

Repeated Games with Bonuses

G. Schoenmakers · J. Flesch · F. Thuijsman ·
O.J. Vrieze

Published online: 5 February 2008
© The Author(s) 2008

Abstract This paper deals with 2-player zero-sum repeated games in which player 1 receives a bonus at stage t if he repeats the action he played at stage $t - 1$. We investigate the optimality of simple strategies for player 1. A simple strategy for player 1 consists of playing the same mixed action at every stage, irrespective of past play. Furthermore, for games in which player 1 has a simple optimal strategy, we characterize the set of stationary optimal strategies for player 2.

Keywords Noncooperative games · Repeated games · Stochastic games · Optimal strategies

1 Introduction

Consider the following little story: Someone is learning to play a card or a board game. In the beginning he is a fanatical player, trying to reach the height of his powers. He plays the game frequently, investing every hour of his spare time practicing. As a result his skills improve dramatically and within a few years he manages to beat all the opponents that beat him in the beginning, thereby becoming the champion. Having reached the ultimate goal, the player is unable to find any new challenges in the game and consequently it starts losing its appeal to him. Our player loses interest in the game and he does not (want to) make time to play and/or practice the game on a regular basis anymore. As a result he gradually starts to lose some of the skills that he obtained, which affects his play and thereby his results. This little story provides a classical example of how people can learn and unlearn certain skills, purely based on exercising or ceasing to exercise them.

Communicated by F. Zirilli.

G. Schoenmakers (✉) · J. Flesch · F. Thuijsman · O.J. Vrieze
Department of Mathematics, Maastricht University, Maastricht, The Netherlands
e-mail: GM.Schoenmakers@MICC.unimaas.nl

These phenomena may be called learning by doing and unlearning by not doing respectively. From the seminal paper of Arrow [1] on, where learning is considered a by-product of doing, rather than an objective, learning (by doing) has become a very popular subject of research in economics and in game-theory. However, on the unlearning side of the story, although very prominent present in real life, very little research has been done, perhaps due to the fact that people often unlearn, or forget, things that are no longer relevant to them. In 1995 Joosten et al. [2] introduced a model of unlearning for infinitely repeated zero-sum games, which was generalized by Schoenmakers [3] to nonzero-sum games. In this paper we discuss a model that deals with learning and unlearning in infinitely repeated games in a slightly different way: We will insert a skill-improvement and -deterioration component into the model of zero-sum repeated games. Suppose that player 1 played action i at stage t . Now at stage t player 1 has learned “right how to handle action i ” and action i becomes a high-skill action at the next stage. Now if player 1 decides to play action i again at stage $t + 1$, then he receives from player 2 a bonus $\xi \geq 0$. However, if at stage $t + 1$ player 1 decides to play action $\hat{i} \neq i$, then he will not receive a bonus and he will forget how to take advantage of action i , which we will call the unlearning of action i , and action i becomes low-skill. At the same time player 1 learns action \hat{i} that thereby becomes high-skill, so if player 1 decides to play action \hat{i} again at stage $t + 2$, then he receives from player 2 the bonus ξ , whereas if he plays action i at stage $t + 2$, then, having unlearned it at stage $t + 1$, he does not receive a bonus.

A realistic interpretation of a zero-sum repeated game with a bonus, in which only player 1 can get the bonus, is considering player 2 to be a computer that always plays the game at a certain fixed level. Hence a player gets a higher reward (i.e. improves his results against the computer) by playing better himself, which is expressed as a bonus on top of the normal payoff. Notice that if $\xi = 0$, i.e. there is no bonus, then the game reduces to an ordinary zero-sum repeated game.

In Sect. 2 the model of the (zero-sum) repeated game with bonus ξ is presented in more detail. In Sect. 3 some results concerning the value of the game are presented. In Sect. 4 we characterize the set of stationary optimal strategies of the generalizations of 2×2 -matrix games. Furthermore we present conditions for a specific stationary strategy to be optimal, namely one that prescribes to play, at each stage, a mixed action that is optimal in the underlying matrix game. Such a strategy will be called a simple strategy. Section 5 deals with $2 \times n$ -games and in Sect. 6 we take a look at games of arbitrary size.

2 Model

A zero-sum repeated game is characterized by a payoff-matrix M ; the corresponding repeated game with bonus ξ is characterized by the same payoff matrix M in combination with the bonus ξ and it proceeds as follows: Take an $(m \times n)$ -matrix M and consider the corresponding matrix game M with action sets $\{1, \dots, m\}$ and $\{1, \dots, n\}$ for players 1 and 2 respectively that is played repeatedly. At each stage the players are assumed to choose actions independently and simultaneously and if player 1 chooses action i and player 2 chooses action j , then player 1 receives an amount of m_{ij} from

player 2, where m_{ij} is the (i, j) -th entry of M . However, if player 1 also selected action i at the previous stage, then he receives $m_{ij} + \xi$.

Definition 2.1 The repeated game with bonus ξ corresponding to the matrix game $M \in \mathbb{R}^{m \times n}$ can be reformulated as a zero-sum stochastic game with finite state and action spaces with the following properties:

- (i) The set of states is $\{1, \dots, m\}$; state s relates to player 1’s action s in M .
- (ii) In each state, players 1 and 2 have action sets $\{1, \dots, m\}$ and $\{1, \dots, n\}$.
- (iii) The payoffs in state s relate in the following way to the payoffs in M : The payoffs in row $r \neq s$ are equal to the corresponding payoffs in M , whereas the payoffs in row s are increased by ξ .
- (iv) The state transition structure is straightforward: Play can start in each of the m states. If, at a certain stage, play is in state s and player 1 plays action s' , then with probability 1 play moves to state s' .

Notation 2.1 Both the repeated game with bonus ξ corresponding to the matrix game M and the stochastic game representation of Definition 2.1 are called M_ξ .

Stochastic games were introduced by Shapley [4] in a more general fashion. For basic results we refer to Thuijsman [5] and Filar and Vrieze [6].

A mixed action a^s (b^s) for player 1 (2) in state s of M_ξ is a probability distribution over $\{1, \dots, m\}$ ($\{1, \dots, n\}$). Each pure action is a mixed action where one of the actions is chosen with probability 1. A strategy for player 1 (2) is a map π (σ) that assigns a mixed action in any present state for any history (the collection of all visited states and chosen actions pairs in the past). A strategy is called stationary if it prescribes to play the same mixed action each time state s is visited. We denote such a strategy by $x := (a^1, a^2, \dots, a^m)$ for player 1 and $y := (b^1, b^2, \dots, b^m)$ for player 2. Furthermore a pure strategy is a strategy that prescribes to play a pure action at each stage. As evaluation criterion for the stream of payoffs generated by the strategy pair (π, σ) , we will use the *limiting average reward*, i.e.

$$\gamma_\xi^s(\pi, \sigma) = \liminf_{T \rightarrow \infty} (1/T) \sum_{t=1}^T E_{\pi, \sigma}^s(R_t),$$

where $E_{\pi, \sigma}^s(R_t)$ denotes the expected payoff to player 1 at stage t given that (π, σ) is being played and that the initial state is state $s \in \{1, \dots, m\}$. Player 1 wants to maximize $\gamma_\xi^s(\pi, \sigma)$, whereas player 2 wants to minimize the same reward. The limiting average reward was introduced by Gillette [7].

3 General Results

Notation 3.1 The unit simplex in R^z is denoted by Δ^z .

Theorem 3.1 (See Von Neumann [8]) *Each matrix game M has a value $v \in \mathbb{R}$, for which*

$$v := \min_{b \in \Delta^n} \max_{a \in \Delta^m} a^T M b = \max_{a \in \Delta^m} \min_{b \in \Delta^n} a^T M b.$$

This implies that there exist (mixed) actions $a^ \in \Delta^m$ and $b^* \in \Delta^n$ such that $a^{*T} M b \geq v \geq a^T M b^*$ for all mixed actions a and b of players 1 and 2 respectively. Such actions a^* and b^* are called optimal.*

Mertens and Neyman [9] proved that each zero-sum stochastic games has a value $v = (v^1, v^2, \dots, v^m)$, which may depend on the initial state. In zero-sum stochastic games a strategy π^* for player 1 is called optimal if $\gamma^s(\pi^*, \sigma) \geq v^s$ for all player 2’s strategies σ and for each initial state s . Analogously for player 2, σ^* is optimal if $\gamma^s(\pi, \sigma^*) \leq v^s$ for all π and for all s . In general, optimal strategies fail to exist, a famous example of which is the so-called Big Match by Gillette [7]. For M_ξ however, not only do optimal strategies exist, but as Theorem 3.2 shows, even stationary optimal strategies exist.

Theorem 3.2 *The game M_ξ has a state-independent value and players 1 and 2 have stationary optimal strategies x^* and y^* respectively.*

Proof According to point (iv) of the model description, the zero-sum game M_ξ has state independent transitions and these are controlled by player 1. For games with these properties it is known that both players possess stationary optimal strategies (cf. Filar [10]). Furthermore Thuijsman [5] proved that for zero-sum games with state independent transitions the average reward value is independent of the initial state. \square

The state independent value of M_ξ will be called v_ξ and, by Theorem 3.2, we have

$$\gamma_\xi^s(x^*, y^*) = v_\xi \tag{1}$$

for each pair of stationary optimal strategies and for each initial state s .

We are especially interested in a specific type of stationary strategies of player 1, the so-called simple strategies, which are defined as follows.

Definition 3.1 A stationary strategy $x = (a^1, a^2, \dots, a^m)$ is called simple, if $a^i = a^j$ for all $i, j \in \{1, \dots, m\}$. The simple strategy that prescribes to play the mixed action $a \in \Delta^m$ in each state and stage is denoted by a' .

Notice that, for each simple strategy a' of player 1 and each stationary strategy y of player 2, as a consequence of the fact that there is only one ergodic class, we have

$$\gamma_\xi^s(a', y) = \gamma_\xi^{s'}(a', y), \quad \text{for all } s, s' \in \{1, \dots, m\}. \tag{2}$$

Therefore w.l.o.g. if player 1 uses a simple strategy, we shall write γ_ξ instead of γ_ξ^s .

Definition 3.2 The carrier of a mixed action a is defined as follows: $\text{car}(a) = \{i \in \{1, \dots, m\} \mid a_i > 0\}$. For the stationary strategy $x = (a^1, a^2, \dots, a^m)$ let $\text{car}(x)$ denote the Cartesian product $\text{car}(a^1) \times \text{car}(a^2) \times \dots \times \text{car}(a^m)$.

Notation 3.2 The variable s will denote an action of player 1, a state or even both at once. The line “state $s \in \text{car}(a)$ ” should be interpreted as the state s that is visited after player 1 plays action $s \in \text{car}(a)$, when player 1 uses the simple strategy a' . Furthermore, a superscript refers to a state, whereas a subscript refers to an action. So a^s is a component of the stationary strategy x , a probability vector prescribing a mixed action in state s , whereas a_s is a component of the mixed action $a \in \Delta^m$, denoting the probability to play action s .

Notation 3.3 Let $B(a)$ denote the set of best replies of player 2 against the (mixed) action a in M and, similarly, let $B_\xi(x)$ denote the set of stationary best replies of player 2 against the stationary strategy x of player 1 in M_ξ . Both sets are clearly nonempty.

Since player 1 controls the transitions, player 2 essentially plays a one-shot game each stage. Therefore in order to determine $B_\xi(x)$ it suffices to consider the sets of one-shot best replies per state. The payoffs in state s of M_ξ only differ from the payoffs in M in row s ; in state s of M_ξ they are exactly an amount ξ higher than in M . Now let e_j be the j th unit vector. Then,

$$\min_{j \in J} \{a^T M e_j + \xi \cdot a_s\} = \min_{j \in J} \{a^T M e_j\} + \xi \cdot a_s \quad \text{for all } s \in \{1, \dots, m\},$$

for each mixed action a and therefore player 2’s set of one-shot best replies against a^s in state s of M_ξ is the same as his set of best replies against a^s in M .

Consider the simple strategy a' . Notice that a' induces a Markov-chain over the set of states, in which state s is visited with frequency a_s . Therefore each state $s \in \text{car}(a)$ will be visited infinitely often, whereas each state outside $\text{car}(a)$ will not be visited at all, except if it happens to be the initial state. Now a stationary strategy $y = (b^1, b^2, \dots, b^m)$ is a best reply against a' if and only if for each state $s \in \text{car}(a)$ we have: $b^s \in B(a)$. Furthermore, each state $s \notin \text{car}(a)$ will not be visited and it makes no difference what player 2 would have played in that state. Hence each $b^s \in \Delta^n$ suffices. The set $B_\xi(a')$ is the Cartesian product of the sets the b^s ’s have to belong to. Hence, for the mixed action a , we have

$$B_\xi(a') = \{(b^1, \dots, b^m) \mid b^s \in B(a) \text{ for all } s \in \text{car}(a), b^s \in \Delta^n \text{ for all } s \notin \text{car}(a)\}. \tag{3}$$

Notation 3.4 The reward that player 1 can guarantee himself by playing the stationary strategy x , is denoted by $\varphi_\xi(x)$, which may depend on the initial state. So $\varphi_\xi^s(x) = \min_y \gamma_\xi^s(x, y)$. Notice that for each simple strategy a' we have that $\varphi_\xi(a')$ is independent of the initial state (cf. (2)). Notice furthermore that, by (1), for any stationary optimal strategy x^* also $\varphi_\xi(x^*) = v_\xi$ is independent of the initial state. Further, let $\varphi_0(a') = \min_{b \in B} a^T M b$.

From (3), we derive the next lemma.

Lemma 3.1 For each $y = (b^1, b^2, \dots, b^m) \in B_\xi(a')$ and $s \in \text{car}(a')$, we have

$$a^T M b^s = \varphi_0(a').$$

Notice that for any optimal mixed action a^* in M we have $\varphi_0(a^*) = v$.

Given the strategy pair (a', y) , with $y \in B_\xi(a')$, if, at a certain stage, state $s \in \text{car}(a')$ is visited, then the expected immediate payoff to player 1 is $\varphi_0(a') + \xi \cdot a_s$. State s is visited with frequency a_s , so

$$\varphi_\xi(a') = \sum_{s=1}^m a_s \cdot (\varphi_0(a') + \xi \cdot a_s) = \varphi_0(a') + \xi \cdot \sum_{s=1}^m a_s^2. \tag{4}$$

Since $a^{*'}$ might not be optimal in M_ξ , for the optimal mixed action a^* in M we have

$$v_\xi \geq \varphi_\xi(a^{*'}) = v + \xi \cdot \sum_{s=1}^m (a_s^*)^2. \tag{5}$$

However, if a' would be a simple optimal strategy, then by definition

$$v_\xi = \varphi_\xi(a') = \varphi_0(a') + \xi \cdot \sum_{s=1}^m a_s^2. \tag{6}$$

Suppose that in M the (mixed) action y is optimal for player 2 and consider the simple strategy y' of player 2 in M_ξ . Then $\gamma_\xi^s(x, y') \leq v + \xi$ for all initial states s and for all strategies x of player 1, since at each stage the expected immediate payoff to player 1 in M_ξ is at most ξ higher than the expected payoff to player 1 in M , which is at most v . Furthermore by using a pure simple strategy \hat{y}' such that $\hat{y}' = \arg \max_{i \in I} \min_{j \in J} m_{ij} + \xi$ player 1 can guarantee a reward of $\max_{i \in I} \min_{j \in J} m_{ij} + \xi$. Combining these observations with (5), we find the following lemma.

Lemma 3.2 For each M_ξ , we have

$$\max \left\{ v + \xi \cdot \sum_{s=1}^m (a_s^*)^2, \max_{i \in I} \min_{j \in J} m_{ij} + \xi \right\} \leq v_\xi \leq v + \xi.$$

Notice that, if a^* is pure, then $\sum_{s=1}^m (a_s^*)^2 = 1$ and $\max_{i \in I} \min_{j \in J} m_{ij} + \xi = v$ and hence $v_\xi = v + \xi$ and $a^{*'}$ is optimal in M_ξ . This means that the following theorem holds.

Theorem 3.3 If a^* is a pure optimal action of player 1 in M , then $a^{*'}$ is a simple optimal strategy for player 1 in M_ξ and $v_\xi = v + \xi$.

This theorem solves the case of pure optimal actions in M . In the following sections we only consider games M_ξ for which player 1 has no pure optimal action in M .

4 2 × 2 Games

In this section we consider repeated games with bonuses, in which the size of the underlying matrix game M is 2×2 . Theorem 3.3 shows what happens if player 1 has a pure optimal action in M . Consequently in this section we only have to consider the case, in which player 1 does not have a pure optimal action in M . Since the size of the game is 2×2 , this means that player 1 has a unique completely mixed optimal action in M . This fact can, without loss of generality, be modelled as follows:

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{R}^{2 \times 2}, \tag{7}$$

with $\min\{a, d\} > \max\{b, c\}$. We use the following representation of M_ξ :

| | | | |
|-----------|-----------|-----------|-----------|
| $a + \xi$ | $b + \xi$ | a | b |
| 1 | 1 | 1 | 1 |
| c | d | $c + \xi$ | $d + \xi$ |
| 2 | 2 | 2 | 2 |
| state 1 | | state 2 | |

(8)

In each cell the number in the upper-left corner denotes the payoff to player 1, whereas the number in the lower-right corner denotes the state number that will be visited next when this particular cell is selected. Notice that, with respect to M , in state 1 respectively state 2 the payoffs in row 1 respectively row 2 are increased with the bonus ξ .

Notation 4.1 In Sect. 4 in M player 1’s unique optimal mixed action is $a^* = (a_1^*, a_2^*)$ with $a_1^* = (d - c)/(a - b + d - c)$ and $a_2^* = (a - b)/(a - b + d - c)$. Furthermore, $v = (ad - bc)/(a - b + d - c)$ and

$$\max\{b, c\} < v < \min\{a, d\}. \tag{9}$$

We have $B(a^*) = \Delta^2$ and, by (3), $B_\xi(a^{*'}) = B(a^*) \times B(a^*) = \Delta^2 \times \Delta^2$. So,

$$\varphi_\xi(a^{*'}) = \gamma_\xi(a^{*'}, y), \quad \text{for all strategies } y \text{ of player 2.} \tag{10}$$

Now, we present a necessary and sufficient condition for $a^{*'}$ to be optimal in M_ξ .

Theorem 4.1 *The following two statements are equivalent:*

- (i) $a^{*'}$ is optimal in M_ξ for player 1.
- (ii) $\varphi_\xi(a^{*'}) \geq \max\{b, c\} + \xi$.

Proof (i) \Rightarrow (ii): If $a^{*'}$ is optimal, then, in M_ξ , it will yield at least as much as $(1, 0)'$ and $(0, 1)'$, which in turn yield at least $b + \xi$ and $c + \xi$ respectively.

(ii) \Rightarrow (i): Let $y := (b^1, b^2)$ with

$$\begin{aligned} b_1^1 &:= (v - 2\xi a_1^* a_2^* - b)/(a - b), & b_2^1 &:= 1 - b_1^1, \\ b_1^2 &:= (d + 2\xi a_1^* a_2^* - v)/(d - c), & b_2^2 &:= 1 - b_1^2. \end{aligned}$$

Given $\varphi_\xi(a^{*'}) \geq \max\{b, c\} + \xi$, it follows easily that $0 \leq b_1^1, b_1^2 \leq 1$.

We show that $\gamma_\xi^s(\pi, y) = \varphi_\xi(a^{*'})$ for all strategies π of player 1 and each initial state s . For this purpose, we first prove that, for any pure stationary strategy x of player 1, it holds that

$$\gamma_\xi^s(x, y) = \varphi_\xi(a^{*'}).$$

Some elementary calculations show that

$$\begin{aligned} (a + \xi)b_1^1 + (b + \xi)(1 - b_1^1) &= (c + \xi)b_1^2 + (d + \xi)(1 - b_1^2) \\ &= v + \xi \cdot ((a_1^*)^2 + (a_2^*)^2) \\ &= \varphi_\xi(a^{*'}) \end{aligned}$$

(cf. (5)). This shows the result for pure stationary strategies $((1, 0), (1, 0)), ((1, 0), (0, 1))$ and $((0, 1), (0, 1))$. Some more calculations also show that

$$1/2 \cdot (cb_1^1 + d(1 - b_1^1)) + 1/2 \cdot (ab_1^2 + b(1 - b_1^2)) = \varphi_\xi(a^{*'}),$$

which proves the result for the remaining pure strategy $((0, 1), (1, 0))$. So all pure stationary strategies yield exactly the same reward against y . Since y is stationary, it is well-known that player 1 has a pure stationary best reply against y (cf. Hordijk, Vrieze and Wanrooij [11]). This holds for a minimizing player 1 as well as for a maximizing player 1 and consequently $\gamma_\xi^s(\pi, y) = \varphi_\xi(a^{*'})$ for all strategies π of player 1 and each initial state s . This means that $a^{*'}$, guaranteeing a reward of at least $\varphi_\xi(a^{*'})$ to player 1, is optimal. □

Theorem 4.2 *Player 1 has a pure simple optimal strategy in M_ξ if and only if $\varphi_\xi(a^{*'}) \leq \max\{b, c\} + \xi$.*

Proof The only if-part of the proof is trivial: $a^{*'}$ guarantees $\varphi_\xi(a^{*'})$ and any pure stationary strategies at most $\max\{b, c\} + \xi$.

For the “if-part”, suppose without loss of generality that $c \geq b$. It can easily be verified that the assumption $\varphi_\xi(a^{*'}) \leq c + \xi$ is equivalent to

$$\xi \geq (v - c)/(2a_1^*a_2^*) = (a - c)/(2a_2^*).$$

Consider the stationary strategy $y = (((c - b)/(a - b), (a - c)/(a - b)), (1, 0))$ for player 2. It can straightforwardly be verified that: (i) the absorbing first action in state 1 for player 1 yields $b + \xi$, (ii) the absorbing second action in state 2 for player 1 yields $c + \xi$, and (iii) the jumping strategy $((0, 1), (1, 0))$ yields

$$1/2[(c(c - b))/(a - b) + d(a - c)/(a - b)] + 1/2 \cdot a = c + (a - c)/(2a_2^*) \leq c + \xi.$$

Hence, the value of M_ξ is at most $c + \xi$. Since the simple strategy $(0, 1)'$ yields at least $c + \xi$ against any strategy of player 2, it follows that $c + \xi$ is the value of the game and that $(0, 1)'$ is optimal for player 1. □

The following corollary follows directly from Theorems 4.1 and 4.2.

Corollary 4.1 *For any $M \in \mathbb{R}^{2 \times 2}$ and $\xi > 0$, player 1 has a simple optimal strategy in M_ξ .*

5 $2 \times n$ Games

Let M be a $2 \times n$ -matrix game with $n \geq 3$. As in the previous section we only consider the case, in which player 1 does not have a pure optimal action in M ; that case has been taken care of in Theorem 3.3.

From the viewpoint of player 2, he only will use columns that are not dominated by any other column. Column j_1 dominates column j_2 whenever $m_{1j_1} < m_{1j_2}$ and $m_{2j_1} < m_{2j_2}$. And without loss of generality we can state that player 2 only needs to use columns that are not weakly dominated, where the inequality signs may be equality signs. So, without loss of generality we may suppose that none of the columns of M are weakly dominated by any other one which means that the payoffs in M can be ordered in the following way:

$$M = \begin{pmatrix} \mathbf{a} & > & \cdots & > & \cdots & > & \mathbf{b} \\ \vee & & & & & & \wedge \\ \mathbf{c} & < & \cdots & < & \cdots & < & \mathbf{d} \end{pmatrix}. \tag{11}$$

If either $\mathbf{a} > \mathbf{c}$ or $\mathbf{d} > \mathbf{b}$ would not be the case, player 1 would have a pure optimal strategy in M , which was excluded.

Let as before $a^* = (a_1^*, a_2^*)$ be an optimal mixed action for player 1 in M . For the simple strategy $a^{*'}$ we have, according to (3), that $B_\xi(a^{*'}) = B(a^*) \times B(a^*)$. Define

$$j_1 := \min\{j; e_j \in B(a^*)\} \quad \text{and} \quad j_2 := \max\{j; e_j \in B(a^*)\}.$$

The next theorem is the generalization of Theorem 4.1.

Theorem 5.1 *The following two statements are equivalent:*

- (i) $a^{*'}$ is optimal in M_ξ for player 1.
- (ii) $\varphi_\xi(a^{*'}) \geq \max\{m_{1j_2}, m_{2j_1}\} + \xi$.

Proof (i) \Rightarrow (ii): If $a^{*'}$ is optimal for player 1 and y is optimal for player 2, then obviously $y \in B_\xi(a^{*'})$ and

$$v_\xi = \varphi_\xi(a^{*'}) = \gamma_\xi(a^{*'}, y) \geq \gamma(x, y),$$

for all stationary strategies x . Suppose that the second statement of the theorem is not true and without loss of generality let $\varphi_\xi(a^{*'}) < m_{2j_1} + \xi$. The definition of j_1 and the representation (11) imply that $m_{2j} \geq m_{2j_1}$ for each j such that $e_j \in B_\xi(a^{*'})$. Hence, for any $y \in B_\xi(a^{*'})$,

$$\gamma_\xi((0, 1)', y) \geq m_{2j_1} + \xi > \varphi_\xi(a^{*'}) = v_\xi.$$

So, y can not be optimal for player 2 and since $y \in B_\xi(a^{*'})$ was arbitrary we have reached a contradiction.

(ii) \Rightarrow (i): This proof is equal to the proof of Theorem 4.1 with a few notational adjustments: a, b, c and ϑ are replaced by $m_{1j_1}, m_{1j_2}, m_{2j_1}$ and m_{2j_2} respectively and player 2's strategy $y = (b^1, b^2)$ that takes care that all pure strategies of player 1 yield exactly $v + \xi((a_1^*)^2 + (a_2^*)^2)$, consists of the following entries:

$$b_{j_1}^1 = (v + \xi a_1^* a_2^* - m_{1j_2}) / (m_{1j_1} - m_{1j_2}), \quad b_{j_2}^1 = 1 - b_{j_1}^1,$$

$$b_{j_1}^2 = (m_{2j_2} + 2\xi a_1^* a_2^* - v) / (m_{2j_2} - m_{2j_1}), \quad b_{j_2}^2 = 1 - b_{j_1}^2,$$

while all other entries are zero. □

In the 2×2 -case we had that failure of the characterizing condition for the existence of a_1^* to be simple optimal, implied the existence of a pure simple optimal strategy. This property is no longer true for the $(2 \times n)$ -case as the following example shows:

Example 5.1 Take $M = \begin{pmatrix} 10 & 1.99 & 0.39 \\ 0 & 0.89 & 1.29 \end{pmatrix}$ and $\xi = 1$. Then, the (unique) optimal mixed action a^* of player 1 in M is $(0.2, 0.8)$. M_1 equals

| | | | |
|---------|------|------|--|
| 11 | 2.99 | 1.39 | |
| 1 | 1 | 1 | |
| 0 | 0.89 | 1.29 | |
| 2 | 2 | 2 | |
| state 1 | | | |
| 10 | 1.99 | 0.39 | |
| 1 | 1 | 1 | |
| 1 | 1.89 | 2.29 | |
| 2 | 2 | 2 | |
| state 2 | | | |

and $\varphi_\xi(a^{*'}) = 1.11 + 0.04 + 0.64 = 1.79$. However, in M_1 for player 1 the unique stationary optimal strategy is to play $x = ((0.2, 0.8), (0.1, 0.9))$. The reward corresponding to this strategy is $\varphi_\xi(x) = 16.51/9 > 1.79$.

However, it is possible to characterize the existence of a pure simple optimal strategy. For this purpose, consider the matrices

$$\tilde{M}_1 := \begin{pmatrix} 1/2(e_1^T M + a1^T) - \xi 1^T \\ e_2^T M \end{pmatrix} \quad \text{and} \quad \tilde{M}_2 := \begin{pmatrix} e_1^T M \\ 1/2(e_2^T M + a1^T) - \xi 1^T \end{pmatrix}.$$

Theorem 5.2 *In M_ξ the simple strategy $(1, 0)'$ is optimal for player 1 if and only if the value of \tilde{M}_1 does not exceed b : $v(\tilde{M}_1) \leq b$. Furthermore the simple strategy $(0, 1)'$ is optimal for player 1 if and only if $v(\tilde{M}_2) \leq c$.*

Proof We prove only the first statement.

“ \Rightarrow ”: Suppose $(1, 0)'$ is optimal and suppose that (b^1, b^2) is an optimal strategy for player 2. Obviously b^1 must be e_n , i.e. all weight on his last column and we see that $(1, 0)'$ assures player 1 an average payoff $b + \xi$. Now $(1, 0)'$ can only be optimal when the other pure strategies of player 1 give, against (b^1, b^2) , an average payoff that is at most $b + \xi$. The average payoff of $((0, 1), (1, 0))$ against (b^1, b^2) is $1/2(e_1^T M b^2 + \vartheta)$, so $1/2(e_1^T M b^2 + \vartheta) \leq b + \xi$ or $1/2(e_1^T M b^2 + \vartheta) - \xi \leq b$. The

average payoff of $(0, 1)'$ against (b^1, b^2) is $e_2^T M b^2 + \xi$, so $e_2^T M b^2 + \xi \leq b + \xi$ or $e_2^T M b^2 \leq b$. Hence we see that the action b^2 of player 2 guarantees player 1 a payoff at most b in \tilde{M}_1 , which shows the “only if” part.

“ \Leftarrow ”: Let $v(\tilde{M}_1) \leq b$, let b^2 be optimal for player 2 in \tilde{M}_1 and consider (e_n, b^2) for player 2 in M_ξ . One can easily check that against any pure strategy of player 1 the average payoff is at most $b + \xi$, while $(1, 0)'$ gives exactly $b + \xi$. Hence $v_\xi = b + \xi$ and $(1, 0)'$ is optimal. \square

6 $m \times n$ Games

In this section, we look at games in which player 1 has more than 2 actions. These games are more complex, since the number of states is now $m > 2$ and player 1 has m actions in each of these states. Examples can be constructed that even in the 3×2 -case player 1 does not necessarily have a simple optimal strategy (see Schoenmakers [3]).

As in the preceding sections, a^* will denote an optimal action in M and $a^{*'}$ the associated simple strategy. For a number of cases we will provide necessary and sufficient conditions for $a^{*'}$ to be optimal.

First, we consider a square matrix game M , for which we assume that both player 1 and player 2 have a unique optimal completely mixed action and without loss of generality we assume that the value of M is unequal to zero. We will give a characterization for the associated strategy $a^{*'}$ of player 1 to be simple optimal. For that purpose, define for each $s \in S$ an action b^s for player 2,

$$b^s := b + \xi M^{-1}[(a^{*T} a^* + a_s^*)\tilde{1} - a^* - e_s], \tag{12}$$

where $b = (M^{-1}\tilde{1})/(\tilde{1}^T M^{-1}\tilde{1})$, e_s is the s th unit vector and $\tilde{1} = (1, \dots, 1) \in \mathbb{R}^m$.

Observe by the Shapley-Snow theorem for matrix games [12] that, as a consequence of the uniqueness of optimal strategies, M is the only kernel and that the optimal actions can be written as $a^{*T} = (\tilde{1}^T M^{-1})/(\tilde{1}^T M^{-1}\tilde{1})$ respectively $b = (M^{-1}\tilde{1})/(\tilde{1}^T M^{-1}\tilde{1})$ while $v = 1/(\tilde{1}^T M^{-1}\tilde{1})$.

Theorem 6.1 *Let M be a square matrix with value unequal to zero for which both players have a unique optimal completely mixed action. Then, $a^{*'}$ is a simple optimal strategy in M_ξ if and only if b^s as defined in (12) is a mixed action for each $s \in S$. Moreover, whenever b^s is a mixed action, $\forall s \in S$, then (b^1, \dots, b^n) is a stationary optimal strategy for player 2.*

Proof First, observe that the entries of each b^s sum up to 1,

$$\begin{aligned} \tilde{1}^T b^s &= \tilde{1}^T b + \xi[(a^{*T} a^* + a_s^*)\tilde{1}^T M^{-1}\tilde{1} - \tilde{1}^T M^{-1} a^* - \tilde{1}^T M^{-1} e_s] \\ &= 1 + \xi[(a^{*T} a^* + a_s^*) - (\tilde{1}^T M^{-1})/(\tilde{1}^T M^{-1}\tilde{1}) \cdot (a^* - e_s)]\tilde{1}^T M^{-1}\tilde{1} = 1. \end{aligned}$$

Hence, the condition in the theorem states that b^s is nonnegative.

“ \Leftarrow ”: Suppose that player 2 plays (b^1, \dots, b^m) as defined in (12). Since $M b^s = M b$, the immediate expected payoffs in state s of M_ξ are

$$v + \xi(a^{*T} a^*), \quad \text{for row } s, \tag{13}$$

$$v + \xi(a^{*T} a^*) + \xi(a_s^* - a_{s'}^*), \quad \text{for row } s' \neq s. \tag{14}$$

If a pure stationary strategy for player 1 is absorbing, obviously the absorbing payoff is $v + \xi(a^{*T} a^*)$. If a pure stationary strategy is nonabsorbing, then the recurrent part must be cyclic and along one cycle the average is $v + \xi(a^{*T} a^*)$, since the terms $\xi(a_s^* - a_{s'}^*)$ cancel out along a cycle. Hence, player 1 can not get more than $v + \xi(a^{*T} a^*)$ and since $a^{s'}$ guarantees this amount it must be optimal. Obviously (b^1, \dots, b^m) is optimal in this case.

“ \Rightarrow ”: When $a^{s'}$ is supposed to be simple optimal and since a^* uses all components with positive probability it must be that any pure stationary strategy of player 1 is a best reply against optimal stationary strategies of player 2 and give exactly $v + \xi(a^{*T} a^*)$ as average reward. So, let (b^1, b^2, \dots, b^m) be optimal for player 2 in M_ξ . From the mentioned requirement it follows that the immediate expected payoffs in state s , when player 2 uses (b^1, b^2, \dots, b^m) in M_ξ must be

$$v + \xi(a^{*T} a^*), \quad \text{for row } s, \tag{15}$$

$$v + \xi(a^{*T} a^*) + \lambda_{s'}^s, \quad \text{for row } s' \neq s, \tag{16}$$

for some $\lambda_{s'}^s \in \mathbb{R}$ and set $\lambda_s^s = 0$. Moreover it is required that, for all $k \in \{1, \dots, m\}$ and any set $\{s_1, s_2, s_3, \dots, s_k\} \subset \{1, 2, \dots, m\}$, it holds that

$$\sum_{i=1}^{k-1} \lambda_{s_{i+1}}^{s_i} + \lambda_{s_1}^{s_k} = 0. \tag{17}$$

Compare the similarity of (15), (16) with (13), (14).

Hence, in vector notation, with $\lambda^s = (\lambda_1^s, \lambda_2^s, \dots, \lambda_m^s)$ (recall that $\lambda_s^s = 0$), we have that $Mb^s = v\tilde{1} + \xi[(a^{*T} a^*)\tilde{1} - e_s] + \lambda^s$ or

$$b^s = vM^{-1}\tilde{1} + \xi M^{-1}[(a^{*T} a^*)\tilde{1} - e_s] + \lambda^s. \tag{18}$$

Since b^s is a mixed action the sum of its components must be one,

$$1 = \tilde{1}^T b^s = v\tilde{1}^T M^{-1}\tilde{1} + \xi(a^{*T} a^*)\tilde{1}^T M^{-1}\tilde{1} - \xi\tilde{1}^T M^{-1}e_s + \tilde{1}^T M^{-1}\lambda^s.$$

Since $v\tilde{1}^T M^{-1}\tilde{1} = 1$, it follows, after division by $\tilde{1}^T M^{-1}\tilde{1}$ that

$$\xi(a^{*T} a^*) - \xi a_s^* + a^{*T} \lambda^s = 0, \quad \forall s \in S.$$

Summarizing, we have the following system of equations in the m^2 variables $(\lambda_1^s, \lambda_2^s, \dots, \lambda_m^s)$, $s \in \{1, 2, \dots, m\}$:

$$\begin{aligned} \xi(a^{*T} a^*) - \xi a_s^* + a^{*T} \lambda^s &= 0, \quad s \in \{1, 2, \dots, m\}, \\ \sum_{i=1}^{k-1} \lambda_{s_{i+1}}^{s_i} + \lambda_{s_1}^{s_k} &= 0, \quad \{s_1, s_2, \dots, s_k\} \subset \{1, 2, \dots, m\}. \end{aligned} \tag{19}$$

If we substitute $\lambda^s = (a_s^* \tilde{1} - a^* + x^s)\xi$, we derive

$$a^{*T} x^s = 0, \quad s \in \{1, 2, \dots, m\}, \tag{20}$$

$$\sum_{i=1}^{k-1} x_{s_{i+1}}^{s_i} + x_{s_1}^{s_k} = 0, \quad \{s_1, s_2, \dots, s_k\} \subset \{1, 2, \dots, m\}. \tag{21}$$

At first glance, the system of (20) and (21) appears to be overdetermined since we have $m + 2^m - 1$ equations for m^2 variables. However, (23) are completely determined by the 1-term, 2-term and certain 3-term equations in the following way:

$$x_s^s = 0, \quad s \in \{1, 2, \dots, m\}, \tag{22}$$

$$x_{s_1}^{s_2} = -x_{s_2}^{s_1}, \quad s_1 \neq s_2, \tag{23}$$

$$x_{s_1}^1 + x_{s_2}^{s_1} + x_1^{s_2} = 0, \quad 1 \neq s_1 < s_2 \neq 1. \tag{24}$$

It can easily be checked that (22), (23) and (24) are independent and imply all the equations (21). In total (22), (23) and (24) are $m + m(m - 1)/2 + (m - 1)(m - 2)/2 = m^2 - m + 1$ equations. Further, (20) delivers m equations, so there is still one equation too much. If we multiply the equation $a^{*T} x^s = 0$ by $a_s^* a_m^*$ for $s = 1, 2, \dots, m - 1$ and invoke (22) and (23), it can be checked that we derive the last equation $a^{*T} x^m = 0$, hence this last equation is linearly dependent of the other ones and can therefore be deleted.

Hence we have a system of m^2 linearly independent equations (namely (20) for $s \in \{1, \dots, m - 1\}$, (22), (23) and (24)) in the m^2 variables (x_1^s, \dots, x_m^s) , $s \in \{1, \dots, m\}$ for which all the right-hand sides are zero. Thus, the unique solution is $x_{s'}^s = 0$ for all $s, s' \in \{1, 2, \dots, m\}$ and so $\lambda^s = a_s^* \tilde{1} - a^*$ is the unique solution that satisfies (19). Hence b^s is unique and by (18) equals expression (12). □

The following corollary is an easy consequence of Theorem 6.1, as b^s is a mixed action for ξ small enough, whenever b is completely mixed.

Corollary 6.1 *If the square matrix M is such that $v \neq 0$ and if both players have a unique optimal completely mixed action in M , then there exists a $\tilde{\xi} > 0$ such that a^{*l} is a simple optimal strategy in the repeated game with bonus ξ for all $\xi \in [0, \tilde{\xi}]$. Moreover if a^{*l} happens to be a simple optimal strategy for a certain $\tilde{\xi}$, then it is optimal for all $\xi \in [0, \tilde{\xi}]$.*

An interesting question concerns the number of simple optimal strategies. It turns out that there can only be a finite number of different simple optimal strategies.

Theorem 6.2 *For the repeated game M_ξ , with $\xi > 0$, if a^{1l} and a^{2l} are both simple optimal strategies with $\text{car}(a^{1l}) = \text{car}(a^{2l})$, then $a^{1l} = a^{2l}$.*

Proof Without loss of generality we assume $\text{car}(a^{1l}) = \{1, 2, \dots, m\}$, otherwise we can take an appropriate subgame. Since all states are recurrent with respect to both

$a^{1'}$ and $a^{2'}$ it can easily be seen that $(\lambda a^1 + (1 - \lambda)a^2)'$ is optimal as well for all $\lambda \in [0, 1]$. Let (b^1, b^2, \dots, b^m) be an optimal strategy for player 2. Then, this strategy applied to respectively $a^{1'}$, $a^{2'}$ and $(\lambda a^1 + (1 - \lambda)a^2)'$ yields

$$\begin{aligned} v_\xi &= \varphi_o(a^1) + \xi a^{1T} a^1 = \varphi_o(a^2) + \xi a^{2T} a^2 \\ &= \varphi_o(\lambda a^1 + (1 - \lambda)a^2) + \xi(\lambda a^1 + (1 - \lambda)a^2)^T (\lambda a^1 + (1 - \lambda)a^2). \end{aligned}$$

On the other hand, by a straightforward calculation,

$$\begin{aligned} &\varphi_o(\lambda a^1 + (1 - \lambda)a^2) + \xi(\lambda a^1 + (1 - \lambda)a^2)^T (\lambda a^1 + (1 - \lambda)a^2) \\ &= \lambda(\varphi_o(a^1) + \xi(a^{1T} a^1)) + (1 - \lambda)(\varphi_o(a^2) + \xi(a^{2T} a^2)) \\ &\quad - \xi(1 - \lambda)(a^{1T} - a^{2T})(a^1 - a^2) \\ &= v_\xi - \xi\lambda(1 - \lambda)(a^{1T} - a^{2T})(a^1 - a^2). \end{aligned}$$

Since $(a^1 - a^2)^T (a^1 - a^2)$ is strictly positive unless $a^1 = a^2$, the result follows. \square

Combination of Corollary 6.1 and Theorem 6.2 shows

Corollary 6.2 *If M is such that $v \neq 0$ and both players have a unique optimal completely mixed action in M , then $a^{s'}$ is the only candidate for a completely mixed simple optimal strategy for the repeated game with bonus ξ for all ξ .*

We conclude this section with a theorem that shows that for large ξ player 1 always has a simple optimal strategy in the repeated game that moreover is even pure.

Theorem 6.3 *For each matrix game M , there exists $\tilde{\xi} \geq 0$ such that both players have a pure simple optimal strategy for the repeated game with bonus ξ for all $\xi \geq \tilde{\xi}$.*

Proof Let j^s be the column such that $m_{sj^s} = \min_j \{m_{sj}\}$ and let s^* be such that $m_{s^*j^s} = \max_s m_{sj^s}$. Take $\tilde{\xi}$ such that $m_{sj^s} + \tilde{\xi} \geq m_{ij^s}, \forall i \neq s, \forall s$. Now if player 2 plays the pure strategy (j^1, j^2, \dots, j^m) , then all possible immediate payoffs, for any $\xi \geq \tilde{\xi}$, are smaller than or equal to $m_{s^*j^s} + \xi$ which implies that player 1 can never get more against this strategy. Hence $v_\xi \leq m_{s^*j^s} + \xi$. On the other hand, the pure simple strategy $s^{*'}$ guarantees player 1 a payoff of at least $m_{s^*j^s} + \xi$. Thus $v_\xi = m_{s^*j^s} + \xi$ and $s^{*'}$ is optimal. \square

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Arrow, K.J.: The economic implications of learning by doing. Rev. Econ. Stud. **29**, 155–173 (1962)
2. Joosten, R., Peters, H., Thuijsman, F.: Unlearning by not doing: Repeated games with vanishing actions. Games Econ. Behav. **9**, 1–7 (1995)

3. Schoenmakers, G.: The profit of skills in repeated and stochastic games. PhD Thesis, Maastricht University, Maastricht, Netherlands (2004)
4. Shapley, L.S.: Stochastic games. Proc. Nat. Acad. Sci. USA **39**, 1095–1100 (1953)
5. Thuijjsman, F.: Optimality and Equilibria in Stochastic Games, vol. 82. Centre for Mathematics and Computer Science, Amsterdam (1992)
6. Filar, J.A., Vrieze, O.J.: Competitive Markov Decision Processes. Springer, New York (1997)
7. Gillette, D.: Stochastic games with zero stop probabilities, contributions to the theory of games III. In: Dresher, M., Tucker, A.W., Wolfe, P. (eds.) Annals of Mathematical Studies, vol. 39, pp. 179–187. Princeton University Press, Princeton (1957)
8. von Neumann, J.: Zur Theorie der Gesellschaftsspiele. Math. Ann. **100**, 295–320 (1928)
9. Mertens, J.F., Neyman, A.: Stochastic games. Int. J. Game Theory **10**, 53–66 (1981)
10. Filar, J.A.: Ordered field property for stochastic games when the player who controls transitions changes from state to state. J. Optim. Theory Appl. **34**, 503–515 (1981)
11. Hordijk, A., Vrieze, O.J., Wanrooij, G.L.: Semi-Markov strategies in stochastic games. Int. J. Game Theory **12**, 81–89 (1983)
12. Shapley, L.S., Snow, R.N.: Basic solutions of discrete games. Ann. Math. Stud. **24**, 27–35 (1950)