0)$; to player i there takes place an immediate payoff $g_i(s_0,a(s_0))$ where $a(s_0) = (a_1(s_0),a_2(s_0),\ldots\ldots)$;

the game moves according to the probability measure $p(.|s_0,a(s_0))$ to a new state s_1, which may be viewed as a new starting state, etc.

Let $M_i(s)$ be the set of all Borel subsets of $A_i(s)$, $\forall s \in S$, $\forall i \in I$. We now define a mixed action for player i in state s as a probability measure $\mu_i(s)$ on $A_i(s)$, defined for each $H \in M_i(s)$.
Let $N_i(s)$ be the set of all mixed actions for player i in state s, $\forall s \in S$, $\forall i \in I$, endowed with the weak topology (see e.g. PARTHASARATHY [7], page 39). Since $A_i(s)$ is compact metric it follows from PARTHASARATHY [6] (theorem 6.4, page 45) that $N_i(s)$ is compact in the weak topology and can be metrized. Let $N(s) = \underset{i \in I}{X} N_i(s)$. From Tychonoff's theorem (e.g. ROYDEN [9] page 166) we see that $N(s)$ is compact in the product topolgy. $N(s)$ is the set of product measures defined on the product σ-algebra on $A(s)$. As $N_i(s)$ is compact and can be metrized we note from KELLEY [6] (theorem 17, page 125) that the weak topology on $N_i(s)$ satisfies the first axiom of countability. But then also $N(s)$ satisfies the first axiom of countability, which ensures that the topology of $N(s)$ can be characterized by sequences (see KELLEY [6], theorem 8, page 72). An element of $N(s)$ will be denoted by $\mu(s) = (\mu_1(s),\mu_2(s),\ldots\ldots)$. If in state s the players joint mixed actions are $\mu(s)$, then the expected immediate payoff $g_i(s,\mu(s))$ to player i equals:

1.1. $$g_i(s,\mu(s)) = \int_{A(s)} g_i(s,a(s)\ d\mu(s) \qquad \forall i \in I.$$

In the same way we can express the expectation of the transition probabilities under $\mu(s) \in N(s)$:

$$p(s'|s,\mu(s)) = \int_{A(s)} p(s'|s,a(s))\ d\mu(s), \qquad \forall s' \in S.$$

A stationary strategy for player i is a mapping μ_i defined on S, such that $\mu_i(s) \in N_i(s)$, i.e. every time the play stays in state s, player i plays mixed action $\mu_i(s)$, $\forall s \in S$.
A joint stationary strategy for the players is a mapping μ defined on S, such that $\mu(s) = (\mu_1(s),\mu_2(s),\ldots\ldots) \in N(s)$, $\forall s \in S$.

Let $N_i = \underset{s\in S}{X} N_i(s)$ and let $N = \underset{s\in S}{X} N(s)$.

In the same way as with the space $N(s)$ above we can show that both N_i and N are compact in the product topology and both topologies satisfy the first axiom of countability, so both topologies are characterized by sequences. An element of N_i represents a stationary strategy for player i and conversely. An element of N represents a joint stationary strategy for the players and conversely.

In section 2 we only consider stationary strategies. In section 3 we also allow non-stationary strategies and they will be defined there.

Let $F_i(s)$ be the set of all finite signed measures on $A_i(s)$, defined for each $H \in M_i(s)$ endowed with the weak topology. In VRIEZE [11], lemma 2.1, it is proven that $F_i(s)$ is a linear topological space.
Let $F(s) = \underset{i\in I}{X} F_i(s)$ and $F = \underset{s\in S}{X} F(s)$, then F is in the product topology again a linear Hausdorff topological space and N is a compact convex subset of F.

In this paper we consider the discounted model, so there is specified a discount factor $\beta \in [0,1)$ i.e. a payoff on the t^{th}-step to player i will be discounted with a factor β^t, $t = 0,1,2,\dots$.
Let $V_i^t(s_0,\mu)$ denote the expected payoff to player i at the t^{th}-step when the play is started in s_0 and the players use stationary strategy $\mu \in N$, $t = 0,1,2,\dots$.
Let $V_i(s_0,\mu) = \sum_{t=0}^{\infty} \beta^t V_i^t(s_0,\mu)$, then $V_i(s_0,\mu)$ represents the total expected discounted payoff to player i when the play starts in s_0 and the players use stationary strategy $\mu \in N$.

Note that $|V_i(s,\mu)|$ is uniform bounded by $\frac{M}{1-\beta}$, for from assumption A_4 it follows that $|V_i^t(s,\mu)| \leq M$, so that $|V_i(s,\mu)| \leq \sum_{t=0}^{\infty} \beta^t \mid V_i^t(s,\mu)| \leq \frac{M}{1-\beta}$.

<u>DEFINITION</u>: An element $\mu^* \in N$ is called an equilibrium point iff $V_i(s,\mu^*) \geq V_i(s,(\mu^{-i^*},\mu_i))$, $\forall \mu_i \in N_i$, $\forall i \in I$, $\forall s \in S$, where (μ^{-i^*},μ_i) denotes the joint stationary strategy with player i playing μ_i and the other players playing according to μ^*.

In section 2 we show the existence of an equilibrium point for the game Γ as specified above, within the class of stationary strategies.

In section 3 we show that this point is also equilibrium point in the most general class of strategies, namely the class of behavioral strategies.

An extension of the model presented here is earlier studied by SOBEL [10] and FEDERGRüN [5]. The extension concerns the state space. They both considered a compact state space. Sobel's set of players is arbitrary and Federgrün's set of players is finite. However they both made a mistake in their proofs of the existence of an equilibrium point and for both it holds that it is very hard to rectify their proofs, if possible. Therefore nowadays it is unclear whether and under which conditions there exists an equilibrium point in a stochastic game model with the state space uncountable. When in the model of Federgrün we take the statespace countable then his proof goes on. When we do the same in Sobel's method then there must be made further restrictions on the other game parameters to come to the existence of an equilibrium point.
One of the main reasons why we present our method is that in our proof we do not need to use a selection theorem. By □ the end of a proof will be denoted.

2. EXISTENCE OF AN EQUILIBRIUM POINT WITHIN THE CLASS N.

We start with two continuity lemma's.

LEMMA 2.1. *For every function* $f(.,.) \in V(SA)$ *the function* $f(s,\mu(s))$ *defined on* $SN = \{(s,\mu(s)) \mid \mu(s) \in N(s)\}$ *as*

$$f(s,\mu(s)) = \int_{A(s)} f(s,a(s))\, d\mu(s) \quad \text{is for each } s \in S$$

a continuous function in $\mu(s)$ *on* $N(s)$.

PROOF. For fixed $s \in S$ the proof is analogue to the proof of Lemma 2.2 in VRIEZE [11], which is an extension of the proof of Lemma 2.1 in PARTHASARATHY & MAITRA [8], which is also used by FEDERGRüN [5].
Therefore the proof will not be repeated here. □

LEMMA 2.2. *The total expected discounted payoff function* $V_i(s,\mu)$ *for player* i *is for fixed* s *a continuous function on* N, $\forall s \in S$, $\forall i \in I$.

PROOF: The proof is quite analogue to the proof of Lemma 2.2 in FEDERGRüN [5]. However, as a referee of his paper pointed out, because their strategy space is an uncountable product of metrisable topological spaces, the product topology of this space need not be characterized by sequences, so for his case the proof of Lemma 2.2 fails.
However, in our case the product topology of the strategy space is characterized by sequences as we already pointed out in section 1, so in our case the proof of his Lemma 2.2 can be applied and will therefore not be repeated here. □

We now define a mapping $T : N \to N$ (analogue to the mapping T in VRIEZE [11]) which is the key to the proof of the existence of an equilibrium point within the class N.

In assumption A_3 we stated that $A_i(s)$ is a compact subset of some metric space so there exists a countable subset $d_i(s) = \{d_{in}(s),\ n = 1,2,\ldots\} \subset A_i(s)$, which lies dense in $A_i(s)$ (cf. ROYDEN [9], proposition 13, page 163).
Let $\lambda_i^s \in N_i(s)$ be the probability measure on $A_i(s)$ such that $\lambda_i^s(d_{in}(s)) = 2^{-n}$, $n = 1,2,\ldots$, so λ_i^s is concentrated on $d_i(s)$.
Choose $\mu \in N$ and let $V_i(s,\mu)$ be the total expected discounted payoff to player i when the game starts in state s.
$V_i(s,\mu)$ satisfies:

2.1. $$V_i(s,\mu) = g_i(s,\mu(s)) + \beta\int_S V_i(s',\mu)p(ds'|\,s,\mu(s)), \quad \forall s \in S$$

This can be seen from e.g. BLACKWELL [2], page 231, DENARDO [3], page 166.
Let

2.2. $$y_i(s,\mu,a_i(s)) = g_i(s,(\mu^{-i}(s),a_i(s)) + \beta\int_S V_i(s',\mu)p(ds'|s,(\mu^{-i}(s),a_i(s)))$$

$$\forall a_i(s) \in A_i(s),\ \forall s \in S,\ \forall i \in I$$

Here $(\mu^{-i}(s),a_i(s))$ is the joint action in state s where player i plays a_i and the other players play according to $\mu(s)$. Note that $y_i(s,\mu,a_i(s)$ is the total expected discounted payoff to player i if the play starts in state s, the joint players action in this starting state is $(\mu^{-i}(s),a_i(s))$ and after the first step the players play according to μ.

From assumption A_5 and Lemma 2.1 we see that $p(s'|s,(\mu^{-1}(s),a_i(s)))$ for fixed s',s and $a_i(s)$ is a continuous function in $\mu(s)$ (in μ) on $N(s)$ (on N). Since $V_i(s,\mu)$ is continuous in μ and uniform bounded we may apply proposition 18, page 232 of ROYDEN [9] on the expression $\int_S V_i(s',\mu)p(ds'|s,(\mu^{-i}(s),a_i(s)))$ to conclude that this expression is continuous in μ on N for fixed $s \in S, a_i(s) \in A_i(s)$.

From assumption A_5 and Lemma 2.1 it follows that $g_i(s,(\mu^{-i}(s),a_i(s)))$ is continuous in μ for fixed $s \in S$ and $a_i(s) \in A_i(s)$, and so we may conclude that $y_i(s,\mu,a_i(s))$ is continuous in μ on N for fixed $s \in S$ and $a_i(s) \in A_i(s)$. In the same way we can show that $y_i(s,\mu,a_i(s))$ is a continuous function in $a_i(s)$ on $A_i(s)$ for fixed $s \in S$ and $\mu \in N$.
Let

2.3. $$\phi_i(s,\mu,a_i(s)) = \max\{0, y_i(s,\mu,a_i(s)) - V_i(s,\mu)\}, \quad \forall s \in S, \ \forall a_i(s) \in A(s), \ \forall i \in I$$

Let

2.4. $$\Phi_i(s,\mu,H) = \int_H \phi_i(s,\mu,a_i(s))d\lambda_i^s = \sum_{d_{in}(s)\in H} \phi_i(s,\mu,d_{in}(s))\lambda_i^s(d_{in}(s)) \quad \forall s \in S, \ \forall H \in M_i(s), \ \forall i \in I$$

The mapping $T : N \to N$ will now be defined as:

2.5. $$(T\mu)_i(s)(H) = \frac{\mu_i(s)(H) + \Phi_i(s,\mu,H)}{1 + \Phi_i(s,\mu,A_i(s))}, \quad \forall s \in S, \forall H \in M_i(s), \forall i \in I.$$

We now succesively prove two properties of this mapping T.

<u>PROPERTY 1</u> : *For every* $\mu \in N$ *we have* $T\mu \in N$.

<u>PROPERTY 2</u> : $T : N \to N$ *is a continuous mapping and possesses at least one fixed point.*

Proof of Property 1: From (2.5) we see that $(T\mu)_i(s)(.)$ is a probability measure on $A_i(s)$ defined for each Borel subset $H \in M_i(s)$, so it follows that $T\mu \in N$. □

Proof of Property 2: We have already shown that $y_i(s,\mu,a_i(s))$ is continuous in μ for fixed $s \in S$ and $a_i(s) \in A_i(s)$ and that $y_i(s,\mu,a_i(s))$ is continuous in $a_i(s)$ for fixed $s \in S$ and $\mu \in N$.
In exactly the same way as the proof of property 1 in VRIEZE [11] it follows that $\phi_i(s,\mu,a_i(s))$ is continuous in μ and $a_i(s)$, next that $\Phi_i(s,\mu,H)$ is continuous in μ for each $s \in S$, $H \in M_i(s)$ and from the definition (2.5) we then see that $(T\mu)_i(s)(.)$ is continuous in the weak topology in μ on N, for $\forall s \in S$, $\forall i \in I$. So if $\mu_n \to \mu_0$ on N in the product topology, then $T\mu_n \to T\mu_0$ on N in the product topology. Applying the Schauder-Tychonoff Theorem (e.g. DUNFORD & SCHWARTZ [4], page 456) gives us then the existence of a fixed point in N for the above defined mapping T.
A more detailed version of this proof one can find in VRIEZE [11] (proof of Property 1). □

THEOREM 2.1. : *The stochastic game* $\Gamma = (I,S,A,g,p)$ *under the assumptions* A_1,A_2,A_3,A_4 and A_5 *possesses an equilibrium point within the class of stationary strategies.*

PROOF : Let μ^* be fixed point of the above defined mapping T. And let $V_i(s,\mu^*)$ be the payoff to player i when play starts in state s.
We are going to prove:

$$T\mu^* = \mu^* \Leftrightarrow \mu^* \text{ is equilibrium point.}$$

Quite analogue to VRIEZE [11] (proof of Property 2) it follows that

$$2.6. \qquad T\mu^* = \mu^* \Leftrightarrow V_i(s,\mu^*) = \max_{a_i(s)\in A_i(s)} \{g_i(s,(\mu^{-i^*}(s),a_i(s))) + \\ + \beta\int_S V_i(s',\mu^*)p(ds'|s,(\mu^{-i^*}(s),a_is)))\} \quad \forall s \in S, \quad \forall i \in I.$$

We should prove:

2.7. μ^* equilibrium point $\Leftrightarrow$

$$V_i(s,\mu^*) = \max_{a_i(s)\in A_i(s)} \{g_i(s,(\mu^{-i^*}(s),a_i(s))) + \beta \int_S V_i(s',\mu^*)p(ds'|s,(\mu^{-i^*}(s),a_i(s)))\} \quad \forall s \in S,\ \forall i \in I.$$

Fix the strategies of all the players on μ^* except for player i. Then we can in the usual way define a Markovian decision problem for player i, namely: Statespace S, pure action space $A_i(s)$ in state $s \in S$, immediate payoff $g_i(s,(\mu^{-i*}(s),a_i(s)))$ in state s when he plays pure action $a_i(s)$ and transition probabilities $p(s'|s,(\mu^{-i*}(s),a_i(s)))$.

It is only a matter of substitution to see that, if player i plays in the above Markovian decision problem stationary strategy $\mu_i \in N_i$, then he yields a payoff $W_i(s,\mu_i)$ which is the same as the payoff $V_i(s,(\mu^{-i*},\mu_i))$, which he would get if in the game Γ player i plays μ_i and the other players playing according to μ^*.

BLACKWELL [2] has shown that for the above mentioned Markovian decision problem a stationary strategy is then and only then optimal for the discounted criterium if his payoff satisfies the optimality criterium.

So here $\mu_i \in N_i$ is optimal for the above Markovian decision problem iff:

2.8.
$$W_i(s,\mu_i) = \max_{a_i(s)\in A_i(s)} \{g_i(s,(\mu^{-i^*},a_i(s))) + \int_S W_i(s,\mu_i)p(ds'|s,(\mu^{-i^*},a_i(s)))\} \quad \forall s \in S$$

As $W_i(s,\mu_i) = V_i(s,(\mu^{-i^*},\mu_i))$ we see from (2.6) and (2.8) that:

2.9. $T\mu^* = \mu^* \Leftrightarrow \mu_i^*$

is optimal for his Markovian decision problem when the other players are fixed at μ^{-i*}, $\forall i \in I$.

So from (2.9) we see:

2.10. $$T\mu^* = \mu^* \Leftrightarrow V_i(s,\mu^*) = V_i(s,(\mu^{-i^*},\mu_i^*)) = W_i(s,\mu_i^*) \geq W_i(s,\mu_i) = $$
$$= V_i(s,(\mu^{-i^*},\mu_i)), \qquad \forall i \in I, \ \forall \mu_i \in N_i.$$

which is the same as

$$T\mu^* = \mu^* \Leftrightarrow \mu^* \text{ equilibrium point. } \square$$

3. EXTENSION TO GENERAL STRATEGY SPACES

In this section we impose milder condition on the parameter of the game $\Gamma = (I,S,A,g,P)$.

B_1 : I is an arbitrary set of players.

B_2 : S is an arbitrary state space, with defined an arbitrary σ-algebra B_S on it.

B_3 : $A = \underset{s\in S}{X} \ \underset{i\in I}{X} A_i(s)$, where $A_i(s)$, the set of pure actions for player i in state s, is an arbitrary space.

B_4 : $g = \{g_i \mid i \in I\}$ where g_i is the payoff function for player i, i.e. if in state $s \in S$ the joint players action is $a(s) \in \underset{i\in I}{X} A_i(s)$, then g_i specifies an immediate payoff $g_i(s,a(s))$ to player i. The only condition on the g_i's is $\sup_{I,S,A} |g_i(s,a(s))| = M$.

B_5 : $P = \{p(H|s,a(s)) \mid \forall H \in B_S, \ \forall s \in S, \ \forall a(s) \in \underset{i\in I}{X} A_i(s)\}$ P is the set of transition probabilities, i.e. if in state s the joint players actions are $a(s) \in \underset{i\in I}{X} A_i(s)$, then P specifies probabilities $p(H \mid s,a(s))$ for each set $H \in B_S$ i.e. $p(H|s,a(s))$ is the probability that the next state belongs to H if in state s joint action a(s) has taken place. No further condition on P is made.

What we intend to do in this section is the following:
Assume that in the stochastic game model specified by B_1,B_2,B_3,B_4 and B_5, we have found in the discounted case an equilibrium point within a later

to be specified subclass of the stationary strategies then we show that this point is also equilibrium point in the most general class of strategies.
So there is specified a discount factor $\beta \in [0,1)$. Let $M_i(s)$ be a σ-algebra on $A_i(s)$, $\forall i \in I$, $\forall s \in S$. Let $A(s) = \underset{i \in I}{X} A_i(s)$ and let $M(s)$ be the product σ-field on $A(s)$. Let $N_i(s)$ be the set of probability measures on $A_i(s)$ defined for each $B \in M_i(s)$. Let $N(s) = \underset{i \in I}{X} N_i(s)$. $N(s)$ is the set of product probability measures defined on the product σ-field $M(s)$.
An element $\mu_i(s) \in N_i(s)$ should be viewed as a randomized action for player i in state s and an element $\mu(s) \in N(s)$ should be viewed as a joint action of the players in state s.

Let h_t be the history of the game until time t, i.e.

$$h_t = (s_0, a_0(s_0), s_1, \ldots, a_{t-1}(s_{t-1})) \quad \text{where } a(s_j) \in \underset{i \in I}{X} A_i(s_j)$$

We are now going to define the most general strategy space for the players. As a result of AUMANN [1] we may restrict ourselves to behavioral strategies, without loss of any generality, when we quite naturaly assume that the play is of perfect recall, i.e. each player remembers at every time t the whole history h_t and knows exactly in what state s_t he is.

A behavioral strategy for player i is a strategy μ_i which specifies for each t, each history h_t and each state s_t an element $\mu_i(t,h_t,s_t) \in N_i(s_t)$.

Let Π_i be the set of all strategies for player i and let $\Pi = \underset{i \in I}{X} \Pi_i$.
Then an element μ of Π should be viewed as a joint strategy for the players and $\mu(t,h_t,s_t) \in N(s_t)$.
If $\mu_i(t,h_t,s_t) \in \Pi_i$ depends for every t through h_t only on s_0, so $\mu_i(t,h_t,s_t) = \mu_i(t,s_0,s_t)$, then we speak of a semi-markov strategy for player i. If $\mu_i(t,h_t,s_t) \in \Pi_i$ depends for every t not on h_t, so $\mu_i(t,h_t,s_t) = \mu_i(t,s_t)$, then we speak of a Markov strategy.
If $\mu_i(t,h_t,s_t) \in \Pi_i$ depends only on s_t, so $\mu_i(t,h_t,s_t) = \mu_i(s_t)$, then we speak of a stationary strategy.
Note that there is a one-to-one correspondence between the set of stationary strategies for player i and the set $\underset{s \in S}{X} N_i(s)$.
It is easy to see that when we take an arbitrary $\mu \in \Pi$ then we get great measurability and integrability difficulties with the calculation of the

expected payoff to player i, $\forall i \in I$. Following DENARDO [3] page 175, we therefore use the following trick. Let T be an arbitrary topological space: let f(.) be an arbitrary function on this space T and let p(.) be a probability measure on this space T, defined for each Borel subset $B \subset T$. Let

$$K(T,f(.),p(.)) = \{v \mid v : T \to \mathbb{R}' \ ; \ f(t) \le v(t), \ \forall t \in T \text{ and } v(.) \text{ is integrable on } T \text{ with respect to } p(.)\}$$

We now define:

$$3.1. \qquad \int_T f(t)\, dp(t) \equiv \inf_{v \in K(T,f(.),p(.))} \int_T v(t)\, dp(t).$$

In what follows all integrals used are meant in the sense of 3.1.

Let $\mu(-(s_0,a_0(s_0)))$ be the joint stationary strategy generated by $\mu \in \Pi$, defined as:

$$\mu_i(-(s_0,a_0(s_0)))(t,h_t,s_t) = \mu_i(t+1,((s_0,a_0(s_0)),h_t),s_t)$$

where $((s_0,a_0(s_0)),h_t)$ is a history consisting of t+1 -states and t+1 -joint actions, $\forall s_0 \in S$, $\forall a_0(s_0) \in \underset{i \in I}{X} A_i(s_0)$, $\quad \forall i \in I$.

Define for each $s \in S$ and each $i \in I$ a sequence of mappings $\{Z_t^{si} : \Pi \to \mathbb{R}', \ t = 0,1,2,\ldots .\}$ as follows:

$$3.2. \qquad Z_0^{si}(\mu) = \int_{A(s)} g_i(s,a(s))\, d\mu_0(0,h_0,s).$$

$$3.3. \qquad Z_t^{si} = \int_{A(s)} \{ \int_S Z_{t-1}^{s'i}(\mu(-(s,a(s)))p(ds' \mid s, a(s))\}\, d\mu(0,h_0,s)$$

If μ is such, that every integral, which appears in the sequence $\{Z_t^{si}(\mu)\}$ can be calculated without making use of the sense (3.1) then we see that $Z_t^{si}(\mu)$ equals the expected payoff to player i at the t^{th} step if the play starts in state s and the players joint strategy is $\mu \in \Pi$.

Now in general we define the expected payoff to player i as
$Z^{si}(\mu) = \lim_{n\to\infty} \sum_{t=0}^{n} \beta^t Z_t^{si}(\mu)$ when the play starts in state $s \in S$ and the joint player's strategy is $\mu \in \Pi$. As

$\sup_{I,S,A} |g_i(s,a(s))| \le M$, it follows that $|Z_t^{si}(\mu)| \le M$ and

$$|Z^{si}(\mu)| \le \frac{M}{1-\beta}, \quad \forall s \in S,\ \forall i \in I,\ \forall \mu \in \Pi.$$

The concept of equilibrium point will be defined in the usual way

DEFINITION : An element $\mu^* \in \Pi$ is called an equilibrium point iff
$Z^{si}(\mu^{-i^*},\mu_i) \le Z^{si}(\mu^*)$, $\forall s \in S$, $\forall i \in I$, $\forall \mu_i \in \Pi_i$,
where (μ^{-i^*},μ_i) is the joint strategy where player i plays μ_i and the other players play according to μ^*.

Now we are able to state our main theorem of this section:

THEOREM 3.1. *Let* Q *be a class of stationary joint strategies for the players with the property that if* $\mu \in Q$, *then also* $(\mu^{-i},a_i) \in Q$, $\forall a_i \in \mathop{X}_{s\in S} A_i(s)$. *If* $\mu^* \in Q$ *is an equilibrium point within this class* Q, *then* μ^* *is an equilibrium point within the most general class of strategies* Π.

Before proving this theorem we prove a useful lemma.

LEMMA 3.1. *Let* $W_i(s) = \sup_{\mu_i \in \Pi_i} Z^{si}(\mu^{-i^*},\mu_i)$, $\forall i \in I$, $\forall s \in S$, *where* μ^* *obeys theorem* 3.1.
Then $W_i(s)$ *satisfies the equation*:

3.4. $$W_i(s) = \sup_{a_i(s)\in A_i(s)} \{ \int_{A(s)} \{g_i(s,a(s)) + \beta \int_S W_i(s')p(ds'|s,a(s))\} d(\mu^{-i^*}(s),a_i(s))\} \quad \forall s \in S,\ \forall i \in I.$$

PROOF: First note that

$$3.5. \quad Z^{si}(\mu^{-i^*},\mu_i) = \int_{A(s)} \{g_i(s,a(s)) + \beta \int_S Z^{s'i}(\mu^{-i^*},\mu_i(-(s,a(s))))p(ds'|s,a(s))\} \, d(\mu^{-i^*}(s),\mu_i(0,h_0,s)) \quad \forall s \in S.$$

From (3.4) we see that for every $\varepsilon > 0$, there exist a $\mu_i^{\varepsilon} \in \Pi_i$ such that

$$3.6 \quad Z^{si}(\mu^{-i^*},\mu_i^{\varepsilon}) \geq W_i(s) - \varepsilon, \qquad \forall s \in S, \ \forall i \in I.$$

Since $\mu_i(-(s,a(s))) \in \Pi_i$ we see from (3.4) that

$$3.7. \quad Z^{s'i}(\mu^{-i^*},\mu_i(-(s,a(s)))) \leq W_i(s'), \qquad \forall \mu_i \in \Pi_i, \ \forall s \in S, \ \forall a(s) \in A(s), \ \forall i \in I.$$

Combining (3.5), (3.6) and (3.7) yields:

$$3.8. \quad W_i(s) - \varepsilon \leq Z^{si}(\mu^{-i^*},\mu_i^{\varepsilon}) \leq \int_{A(s)} \{g_i(s,a(s)) + \beta \int_S W_i(s')p(ds'|s,a(s))\} \, d(\mu^{-i^*}(s),\mu_i^{\varepsilon}(0,h_0,s))$$

$$\leq \sup_{a_i(s) \in A_i(s)} \left\{ \int_{A(s)} \{g_i(s,a(s)) + \beta \int_S W_i(s')p(ds'|s,a(s))\} d(\mu^{-i^*}(s),a_i(s)) \right\} \quad \forall s \in S, \ \forall i \in I$$

So one part of the lemma is proved.

Consider now the expression:

$$3.9.\quad Z^{si}(\mu^{-i^*},(\mu_i,a_i(s))) = \int_{A(s)} \{g_i(s,a(s)) +$$

$$\beta\int_B Z^{s'i}(\mu^{-i^*},\mu_i(-(s,a(s))))\ p(ds'|s,a(s))\}d(\mu^{-i^*}(s),a_i(s))$$

Then $Z^{si}(\mu^{-i^*},(\mu_i,a_i(s)))$ represents the expected payoff to player i when play starts in state s, player i plays in the first step pure action $a_i(s)$ and after the first step he plays μ_i, while the other players play according to μ^* during the whole game. So by (3.4).

$$3.10.\quad Z^{si}(\mu^{-i^*},(\mu_i,a_i(s)) \le W_i(s), \qquad \forall a_i(s) \in A_i(s),\ \ \forall\mu_i \in \Pi_i.$$

Choose now μ_i^ε such that:

$$3.11.\quad Z^{s'i}(\mu^{-i^*},(\mu_i^\varepsilon(-(s,a(s)))) \ge W_i(s') - \varepsilon,\ \ \forall s' \in S,\quad \forall s \in S,$$
$$\forall a(s) \in A(s),\ \ \forall i \in I.$$

Combining (3.9), (3.10) and (3.11) yields:

$$3.12.\quad W_i(s) \ge Z^{si}(\mu^{-i^*},(\mu_i^\varepsilon,a_i(s))) \ge \int_{A(s)} \{g_i(s,a(s)) +$$

$$+ \beta\int_S (W_i(s')-\varepsilon)p(ds'|s,a(s))\}d(\mu^{-i^*}(s),a_i(s)$$

$$\ge \int_{A(s)} \{g_i(s,a(s) + \beta\int_S W_i(s')p(ds'|s,a(s))\}d(\mu^{-i^*}(s),a_i(s)) - \varepsilon$$

$$\forall a_i(s) \in A(s),\ \forall s \in S,\ \ \forall i \in I.$$

So from (3.12) we can conclude

$$3.13.\quad W_i(s) + \varepsilon \ge \sup_{a_i(s)\in A(s)} \{\int_{A(s)} \{g_i(s,a(s)) +$$

$$\beta\int_S W_i(s')p(ds'|s,a(s))\}\ d(\mu^{-i^*}(s),a_i(s))\}, \qquad \forall s \in S, \qquad \forall i \in I.$$

Combining (3.8) and (3.13) proofs the lemma. □

PROOF OF THEOREM 3.1 : Fix player i and choose $\varepsilon > 0$.
From Lemma 3.1. we see that for each s there exists an $a_i(s) \in A_i(s)$ such that:

$$3.14. \quad \int_{A(s)} \{g_i(s,a(s)) + \beta\int_S W_i(s')p(ds'|s,a(s))\}d(\mu^{-i^*}(s),a_i(s)) \geq W_i(s) - \varepsilon(1-\beta).$$

Let $Q_i = \{\mu_i \mid \mu_i \in N_i \text{ and } (\mu^{-i^*},\mu_i) \in Q\}$, so by the assumption on Q in Theorem 3.1. we have $Q_i \supset \underset{s\in S}{X} A_i(s)$.

Let μ_i^ε be the stationary strategy for player i which prescribes to play action $a_i(s)$ in state s, where $a_i(s)$ obeys (3.14), so $\mu_i^\varepsilon \in Q_i$.

Then from (3.14) and (3.2):

$$3.15. \quad Z_0^{si}(\mu^{-i^*},\mu_i^\varepsilon) + \beta\int_{A(s)}\int_S W_i(s')p(ds'|s,a(s))d(\mu^{-i^*}(s),\mu_i^\varepsilon(s)) + \varepsilon(1-\beta) \geq W_i(s), \qquad \forall s \in S.$$

Substitution of the left side of (3.15) for $W_i(s')$ in the left side of (3.14) yields :

$$3.16. \quad Z_0^{si}(\mu^{-i^*},\mu_i^\varepsilon) + \beta\int_{A(s)}\int_S \{Z_0^{s'i}(\mu^{-i^*},\mu_i^\varepsilon) + \beta\int_{A(s)}\int_S W_i(s'')p(ds''|s',a(s'))d(\mu^{-i^*}(s'),\mu_i^\varepsilon(s') + \varepsilon(1-\beta)\}\, p(ds'|s,a(s))d(\mu^{-i^*}(s),\mu_i^\varepsilon(s)) \geq W_i(s) - \varepsilon(1-\beta).$$

Or using (3.3):

$$3.17. \quad Z_0^{si}(\mu^{-i^*},\mu_i^\varepsilon) + \beta Z_1^{si}(\mu^{-i^*},\mu_i^\varepsilon) + \beta^2 I^2(W_i)(s) \geq W_i(s)-\varepsilon(1-\beta)-\varepsilon(1-\beta)\beta, \qquad \forall s \in S,$$

where

$$I^2(W_i)(s) \int_{A(s)} \int_S \int_{A(s)} \int_S W_i(s'')p(ds''|s',a(s'))d(\mu^{-i^*}(s'),\mu_i^\varepsilon(s'))\, p(ds'|s,a(s))d(\mu^{-i^*}(s),\mu_i^\varepsilon(s))$$

In the same way it is easy to prove by induction that:

3.18. $$\sum_{t=0}^{k} \beta^t Z_t^{si}(\mu^{-i^*},\mu_i^\varepsilon) + \beta^{k+1} I^{k+1}(W_i)(s) \geq W_i(s) - \varepsilon(1-\beta) \sum_{t=0}^{k} \beta^t \quad \forall s \in S.$$

Since $|W_i(s)| \leq M$, it follows $|I^{k+1}(W_i)(s)| \leq M$, $\forall s \in S$, $\forall k$.

As
$$\lim_{k\to\infty} \sum_{t=0}^{k} \beta^t Z_t^{si}(\mu^{-i^*},\mu_i^\varepsilon) = Z^{si}(\mu^{-i^*},\mu_i^\varepsilon),$$

we see from 3.18 that

3.19. $$Z^{si}(\mu^{-i^*},\mu_i^\varepsilon) \geq W_i(s) - \varepsilon, \quad \forall s \in S,$$

So for every $\varepsilon > 0$ there exists a $\mu_i^\varepsilon \in Q_i$ such that (3.19) holds.
Therefore we may conclude:

3.20. $$W_i(s) = \sup_{\mu_i \in \Pi_i} Z^{si}(\mu^{-i^*},\mu_i) = \sup_{\mu_i \in Q_i} Z^{si}(\mu^{-i^*},\mu_i) = Z^{si}(\mu^*), \quad \forall s \in S.$$

As i was arbitrary chosen (3.20) holds for every $i \in I$, so we see that μ^* is also equilibrium point within the class Π. □

THEOREM 3.2. *The equilibrium point* $\mu^* \in N$ *as found in section 2 for the game* $\Gamma = (I,S,A,g,P)$ *under the conditions* A_1,A_2,A_3,A_4 *and* A_5 *within the class* N, *is also equilibrium point in the most general strategy class for that game.*

PROOF: As the conditions B_1,B_2,B_3,B_4 and B_5 are all weaker than respectively A_1,A_2,A_3,A_4 and A_5 and as N satisfies the condition on Q in Theorem 3.1, we may apply theorem 3.1. □

ACKNOWLEDGEMENT

We would like to thank Arie Hordijk and Gerard Wanrooij for many useful comments. Actually the first stimulating discussion about the statements in section 3 is held together.

REFERENCES

[1] AUMANN, R.J., *Mixed and behavioral strategies in infinite extensive games*, Advances in Game Theorey, Ann. of Math. Stud. 52, Princeton University Press, Princeton, 1964, (627-651).

[2] BLACKWELL, D., *Discounted dynamic programming*, Annals of Math. Stat. 36, 1965, (226-235).

[3] DENARDO, E.V., *Contraction Mappings in the theory underlying dynamic programming*, SIAM Rev. 9, 1967, (165-177).

[4] DUNFORD, N. & T. SCHWARTZ, *Linear Operations, part I, General Theory*, Interscience Publishers Inc., New York, 1967.

[5] FEDERGRÜN, A., in cooperation with VRIEZE, O.J. & G.L. WANROOIJ, *On the existence of discounted and average return equilibrium policies in* N*-person games*, Report BW 57/75, Mathematisch Centrum, Amsterdam, 1975.

[6] KELLEY, J.L., *General Topology*, Van Nostrand, New York, 1967.

[7] PARTHASARATHY, K.R., *Probability measures on metric spaces*, Acad. Press, New York - London, 1967.

[8] PARTHASARATHY, T. & A. MAITRA, *On Stochastic games*, Journal of Opt. Theory and Appl. 5, 1970, (289-300).

[9] ROYDEN, H.L., *Real Analysis*, sec.ed. The McMillan Comp., New York, 1968.

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

[11] VRIEZE, O.J., *Non-cooperative countable-person games with compact action spaces*, Report BW 65/76, Mathematisch Centrum, Amsterdam, 1976.