REPEATED GAMES WITH ABSORBING STATES

FRANK THUIJSMAN
Maastricht University
Maastricht, The Netherlands

1. Introduction

In this paper we shall present a proof for the existence of limiting average ε-equilibria in non-zero-sum repeated games with absorbing states, i.e., stochastic games in which all states but one are absorbing. We assume that the action spaces shall be finite; hence there are only finitely many absorbing states. A limiting average ε-equilibrium is a pair of strategies $(\sigma_\varepsilon, \tau_\varepsilon)$, with $\varepsilon > 0$, such that for all σ and τ we have $\gamma_1(\sigma, \tau_\varepsilon) \leq \gamma_1(\sigma_\varepsilon, \tau_\varepsilon) + \varepsilon$ and $\gamma_2(\sigma_\varepsilon, \tau) \leq \gamma_2(\sigma_\varepsilon, \tau_\varepsilon) + \varepsilon$. The proof presented in this chapter is based on the publications by Vrieze and Thuijsman [7] and by Thuijsman [5]. Several examples will illustrate the proof.

Vrieze and Thuijsman [7] derived existence of limiting average ε-equilibria in non-zero-sum repeated games with absorbing states after an inspiring study on the Paris Match examined by Sorin [4]. Despite Sorin's correct observation of a gap between the set of limiting average equilibrium rewards and the set of the λ-discounted equilibrium rewards, Vrieze and Thuijsman showed that a limiting average ε-equilibrium can be derived from any arbitrary sequence of stationary λ-discounted equilibria, converging for λ going to 0. The limiting average equilibrium strategies involved are history-dependent "Big Match strategies," where a player has to adjust his mixed actions at all stages in response to the behavior of his opponent. In Thuijsman [5] a simpler class of strategies is used for the limiting average equilibria, the so-called "almost stationary strategies." An almost stationary strategy essentially consists of a stationary strategy that is played as long as no deviation by the opponent has been observed, and a (possibly history-dependent) retaliation strategy to punish the opponent in case a deviation is detected. The strategies are called almost stationary to stress the fact that if the players refrain from deviations, then, with probability close to 1, stationary strategies are used throughout the whole play. The retaliation strategies are typically taken to be ε-optimal strategies that minimize (up to ε) the limiting average reward of the opponent. For the existence of these retaliation strategies the proof relies on the existence of the limiting average value for zero-sum repeated games with absorbing states established by Kohlberg [3].

The observations for non-zero-sum repeated games with absorbing states by Vrieze and Thuijsman [7] were generalized in [6] and [5], and have led to proofs for existence of limiting average ε-equilibria in several special classes of stochastic games.

In this paper we shall examine a few particular examples of repeated games in Section 1, which will clarify how to derive ε-equilibria in the general case. The latter will be done in Section 2.

2. Examples

In the theory of repeated games the Folk theorem states that any feasible and individually rational reward can be obtained as an equilibrium reward (see, e.g., [1]). In our first example we show how this is usually achieved by means of threats.

2.1. EXAMPLE 1

We examine the repeated game:

	1	2
1	1,0	0,1
2	0,2	1,0

In this game player 1 can be sure to get a (limiting average) reward of at least $\frac{1}{2}$ by playing the stationary strategy $(\frac{1}{2}, \frac{1}{2})$ while player 2 can make sure by playing $(\frac{1}{2}, \frac{1}{2})$ that player 1's reward will not exceed $\frac{1}{2}$. Similarly, player 2 can guarantee himself at least $\frac{2}{3}$ by playing $(\frac{1}{3}, \frac{2}{3})$ while player 1 can make sure by playing $(\frac{2}{3}, \frac{1}{3})$ that player 2's reward will not exceed $\frac{2}{3}$. Hence $(\frac{1}{2}, \frac{2}{3})$ is the pair of values (v^1, v^2) of the zero-sum games for players 1 and 2 respectively. Clearly, for any equilibrium, each player should receive at least the value of "his" zero-sum game, i.e., the zero-sum game where he is maximizing his reward while his opponent is trying to minimize it. Therefore (v^1, v^2) is the vector of individually rational levels for the players. The set of feasible individually rational rewards is the small triangle in the following picture, i.e., it is the convex hull of $\{\ (\frac{1}{2}, \frac{2}{3}), (\frac{1}{2}, 1), (\frac{2}{3}, \frac{2}{3})\ \}$.

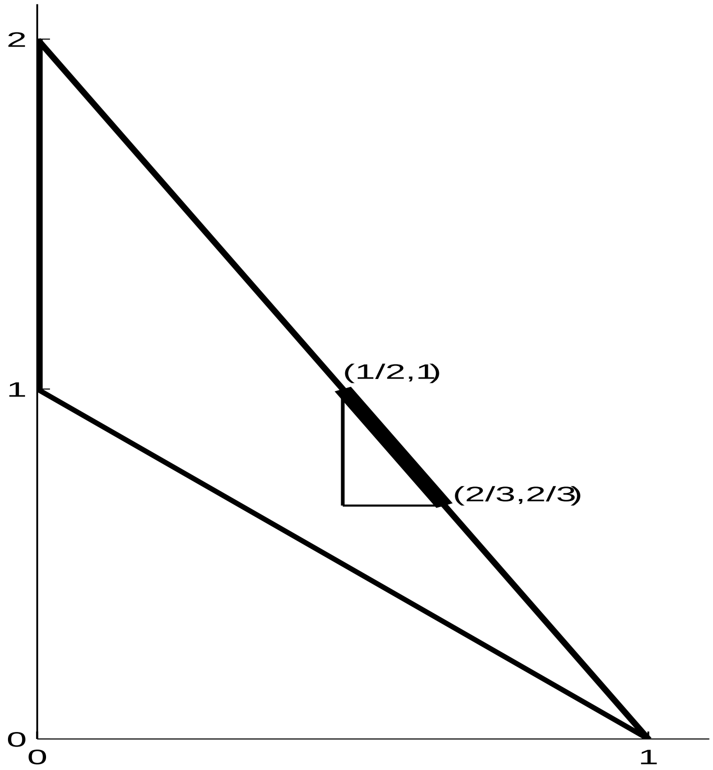

Figure 1.

The Folk theorem says that any reward (a, b) in this small triangle can be achieved as an equilibrium reward, i.e., for all $\varepsilon > 0$ there is an ε-equilibrium $(\sigma_\varepsilon, \tau_\varepsilon)$ such that $\| \ (a, b) - \gamma(\sigma_\varepsilon, \tau_\varepsilon) \ \|_\infty < \varepsilon$. To see why this is true we consider the reward $(\frac{7}{12}, \frac{10}{12})$ which can be achieved by playing the action sequence $(1, 1), (1, 1), (1, 1), (1, 1), (1, 1), (2, 1), (2, 1), (2, 1), (2, 1), (2, 1),(1, 2), (1, 2)$ repeatedly. So, if we define f to be the Markov strategy consisting of repeatedly playing $1, 1, 1, 1, 1, 2, 2, 2, 2, 2,1, 1$ and if we define g as consisting of repeated play of $1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2$, then $\gamma(f, g) = (\frac{7}{12}, \frac{10}{12})$. Now (f, g) is no equilibrium since player 1 could obtain a reward of $\frac{10}{12}$ by playing action 1 all the time, instead of f. However, player 2 can prevent player 1 from doing so by adding a threat to g to obtain g^* defined by: play g as long as player 1 has acted according to f, otherwise play $(\frac{1}{2}, \frac{1}{2})$. Similarly, player 1 can prevent player 2 from deviating from g by playing f^*: play f as long as player 2 has acted according to g, otherwise play $(\frac{2}{3}, \frac{1}{3})$. Clearly, (f^*, g^*) is an equilibrium and $\gamma(f^*, g^*) = (\frac{7}{12}, \frac{10}{12})$.

2.2. EXAMPLE 2

We examine the following repeated game with absorbing states.

	1	2
1	1,0 *	0,1 *
2	0,2	1,0

Notice that the zero-sum game determined by the payoffs for player 1 alone is precisely the Big Match. Obviously, the zero-sum game determined by the payoffs for player 2 is very similar to the Big Match. It can be verified that for this game we still have $(v^1, v^2) = (\frac{1}{2}, \frac{2}{3})$ and the set of feasible individually rational rewards is still represented by the small triangle depicted in Figure 1. Again we have that $(\frac{7}{12}, \frac{10}{12})$ is in this set, but now it cannot be achieved by (f, g) because of the absorbing entries (the ones with $*$). However, we could still achieve this reward by repeatedly playing $(2,1),(2,1),(2,1),(2,1),(2,1),(2,2),(2,2),(2,2),\ (2,2),(2,2),(2,2),(2,2)$. Unfortunately, this sequence cannot be used to "cook up" an equilibrium as in the previous example because player 1 could deviate to get 1 by playing 1 at the very first stage and play would be over due to absorption. Therefore we have to approach the matter in a more subtle way, so that player 1 does not have any profitable absorbing deviations. Yet if player 1 plays some non-absorbing strategy then player 2 prefers to play action 1 all the time, thereby creating a possibility for player 1 to deviate in an absorbing way. Although this seems to be a dilemma, we can create an ε-equilibrium by observing that if players 1 and 2 are playing the stationary strategies $\alpha = (0,1)^\infty$ and $\beta = (\beta_1, \beta_2)^\infty$ respectively, then player 2's action frequencies should converge to (β_1, β_2). To put it more precisely: if Y_n is the random variable denoting the action frequencies of player 2 playing (β_1, β_2) up to stage n, then

$$\forall\, \delta > 0\ \exists\, N_\delta\ : \Pr_\beta\{\ ||\ Y_n - (\beta_1, \beta_2)\ ||_\infty > \delta \ \text{ for any } n \geqslant N_\delta\} < \delta.$$

We shall use this observation to create an ε-equilibrium that yields the reward $(\frac{5}{12}, \frac{10}{12})$. Consider the stationary strategy $\beta = (\frac{5}{12}, \frac{7}{12})^\infty$ for player 2 and note that against β player 1 would prefer to play his non-absorbing action 2 at all stages. Define for player 1 a strategy α^*_δ by: play action 2 unless for some $n \geqslant N_\delta$ it has turned out that $||\ y_n - (\beta_1, \beta_2)\ ||_\infty > \delta$, where y_n is the realization of Y_n, then play, from that first moment onwards as a minimizing player 1, a δ-optimal strategy in the stochastic game:

	1	2
1	0 *	1 *
2	2	0

By doing so in case a deviation by player 2 is detected, player 1 can make sure that player 2's limiting average reward will be at most $\frac{2}{3} + \delta$, and thus player 1 has an effective threat to counter possible deviations by player 1. By choice of strategy the probability of an unjustified punishment by player 1 is less than δ. It can now be verified that (α^*_δ, β) is a limiting average ε-equilibrium for δ sufficiently small.

2.3. EXAMPLE 3

Finally, for this section we examine what happens if the second row is absorbing instead of the first one.

	1	2
1	1,0	0,1
2	0,2 *	1,0 *

As in the previous cases we have that $(v^1, v^2) = (\frac{1}{2}, \frac{2}{3})$ and the set of feasible individually rational rewards is still represented by the small triangle depicted in Figure 1. This time, however, we cannot apply the same approach as in the previous example because the non-absorbing rewards are not individually rational for at least one of the players. Although $(\frac{7}{12}, \frac{10}{12})$ is still feasible and individually rational, we cannot use $((0,1)^\infty, (\frac{5}{12}, \frac{7}{12})^\infty)$ to achieve this point as an equilibrium reward, because player 2 would deviate at the very first stage. However, for any $\mu \in (0,1)$ we have that $\gamma((1-\mu, \mu)^\infty, (\frac{5}{12}, \frac{7}{12})^\infty) = (\frac{7}{12}, \frac{10}{12})$. Let the action frequencies Y_n, y_n, as well as the number N_δ, be as defined in the previous example. Define now for player 1 the strategy $\alpha^*_{\mu\delta}$ by: play according to $(1-\mu, \mu)^\infty$ as long as for all $n \geqslant N_\delta$ you have found $|| \, y_n - (\frac{5}{12}, \frac{7}{12}) \, ||_\infty < \delta$, otherwise punish player 2 by playing, from that first moment onwards as a minimizing player 1, a δ-optimal strategy in the stochastic game.

	1	2
1	0	1
2	2 *	0 *

Then, for μ and δ sufficiently small, $(\alpha^*_{\mu\delta}, (\frac{5}{12}, \frac{7}{12})^\infty)$ is an ε-equilibrium with $\| \gamma(\alpha^*_{\mu\delta}, (\frac{5}{12}, \frac{7}{12})^\infty) - (\frac{7}{12}, \frac{10}{12}) \|_\infty < \varepsilon$.

3. General Solution

In this section we generalize the approach developed in the examples of the previous section. We shall use the following notations. For actions a and b for players 1 and 2 respectively, we shall write u_{ab} and w_{ab} for the non-absorbing payoffs for the respective players, while u^*_{ab} and w^*_{ab} will denote the absorbing payoffs, i.e., the rewards in case absorption occurs in entry (a, b). If entry (a, b) is selected by the players, then absorption will occur with probability p_{ab}. We shall call a pair of stationary strategies (α, β) absorbing in case these strategies yield absorption with probability 1, i.e., $\sum_a \sum_b \ \alpha_a p_{ab} \beta_b > 0$.

Using these notations we can derive the following.

Lemma 1 *For stationary strategies α and β we have:*

$$\gamma^1_\lambda(\alpha, \beta) = \frac{\lambda \sum_a \sum_b \ \alpha_a u_{ab} \beta_b + (1-\lambda) \sum_a \sum_b \ \alpha_a p_{ab} u^*_{ab} \beta_b}{\lambda + (1-\lambda) \sum_a \sum_b \ \alpha_a p_{ab} \beta_b} \quad \forall \ \lambda \in (0,1);$$

$$\gamma^1(\alpha, \beta) = \sum_a \sum_b \ \alpha_a u_{ab} \beta_b \quad \textit{if } (\alpha, \beta) \quad \textit{is non-absorbing;}$$

$$\gamma^1(\alpha, \beta) = \frac{\sum_a \sum_b \ \alpha_a p_{ab} u^*_{ab} \beta_b}{\sum_a \sum_b \ \alpha_a p_{ab} \beta_b} \quad \textit{if } (\alpha, \beta) \quad \textit{is absorbing.}$$

Here the discounted reward follows straightforwardly from the Shapley equation, while the average rewards are immediate.

It is well known from Fink [2] that stationary λ-discounted equilibria exist in any (n-person) stochastic game. We shall now examine properties of a sequence of stationary λ_n-discounted equilibria $(\alpha^{\lambda_n}, \beta^{\lambda_n})$, where we assume, without loss of generality since one can always take a subsequence, that for all n the strategies α^{λ_n} all have the same carrier, while the same holds for the strategies β^{λ_n}; moreover, the sequences are assumed to converge and $\lim_{n\to\infty} \ \lambda_n = 0$, $\lim_{n\to\infty} \ (\alpha^{\lambda_n}, \beta^{\lambda_n}) \equiv (\alpha^0, \beta^0)$ and $\lim_{n\to\infty} \ \gamma_{\lambda_n}(\alpha^{\lambda_n}, \beta^{\lambda_n}) \equiv (V^1, V^2)$. In order to keep notations simple we shall drop the subscripts n and write, e.g., $\lim_{\lambda\downarrow 0} \ \gamma_\lambda(\alpha^\lambda, \beta^\lambda)$ instead of $\lim_{n\to\infty} \ \gamma_{\lambda_n}(\alpha^{\lambda_n}, \beta^{\lambda_n})$.

Now note that the following observations apply:

a. If (α^0, β^0) is absorbing, then $(\alpha^\lambda, \beta^\lambda)$ is absorbing;

b. $\gamma^1_\lambda(\alpha^0, \beta^\lambda) = \gamma^1_\lambda(\alpha^\lambda, \beta^\lambda)$ for λ near 0, because each action in the carrier of α^0 is in the carrier of α^λ and therefore a λ-discounted best reply to β^λ;

c. If either (α^0, β^0) is absorbing or $(\alpha^\lambda, \beta^\lambda)$ is non-absorbing, then $\gamma(\alpha^0, \beta^0)$ $=\lim_{\lambda\downarrow 0}\ \gamma_\lambda(\alpha^\lambda, \beta^\lambda)$. This follows straightforwardly from Lemma 1;

d. If β is such that (α^0, β) is absorbing, then $\gamma^2(\alpha^0, \beta) \leqslant V^2$, because

$$\gamma^2(\alpha^0, \beta) = \lim_{\lambda\downarrow 0}\ \gamma^2_\lambda(\alpha^\lambda, \beta) \leqslant \lim_{\lambda\downarrow 0}\ \gamma^2_\lambda(\alpha^\lambda, \beta^\lambda) = V^2;$$

where the first equality again follows from Lemma 1 and the inequality follows from the fact that the strategies β^λ are λ-discounted best replies to α^λ (since $(\alpha^\lambda, \beta^\lambda)$ is a λ-discounted equilibrium);

e. $v = \lim_{\lambda\downarrow 0}\ v_\lambda \leqslant \lim_{\lambda\downarrow 0}\ \gamma_\lambda(\alpha^\lambda, \beta^\lambda) = V$.

Theorem 2 *Limiting average ε-equilibria can be derived from the sequence* $\{\ (\alpha^\lambda, \beta^\lambda) : \lambda \in (0,1)\ \}$.

Proof. We distinguish two cases: (A) with $\gamma^i(\alpha^0, \beta^0) \geqslant V^i$ for $i = 1, 2$, and (B) with $\gamma^i(\alpha^0, \beta^0) < V^i$ for $i = 1$ or for $i = 2$.

A. If $\gamma^i(\alpha^0, \beta^0) \geqslant V^i$ for $i = 1, 2$, then neither player can improve his reward by an absorbing deviation because of observation (d). Thus, the only deviations that could be profitable for a player are necessarily non-absorbing. However, non-absorbing deviations can be observed. To see this, suppose that player 2 deviates in a non-absorbing way. Then either player 2 chooses some action outside the carrier of β^0, which will be observed by player 1 immediately, or player 2's action frequencies do not converge to β^0, which will eventually be observed by player 1. If player 1 observes a deviation by player 1, then he can make sure that player 2's limiting average reward will be at most $v^2 + \delta$, by playing some δ-optimal strategy that minimizes player 2's reward. By observation (e) we have that $v^2 + \delta \leqslant V^2 + \delta \leqslant \gamma^2(\alpha^0, \beta^0)$; hence player 1 can effectively threaten to retaliate player 2 in case of a deviation, to prevent non-absorbing deviations. Of course player 2 can threaten player 1 in a similar way. Therefore, we can modify α^0 and β^0 with such δ-threats to establish ε-equilibria $(\alpha^0_\delta, \beta^0_\delta)$ for δ sufficiently small, just as we did in Examples 1 and 2.

B. If, without loss of generality, we have $\gamma^2(\alpha^0, \beta^0) < V^2$, then we must necessarily have, by observation (c), that (α^0, β^0) is absorbing while $(\alpha^\lambda, \beta^0)$ is non-absorbing for all λ. Hence $C \equiv \{a \in A : \sum_b\ p_{ab}\beta_b = 0\} \neq \emptyset$ and also $D \equiv \{a \in A : \sum_b\ p_{ab}\beta_b > 0\} \neq \emptyset$. Now define $\alpha^{\lambda\prime}_a = \frac{\alpha^\lambda_a}{\sum_{e\in C}\ \alpha^\lambda_e}$ and define $\alpha^{\lambda *}_a = \frac{\alpha^\lambda_a}{\sum_{e\in D}\ \alpha^\lambda_e}$. Then $\lim_{\lambda\downarrow 0}\ \alpha^{\lambda\prime} = \alpha^0$ and we can assume that $\lim_{\lambda\downarrow 0}\ \alpha^{\lambda *}$ also exists and equals, say, α^*.

Using Lemma 1 it can be shown that

$$V^2 = \omega \cdot \gamma^2(\alpha^0,\beta^0) + (1-\omega)\cdot\gamma^2(\alpha^*,\beta^0)$$

where $\omega = \lim_{\lambda\downarrow 0} \frac{\lambda}{\lambda+(1-\lambda)\sum_a\sum_b\ \alpha_a^\lambda p_{ab}\beta_b^0} \in [0,1]$. Since $\gamma^2(\alpha^0,\beta^0) < V^2$, we must have $\omega < 1$ and $\gamma^2(\alpha^*,\beta^0) \geqslant V^2$.

Because (α^*,β^0) is absorbing and because the carrier of α^* is a subset of the carrier of α^λ we also have $\gamma^1(\alpha^*,\beta^0) = \lim_{\lambda\downarrow 0}\ \gamma^1_\lambda(\alpha^*,\beta^\lambda) = \lim_{\lambda\downarrow 0}\ \gamma^1_\lambda(\alpha^\lambda,\beta^\lambda) = V^1$.

Now for $\mu \in (0,1)$ define $\alpha_\mu = (1-\mu)\cdot\alpha^0 + \mu\cdot\alpha^*$. Then for all μ we have that $\gamma(\alpha_\mu,\beta^0) = \gamma(\alpha^*,\beta^0)$. For $\delta > 0$ sufficiently small there is N_δ such that

$$\Pr_{(\alpha^0,\beta^0)}\{\ ||Y_n(\beta^0) - (\beta^0)||_\infty > \delta \ \text{ for any }\ n \geqslant N_\delta\ \} < \delta$$

and also

$$\Pr_{(\alpha^0,\beta^0)}\{\ ||X_n(\alpha^0) - (\alpha^0)||_\infty > \delta \ \text{ for any }\ n \geqslant N_\delta\ \} < \delta.$$

Next take $\mu > 0$ sufficiently small to have that the probability of absorption before N_δ with (α_μ,β^0) is less than δ and modify the strategies α_μ and β^0 by adding threats for punishment, as we did in Example 3, to get strategies $\alpha^*_{\mu\delta}$ and β^{0*}_δ that yield an ε-equilibrium for δ sufficiently small. The threat player 2 can use to prevent player 1 from not playing α^* with positive probability can be based on a number $M_{\mu\delta} > N_\delta$ with the property that for (α_μ,β^0) play will absorb before stage $M_{\mu\delta}$ with probability at least $1-\delta$; if absorption does not occur before stage $M_{\mu\delta}$, then player 2 will punish player 1. Besides player 2 should also check whether or not player 1 always takes actions from within the carrier of α_μ. ∎

3.1. EXAMPLE 3 REVISITED

For Example 3 we have that $(\alpha^\lambda,\beta^\lambda) = ((\frac{2}{2+\lambda},\frac{\lambda}{2+\lambda})^\infty,(\frac{1}{2},\frac{1}{2})^\infty)$ and $\gamma(\alpha^\lambda,\beta^\lambda) = v_\lambda = v = V = (\frac{1}{2},\frac{2}{3})$ for all $\lambda \in (0,1)$. We find that $(\alpha^0,\beta^0) = ((1,0)^\infty,(\frac{1}{2},\frac{1}{2})^\infty)$ and $\gamma(\alpha^0,\beta^0) = (\frac{1}{2},\frac{1}{2})$, so $\gamma^2(\alpha^0,\beta^0) = \frac{1}{2} < \frac{2}{3} = V^2$. Following the proof in case (B) of the previous theorem we find that $\alpha^* = (0,1)^\infty$. Notice that $\gamma^1(\alpha^*,\beta^0) = \frac{1}{2} = V^1$ and $\gamma^2(\alpha^*,\beta^0) = 1 > \frac{2}{3} = V^2$. Thus the equilibrium constructed in part (B) is very similar to the one presented in the discussion of Example 3. The only difference is that in the example we do not need to check whether player 1 is really playing action 2 with positive probability, since it would not be profitable for player 1 not to do so.

References

1. Aumann, R.J. (1981) Survey of repeated games, in *Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern*, Bibliographisches Institüt, Mannheim, pp. 11–42.
2. Fink, A.M. (1964) Equilibrium in a stochastic n-person game, *Journal of Science of Hiroshima University, Series* A-I **28**, 89–93.
3. Kohlberg, E. (1974) Repeated games with absorbing states, *Annals of Statistics* **2**, 724–738.
4. Sorin, S. (1986) Asymptotic properties of a non-zero-sum stochastic game, *International Journal of Game Theory* **15**, 101–107.
5. Thuijsman, F. (1992) Optimality and equilibria in stochastic games, CWI-Tract 82, Center for Mathematics and Computer Science, Amsterdam.
6. Thuijsman, F. and Vrieze, O.J. (1991) Easy initial states in stochastic games, in T.E.S. Raghavan, T.S. Ferguson, T. Parthasarathy, and O.J. Vrieze (eds.), *Stochastic Games and Related Topics*, Kluwer Academic Publishers, Dordrecht, pp. 85–100.
7. Vrieze, O.J. and Thuijsman, F. (1989) On equilibria in repeated games with absorbing states, *International Journal of Game Theory* **18**, 293–310.