

# Stochastic games with non-observable actions

## Citation for published version (APA):

Flesch, J., Thuijsman, F., & Vrieze, OJ. (2003). Stochastic games with non-observable actions. *Mathematical Methods of Operations Research*, 58(3), 459-475. <https://doi.org/10.1007/s001860300293>

## Document status and date:

Published: 01/12/2003

## DOI:

[10.1007/s001860300293](https://doi.org/10.1007/s001860300293)

## Document Version:

Publisher's PDF, also known as Version of record

## Document license:

Taverne

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

## Stochastic games with non-observable actions

J. Flesch, F. Thuijsman, O.J. Vrieze

Department of Mathematics, Maastricht University, P.O.Box 616, 6200 MD Maastricht,  
The Netherlands

Manuscript received: December 2002/Final version received: April 2003

**Abstract.** We examine  $n$ -player stochastic games. These are dynamic games where a play evolves in stages along a finite set of states; at each stage players independently have to choose actions in the present state and these choices determine a stage payoff to each player as well as a transition to a new state where actions have to be chosen at the next stage. For each player the infinite sequence of his stage payoffs is evaluated by taking the limiting average. Normally stochastic games are examined under the condition of full monitoring, i.e. at any stage each player observes the present state and the actions chosen by all players. This paper is a first attempt towards understanding under what circumstances equilibria could exist in  $n$ -player stochastic games without full monitoring. We demonstrate the non-existence of  $\varepsilon$ -equilibria in  $n$ -player stochastic games, with respect to the average reward, when at each stage each player is able to observe the present state, his own action, his own payoff, and the payoffs of the other players, but is unable to observe the actions of them. For this purpose, we present and examine a counterexample with 3 players. If we further drop the assumption that the players can observe the payoffs of the others, then counterexamples already exist in games with only 2 players.

### 1 Introduction

Stochastic games can be seen as multi-player Markov decision processes. A stochastic game can be described by (1) a nonempty and finite set of players  $I$ , (2) a nonempty and finite set of states  $S$ , (3) for each state  $s$ , a nonempty and finite set of actions  $A^i(s)$  for each player  $i$ , (4) for each state  $s$  and each joint action  $a \in \times_{i \in I} A^i(s)$ , a payoff  $r^i(s, a) \in \mathbb{R}$  to each player  $i$ , (5) for each state  $s$  and each joint action  $a \in \times_{i \in I} A^i(s)$ , a transition probability vector  $p(s, a) = (p(t|s, a))_{t \in S}$ . The game is to be played at stages in  $\mathbb{N}$  in the following way. The play starts at

stage 1 in an initial state, say in state  $s_1 \in S$ , where, simultaneously and independently, each player  $i$  is to choose an action  $a_1^i \in A^i(s_1)$ . These choices induce an immediate payoff  $r^i(s_1, (a_1^j)_{j \in I})$  to player  $i$ , and next, the play moves to a new state according to the probability vector  $p(s_1, (a_1^j)_{j \in I})$ , say to state  $s_2$ . At stage 2 a new action  $a_2^i \in A^i(s_2)$  is to be chosen by each player  $i$  in state  $s_2$ . Then player  $i$  receives payoff  $r^i(s_2, (a_2^j)_{j \in I})$  and the play moves to some state  $s_3$  according to the probability vector  $p(s_2, (a_2^j)_{j \in I})$ , and so on.

A stochastic game is usually played under the assumption of full monitoring:

**Assumption of full monitoring:** each player is able to observe the present state, his own action and the actions of the other players. So, a history for player  $i \in I$  up to stage  $n$  is of the form:

$$h^i(n) = (s_1, (a_1^j)_{j \in I}, \dots, s_n, (a_n^j)_{j \in I}).$$

Notice that, under the assumption of full monitoring, the histories of the players coincide.

A mixed action  $x^i(s)$  for player  $i$  in state  $s$  is a probability distribution on  $A^i(s)$ . The set of mixed actions for player  $i$  in state  $s$  is denoted by  $X^i(s)$ . A (history dependent or behavior) strategy  $\sigma^i$  for player  $i$  is a decision rule that, for the choice of action, prescribes a mixed action  $\sigma^i(s, h^i) \in X^i(s)$  in the present state  $s$  depending on his past history  $h^i$ . If the mixed actions prescribed by a strategy only depend on the present state and stage then the strategy is called Markov, while if they only depend on the present state then the strategy is called stationary.

A joint strategy  $\sigma = (\sigma^i)_{i \in I}$  together with an initial state  $s \in S$  determines a stochastic process on the payoffs. The sequences of payoffs are evaluated by the average reward, which is given for player  $i \in I$  by

$$\gamma^i(s, \sigma) := \liminf_{N \rightarrow \infty} \mathbb{E}_{s\sigma} \left( \frac{1}{N} \sum_{n=1}^N R_n^i \right) = \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mathbb{E}_{s\sigma} (R_n^i),$$

where  $\mathbb{E}_{s\sigma}$  stands for expectation and where  $R_n^i$  is a random variable for the payoff for player  $i$  at stage  $n$ .

A joint strategy  $\sigma = (\sigma^i)_{i \in I}$  is called an  $\varepsilon$ -equilibrium,  $\varepsilon \geq 0$ , with respect to the average reward, if for each player  $i \in I$  and initial state  $s \in S$

$$\gamma^i \left( s, \bar{\sigma}^i, (\sigma^j)_{j \in I \setminus \{i\}} \right) \leq \gamma^i(s, \sigma) + \varepsilon \quad \forall \bar{\sigma}^i,$$

which means that no player can gain more than  $\varepsilon$  by a unilateral deviation, for any initial state  $s \in S$ . Hence, for small  $\varepsilon$ , the rewards corresponding to an  $\varepsilon$ -equilibrium are an appealing solution of the stochastic game.

Under the assumption of full monitoring, the famous game called the Big Match, in Gillette (1957) and Blackwell & Ferguson (1968), showed that 0-equilibria do not always exist and history dependent strategies are indispensable for obtaining  $\varepsilon$ -equilibria, with  $\varepsilon > 0$ . However the existence of  $\varepsilon$ -equilibria for all  $\varepsilon > 0$  is not yet known in general and is the most challenging open problem these days, even though the existence problem has been answered in the affirmative for several special classes, such as for games with only two players (cf. Vieille (2000, I and II)).

As far as we know, non zero-sum stochastic games have always been studied with the assumption of full monitoring. So the question arises what happens if the assumption of full monitoring is dropped. We shall therefore examine the following two weaker assumptions:

**Assumption 1:** each player is able to observe the present state, his own action, his own payoff, and the payoffs of the other players. So, a history for player  $i \in I$  up to stage  $n$  is of the form

$$h^i(n) = (s_1, a_1^i, (u_1^j)_{j \in I}, \dots, s_n, a_n^i, (u_n^j)_{j \in I}),$$

where  $u_k^j$  denotes the payoff for player  $j$  at stage  $k$ .

**Assumption 2:** each player is able to observe the present state, his own action and his own payoff. So, a history for player  $i \in I$  up to stage  $n$  is of the form

$$h^i(n) = (s_1, a_1^i, u_1^i, \dots, s_n, a_n^i, u_n^i),$$

where  $u_k^i$  denotes the payoff for player  $i$  at stage  $k$ .

Notice that Assumption 1 is somewhat weaker than the assumption of full monitoring, because if a player can observe all the actions then he can conclude all the payoffs. Vieille (2000, III) recently managed to show that, in this case, each two-player stochastic game has at least one initial state for which  $\varepsilon$ -equilibria exist for all  $\varepsilon > 0$ . Despite this result, we will show that stochastic games do not always admit  $\varepsilon$ -equilibria for all initial states, which shall be demonstrated by a counterexample with 3 players. Consequently, for obtaining  $\varepsilon$ -equilibria, the players should be able to gain more information during the play of the game than under Assumption 1. It is not clear whether or not a counterexample exists with only 2 players.

Obviously, Assumption 2 is the weakest amongst the assumptions above. Under this assumption counterexamples already exist in games with only 2 players. Such an example shall also be analyzed below.

The idea that the players cannot fully monitor the behavior of their opponents is not new. However, the earlier studies on stochastic games without full monitoring always focused on the zero-sum case, see for instance Coulomb (1992, 1999). Thus, our goal is to provide a first step towards understanding under what circumstances equilibria could exist in  $n$ -player stochastic games with non-observable actions.

**Absorbing games.** Both counterexamples we will present below are so-called absorbing games. These are very special stochastic games with the properties that all the states but one are absorbing (play remains there forever, regardless the actions chosen by the players) and in any absorbing state each player has only one action. Note that if the play moves to an absorbing state  $s$  (absorption occurs in state  $s$ ) then the play is strategically over and the average rewards of the players will equal the payoffs in state  $s$ .

In these absorbing games, since the non-absorbing state is the only non-trivial state, we will assume that it is the initial state and we will suppress the states in the notations. For a (history dependent) strategy  $\sigma^i$  for player  $i$ , the prescribed mixed action after past history  $h^i$  (if no absorption has occurred) is denoted by  $\sigma^i(h^i)$ , while for a Markov strategy  $\sigma^i$ , the mixed action for stage  $n$  is  $\sigma^i(n)$ . The probability on action  $a^i$  is denoted by  $\sigma^i(h^i, a^i)$  or  $\sigma^i(n, a^i)$ , respectively.

For convenience, let  $t^*$  denote the random variable for the stage of absorption; if no absorption occurs at all, then let  $t^* = \infty$ . Moreover, with regard to a joint Markov strategy  $\sigma = (\sigma^j)_{j \in I}$ , we use the notation  $r^{i*}(\sigma(n))$  for the expected payoff of player  $i$  with respect to the joint mixed action  $\sigma(n)$  given absorption occurs.

**2 A counterexample under Assumption 2**

**Example 1.** Under Assumption 2, consider the following game with 2 players:

	Left	Right
Top	0,1	0,0
Bottom	1,-1*	-1,1*

In this game, player 1’s actions are the rows (Top and Bottom) and player 2’s actions are the columns (Left and Right). Entries (Top, Left) and (Top, Right) are non-absorbing, meaning that after the players receive the payoffs the play remains in the non-absorbing state. On the other hand, entries (Bottom, Left) and (Bottom, Right) are absorbing with probability 1 (indicated by \*), meaning that the play moves to an absorbing state whose payoffs are given in this entry. (This game shows some similarity with the Big Match, cf. Gillette (1957).)

Notice that player 1 can determine the time of absorption, but it is player 2 who determines the place of it. Also, if player 1 plays Top then no absorption occurs and due to Assumption 2, player 2’s behavior is completely invisible to player 1; while as soon as player 1 plays Bottom, absorption occurs with probability 1.

Based on the above observations we will now show the following lemma, which says that player 1 has only Markov strategies at his disposal, while player 2 can also restrict himself to using Markov strategies, in a certain sense.

**Lemma 1.** Consider example 1 (under Assumption 2).

1. Any strategy  $\sigma^1$  of player 1 is simply a Markov strategy.
2. For any strategy  $\sigma^2$  of player 2, there exists a Markov strategy  $\bar{\sigma}^2$  so that for any strategy  $\sigma^1$  of player 1

$$\gamma^i(\sigma^1, \sigma^2) = \gamma^i(\sigma^1, \bar{\sigma}^2) \quad \forall i \in I = \{1, 2\}.$$

*Proof. Part 1.* If the play is at stage  $n$  and no absorption has occurred, then player 1’s history must be

$$h^1(n - 1) = (\text{Top}, 0, \dots, \text{Top}, 0).$$

Hence, the only information such a history carries is its “length”  $n - 1$ , or equivalently that the present stage is  $n$ . This means that player 1 can only have Markov strategies, indeed.

**Part 2.** Take an arbitrary strategy  $\sigma^2$  for player 2. Let  $q_n$  be the probability of playing Left at stage  $n$  with respect to  $\sigma^2$ , given that no absorption has occurred. Note that this probability is independent of the strategy of player 1, since, as long as no absorption occurs, player 1’s actions have been Top at all stages. Let  $\bar{\sigma}^2$  be the Markov strategy where player 2 plays  $(q_n, 1 - q_n)$  at stage  $n$ . Then, for any strategy  $\sigma^1$  of player 1, we have for each stage  $n$  and for each player  $i$ :

$$\mathbb{E}_{(\sigma^1, \bar{\sigma}^2)}(R_n^i) = \mathbb{E}_{(\sigma^1, \sigma^2)}(R_n^i).$$

Hence the average rewards are the same as well. □

The following corollary will make it sufficient to deal with Markov  $\varepsilon$ -equilibria.

**Corollary 2.** *Consider example 1 (under Assumption 2). Suppose there exists an  $\varepsilon$ -equilibrium for some  $\varepsilon \geq 0$ . Then, there must also exist a Markov  $\varepsilon$ -equilibrium.*

*Proof.* Suppose  $(\sigma^1, \sigma^2)$  is an  $\varepsilon$ -equilibrium for some  $\varepsilon \geq 0$ . For  $\sigma^2$ , take a Markov strategy  $\bar{\sigma}^2$  as in part 2 of lemma 1. Then by lemma 1, the pair  $(\sigma^1, \bar{\sigma}^2)$  must be a Markov  $\varepsilon$ -equilibrium. □

The main result of this section is the following theorem.

**Theorem 3.** *Consider example 1 (under Assumption 2). There exists no  $\varepsilon$ -equilibrium for  $\varepsilon \in [0, \frac{1}{8})$ .*

*Proof.* Suppose the opposite. Then by corollary 2 we should have a Markov  $\varepsilon$ -equilibrium  $(\sigma^1, \sigma^2)$  for some  $\varepsilon \in [0, \frac{1}{8})$ . We will subsequently derive a contradiction in several steps.

• **Step 1.**  $\gamma^1(\sigma^1, \sigma^2) \geq -\varepsilon$

Observe that player 1 can guarantee 0 by playing the stationary strategy  $x = (1, 0)$ , which prescribes action Top for each stage. Since  $(\sigma^1, \sigma^2)$  is an  $\varepsilon$ -equilibrium, we obtain

$$\gamma^1(\sigma^1, \sigma^2) \geq \gamma^1(x, \sigma^2) - \varepsilon = -\varepsilon.$$

• **Step 2.**  $\gamma^2(\sigma^1, \sigma^2) \geq 1 - \varepsilon$

(The arguments below are similar to ones used in the study of the Big Match by Gillette (1957) and by Blackwell & Ferguson (1968).)

Since

$$\mathbb{P}_{\sigma^1}(t^* < \infty) = \lim_{n \rightarrow \infty} \mathbb{P}_{\sigma^1}(t^* \leq n),$$

we have

$$\lim_{n \rightarrow \infty} \mathbb{P}_{\sigma^1}(n < t^* < \infty) = \lim_{n \rightarrow \infty} [\mathbb{P}_{\sigma^1}(t^* < \infty) - \mathbb{P}_{\sigma^1}(t^* \leq n)] = 0. \tag{1}$$

Let  $\delta > 0$  be arbitrary. By (1), there must exist a stage  $n_\delta$  such that

$$\mathbb{P}_{\sigma^1}(n_\delta < t^* < \infty) \leq \delta.$$

Consider the Markov strategy  $\tau_\delta^2$  for player 2 which prescribes action Right for stages  $1, \dots, n_\delta$  and action Left at all further stages.

Then, with respect to  $(\sigma^1, \tau_\delta^2)$ , the following 3 events can occur:

- (i) Absorption takes place in entry (Bottom, Right) at some stage in  $1, \dots, n_\delta$ ;
- (ii) Absorption takes place in entry (Bottom, Left) at some stage in  $n_\delta + 1, n_\delta + 2, \dots$ ;
- (iii) No absorption occurs at all, and entry (Top, Left) is played at all stages in  $n_\delta + 1, n_\delta + 2, \dots$

By the choice of  $n_\delta$ , event (ii) has probability at most  $\delta$ , hence  $\gamma^2(\sigma^1, \tau_\delta^2) \geq 1 - \delta$ . Due to the fact that  $(\sigma^1, \sigma^2)$  is an  $\varepsilon$ -equilibrium, we have

$$\gamma^2(\sigma^1, \sigma^2) \geq \gamma^2(\sigma^1, \tau_\delta^2) - \varepsilon \geq 1 - \delta - \varepsilon.$$

Since  $\delta > 0$  was arbitrary, the proof of step 2 is complete.

• **Step 3.** the probability of absorption in entry (Bottom, Left) is at most  $\varepsilon$  with respect to  $(\sigma^1, \sigma^2)$

It follows immediately from step 2.

• **Step 4.**  $\gamma^1(\sigma^1, \sigma^2) \leq \varepsilon$

It is a consequence of step 3.

• **Step 5.** the probability of absorption in cell (Bottom, Right) is at most  $2\varepsilon$  with respect to  $(\sigma^1, \sigma^2)$

It is an easy implication of steps 1 and 3.

• **Step 6.** there exists a stage  $n$  for which  $\sigma^2(n, \text{Left}) \geq 1 - 3\varepsilon$

Suppose by way of contradiction that  $\sigma^2(n, \text{Left}) < 1 - 3\varepsilon$  for all  $n \in \mathbb{N}$ . Let  $w$  denote the overall probability of absorption in entry (Bottom, Right), with respect to  $(\sigma^1, \sigma^2)$ . Then by using step 5, we must have  $w \leq 2\varepsilon$ . Hence

$$\gamma^2(\sigma^1, \sigma^2) < w \cdot 1 + (1 - w) \cdot (1 - 3\varepsilon) \leq 2\varepsilon + (1 - 3\varepsilon) = 1 - \varepsilon,$$

which is in contradiction with step 2.

• **Step 7.** Deriving a contradiction

Take a stage  $n$  as in step 6. Consider a Markov strategy  $\tau^1$  for player 1 which prescribes action Top at stages  $1, \dots, n - 1$  and action Bottom at stage  $n$ .

Then

$$\begin{aligned} \gamma^1(\tau^1, \sigma^2) &= 1 \cdot \sigma^2(n, \text{Left}) - 1 \cdot \sigma^2(n, \text{Right}) \\ &= 2 \cdot \sigma^2(n, \text{Left}) - 1 \\ &\geq 2 \cdot (1 - 3\varepsilon) - 1 \\ &= 1 - 6\varepsilon. \end{aligned}$$

By using  $\varepsilon \in [0, \frac{1}{6})$  and step 4, this implies

$$\gamma^1(\tau^1, \sigma^2) \geq 1 - 6\varepsilon > \frac{1}{4} > 2\varepsilon \geq \gamma^1(\sigma^1, \sigma^2) + \varepsilon,$$

which is a contradiction, as  $(\sigma^1, \sigma^2)$  should be an  $\varepsilon$ -equilibrium. Therefore, the proof is complete.  $\square$

**Remark.** Consider now example 1 under the stronger Assumption 1 (or the assumption of full monitoring). Notice that, in this case, player 1 can conclude the past actions of player 2, since player 2's payoffs are different in entries (Top, Left) and (Top, Right). Therefore, the approach in Vrieze & Thuijsman (1989) is directly applicable to proving the existence of  $\varepsilon$ -equilibria, for all  $\varepsilon > 0$ .

### 3 A counterexample under Assumption 1

**Example 2.** Under Assumption 1, consider the following game with 3 players:

0,0,0	0,0,0	0,1,3*	0,0,0	0,0,0	3,1,1*	3,0,1*	3, $\frac{1}{2}$ ,0*	3,1,0*
0,0,0	0,0,0	$\frac{1}{2}$ ,0,3*	0,0,0	0,0,0	1,1,1*	1,3,1*	1,1,1*	1,1,1*
1,3,0*	1,1,3*	1,0,3*	0,3, $\frac{1}{2}$ *	1,1,1*	1,1,1*	0,3,1*	1,1,1*	1,1,1*

This is in fact a cubic 3-player game, in which each player has 3 actions: the actions of player 1 are the rows, the actions of player 2 are the columns and the actions of player 3 are the blocks. The interpretation is further similar to that of example 1.

Observe that the payoff and the transition structures are cyclically symmetric, namely it holds for any entry  $(a^1, a^2, a^3) \in \{1, 2, 3\}^3$  that

- (i)  $r^1(a^1, a^2, a^3) = r^2(a^3, a^1, a^2) = r^3(a^2, a^3, a^1)$
- (ii) entries  $(a^1, a^2, a^3), (a^3, a^1, a^2), (a^2, a^3, a^1)$  are all absorbing or non-absorbing simultaneously.

However, we wish to emphasize that we have only introduced this cyclic symmetry to make the analysis of this game clearer.

In the above game, as long as all the players choose their first or second action all the payoffs equal 0 and no absorption occurs; while as soon as at least one player chooses his third action, absorption occurs with probability 1.

We will show for example 2 under Assumption 1 that there exist no  $\varepsilon$ -equilibria for small  $\varepsilon > 0$  (cf. theorem 10). We will now convey the main ideas lying behind the proof.

*Idea of the proof.* For the sake of simplicity we only explain why 0-equilibria fail to exist. Then the general proof for  $\varepsilon$ -equilibria goes along similar lines (even though it is more technical). We shall now sketch the steps and the arguments and give the formal proofs thereafter.



Suppose by way of contradiction that there exists a 0-equilibrium. We will show that this would imply the existence of a Markov 0-equilibrium (cf. corollary 5). The advantage of dealing with Markov strategies only is, that it is much easier to examine deviations, because such do not effect the behavior of the other players. With respect to a Markov 0-equilibrium we derive a number of steps, which will finally lead to a contradiction.

Because the absorbing payoffs are mainly positive, the first two steps cause no real difficulties.

- **Step 1.** Absorption must occur with probability 1 (cf. lemma 7).
- **Step 2.** Each player's average reward is positive (cf. part 1 of lemma 8).

Notice that absorption cannot occur in entries with payoffs (1,1,1) with probability 1, because at least one player could always improve his reward by switching to action 1 and receive payoff 3 with a good chance. Hence, step 1 implies that absorption must have a positive probability in at least one other absorbing entry. By looking at the sum of the payoffs in each absorbing entry, this implies that the sum of the average rewards must be larger than 3. This observation brings us to the following conclusion.

- **Step 3.** The average reward for at least one player is larger than 1 (cf. part 2 of lemma 8).

Due to symmetry, we may now assume that player 3's reward is more than 1. It is therefore also reasonable to assume that at stage 1 absorption occurs with positive probability and player 3's reward given absorption at stage 1 is larger than 1 (if it is not the case for stage 1, then we could do the same with the first stage which satisfies this property). Let  $x, y$  and  $z$  denote the mixed actions for stage 1. Under the two previous assumptions we proceed as follows. Player 3 surely does not use action 3, as it offers a reward at most 1. Also, action 2 is worse than action 1 for player 3, because his reward given absorption is at most 1 if he selects action 2. Hence the following conclusion.

- **Step 4.**  $z(1) = 1, z(2) = z(3) = 0$  (cf. part 1 of lemma 9).

Now, we only need to look at the first block of the game. We will first derive that  $y(3) = 0$ . Assume, by way of contradiction, that  $y(3) > 0$ . Then, because of step 4, action 1 is worse than action 2 for player 1, hence  $x(1) = 0$ . But then step 2 would imply  $y(3) = 0$ , which is a contradiction. So  $y(3) = 0$  indeed.

Since we assume that the probability of absorption is positive, we conclude that  $x(3) > 0$ , and therefore  $y(2) = 0$  since action 1 gives player 2 a higher absorbing payoff. Thus, from steps 2 and 4, we have derived:

- **Step 5.**  $y(1) = 1, y(2) = y(3) = 0$  (cf. part 2 of lemma 9).

To conclude the sketch of the proof, the final argument is as follows (cf. proof of theorem 10). Recall the assumption that player 3's reward given absorption at stage 1 is more than 1. Then notice that steps 4 and 5 only allow absorption in entry (3,1,1) with payoff 0 for player 3, which is a contradiction. Therefore no 0-equilibrium exists for example 2 under Assumption 1.

We will now turn to the formal proofs. We will often deal with sequences in compact spaces. We shall assume that these sequences are convergent, because we may restrict to a convergent subsequence otherwise.

Based on the property that all the payoffs in the non-absorbing entries are the same (namely 0), one can show the following lemma similarly to part 2 of lemma 1. It says that, essentially, it is sufficient to deal with Markov strategies.

**Lemma 4.** *Consider example 2 (under Assumption 1). For any strategy  $\sigma^1$  of player 1, there exists a Markov strategy  $\bar{\sigma}^1$  so that for any strategies  $\sigma^2$  and  $\sigma^3$  of players 2 and 3*

$$\gamma^i(\sigma^1, \sigma^2, \sigma^3) = \gamma^i(\bar{\sigma}^1, \sigma^2, \sigma^3) \quad \forall i \in I = \{1, 2, 3\}.$$

*A similar statement holds for the other players, as well.*

The next corollary can be shown analogously to corollary 2.

**Corollary 5.** *Consider example 2 (under Assumption 1). Suppose there exists an  $\varepsilon$ -equilibrium for some  $\varepsilon \geq 0$ . Then, there must also exist a Markov  $\varepsilon$ -equilibrium.*

The following lemma deals with the tails of Markov  $\varepsilon$ -equilibria.

**Lemma 6.** *Consider example 2 (under Assumption 1). Let  $(\sigma^1, \sigma^2, \sigma^3)$  be a Markov  $\varepsilon$ -equilibrium for some  $\varepsilon \geq 0$ . Let  $n \in \mathbb{N}$  and suppose that stage  $n$  is reached with positive probability, i.e.*

$$\mathbb{P}_{(\sigma^1, \sigma^2, \sigma^3)}(t^* \geq n) > 0.$$

*For any player  $i$ , let  $\sigma^i|_n$  be the Markov strategy which equals the tail of  $\sigma^i$  starting from stage  $n$  :*

$$\sigma^i|_n(m) := \sigma^i(m + n - 1) \quad \forall m \in \mathbb{N}.$$

*Then  $(\sigma^1|_n, \sigma^2|_n, \sigma^3|_n)$  is a Markov  $\delta$ -equilibrium with*

$$\delta = \frac{\varepsilon}{\mathbb{P}_{(\sigma^1, \sigma^2, \sigma^3)}(t^* \geq n)}.$$

*Proof.* The proof is quite straightforward. Assume by way of contradiction that  $(\sigma^1|_n, \sigma^2|_n, \sigma^3|_n)$  is no  $\delta$ -equilibrium. Then, there must exist a player, say player 1, with a strategy  $\tau^1$  such that

$$\gamma^1(\tau^1, \sigma^2|_n, \sigma^3|_n) - \gamma^1(\sigma^1|_n, \sigma^2|_n, \sigma^3|_n) > \delta.$$

Now consider the Markov strategy  $\bar{\sigma}^1$  for player 1 which coincides with  $\sigma^1$  up to stage  $n - 1$  and starts to play  $\tau^1$  afterwards. Then we have

$$\begin{aligned} \gamma^1(\bar{\sigma}^1, \sigma^2, \sigma^3) - \gamma^1(\sigma^1, \sigma^2, \sigma^3) &= \mathbb{P}_{(\sigma^1, \sigma^2, \sigma^3)}(t^* \geq n) \cdot [\gamma^1(\tau^1, \sigma^2|_n, \sigma^3|_n) \\ &\quad - \gamma^1(\sigma^1|_n, \sigma^2|_n, \sigma^3|_n)] > \mathbb{P}_{(\sigma^1, \sigma^2, \sigma^3)}(t^* \geq n) \cdot \delta \\ &= \varepsilon, \end{aligned}$$

which contradicts the assumption that  $(\sigma^1, \sigma^2, \sigma^3)$  is an  $\varepsilon$ -equilibrium.  $\square$

Next, we observe that absorption should occur almost certainly with respect to Markov  $\varepsilon$ -equilibria, with small  $\varepsilon > 0$ .

**Lemma 7.** *Consider example 2 (under Assumption 1). Let  $(\sigma^1, \sigma^2, \sigma^3)$  be a Markov  $\varepsilon$ -equilibrium for  $\varepsilon \in [0, \frac{1}{3}]$ . Then the overall probability of absorption is more than  $1 - \sqrt{\varepsilon}$ , i.e.:*

$$\mathbb{P}_{(\sigma^1, \sigma^2, \sigma^3)}(t^* < \infty) > 1 - \sqrt{\varepsilon} \quad \text{or equivalently} \quad \mathbb{P}_{(\sigma^1, \sigma^2, \sigma^3)}(t^* = \infty) < \sqrt{\varepsilon}.$$

*Proof.* It can be checked that, with respect to any joint mixed action, there is at least one player who can get at least  $\frac{1}{2}$  by playing action 3. (Take player 1 if player 3 plays action 1 with probability at least  $\frac{1}{2}$ , and take player 2 otherwise.) Therefore, there is always one player who would rather stop play than to continue forever. Therefore we should have  $\mathbb{P}_{(\sigma^1, \sigma^2, \sigma^3)}(t^* = \infty) < \sqrt{\varepsilon}$ , for otherwise the deviation would yield at least  $\frac{1}{2}\sqrt{\varepsilon} > \varepsilon$ , contradicting that  $(\sigma^1, \sigma^2, \sigma^3)$  is an  $\varepsilon$ -equilibrium. Hence the result.  $\square$

The next lemma derives some bounds for rewards corresponding to Markov  $\varepsilon$ -equilibria.

**Lemma 8.** *Consider example 2 (under Assumption 1). There exist  $\bar{\varepsilon} > 0$  and  $M > 0$  such that for any  $\varepsilon \in [0, \bar{\varepsilon}]$  and any Markov  $\varepsilon$ -equilibrium  $(\sigma^1, \sigma^2, \sigma^3)$  we have the following two properties:*

1. for all players  $i \in \{1, 2, 3\}$

$$\gamma^i(\sigma^1, \sigma^2, \sigma^3) \geq 2M,$$

2. there is a player  $i \in \{1, 2, 3\}$  such that

$$\gamma^i(\sigma^1, \sigma^2, \sigma^3) \geq 1 + 2M.$$

*Proof.* Let  $(\sigma_k^1, \sigma_k^2, \sigma_k^3)$  be an arbitrary sequence of Markov  $\varepsilon_k$ -equilibria, for a positive sequence  $\varepsilon_k$  converging to 0. First we derive some general observations.

For any  $(a^1, a^2, a^3) \in \{1, 2, 3\}^3$ , let  $w_k(a^1, a^2, a^3)$  denote the overall probability, with respect to  $(\sigma_k^1, \sigma_k^2, \sigma_k^3)$ , that absorption occurs in entry  $(a^1, a^2, a^3)$ , and let

$$w(a^1, a^2, a^3) = \lim_{k \rightarrow \infty} w_k(a^1, a^2, a^3)$$

(by taking a subsequence, the above limits exist).

Now, let  $E^*$  denote the set of absorbing entries of the game, i.e.  $E^* = \{1, 2, 3\}^3 - \{1, 2\}^3$ .

Clearly, for any  $k \in \mathbb{N}$

$$\mathbb{P}_{(\sigma_k^1, \sigma_k^2, \sigma_k^3)}(t^* < \infty) = \sum_{(a^1, a^2, a^3) \in E^*} w_k(a^1, a^2, a^3)$$

(note that summation is taken over absorbing entries only). Therefore, by lemma 7

$$\begin{aligned}
 \sum_{(a^1, a^2, a^3) \in E^*} w(a^1, a^2, a^3) &= \lim_{k \rightarrow \infty} \sum_{(a^1, a^2, a^3) \in E^*} w_k(a^1, a^2, a^3) \\
 &= \lim_{k \rightarrow \infty} \mathbb{P}_{(\sigma_k^1, \sigma_k^2, \sigma_k^3)}(t^* < \infty) \\
 &= 1.
 \end{aligned}
 \tag{2}$$

As all the payoffs in the non-absorbing entries equal 0, we have for all players  $i \in \{1, 2, 3\}$

$$\begin{aligned}
 \gamma^i(\sigma_k^1, \sigma_k^2, \sigma_k^3) &= \sum_{(a^1, a^2, a^3) \in E^*} w_k(a^1, a^2, a^3) r^{i*}(a^1, a^2, a^3) \\
 \lim_{k \rightarrow \infty} \gamma^i(\sigma_k^1, \sigma_k^2, \sigma_k^3) &= \sum_{(a^1, a^2, a^3) \in E^*} w(a^1, a^2, a^3) r^{i*}(a^1, a^2, a^3).
 \end{aligned}
 \tag{3}$$

Obviously, it suffices to find an  $\bar{\varepsilon}$  and an  $M$  for properties 1 and 2 separately. We will first show statement (2) of the lemma and then continue with statement (1).

*Proof of statement 2.* Suppose this statement is false. Then, there exists a positive sequence  $\varepsilon_k$  converging to 0 and a sequence of Markov  $\varepsilon_k$ -equilibria  $(\sigma_k^1, \sigma_k^2, \sigma_k^3)$  so that

$$\lim_{k \rightarrow \infty} \gamma^i(\sigma_k^1, \sigma_k^2, \sigma_k^3) \leq 1 \tag{4}$$

for all players  $i \in \{1, 2, 3\}$ . Hence

$$\sum_{i \in \{1, 2, 3\}} \lim_{k \rightarrow \infty} \gamma^i(\sigma_k^1, \sigma_k^2, \sigma_k^3) \leq 3.$$

Therefore, (3) yields

$$\sum_{(a^1, a^2, a^3) \in E^*} w(a^1, a^2, a^3) \left[ \sum_{i \in \{1, 2, 3\}} r^{i*}(a^1, a^2, a^3) \right] \leq 3.$$

Observe that the sum of the payoffs is at least 3 in any absorbing entry  $(a^1, a^2, a^3)$ , namely

$$\sum_{i \in \{1, 2, 3\}} r^{i*}(a^1, a^2, a^3) \geq 3.$$

Hence (2) yields

$$\sum_{(a^1, a^2, a^3) \in [\{2, 3\}^3 - \{(2, 2, 2)\}]} w(a^1, a^2, a^3) = 1 \tag{5}$$

(here the summation is taken over all absorbing entries in which the sum of the payoffs equals 3).

We distinguish two cases. In each case we derive a contradiction by finding a player with a profitable deviation.

- **Case 1.**  $w(2, 3, 2) + w(2, 2, 3) + w(2, 3, 3) > 0$  or  $w(3, 2, 2) + w(2, 2, 3) + w(3, 2, 3) > 0$  or  $w(3, 2, 2) + w(2, 3, 2) + w(3, 3, 2) > 0$

By symmetry, we may assume without loss of generality that

$$w(2, 3, 2) + w(2, 2, 3) + w(2, 3, 3) > 0.$$

Clearly (5) yields  $w(2, 3, 1) = 0$ . Note that the payoffs for player 1 in entries  $(1, 3, 2), (1, 2, 3), (1, 3, 3)$  are all equal to 3, while the payoffs in entries  $(2, 3, 2), (2, 2, 3), (2, 3, 3)$  are all equal to 1. This implies that, for  $k$  sufficiently large (or equivalently  $\varepsilon_k$  sufficiently small), player 1 allows for absorption through his second action while absorption through his first action would yield a higher payoff, namely 3 instead of 1. This contradicts the fact that the triples  $(\sigma_k^1, \sigma_k^2, \sigma_k^3)$  are  $\varepsilon_k$ -equilibria.

• **Case 2.**  $w(3, 3, 3) = 1$

This can only happen if, for  $k$  sufficiently large (or equivalently  $\varepsilon_k$  sufficiently small), there exists a stage such that the players play their third action simultaneously with probability almost 1. Then by using that the payoff for player 1 is higher in entry  $(1, 3, 3)$  than in  $(3, 3, 3)$ , we again derive a contradiction with the fact that the triples  $(\sigma_k^1, \sigma_k^2, \sigma_k^3)$  are  $\varepsilon_k$ -equilibria.

*Proof of statement 1.* Suppose this statement is false. Then, there exists a positive sequence  $\varepsilon_k$  converging to 0 and a sequence of Markov  $\varepsilon_k$ -equilibria  $(\sigma_k^1, \sigma_k^2, \sigma_k^3)$  so that

$$\lim_{k \rightarrow \infty} \gamma^i(\sigma_k^1, \sigma_k^2, \sigma_k^3) = 0$$

for some player  $i \in \{1, 2, 3\}$ . Due to symmetry, we may assume without loss of generality that

$$\lim_{k \rightarrow \infty} \gamma^1(\sigma_k^1, \sigma_k^2, \sigma_k^3) = 0. \tag{6}$$

Then just like for (5) in the proof of part 1, one can show that

$$w(1, 3, 1) + w(3, 1, 2) + w(3, 1, 3) = 1.$$

We distinguish three cases. In each case we derive a contradiction by finding a player with a profitable deviation.

• **Case 1.**  $w(1, 3, 1) > 0$

Clearly impossible because player 1 could get a positive reward by shifting all the weights from action 1 to action 2, for which all absorbing entries give positive payoffs. Especially entry  $(2, 3, 1)$  will appear with positive probability, giving player 1 a payoff  $\frac{1}{2}$ .

• **Case 2.**  $w(3, 1, 3) > 0$

The proof is similar to case 1, but now player 1 has to shift weight from action 3 to action 2.

• **Case 3.**  $w(3, 1, 2) = 1$

This condition implies

$$\lim_{k \rightarrow \infty} \gamma^3(\sigma_k^1, \sigma_k^2, \sigma_k^3) = \frac{1}{2}$$

and also that there must be a stage at which player 2 plays his first action with probability almost 1. It is now player 3 who can deviate in a profitable way by playing action 3 at that stage, which gives him 1. □

Next we derive for Markov  $\varepsilon$ -equilibria some properties on the overall probability of absorption in certain entries. Take  $\bar{\varepsilon}$  and  $M$  as in lemma 8. From now on let  $\varepsilon_k \in [0, \min\{\bar{\varepsilon}^2, \frac{1}{3}\}]$  be a sequence converging to 0 and let  $(\sigma_k^1, \sigma_k^2, \sigma_k^3)$  be Markov  $\varepsilon_k$ -equilibria. For any player  $i$  and  $k \in \mathbb{N}$ , let  $\sigma_k^i|_n$  be the Markov strategy which equals the tail of  $\sigma_k^i$  starting from stage  $n$ :

$$\sigma_k^i|_n(m) := \sigma_k^i(m + n - 1) \quad \forall m \in \mathbb{N}.$$

Define for all players  $i$  and for all  $k, n \in \mathbb{N}$  the average reward conditioned on starting at stage  $n$  by:

$$g_k^i(n) := \gamma^i(\sigma_k^1|_n, \sigma_k^2|_n, \sigma_k^3|_n).$$

In view of lemma 8 we assume, without loss of generality, that, similar to the 0-equilibrium case, player 3's average reward is essentially larger than 1, i.e.

$$g_k^3(1) = \gamma^3(\sigma_k^1, \sigma_k^2, \sigma_k^3) \geq 1 + 2M \quad \forall k \in \mathbb{N}.$$

For any  $k \in \mathbb{N}$ , let  $B_k$  be the set of stages which are reached with probability at least  $\sqrt{\varepsilon_k}$  with respect to  $(\sigma_k^1, \sigma_k^2, \sigma_k^3)$ :

$$B_k = \{n \in \mathbb{N} \mid \mathbb{P}(\sigma_k^1, \sigma_k^2, \sigma_k^3)(t^* \geq n) \geq \sqrt{\varepsilon_k}\}.$$

Then  $B_k$  is a connected set of stages that contains at least the initial stage 1. By lemma 7 the set  $B_k$  is finite and the probability of absorption in  $B_k$  is almost 1:

$$\begin{aligned} \mathbb{P}(\sigma_k^1, \sigma_k^2, \sigma_k^3)(t^* \in B_k) &= \mathbb{P}(\sigma_k^1, \sigma_k^2, \sigma_k^3)(t^* < \infty) - \mathbb{P}(\sigma_k^1, \sigma_k^2, \sigma_k^3)(t^* < \infty \text{ and } t^* \notin B_k) \\ &\geq (1 - \sqrt{\varepsilon_k}) - \sqrt{\varepsilon_k} = 1 - 2\sqrt{\varepsilon_k}. \end{aligned} \tag{7}$$

Hence the average reward is mainly determined on  $B_k$ .

Also notice that for each  $n \in B_k$  the strategy triple  $(\sigma_k^1|_n, \sigma_k^2|_n, \sigma_k^3|_n)$  is a  $\delta_{k,n}$ -equilibrium, where by using lemma 6:

$$\delta_{k,n} = \frac{\varepsilon_k}{\mathbb{P}(\sigma_k^1, \sigma_k^2, \sigma_k^3)(t^* \geq n)} \leq \sqrt{\varepsilon_k} \leq \bar{\varepsilon}.$$

Therefore part 1 of lemma 8 is applicable for the triple  $(\sigma_k^1|_n, \sigma_k^2|_n, \sigma_k^3|_n)$  and yields for each player  $i$ :

$$\gamma^i(\sigma_k^1|_n, \sigma_k^2|_n, \sigma_k^3|_n) \geq 2M. \tag{8}$$

Let  $n_k$  be the *last* stage of the first block in  $B_k$  at which player 3's future reward is above  $1 + M$ , i.e.

$$g_k^3(1) \geq 1 + M, \dots, g_k^3(n_k) \geq 1 + M.$$

Define this first block by  $B_k^3$ , so  $B_k^3 = \{1, 2, \dots, n_k\}$ . Notice that either  $B_k^3 = B_k$  or  $g_k^3(n_k + 1) < 1 + M$ . Moreover, let  $B_k^{3*}$  consist of those stages  $n$  in  $B_k^3$  for which

$$\mathbb{P}(\sigma_k^1, \sigma_k^2, \sigma_k^3)(t^* = n) > 0 \quad \text{and} \quad r^{3*}(\sigma_k^1(n), \sigma_k^2(n), \sigma_k^3(n)) \geq 1 + M. \tag{9}$$

For any  $(a^1, a^2, a^3) \in \{1, 2, 3\}^3$ , let  $w_k(a^1, a^2, a^3)$  denote the overall probability, with respect to  $(\sigma_k^1, \sigma_k^2, \sigma_k^3)$ , that absorption occurs in entry  $(a^1, a^2, a^3)$  at some stage in  $B_k^{3*}$ , and let

$$w(a^1, a^2, a^3) = \lim_{k \rightarrow \infty} w_k(a^1, a^2, a^3)$$

(due to compactness, it suffices to deal with convergent subsequences). Therefore, similarly to steps 4 and 5 in the 0-equilibrium case, players 2 and 3 will essentially refrain from using actions 2 and 3.

**Lemma 9.**  $w(\cdot, \cdot, 2) = w(\cdot, \cdot, 3) = 0$  and  $w(\cdot, 2, \cdot) = w(\cdot, 3, \cdot) = 0$ .

*Proof.* Notice that for any  $n \in B_k^{3*}$  we have

$$r^{3*}(\sigma_k^1(n), \sigma_k^2(n), 2) \leq 1, \quad r^{3*}(\sigma_k^1(n), \sigma_k^2(n), 3) \leq 1, \tag{10}$$

hence, by (9), we must also have

$$r^{3*}(\sigma_k^1(n), \sigma_k^2(n), 1) \geq 1 + M. \tag{11}$$

We divide the proof into several steps. Below, we will consider several deviations of the players at stages in  $B_k^{3*}$ , for large  $k \in \mathbb{N}$ . In most cases, the deviations consist of shifting probability from action 2 to action 1 or from action 1 to action 2. For these deviations, the distribution of  $t^*$  (time of absorption) does not change because of the transition structure of the game. (Note that the place of absorption may actually change.) However, we also need to consider deviations when the probabilities are shifted from action 3 (to either action 1 or action 2). Since here the distribution of  $t^*$  will change, we need to examine these deviations in more detail.

• **Step 1.**  $w(\cdot, \cdot, 3) = 0$

Take an entry  $(a^1, a^2, 3)$  for some  $a^1, a^2 \in \{1, 2, 3\}$ . We will show for any  $k \in \mathbb{N}$  that there exist a Markov strategy  $\tau_k^3$  for player 3 for which

$$\gamma^3(\sigma_k^1, \sigma_k^2, \tau_k^3) - \gamma^3(\sigma_k^1, \sigma_k^2, \sigma_k^3) \geq w_k(a^1, a^2, 3) \cdot M. \tag{12}$$

Since  $(\sigma_k^1, \sigma_k^2, \sigma_k^3)$  is an  $\varepsilon_k$ -equilibrium, the left-hand side must be at most  $\varepsilon_k$ . Taking the limit for  $k$  to  $\infty$  completes the proof.

Let  $\tilde{t}$  denote the time of absorption in entry  $(a^1, a^2, 3)$ ; if no absorption occurs in this entry at all then let  $\tilde{t} = \infty$ . So

$$\mathbb{P}_{(\sigma_k^1, \sigma_k^2, \sigma_k^3)}(\tilde{t} \in B_k^{3*}) = w_k(a^1, a^2, 3). \tag{13}$$

Suppose  $m_k^1$  is the first stage of  $B_k^{3*}$  and consider player 3 deviating from  $\sigma_k^3$  by only shifting at stage  $m_k^1$ , the probability from action 3 to action 1. Denote this strategy by  $\tau_k^{3,1}$ . Because of (10) and (11) action 3 gives at most 1 to player 3, and action 1 dominates action 2 at stage  $m_k^1$ . Since  $m_k^1 \in B_k^{3*} \subset B_k^3$  we must have that at stage  $m_k^1$  action 1 yields an expected reward of at least  $1 + M$ . Therefore,

$$\begin{aligned} \gamma^3(\sigma_k^1, \sigma_k^2, \tau_k^{3,1}) - \gamma^3(\sigma_k^1, \sigma_k^2, \sigma_k^3) &\geq \mathbb{P}_{(\sigma_k^1, \sigma_k^2, \sigma_k^3)}(t^* \geq m_k^1) \cdot ((1 + M) - 1) \\ &\geq \mathbb{P}_{(\sigma_k^1, \sigma_k^2, \sigma_k^3)}(\tilde{t} = m_k^1) \cdot M. \end{aligned}$$

Let  $m_k^2$  be the second stage of  $B_k^{3*}$  and consider player 3 deviating from  $\sigma_k^3$  by only shifting at stages  $m_k^1$  and  $m_k^2$ , the probability from action 3 to action 1. Denote this strategy by  $\tau_k^{3,2}$ . Then, similarly:

$$\begin{aligned} \gamma^3(\sigma_k^1, \sigma_k^2, \tau_k^{3,2}) - \gamma^3(\sigma_k^1, \sigma_k^2, \tau_k^{3,1}) &\geq \mathbb{P}(\sigma_k^1, \sigma_k^2, \tau_k^{3,1})(t^* \geq m_k^2) \cdot ((1 + M) - 1) \\ &\geq \mathbb{P}(\sigma_k^1, \sigma_k^2, \tau_k^{3,1})(\tilde{t} = m_k^2) \cdot M. \\ &\geq \mathbb{P}(\sigma_k^1, \sigma_k^2, \sigma_k^3)(\tilde{t} = m_k^2) \cdot M. \end{aligned}$$

We can continue like this for all stages in the finite set  $B_k^{3*}$  and, reaching the final stage of that set, we have a strategy  $\tau_k^3$  which has player 3 shifting weight from action 3 to action 1 at all stages in  $B_k^{3*}$ . Then, by (13) and identifying  $\tau_k^{3,0}$  with  $\sigma_k^3$ , we get

$$\begin{aligned} \gamma^3(\sigma_k^1, \sigma_k^2, \tau_k^3) - \gamma^3(\sigma_k^1, \sigma_k^2, \sigma_k^3) &= \sum_d \left( \gamma^3(\sigma_k^1, \sigma_k^2, \tau_k^{3,d}) - \gamma^3(\sigma_k^1, \sigma_k^2, \tau_k^{3,d-1}) \right) \\ &\geq \sum_d \mathbb{P}(\sigma_k^1, \sigma_k^2, \sigma_k^3)(\tilde{t} = m_k^d) \cdot M. \\ &= w_k(a^1, a^2, 3) \cdot M. \end{aligned}$$

Hence, step 1 follows.

• **Step 2.**  $w(\cdot, \cdot, 2) = 0$

Take an absorbing entry  $(a^1, a^2, 2)$  for some  $a^1, a^2 \in \{1, 2, 3\}$ . Consider a deviation by player 3 from  $\sigma_k^2$  that consists of shifting the probability from action 2 to action 1 at all stages in  $B_k^{3*}$ . Then his reward would increase, in view of (10) and (11), by at least

$$w_k(a^1, a^2, 2) \cdot [(1 + M) - 1].$$

Since  $(\sigma_k^1, \sigma_k^2, \sigma_k^3)$  is an  $\varepsilon_k$ -equilibrium, this expression must be at most  $\varepsilon_k$ . Taking the limit for  $k$  to  $\infty$  completes the proof.

• **Step 3.**  $w(\cdot, 3, \cdot) = 0$

Notice that it is enough to show  $w(\cdot, 3, 1) = 0$ , because of steps 1 and 2.

Consider a deviation by player 1 from  $\sigma_k^1$  that consists of shifting the probability from action 1 to action 2 at all stages in  $B_k^{3*}$ . Then he would gain at least

$$w_k(1, 3, 1) \cdot \left[ \frac{1}{2} - 0 \right].$$

Since  $(\sigma_k^1, \sigma_k^2, \sigma_k^3)$  is an  $\varepsilon_k$ -equilibrium, this expression must be at most  $\varepsilon_k$ . Taking the limit for  $k$  to  $\infty$  yields  $w(1, 3, 1) = 0$ .

Consider player 2 deviating by shifting the probability from action 3 to action 1 at all stages in  $B_k^{3*}$ . In fact, by using that  $w(1, 3, 1) = 0$ , and the approach of the proof of step 1, it may be seen that player 2 could gain at least

$$\frac{1}{2} [w_k(2, 3, 1) + w_k(3, 3, 1)] \cdot [M - 0].$$

Since  $(\sigma_k^1, \sigma_k^2, \sigma_k^3)$  is an  $\varepsilon_k$ -equilibrium, this expression must be at most  $\varepsilon_k$ . Taking the limit for  $k$  to  $\infty$  yields  $w_k(2, 3, 1) = w_k(3, 3, 1) = 0$ , which completes the proof.



• **Step 4.**  $w(\cdot, 2, \cdot) = 0$

Notice that it is enough to show  $w(3, 2, 1) = 0$ , because of steps 1 and 2.

Consider player 2 deviating by shifting the probability from action 2 to action 1 at all stages in  $B_k^{3*}$ . By doing so he could gain at least

$$w_k(3, 2, 1) \cdot [3 - 1].$$

Since  $(\sigma_k^1, \sigma_k^2, \sigma_k^3)$  is an  $\varepsilon_k$ -equilibrium, this expression must be at most  $\varepsilon_k$ . Taking the limit for  $k$  to  $\infty$  completes the proof.  $\square$

Now we are ready to prove the main theorem of this section.

**Theorem 10.** *Consider example 2 (under Assumption 1). There exist no  $\varepsilon$ -equilibria for small  $\varepsilon > 0$ .*

*Proof.* Suppose the opposite. Let  $\varepsilon_k$  be a positive sequence converging to 0. Then by corollary 5 there exists a Markov  $\varepsilon_k$ -equilibrium  $(\sigma_k^1, \sigma_k^2, \sigma_k^3)$  for all  $k \in \mathbb{N}$ .

We will now use the notations of lemma 9. By disregarding symmetric cases and taking a subsequence, in view of lemma 8, we may assume that  $\varepsilon_k \in [0, \min\{\bar{\varepsilon}^2, \frac{1}{5}\}]$  and

$$g_k^3(1) = \gamma^3(\sigma_k^1, \sigma_k^2, \sigma_k^3) \geq 1 + 2M \text{ for all } k \in \mathbb{N}. \tag{14}$$

First we would like to show that the probability of absorption in  $B_k^{3*}$  goes to 0:

$$\lim_{k \rightarrow \infty} \mathbb{P}_{(\sigma_k^1, \sigma_k^2, \sigma_k^3)}(t^* \in B_k^{3*}) = 0.$$

Suppose not, then, because of

$$\mathbb{P}_{(\sigma_k^1, \sigma_k^2, \sigma_k^3)}(t^* \in B_k^{3*}) = \sum_{(a^1, a^2, a^3)} w_k(a^1, a^2, a^3),$$

we have that there is an absorbing entry  $(a^1, a^2, a^3)$  for which  $w(a^1, a^2, a^3) > 0$ . By lemma 9 we conclude that  $(a^1, a^2, a^3) = (3, 1, 1)$  since  $w$  is 0 in all other entries. Therefore, player 3's average reward conditioned on absorption in  $B_k^{3*}$  goes to 0. This contradicts the fact that, given absorption at any stage in  $B_k^{3*}$  player 3 would get at least  $1 + M$ .

Observe that

$$g_k^3(1) \leq \mathbb{P}_{(\sigma_k^1, \sigma_k^2, \sigma_k^3)}(t^* \in B_k^{3*}) \cdot 3 + \mathbb{P}_{(\sigma_k^1, \sigma_k^2, \sigma_k^3)}(t^* \in B_k^3 - B_k^{3*}) \cdot (1 + M) + \mathbb{P}_{(\sigma_k^1, \sigma_k^2, \sigma_k^3)}(t^* \notin B_k^3) \cdot g_k^3(n_k + 1).$$

Recall that either  $g_k^3(n_k + 1) < 1 + M$  or  $B_k^3 = B_k$ . In the latter case, recall that  $\mathbb{P}_{(\sigma_k^1, \sigma_k^2, \sigma_k^3)}(t^* \notin B_k^3)$  goes to 0 by (7). Therefore, by taking limits we always have:

$$\lim_{k \rightarrow \infty} g_k^3(1) \leq 1 + M,$$

which contradicts (14). So, the proof is complete.  $\square$

**Remark.** For 3-player absorbing games under the assumption of full monitoring, Solan (1999) has shown the existence of  $\varepsilon$ -equilibria for all  $\varepsilon > 0$ . However, it is crucial that the players are able to observe the actions of the other players.

We wish to remark that example 2 has some similarity with an example examined in Flesch et al. (1997), but the latter example can not be used for the purposes of this paper.

Finally, it should be noted that the situation for 2-player stochastic games under Assumption 1 remains unclear and subject to further study.

## References

- [1] Blackwell D, Ferguson TS (1968) The big match. *Annals of Mathematical Statistics* 39:159–163
- [2] Coulomb JM (1992) Repeated games with absorbing states and no signals. *International Journal of Game Theory* 21:161–174
- [3] Coulomb JM (1999) Generalized “Big Match”. *Mathematics of Operations Research* 4:795–816
- [4] Flesch J, Thuijsman F, Vrieze OJ (1997) Cyclic Markov equilibria in a cubic game. *International Journal of Game Theory* 26:303–314
- [5] Gillette D (1957) Stochastic games with zero stop probabilities. In: Dresher M, Tucker AW, Wolfe P (eds) *Contributions to the Theory of Games III*. *Annals of Mathematical Studies* 39, Princeton University Press, pp. 179–187
- [6] Solan E (1999) Three-person absorbing games. *Mathematics of Operations Research* 24:669–698
- [7] Vieille N (2000, I) 2-person stochastic games I: A reduction. *Israel Journal of Mathematics* 119:55–91
- [8] Vieille N (2000, II) 2-person stochastic games II: The case of recursive games. *Israel Journal of Mathematics* 119:93–126
- [9] Vieille N (2000, III) Solvable States in n-Player Stochastic Games. *SIAM Journal on Control and Optimization* 38:1794–1804
- [10] Vrieze OJ, Thuijsman F (1989) On equilibria in repeated games with absorbing states. *International Journal of Game Theory* 18:293–310