THE BIG MATCH AND THE PARIS MATCH

FRANK THUIJSMAN
Maastricht University
Maastricht, The Netherlands

1. Introduction

In this paper we discuss the most typical and illustrative examples in stochastic game theory: the Big Match and a non-zero-sum extension thereof. These Big Match examples have played key roles in the development of stochastic game theory and we shall therefore examine them in some detail.

The Big Match is a zero-sum stochastic game that was introduced in 1957 by Gillette [3], who made some observations from which it was unclear whether or not the limiting average value would exist for this specific game. The matter was finally settled in 1968 by Blackwell and Ferguson [1], who presented limiting average ε-optimal strategies, which were history-dependent for one of the players. Their results were generalized by Kohlberg [4] for arbitrary repeated games with absorbing states, i.e., stochastic games in which all states but one are absorbing, and finally, in 1981, led to the breakthrough by Mertens and Neyman [5] with their general existence result for the limiting average value. It turned out that for zero-sum stochastic games the limiting average model could be solved by an asymptotic approach, from either the λ-discounted model or the N-stage model, through the observation that $v = \lim_{\lambda \downarrow 0} v_\lambda = \lim_{N \to \infty} v_N$.

For non-zero-sum stochastic games such an approach turned out to be impossible, as was illustrated by Sorin [6]. He analyzed a non-zero-sum extension of the Big Match, which we call the "Paris Match" due to its French origin. For this Paris Match he showed that the set of limiting average equilibrium rewards can neither be approached asymptotically from the set of λ-discounted equilibrium rewards nor from the set of N-stage equilibrium rewards. Vrieze and Thuijsman [7] derived the existence of limiting average ε-equilibria in non-zero-sum repeated games with absorbing states (with finite state and action spaces) after an inspiring study on the Paris Match.

2. The Big Match

Blackwell and Ferguson [1] present the Big Match as follows:

> “Every day player 2 chooses a number, 1 or 2, and player 1 tries to predict 2’s choice, winning a point if he is correct. This continues as long as player 1 predicts 1. But if he ever predicts 2, all future choices for both players are required to be the same as that day’s choices: if player 1 is correct on that day, he wins a point every day thereafter; if he is wrong on that day, he wins zero every day thereafter. The payoff to 1 is $\liminf_{n\to\infty}(a_1 + \cdots + a_n)/n$.”[1]

This game can be represented in the following way, where the payoffs presented are those by player 2 to player 1 and where the asterisks denote transitions to trivial absorbing states with the corresponding payoffs.

	1	2
1	1	0
2	0 *	1 *

The Big Match was introduced by Gillette [3], who made the following observations. First of all, if we consider only stationary strategies α for player 1 and stationary strategies β for player 2, then we have:

Lemma 1 $\max_\alpha \min_\beta \gamma(\alpha,\beta) = 0 < \frac{1}{2} = \min_\beta \max_\alpha \gamma(\alpha,\beta)$.
In other words: the Big Match does not have a limiting average value when only stationary strategies are considered.

Proof. To show correctness of the equality on the left-hand side we distinguish two cases for an arbitrary stationary strategy $\alpha = (\alpha_1,\alpha_2)^\infty$ by player 1: if $\alpha_2 > 0$, then $\gamma(\alpha,(1,0)^\infty) = 0$; if $\alpha_2 = 0$, then $\gamma(\alpha,(0,1)^\infty) = 0$.

As for the equality on the right-hand side: if $\beta = (\frac{1}{2},\frac{1}{2})^\infty$, then $\gamma(\alpha,\beta) = \frac{1}{2}$ for all α; if $\beta \neq (\frac{1}{2},\frac{1}{2})^\infty$, then either $\gamma((1,0)^\infty,\beta) > \frac{1}{2}$ or $\gamma((0,1)^\infty,\beta) > \frac{1}{2}$. ∎

More generally, considering Markov strategies f for player 1 and Markov strategies g for player 2, one also has:

Lemma 2 $\sup_f \inf_g \gamma(f,g) = 0 < \frac{1}{2} = \inf_g \sup_f \gamma(f,g)$.
In other words: the Big Match does not have a limiting average value when only Markov strategies are considered.

[1] In the original paper Blackwell and Ferguson number the actions 0 and 1 and use lim sup instead of lim inf. The results are not affected by these choices, however.

Proof. Let $f = (x^1, x^2, x^3, \ldots)$ be a Markov strategy for player 1, where $x^n = (x_1^n, x_2^n)$ is the mixed action to be played by player 1 at stage n, when play is still in the initial state. Similarly, let $g = (y^1, y^2, y^3, \ldots)$ be a Markov strategy for player 2. If $\Pr_f$[Player 1 will ever choose action 2] $= 1$, then $\gamma(f, (1,0)^\infty) = 0$. Otherwise we have that for each $\varepsilon > 0$ we can take some N_ε sufficiently large to have $\Pr_f$[Player 1 will ever choose action 2 after stage N_ε] $< \varepsilon$. Then define $g_\varepsilon = (\underbrace{(1,0),(1,0),\ldots,(1,0)}_{N_\varepsilon},(0,1),(0,1),\ldots)$ and notice that $\gamma(f, g_\varepsilon) < \varepsilon$.

As for the other side, first observe that $\gamma(f, (\frac{1}{2}, \frac{1}{2})^\infty) = \frac{1}{2}$ for all f. If $g = (y^1, y^2, y^3, \ldots)$ and $y_2^n \leq \frac{1}{2}$ for all n, then we clearly have that $\gamma((1,0)^\infty, g) \geq \frac{1}{2}$. If at some stage n we would have that $y_2^n > \frac{1}{2}$ for the first time, then the strategy $f_n = (\underbrace{(1,0),(1,0),\ldots,(1,0)}_{n},(0,1))$ would give $\gamma(f_n, g) > \frac{1}{2}$. ∎

So, neither in terms of stationary strategies, nor in terms of Markov strategies, does the limiting average value exist for the Big Match. The question about the existence of the value in terms of general strategies, raised in 1957, was finally answered affirmatively in 1968 by Blackwell and Ferguson [1]. Their paper was a real breakthrough in the theory of stochastic games for it was shown that behavioral strategies are indispensable in achieving limiting average ε-optimality in stochastic games. To emphasize this point: they showed that player 1 can guarantee in the Big Match a limiting average reward as close to $\frac{1}{2}$ as he likes, by carefully taking into account the opponent's behavior, i.e., his past actions, in the process of choosing his own actions. Here we write "can guarantee as close to $\frac{1}{2}$ as he likes," because there is no way that player 1 can guarantee $\frac{1}{2}$, as was also pointed out in their paper:

Lemma 3 *For each strategy σ for player 1 there exists a Markov strategy g for player 2 such that $\gamma(\sigma, g) < \frac{1}{2}$.*

Proof. Let σ be a strategy for player 1. If against player 2's strategy $(0,1)^\infty$ player 1 using σ never plays action 2 with positive probability, then clearly $\gamma(\sigma, (0,1)^\infty) = 0$. Otherwise, suppose that n is the first stage at which player 1 is going to play action 2 with positive probability, say ε, against $(0,1)^\infty$. Then player 2 can counter σ by playing action 2 for the first $n-1$ stages, playing action 1 at stage n, and playing $(\frac{1}{2}, \frac{1}{2})$ at every stage thereafter, giving player 1 an expected reward of $\frac{1}{2} - \frac{1}{2}\varepsilon$. ∎

We finally get to the main result on the Big Match and we shall sketch its original proof, in which history-dependent ε-optimal strategies are provided for player 1.

Theorem 4 (Blackwell and Ferguson [1]) *The limiting average value of the Big Match equals* $\frac{1}{2}$.

Sketch of Proof. Let $b_m \in \{1, 2\}$ be the action chosen by player 2 at stage $m \in \mathbb{N}$. Then we shall call $h_n = (b_1, b_2, \ldots, b_n)$ the history up to stage $n+1$. Define $k_0 = 0$ and for $n > 0$, define $k_n = \#1\text{'s} - \#2\text{'s}$ in h_n, i.e., k_n is the difference between the number of times player 2 chooses 1 (Left) and the number of times he chooses 2 (Right). Then we define σ_N for player 1 as the strategy where he chooses action 2 at stage $n+1$ with probability $\frac{1}{(k_n+N+1)^2}$. Note that when, eventually at some stage n, we would have $k_n = -N$, then player 1 would play action 2 with probability 1 at stage $n+1$.

Let T denote the number of stages after which player 1 plays action 2 (at stage $T+1$ and play is essentially over).

Let $T(m)$ denote the event $[T > m$, or $T < m$ and $b_{T+1} = 2]$.

Now we shall consider only pure (Markov) strategies by player 2: pure strategies because if σ_N works fine for those, then it works fine for others as well; Markov strategies because there is no history for player 2 to relate to, since he has observed only player 1 choosing 1 as long as the play is in the initial state. We distinguish two types of pure strategies: one for which k_n would eventually equal $-N$, and one for which it would not. (To distinguish these types it is assumed that player 1 would choose action 1 all the time.)

Case A. Let τ be a pure strategy for which we eventually have $k_n = -N$ for some n. By induction on m one can show that $\Pr_{\sigma_N \tau}[T(m)] \geq \frac{N}{2(N+1)}$ for all m. Since for (σ_N, τ) we have that $T < \infty$ with probability 1, we get

$$\Pr_{\sigma_N \tau}[b_{T+1} = 2] = \lim_{m \to \infty} \Pr_{\sigma_N \tau}[T(m)] \geq \frac{N}{2(N+1)}.$$

Case B. Let τ' be a pure strategy for which $k_n > -N$ for all n. For $a = 1, 2$ let $\mu_a(m) = \Pr_{\sigma_N \tau}[T < m$ and $b_{T+1} = a]$ and let $\mu_a = \lim_{m \to \infty} \mu_a(m)$. Also define $\tau'_m = (\tau_1, \tau_2, \ldots, \tau_{m,} (\frac{1}{2}, \frac{1}{2}), (\frac{1}{2}, \frac{1}{2}), \ldots)$. Then, for each m, we have that τ'_m is a strategy of the type that is considered in Case A. Now observe that

$$\begin{aligned}
\gamma(\sigma_N, \tau') &\geq \mu_2 + \frac{1}{2}(1 - \mu_1 - \mu_2) \\
&= \lim_{m \to \infty} [\mu_2(m) + \frac{1}{2}(1 - \mu_1(m) - \mu_2(m))] \\
&= \lim_{m \to \infty} \gamma(\sigma_N, \tau'_m) \geq \frac{N}{2(N+1)},
\end{aligned}$$

where the first inequality follows from the fact that $k_n > -N$ for all n, which implies that player 1 should get at least $\frac{1}{2}$ if play does not absorb, where the equality signs are straightforward, and where we have used the result of Case A for the last inequality.

Cases A and B together imply that by playing σ_N player 1 can guarantee himself at least $\frac{N}{2(N+1)}$. So, for every $\varepsilon > 0$, by taking N sufficiently large player 1 can guarantee himself a limiting average reward of at least $\frac{1}{2} - \varepsilon$. By playing $(\frac{1}{2}, \frac{1}{2})^\infty$ it is clear that player 2 can guarantee himself a limiting average reward of (at most) $\frac{1}{2}$ to player 1. Hence the result. ∎

In this proof we have seen that player 1 can guarantee himself the value v up to some ε, i.e., for all $\varepsilon > 0$ there is a strategy σ_ε such that for all strategies τ we have $\gamma(\sigma_\varepsilon, \tau) \geq v - \varepsilon$. Such a strategy σ_ε is called an ε-optimal strategy. Lemma 3 shows that, generally, ε cannot be taken to be equal to 0.

This work of Blackwell and Ferguson [1] was generalized by Kohlberg [4] to the class of zero-sum repeated games with absorbing states.

Theorem 5 (Kohlberg [4]) *The limiting average value v exists for every zero-sum repeated game with absorbing states. Moreover, $v = \lim_{n\to\infty} v_n$, the limit of the average values of the n-stage games.*

In his paper Kohlberg employs a slightly different type of ε-optimal strategy, which for the case of the Big Match would tell player 1 at stage $n+1$ to play action 2 with probability ε^2 if $k_n < 0$; and with probability $\varepsilon^2(1-\varepsilon)^{k_n}$ otherwise, where, as above, k_n denotes the excess of 1's over 2's among the first n choices of player 2.

Yet another approach to solve the Big Match can be found in Coulomb [2].

Finally Mertens and Neyman [5] further generalized Kohlberg's [4] result to cover all zero-sum stochastic games with finitely many states and actions.

Theorem 6 (Mertens and Neyman [5]) *The limiting average value v exists for every zero-sum stochastic game. Moreover, $v = \lim_{n\to\infty} v_n = \lim_{\lambda\downarrow 0} v_\lambda$, so the limiting average value, the limit of the average values of the n-stage games and the limit of the λ-discounted games are all equal.*

To illustrate this for the Big Match: there we have that $v_n = v_\lambda = \frac{1}{2}$ for all n and for all λ. The unique optimal strategy for player 2 is $(\frac{1}{2}, \frac{1}{2})^\infty$ for all n-stage games and for all λ-discounted games as well. For player 1 the stationary λ-discounted optimal strategy is $(\frac{1}{1+\lambda}, \frac{\lambda}{1+\lambda})^\infty$, while an optimal Markov strategy for player 1 in the n-stage game is given by playing $(\frac{1+m}{2+m}, \frac{1}{2+m})$ at stage $n-m$, for $m = 1, 2, \ldots, n-1$.

Further generalizations of this result by weakening the finiteness assumptions can be found in other chapters of this volume. In the next section, however, we shall focus on an extension of the zero-sum Big Match to a non-zero-sum situation.

3. The Paris Match

The non-zero-sum extension that we shall discuss here was introduced, and examined in detail, by Sorin [6]. Because of the French origin of the author we shall call it the Paris Match. The structure is essentially the same as in the Big Match, but now players are no longer paying each other and do not have completely opposite interests. The Paris Match is again a repeated game with absorbing states and we can represent it in the following matrix notation, where the asterisks again denote transitions to absorbing states.

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

As we have seen in the previous section the limiting average value v turns out to be equal to the limits of both the finite horizon average values v_n and the λ-discounted values v_λ; we have $v = \lim_{n\to\infty} v_n = \lim_{\lambda\downarrow 0} v_\lambda$. The Paris Match showed that such asymptotic properties are not valid for non-zero-sum stochastic games. Actually, Sorin [6] shows that for the Paris Match there is a gap between the set of limiting average equilibrium rewards E_∞ on the one side and the set of finite horizon equilibrium rewards E_n and the set of λ-discounted equilibrium rewards E_λ on the other side. So $E_\infty \neq \lim_{n\to\infty} E_n$ and $E_\infty \neq \lim_{\lambda\downarrow 0} E_\lambda$. Even worse, as we shall see below, E_n and E_λ do not even get close to E_∞. More precisely, for the Paris Match we have:

Theorem 7 (Sorin [6])
a) $E_\lambda = \{(\frac{1}{2}, \frac{2}{3})\}$ *for all* λ.
b) $E_n = \{(\frac{1}{2}, \frac{2}{3})\}$ *for all* n.
c) $E_\infty = \mathrm{conv}\{(\frac{1}{2}, 1), (\frac{2}{3}, \frac{2}{3})\}$, *where* conv *stands for convex hull.*

This is illustrated in Figure 1 by a graph of the reward space for this game.

We shall sketch the proof for part (a) and part (c) of this theorem. Since the proof for part (b) goes along lines that are roughly similar to those for part (a), we skip this part.

Sketch of proof for part (a). We start our observation by noting that

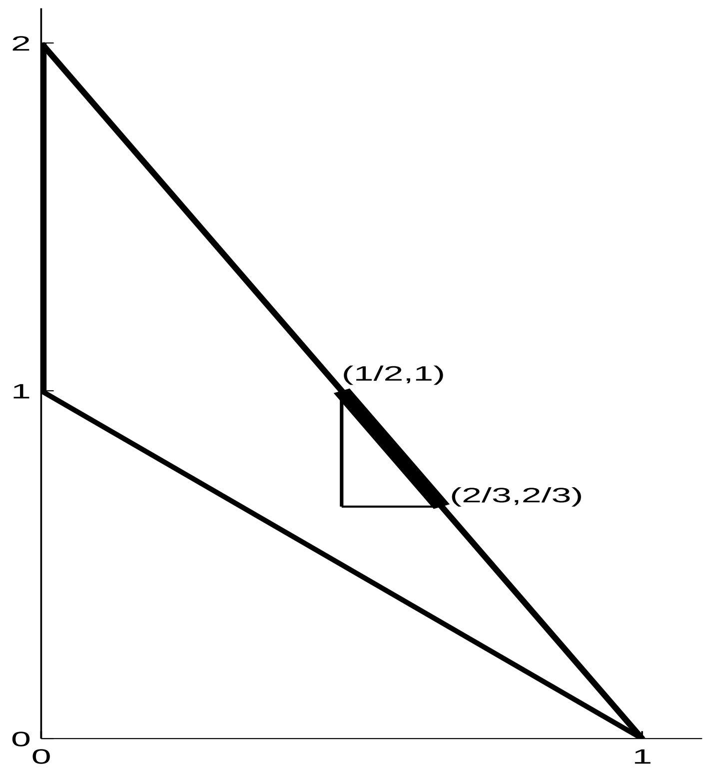

Figure 1.

any equilibrium reward should give the players at least $(\frac{1}{2}, \frac{2}{3})$, because the players can guarantee themselves those rewards by their own effort. To put it more precisely: player 2 can guarantee himself at least $\frac{2}{3}$ by playing $(\frac{1}{3}, \frac{2}{3})^\infty$, and therefore we should also have that any equilibrium reward to player 2 yields at least $\frac{2}{3}$; similarly, player 1 should get at least $\frac{1}{2}$, because his zero-sum situation is exactly the same as it was in the Big Match examined in Section 1.

Let w be the maximal λ-discounted equilibrium reward for player 2 and suppose that (σ, τ) is a λ-discounted equilibrium with $\gamma_\lambda^2(\sigma, \tau) = w$. Let furthermore w_1 and w_2 be the normalized λ-discounted rewards for (σ, τ) on condition that at stage 1 the action pair $(1, 1)$, respectively $(1, 2)$, was played. Then w_1 and w_2 should also be λ-discounted equilibrium rewards, for otherwise players could deviate at stage 2 or later. Hence we must have

$$w_1 \leq w \quad \text{and} \quad w_2 \leq w.$$

Now let p be the probability by which player 1 plays action 2 at stage 1 using σ, and also let q be the probability by which player 2 plays action 2 at stage 1 using τ. One can check straightforwardly that for (σ, τ) to be an equilibrium we must have that $0 < p < 1$ and also $0 < q < 1$. Next we observe by examining the equilibrium conditions at stage 1 that player 2 should be indifferent between action 1 and action 2 at stage 1, and therefore we must have that both actions yield player 2 the same λ-discounted reward, that is:

$$w = \underbrace{2p + (1-p)(1-\lambda)w_1}_{\text{reward for action 1}} = \underbrace{0p + (1-p)(\lambda + (1-\lambda)w_2)}_{\text{reward for action 2}}.$$

Since $w_1 \leq w$ and $w_2 \leq w$, we derive

$$w \leq (1-p)(\lambda + (1-\lambda)w) \quad \text{and} \quad w \leq 2p + (1-p)(1-\lambda)w.$$

The last inequality can be rewritten as

$$2 - w \geq (1-p)(2 - (1-\lambda)w),$$

which together with the first one gives

$$(2-w)(1-p)(\lambda + (1-\lambda)w) \geq (1-p)(2-(1-\lambda)w)w.$$

By removing the brackets and cancelling terms against one another straightforward calculation leads to $w \leq \frac{2}{3}$. So, by our initial observation in this proof, we find that $w = \frac{2}{3}$.

Now for the first player: assume that u is player 1's maximal λ-discounted reward and that u_1 and u_2 are defined similarly to the above as normalized rewards to player 1 conditioned on 2 choosing action 1 or action 2 respectively at stage 1. Again, we must have that u_1 and u_2 are equilibrium rewards for player 1 as well, so

$$u_1 \leq u \qquad \text{and} \qquad u_2 \leq u$$

and, since player 1 is playing both actions with positive probability at stage 1, we should also have:

$$u = \underbrace{(1-q)(\lambda + (1-\lambda)u_1 + q(1-\lambda)u_2)}_{\text{reward for action 1}} = \underbrace{q}_{\text{reward for action 2}}.$$

Therefore

$$u \leq u^2(1-\lambda) + (1-u)(\lambda + (1-\lambda)u)$$

which leads to $u \leq \frac{1}{2}$.

Hence, by our initial observation in this proof, the only possibility is to have $u = \frac{1}{2}$.

Putting these things together we have shown that $E_\lambda = \{(\frac{1}{2}, \frac{2}{3})\}$. ∎

It can be verified that for the Paris Match the unique *stationary* λ-discounted equilibrium is the pair $((\frac{2}{2+\lambda}, \frac{\lambda}{2+\lambda}), (\frac{1}{2}, \frac{1}{2}))$. In this equilibrium strategy pair, each player minimizes his opponent's λ-discounted reward. Although λ-discounted equilibria always exist, we would like to remark that the Big Match, discussed above, shows that, like optimal strategies in the zero-sum case, limiting average equilibria generally fail to exist in the general-sum case. Therefore we have to introduce the concept of ε-equilibria. A pair of strategies $(\sigma_\varepsilon, \tau_\varepsilon)$ is a limiting average ε-equilibrium $(\varepsilon > 0)$ if for all σ and τ we have $\gamma_1(\sigma, \tau_\varepsilon) \leq \gamma_1(\sigma_\varepsilon, \tau_\varepsilon) + \varepsilon$ and $\gamma_2(\sigma_\varepsilon, \tau) \leq \gamma_2(\sigma_\varepsilon, \tau_\varepsilon) + \varepsilon$. Thus σ_ε and τ_ε are ε-best replies to each other.

Sketch of proof for part (c). To illustrate why $E_\infty \subseteq \mathrm{conv}\{(\frac{1}{2}, 1), (\frac{2}{3}, \frac{2}{3})\}$ we quote Sorin [6]:

> "The idea of the proof is very simple: if the probability of getting an absorbing payoff on the equilibrium path is less than 1, then after some time player 1 is essentially playing action 1; the corresponding feasible rewards from this stage on are not individually rational, hence a contradiction."

We refer to the original paper by Sorin [6] for a mathematically sound translation of this argument.

As for the converse, $E_\infty \supseteq \mathrm{conv}\{(\frac{1}{2}, 1), (\frac{2}{3}, \frac{2}{3})\}$, the argument may best be seen by an example. Take for instance the reward $(\frac{7}{12}, \frac{10}{12})$, which is a point in $\mathrm{conv}\{(\frac{1}{2}, 1), (\frac{2}{3}, \frac{2}{3})\}$; we shall explain the method of Sorin [6] to construct ε-equilibria that correspond to this reward. Consider the auxiliary zero-sum repeated game with absorbing states presented by:

	1	2
1	$\frac{7}{12}$	$-\frac{5}{12}$
2	$-\frac{7}{12}$ *	$\frac{5}{12}$ *

Let σ_ε be a limiting average ε-optimal strategy for player 1 in this auxiliary game, and let $\beta = (\frac{5}{12}, \frac{7}{12})^\infty$. Then for $(\sigma_\varepsilon, \beta)$ play absorbs with probability 1, yielding $\gamma(\sigma_\varepsilon, \beta) = (\frac{7}{12}, \frac{10}{12})$. Hence σ_ε is a limiting average best reply for player 1 against β in the original game, because no absorption would yield a limiting average reward of only $\frac{5}{12}$ to player 1. It can also be verified along "Big Match-like" arguments, that β is an ε-best reply for player 2 against σ_ε. ∎

The type of limiting average ε-equilibrium strategies developed by Sorin [6] for the Paris Match enabled Vrieze and Thuijsman [7] to show the existence of ε-equilibria for arbitrary non-zero-sum repeated games with absorbing states, for the case of finite action spaces. The latter approach used the equilibrium structure developed by Sorin [6] and the existence of the limiting average value of zero-sum repeated games as shown by Kohlberg [4]. In the next chapter we shall take a closer look at the class of non-zero-sum repeated games with absorbing states, and we shall provide a simple proof for the existence of ε-equilibria. Several examples will illustrate the solution.

References

1. Blackwell, D. and Ferguson, T. S. (1968) The big match, *Annals of Mathematical Statistics* **39**, 159–163.
2. Coulomb, J. M. (1996) A note on "Big-Match," *ESAIM: Probability and Statistics* **1**, 89–93.
3. Gillette, D. (1957) Stochastic games with zero stop probabilities, in M. Dresher, A. W. Tucker, and P. Wolfe (eds.), *Contributions to the Theory of Games, Vol. III*, Annals of Mathematics Studies 39, Princeton University Press, Princeton, NJ, pp. 179–187.
4. Kohlberg, E. (1974) Repeated games with absorbing states, *Annals of Statistics* **2**, 724–738.
5. Mertens, J.-F. and Neyman, A. (1981) Stochastic games, *International Journal of Game Theory* **10**, 53–66.
6. Sorin, S. (1986) Asymptotic properties of a non-zero-sum stochastic game, *International Journal of Game Theory* **15**, 101–107.
7. Vrieze, O. J. and Thuijsman, F. (1989) On equilibria in repeated games with absorbing states, *International Journal of Game Theory* **18**, 293–310.