

Loss of skills in coordination games

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

Accepted: 1 December 2010 / Published online: 4 February 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract This paper deals with 2-player coordination games with vanishing actions, which are repeated games where all diagonal payoffs are strictly positive and all non-diagonal payoffs are zero with the following additional property: At any stage beyond r , if a player has not played a certain action for the last r stages, then he unlearns this action and it disappears from his action set. Such a game is called an r -restricted game. To evaluate the stream of payoffs we use the average reward. For $r = 1$ the game strategically reduces to a one-shot game and for $r \geq 3$ in Schoenmakers (Int Game Theory Rev 4:119–126, 2002) it is shown that all payoffs in the convex hull of the diagonal payoffs are equilibrium rewards. In this paper for the case $r = 2$ we provide a characterization of the set of equilibrium rewards for 2×2 games of this type and a technique to find the equilibrium rewards in $m \times m$ games. We also discuss subgame perfection.

Keywords Game theory · Repeated games · Coordination games · Nash equilibrium · Subgame perfect equilibrium

1 Introduction

A 2-player coordination game is a repeated game where both players have the same number of actions, where all diagonal payoffs are strictly positive and where all

J. Flesch
Department of Quantitative Economics, Maastricht University, PO Box 616,
Maastricht, 6200 MD, The Netherlands

G. Schoenmakers (✉) · O. J. Vrieze
Department of Knowledge Engineering, Maastricht University, PO Box 616,
Maastricht, 6200 MD, The Netherlands
e-mail: gm.schoenmakers@maastrichtuniversity.nl

non-diagonal payoffs are 0. In this paper we use the idea of vanishing actions introduced by Joosten et al. (1995) for zero-sum repeated games. The idea behind vanishing actions is forgetfulness or unlearning of those actions that have not been used for a specified number of stages, as a counterpart of learning by doing, as introduced by Arrow (1962). One can think of a surgeon who unlearns certain surgical skills when not exercised at some, more or less, regular basis. More precisely, a number r is specified which is to be interpreted as follows: at any stage $t > r$, an action i that has not yet vanished, will vanish from a player's action set, when this player has not played i at any of the previous r stages. Actions that have vanished are no longer available in the remainder of the play. Thus, at any stage of play, players have to balance their interest between achieving a high payoff and keeping the right actions available. Thus we get a so-called r -restricted coordination game.

For $r = 1$ the game strategically reduces to a one-shot game and for $r \geq 3$ in Schoenmakers et al. (2002) it is shown that all payoffs in the convex hull of the diagonal payoffs are equilibrium rewards. In this paper we examine the case $r = 2$ and we provide a characterization of the set of equilibrium rewards for 2×2 games of this type and a technique to find the equilibrium rewards in $m \times m$ games. We also discuss subgame perfection.

An r -restricted coordination game can also be interpreted as a stochastic game with finite state and action spaces, introduced by Shapley (1953). In general n -player stochastic games, the existence of approximate-equilibria is still unclear, but for certain specially structured classes, e.g., zero-sum games (Mertens and Neyman 1981) and general-sum 2-player games (Vieille 2000a,b), they have been shown to exist.

In Sect. 2 we present the model in more detail. In Sect. 3 we investigate 2-restricted 2×2 coordination games by means of two examples, characterizing the set of equilibrium rewards. In Sect. 4 we deal with subgame perfection in 2-restricted 2×2 games. In Sect. 5 we show how to find equilibria and subgame perfect equilibria in games where the players have more than 2 actions. Finally, we discuss possible extensions of our results.

2 The model

A 2-player r -restricted coordination game is determined by the following parameters:

1. $K = \{1, 2\}$ is the set of players;
2. $I^k = \{1, 2, \dots, m\}$ is the initial set of pure actions for each player $k \in K$, with $m \geq 2$;
3. $R^k : I^1 \times I^2 \rightarrow \mathbb{R}$ is the payoff function for each player $k \in K$; we assume that $R^k(i^1, i^2) > 0$ if $i^1 = i^2$ and $R^k(i^1, i^2) = 0$ otherwise;
4. $r \in \mathbb{N}$ is the level of unlearning for the players.

The game is played at stages $1, 2, \dots$, where each time simultaneously and independently players 1 and 2 each have to choose one of their available actions i^1 respectively i^2 . The action choices are announced and each player k receives a payoff $R^k(i^1, i^2)$. Then play moves to the next stage, where actions have to be chosen again. However, actions may disappear according to the following rule: At any stage $t > r$,

an action that has not vanished so far, vanishes from a player’s action set if it was not played at any of the stages $t - r, t - r + 1, \dots, t - 1$. Thus the number of available actions may decrease during the course of play.

The history of play at stage t is the sequence of action pairs chosen at stages $1, \dots, t - 1$. Let H be the set of all possible histories. For any history $h \in H$, we denote by $I^k(h)$ the set of available or not (yet) unlearned pure actions of player k after history h . A probability distribution on $I^k(h)$ is called a mixed action for player k after history $h \in H$. A strategy for player k is a mapping which assigns a mixed action on $I^k(h)$ to every history h . We denote a strategy by π for player 1 and by σ for player 2. Each player k is assumed to maximize his limiting average reward, i.e.,

$$\gamma^k(\pi, \sigma) = \liminf_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E_{\pi\sigma}(R_t^k),$$

where $E_{\pi\sigma}(R_t^k)$ denotes the expected payoff to player k at stage t given that (π, σ) is being played. A strategy pair (π, σ) is called an *equilibrium* if $\gamma^1(\pi, \sigma) \geq \gamma^1(\pi', \sigma)$ for all π' and $\gamma^2(\pi, \sigma) \geq \gamma^2(\pi, \sigma')$ for all σ' , i.e., each player is playing a best reply against the strategy of the other player.

Given a history h occurs, we obtain a subgame starting after h . In this subgame, we can define strategies and equilibria in a similar fashion. We will call them follow-up strategies and follow-up equilibria after history h . Every strategy induces a follow-up strategy in all subgames, and every equilibrium induces a follow-up equilibrium in those subgames which occur with a positive probability with respect to this equilibrium. A strategy pair (π, σ) in the original game is called a *subgame perfect equilibrium* if it induces a follow-up equilibrium in every subgame.

3 Equilibrium rewards in 2×2 coordination games

In this section we analyze 2-restricted 2×2 coordination games. A 2×2 coordination game can be described as follows:

$$\begin{array}{cc}
 & \begin{array}{cc} L & R \end{array} \\
 \begin{array}{c} T \\ B \end{array} & \left(\begin{array}{cc} a_1, b_1 & 0, 0 \\ 0, 0 & a_2, b_2 \end{array} \right),
 \end{array}$$

where $a_1, a_2, b_1, b_2 > 0$. Here T stands for top, B for bottom, L for left and R for right. Analogously the entries of such a game will be called TL, TR, BL and BR . Due to symmetry, it suffices to distinguish between the following 2 types of games. For type 1 we have $a_1 \geq a_2$ and $b_1 \leq b_2$, whereas for type 2 we have $a_1 > a_2$ and $b_1 > b_2$. In the following subsections we characterize the set of equilibrium rewards for both types.

3.1 Games of type 1

In this section we characterize the set of equilibrium rewards of 2-restricted 2×2 games of type 1. We do so by thoroughly investigating the following type 1 game \mathcal{G} :

$$\mathcal{G} = \begin{matrix} & \begin{matrix} L & R \end{matrix} \\ \begin{matrix} T \\ B \end{matrix} & \begin{pmatrix} 2, 1 & 0, 0 \\ 0, 0 & 1, 3 \end{pmatrix} \end{matrix}$$

We will prove the following statement:

Theorem 1 *The set of equilibrium rewards in \mathcal{G} equals the set of equilibrium rewards of the following one-shot games:*

$$\begin{pmatrix} V_1 & W_1 \\ W_2 & V_2 \end{pmatrix},$$

where $V_1, V_2 \in \{(2, 1), (\frac{3}{2}, 2), (1, 3)\}$ and $W_1, W_2 \in \mathbb{U}$. Here \mathbb{U} is the union of the following 3 sets:

$$\begin{aligned} &\{(a, 1) \mid 1 \leq a \leq 2\}, \\ &\{(1, b) \mid 1 \leq b \leq 3\} \end{aligned}$$

and

$$\left\{ \left(1 - \frac{1}{2^{q+1} - 1}, 1 - \frac{2}{3^{q+1} - 1} \right) \mid q \in \{0, 1, 2, \dots\} \right\}.$$

The set of equilibrium rewards in \mathcal{G} is depicted in Fig. 1.

We introduce four auxiliary bimatrices which will help us analyse the set of possible equilibrium rewards in \mathcal{G} . Suppose that the play is at stage $t \geq 2$, and that all four actions are still alive. Then, either the two diagonal or the two off-diagonal entries have been chosen in turn before stage t . Let $(\hat{\pi}, \hat{\sigma})$ denote a follow-up equilibrium from stage t on. Given that the entry chosen at stage $t - 1$ was TL (i.e., entries TL and BR have been chosen in turn before stage t), we now describe the follow-up rewards from stage $t + 1$ depending on which entry is chosen at stage t .

If entry TL is selected at stage t as well: In this case the players unlearn actions B and R , and entry TL will be played at every further stage, leading to reward $(2, 1)$.

If entry TR is selected at stage t : In this case player 1 unlearns action B . We distinguish two subcases. If entry TR had a positive probability according to $(\hat{\pi}, \hat{\sigma})$, then the pair of follow-up strategies from stage $t + 1$ must form a follow-up equilibrium, and therefore player 2 will play action L with long-term frequency 1, leading to reward $(2, 1)$. On the other hand, if entry TR had probability zero according to $(\hat{\pi}, \hat{\sigma})$ and was only played due to a deviation, then the follow-up reward from stage $t + 1$ is either $(2, 1)$, $(0, 0)$, or $(1, \frac{1}{2})$; the last reward arises when player 2 plays L and R from stage $t + 1$ on alternately.

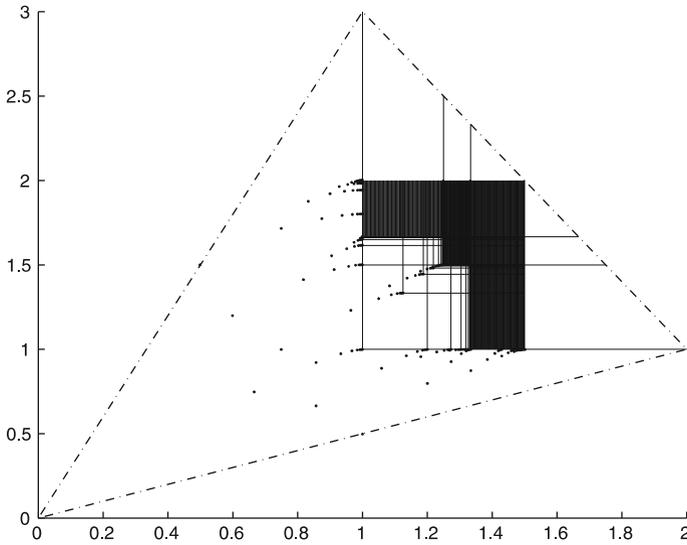


Fig. 1 Equilibrium rewards in \mathcal{G}

If entry BL is selected at stage t : In this case player 2 unlearns action R , and we find similarly to the previous case that if entry BL had a positive probability according to $(\hat{\pi}, \hat{\sigma})$ then the follow-up reward from stage $t + 1$ is $(2, 1)$, and that otherwise it is either $(2, 1)$, $(0, 0)$, or $(1, \frac{1}{2})$.

Finally, if entry BR is selected at stage t : In this case all actions remain alive, with some follow-up reward $\gamma_{BR}^t = (\gamma_{BR}^{t,1}, \gamma_{BR}^{t,2})$.

This situation for stage t is summarized by the following auxiliary bimatrix:

$$\mathcal{G}_{TL}^t = \begin{pmatrix} 2, 1 & 2, 1^\diamond \\ 2, 1^\diamond & \gamma_{BR}^t \end{pmatrix}.$$

Here, the diamonds in entries TR and BL indicate that the corresponding payoffs are the follow-up rewards from stage $t + 1$ only if at stage t these entries occur with a positive probability according to $(\hat{\pi}, \hat{\sigma})$.

In a similar way, when entry BR was played at stage $t - 1$, the situation for stage t is summarized by the auxiliary bimatrix

$$\mathcal{G}_{BR}^t = \begin{pmatrix} \gamma_{TL}^t & 1, 3^\diamond \\ 1, 3^\diamond & 1, 3 \end{pmatrix}.$$

The other two auxiliary bimatrices corresponding to entries TR and BL are

$$\mathcal{G}_{TR}^t = \begin{pmatrix} 2, 1^\diamond & 0, 0 \\ \gamma_{BL}^t & 1, 3^\diamond \end{pmatrix} \quad \mathcal{G}_{BL}^t = \begin{pmatrix} 2, 1^\diamond & \gamma_{TR}^t \\ 0, 0 & 1, 3^\diamond \end{pmatrix}.$$

Note that these four auxiliary bimatrices above have in common that an entry contains a diamond if and only if after playing this entry, precisely one of the players loses an action. Moreover, when an entry with a diamond has probability zero according to the follow-up equilibrium $(\hat{\pi}, \hat{\sigma})$ and nevertheless it is played due to a deviation, then $(0, 0)$ is a possible follow-up reward.

The following lemma regarding the auxiliary bimatrices \mathcal{G}_{TR}^t and \mathcal{G}_{BL}^t is immediate.

Lemma 2 *Let $(\hat{\pi}, \hat{\sigma})$ be a follow-up equilibrium from stage $t \geq 2$ after at stages $1, 2, \dots, t - 1$ the entries TR and BL are selected alternately. If entry BL is the last one at stage $t - 1$, then based on \mathcal{G}_{BL}^t , the pair $(\hat{\pi}, \hat{\sigma})$ prescribes for stage t one of the following:*

- *The players play a completely mixed equilibrium in \mathcal{G}_{BL}^t , i.e., an equilibrium in which all actions are played with a positive probability. In this case, both players receive a reward strictly less than 1.*
- *Player 1 plays T with probability 1 and player 2 uses both L and R with a positive probability, or player 2 plays R with probability 1 and player 1 uses both T and B with a positive probability. In the former case player 2 receives reward 1, whereas in the latter case player 1 receives reward 1.*
- *Both players play a pure action and they play one of the entries TL, TR, BR .*

When entry TR is played last at stage $t - 1$, then based on \mathcal{G}_{TR}^t , the pair $(\hat{\pi}, \hat{\sigma})$ either induces $(2, 1)$ or $(1, 3)$ as a reward, or prescribes for stage t to play entry BL with probability 1 or to play a completely mixed equilibrium in \mathcal{G}_{TR}^t .

We start the analysis of the set of possible equilibrium rewards for the game \mathcal{G} by discussing what happens if at stage 1 one of the entries TL or BR is selected.

Theorem 3 *If at stage 1 one of the entries TL or BR is selected, then the set of possible follow-up equilibrium rewards is $\{(2, 1), (1, 3), (\frac{3}{2}, 2)\}$.*

Proof Suppose that at stage 1 entry TL is selected; the proof is similar for entry BR . Clearly, playing TL at every further stage is a follow-up equilibrium with reward $(2, 1)$, and playing entry BR at every further stage is a follow-up equilibrium with reward $(1, 3)$. Moreover, playing BR and TL alternately, starting with BR at stage 2, also yields a follow-up equilibrium, with reward $\frac{1}{2} \cdot (2, 1) + \frac{1}{2} \cdot (1, 3) = (\frac{3}{2}, 2)$.

Let $(\hat{\pi}, \hat{\sigma})$ be an arbitrary follow-up equilibrium after at stage 1 entry TL is selected. We will now show that $(\hat{\pi}, \hat{\sigma})$ induces one of the three rewards above. Due to the structure of \mathcal{G}_{TL}^1 and \mathcal{G}_{BR}^1 , there are two possibilities for the play induced by $(\hat{\pi}, \hat{\sigma})$, after entry TL is played at stage 1: either entries TL and BR are played alternately during the whole of play, or at the first stage when the play differs from this alternating sequence, the chosen entry leads to reward $(2, 1)$ or to reward $(1, 3)$. Hence, the reward induced by $(\hat{\pi}, \hat{\sigma})$ is a convex combination of $(2, 1)$ and $(1, 3)$.

If this convex combination is neither $(2, 1)$ nor $(1, 3)$, then according to \mathcal{G}_{TL}^2 entry BR is played at stage 2 with probability 1. Indeed, because player 2's reward is more than 1 he will play R with probability 1, and therefore player 1, whose reward is less than 2, cannot use both actions with a positive probability and will play B with probability 1 (in this case, the follow-up reward if player 1 deviates to T is not $(2, 1)$).

We find similarly that at stage 3 entry TL has to be played with probability 1, and continuing this way we obtain that $(\hat{\pi}, \hat{\sigma})$ induces the alternating sequence with TL and BR with probability 1. This however leads to reward $(\frac{3}{2}, 2)$, which completes the proof. \square

The next part of the analysis concerns the case when at stage 1 one of the entries TR or BL is selected. We will make use of the following type of follow-up strategies to retaliate deviations after which the deviator loses an action.

Definition 4 Let h be a history after which player k still has both his actions available whereas his opponent has only one action left. The $(0, 0)$ -threat is the follow-up strategy for player k which, at every future stage, plays the action which yields payoffs $(0, 0)$.

By applying the $(0, 0)$ -threat we obtain exactly the same set of follow-up equilibrium rewards after entry TR as after entry BL .

Lemma 5 Every follow-up equilibrium reward after at stage 1 entry TR is selected is also a follow-up equilibrium reward after at stage 1 entry BL is selected, and vice versa.

Proof Let $(\hat{\pi}_{TR}, \hat{\sigma}_{TR})$ be a follow-up equilibrium after at stage 1 entry TR is chosen. Now consider the pair of follow-up strategies $(\hat{\pi}_{BL}, \hat{\sigma}_{BL})$, after at stage 1 entry BL is chosen, which prescribes to play as follows: At stage 2 the players play entry TR with probability 1, and afterwards from stage 3 on, they execute $(\hat{\pi}_{TR}, \hat{\sigma}_{TR})$ by ignoring stage 1 and thinking of stage 2 as the initial stage. In addition, if a player deviates at stage 2, he unlearns an action and then the other player punishes him by carrying out the $(0, 0)$ -threat. The pair $(\hat{\pi}_{BL}, \hat{\sigma}_{BL})$ is clearly a follow-up equilibrium, and induces the same reward as $(\hat{\pi}_{TR}, \hat{\sigma}_{TR})$. Note that stage 2 is used to shift the game from “the last move is BL ” to “the last move is TR ”. The converse statement can be shown similarly. \square

The next two lemmas examine the possible follow-up equilibrium rewards, after entry TR or BL , in which at least one player receives a reward of at least 1.

Lemma 6 Let (u_1, u_2) be a follow-up equilibrium reward after at stage 1 entry TR or entry BL is chosen. Then, (u_1, u_2) cannot satisfy any of the following:

1. $u_1 > 1$ and $u_2 > 1$,
2. $u_1 < 1$ and $u_2 \geq 1$,
3. $u_1 \geq 1$ and $u_2 < 1$.

Proof We only prove statement 1; the other statements can be proved in a similar fashion. By lemma 5 we may assume that at stage 1 entry BL was chosen. Suppose, by way of contradiction, that a follow-up equilibrium $(\hat{\pi}, \hat{\sigma})$ from stage 2, after entry BL at stage 1, induces reward (u_1, u_2) with $u_1 > 1$ and $u_2 > 1$. To indicate that (u_1, u_2) arises after entry BL at stage 1, it will be convenient to use the notation $\gamma_{BL}^1 = (u_1, u_2)$. It follows by lemma 2 based on \mathcal{G}_{BL}^2 that $(\hat{\pi}, \hat{\sigma})$ selects entry TR at stage 2 with probability 1 and therefore $\gamma_{TR}^2 = \gamma_{BL}^1$. But then by lemma 2 based on \mathcal{G}_{TR}^3 , there are two possibilities for stage 3:

- (i) $(\hat{\pi}, \hat{\sigma})$ selects entry BL at stage 3 with probability 1 and therefore $\gamma_{BL}^3 = \gamma_{BL}^1$. Now at stage 4 the players face the same situation as at stage 2 and jumping back and forth between TR and BL can not occur ad infinitum, since that would lead to a reward of $(0, 0)$. Hence now or later situation (ii) must occur.
- (ii) $(\hat{\pi}, \hat{\sigma})$ prescribes a completely mixed equilibrium in \mathcal{G}_{TR}^3 for stage 3 with reward γ_{BL}^1 . Some straightforward calculations show that then

$$\gamma_{BL}^{3,1} = 3 - \frac{2}{\gamma_{BL}^{1,1}} > \gamma_{BL}^{1,1} \quad \text{and} \quad \gamma_{BL}^{3,2} = 4 - \frac{3}{\gamma_{BL}^{1,2}} > \gamma_{BL}^{1,2}.$$

Continuing this way it follows that γ_{BL}^t converges to $(2, 3)$ when t tends to infinity over the set of odd numbers, which is obviously impossible as there is no feasible average reward close to $(2, 3)$. □

Lemma 7 *Every pair $(u_1, 1)$, for $u_1 \in [1, 2]$, and every pair $(1, u_2)$, for $u_2 \in [1, 3]$, are follow-up equilibrium rewards after at stage 1 one of the entries TR or BL is selected. Moreover, these are the only follow-up rewards in which at least one player receives a reward of at least 1.*

Proof By lemma 5 we may assume that at stage 1 entry BL is chosen. We prove that $(u_1, 1)$, for $u_1 \in [1, 2]$, is a follow-up equilibrium reward after at stage 1 entry BL is selected. The proof for $(1, u_2)$ with $u_2 \in [1, 3]$ is similar, and then the final part of the lemma follows from lemma 6.

We first show that $(1, 1)$ is a follow-up equilibrium reward from stage 3 after at stage 1 entry BL is chosen and at stage 2 entry TR is chosen. We define a pair of follow-up strategies $(\hat{\pi}, \hat{\sigma})$ from stage 3 as follows: At stage $t \geq 3$, if all four actions are still alive, $\hat{\pi}$ prescribes the mixed action $(\frac{2}{3}, \frac{1}{3})$ and $\hat{\sigma}$ prescribes the mixed action $(\frac{1}{2}, \frac{1}{2})$ when t is odd, whereas they prescribe to play entry TR when t is even. In addition, as soon as an entry is chosen by which a player unlearns an action then $(\hat{\pi}, \hat{\sigma})$ prescribes to repeat this entry at all further stages.

Notice that, at an odd stage $t \geq 3$ when all four actions are still alive, the pair of prescribed mixed actions $((\frac{2}{3}, \frac{1}{3}), (\frac{1}{2}, \frac{1}{2}))$ forms an equilibrium in the game \mathcal{G}_{TR}^t if $\gamma_{BL}^t = (1, 1)$, with reward $(1, 1)$. From this observation it follows easily that $(\hat{\pi}, \hat{\sigma})$ is a follow-up equilibrium from stage 3, with reward $(1, 1)$.

Now consider a $u_1 \in [1, 2]$. After entry BL at stage 1, let player 1 play T and player 2 play the mixed action $(u_1 - 1, 2 - u_1)$ at stage 2, and if by doing so entry TL is chosen at stage 2 then repeat entry TL at all further stages, whereas if entry TR is chosen at stage 2 then proceed with the above follow-up equilibrium $(\hat{\pi}, \hat{\sigma})$ from stage 3 giving follow-up reward $(1, 1)$. In addition, if player 1 deviates to action B at stage 2 then the entry chosen at stage 2 should be repeated at all further stages. This is obviously a follow-up equilibrium from stage 2 with reward $(u_1 - 1) \cdot (2, 1) + (2 - u_1) \cdot (1, 1) = (u_1, 1)$. This completes the proof. □

The next three lemmas examine the possible follow-up rewards, after at stage 1 one of the entries TR or BL is selected, in which both players receive strictly less than 1.

Lemma 8 *$(0, 0)$ is a follow-up equilibrium reward from stage $t \geq 2$ after at stages $1, 2, \dots, t - 1$ the entries TR and BL are selected alternately.*

Proof Suppose that at stage $t - 1$ entry TR is chosen; the proof is similar for entry BL . Let $(\hat{\pi}, \hat{\sigma})$ be the following pair of follow-up strategies: From stage t on $(\hat{\pi}, \hat{\sigma})$ prescribes to play the action pairs BL and TR in turn, starting with BL at stage t , as long as neither player deviates. If a player deviates at some stage, he immediately unlearns an action and then the other player punishes him by carrying out the $(0, 0)$ -threat. This is a follow-up equilibrium, and the corresponding reward is $(0, 0)$. \square

Lemma 9 *Let $(\hat{\pi}, \hat{\sigma})$ be a follow-up equilibrium after at stage 1 one of the entries TR or BL is selected, with reward $\gamma(\hat{\pi}, \hat{\sigma}) < (1, 1)$. Let $t \geq 2$ and let h^t denote the alternating sequence with TR and BL of length $t - 1$, starting with the one which was selected at stage 1. Suppose that, with respect to $(\hat{\pi}, \hat{\sigma})$, the history at stage t is h^t with a positive probability. Then, given that this history h^t occurs, we necessarily have that:*

- (i) *The follow-up equilibrium reward from stage t is strictly smaller than $(1, 1)$.*
- (ii) *At stage t either both players randomize or neither player randomizes.*
- (iii) *If from stage t on no more randomization takes place, then the follow-up reward from stage t is $(0, 0)$.*

Proof First we prove part (i). Suppose by way of contradiction that player 1’s follow-up equilibrium reward from stage t is at least 1. Then against $\hat{\sigma}$, starting at stage 2, player 1 can assure a reward of at least 1. Indeed, player 1 should play T and B alternately at stages 2, \dots , $t - 1$, starting with the action that he did not use at stage 1, so that he keeps both actions alive. Subsequently, if player 2 has only one action left at stage t then player 1 can obtain a reward of at least 1 through the entries TL and BR . On the other hand, if player 2 has both actions alive at stage t , which means precisely that history h^t occurs, then player 1 obtains at least 1 according to the follow-up equilibrium from stage t . In either case, player 1 receives a reward of at least 1, which is a contradiction.

Part (ii) follows from part (i) and lemma 2. As for part (iii), if from stage t on no more randomization takes place, then by parts (i) and (ii) and by lemma 2, we obtain an infinite alternating sequence with entries TR and BL , which yields reward $(0, 0)$. \square

Remark 10 Notice that for $(\lambda, \mu) < (1, 1)$ the unique completely mixed equilibrium in the one-shot game

$$G_{BL} = \begin{pmatrix} 2, 1 & \lambda, \mu \\ 0, 0 & 1, 3 \end{pmatrix}$$

provides a reward of $(\frac{2}{3-\lambda}, \frac{3}{4-\mu}) < (1, 1)$ and that this is also the reward that is yielded by the unique completely mixed equilibrium in the one-shot game

$$G_{TR} = \begin{pmatrix} 2, 1 & 0, 0 \\ \lambda, \mu & 1, 3 \end{pmatrix}.$$

We say that a pair of (follow-up) strategies (π, σ) randomizes q times if q is the largest number $m \in \mathbb{N} \cup \{0, \infty\}$ for which there is a history h such that, with respect

to (π, σ) , history h occurs with a positive probability and there are m stages at which both players randomized along h .

Let $(\hat{\pi}_q, \hat{\sigma}_q)$ be a follow-up equilibrium after at stage 1 one of the entries TR or BL is selected, with reward $\gamma(\hat{\pi}, \hat{\sigma}) < (1, 1)$, which randomizes q times (after stage 1). Note that we do not have to consider randomization by only one player at a stage. Indeed, this follows from part (ii) of lemma 9 when no action has vanished yet, and from the uniqueness of the follow-up equilibrium reward when a player unlearns an action.

Assume that q is finite; we will in fact show in the proof of lemma 11 that infinite randomization cannot occur (it would lead to reward at least 1 for both players). In view of remark 10, it makes no difference for the reward induced by $(\hat{\pi}_q, \hat{\sigma}_q)$ when these q randomizations take place. Hence, we can assume without loss of generality that $(\hat{\pi}, \hat{\sigma})$ randomizes at stages $\{2, 3, \dots, q + 1\}$, as long as all four actions are alive (i.e., entries TR and BL have been chosen in turn at the previous stages). So after at stage 1 entry TR or BL is selected, $(\hat{\pi}_q, \hat{\sigma}_q)$ prescribes to play as follows:

1. At stages $t \in \{2, 3, \dots, q + 1\}$, if all actions are alive then play the completely mixed equilibrium in \mathcal{G}_{BL}^t with $\gamma_{TR}^t = u(q + 1 - t)$ or in \mathcal{G}_{TR}^t with $\gamma_{BL}^t = u(q + 1 - t)$, depending on the entry selected at stage $t - 1$, where $u(z) = (1 - \frac{1}{2^{z+1}-1}, 1 - \frac{2}{3^{z+1}-1})$. Note that $q + 1 - t$ is the number of randomizations left after stage t , and that $u(0) = (0, 0)$.
2. If at stage $q + 2$ all actions are alive, then from stage $q + 2$ on, play the follow-up equilibrium as in lemma 8, yielding reward $(0, 0)$ (cf. part (iii) of lemma 9).
3. If at some stage $t \in \{2, 3, \dots, q + 1\}$ a player unlearns an action (an this happens for the first time) then play the action pair that is selected at stage t at all further stages.

Lemma 11 *For each $q \in \{0, 1, 2, \dots\}$ the pair $(\hat{\pi}_q, \hat{\sigma}_q)$ is a follow-up equilibrium after at stage 1 entry TR or entry BL is selected, and it induces reward $u(q) = (1 - \frac{1}{2^{q+1}-1}, 1 - \frac{2}{3^{q+1}-1})$. Moreover, these are the only follow-up rewards that are strictly smaller than $(1, 1)$.*

Proof From stage $q + 2$ on, $(\hat{\pi}_q, \hat{\sigma}_q)$ yields a follow-up equilibrium with reward $u(0) = (0, 0)$. At stage $q + 1$, at the last randomization, the players play the completely mixed equilibrium in \mathcal{G}_{BL}^{q+1} with $\gamma_{TR}^{q+1} = u(0)$ or in \mathcal{G}_{TR}^{q+1} with $\gamma_{BL}^{q+1} = u(0)$, depending on the entry selected at stage q . By remark 10, this induces follow-up reward $u(1) = (\frac{2}{3}, \frac{3}{4})$ from stage $q + 1$ on. Continuing this way we find that $(\hat{\pi}_q, \hat{\sigma}_q)$ is a follow-up equilibrium from stage 2 on with reward $u(q)$.

We will now show that there are no other follow-up rewards that are strictly smaller than $(1, 1)$. To this end, we argue that infinitely many randomization leads to reward at least 1. Suppose that, after at stage 1 entry TR or entry BL is selected, $(\hat{\pi}_\infty, \hat{\sigma}_\infty)$ is a follow-up equilibrium which for every stage $t \geq 2$ prescribes to play the completely mixed equilibrium in \mathcal{G}_{BL}^t or in \mathcal{G}_{TR}^t , as long as all actions are still alive, such that γ_{TR}^t or respectively γ_{BL}^t are precisely equal to the follow-up reward from stage $t + 1$ on. Consider any stage $q \geq 2$. When all actions are alive at stage q , the follow-up reward after stage q is at least $(0, 0)$ and, moreover, there have been $q - 2$ randomizations before stage q . By the first part of the lemma, we obtain $\gamma(\hat{\pi}_\infty, \hat{\sigma}_\infty) \geq u(q - 2)$.

Since $q \geq 2$ was arbitrary, we must have $\gamma(\hat{\pi}_\infty, \hat{\sigma}_\infty) \geq (1, 1)$, which completes the proof. \square

The following Theorem, which follows from the combination of lemmas 7 and 11, summarizes our results for the case when at stage 1 one of the entries TR or BL is selected.

Theorem 12 *If at stage 1 one of the entries TR or BL is selected, then the set of possible follow-up equilibrium rewards is the union of the following 3 sets:*

$$\{(a, 1) \mid 1 \leq a \leq 2\},$$

$$\{(1, b) \mid 1 \leq b \leq 3\}$$

and

$$\left\{ \left(1 - \frac{1}{2^{q+1} - 1}, 1 - \frac{2}{3^{q+1} - 1} \right) \mid q \in \{0, 1, 2, \dots\} \right\}.$$

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

Proof of Theorem 1 Due to Theorems 3 and 12, it is clear that all equilibrium rewards in the one-shot games given in Theorem 1 are equilibrium rewards in \mathcal{G} . It remains to show the converse. To this end, suppose that (π, σ) is an equilibrium in \mathcal{G} . We denote by γ_{TL} the reward that (π, σ) induces after at stage 1 entry TL is chosen. Note that if entry TL is chosen at stage 1 with a positive probability, then γ_{TL} must be a follow-up equilibrium reward, and Theorem 3 is applicable. A similar notation and remark hold for the other entries too.

Let (x^1, x^2) denote the pair of mixed actions prescribed by (π, σ) for stage 1. Depending on which entries are played with a positive probability at stage 1, we distinguish different cases, and show in each of them that the reward $\gamma(\pi, \sigma)$ is also an equilibrium reward in one of the one-shot games given in Theorem 1:

- Case 1: x^1 and x^2 use all actions with a positive probability. In this case, (x^1, x^2) is an equilibrium in the one-shot game with $V_1 = \gamma_{TL}, W_1 = \gamma_{TR}, W_2 = \gamma_{BL}, V_2 = \gamma_{BR}$, with reward $\gamma(\pi, \sigma)$.
- Case 2: x^1 uses both actions T and B with a positive probability, whereas x^2 places probability 1 on one action. Suppose that this action is L ; the proof is similar for R . This means that player 1 is indifferent between entries TL and BL . Due to Theorems 3 and 12, either $\gamma(\pi, \sigma) = \gamma_{TL} = \gamma_{BL} = (2, 1)$ or $\gamma(\pi, \sigma) = \gamma_{TL} = \gamma_{BL} = (1, 3)$. Notice that, if $\gamma(\pi, \sigma) = (2, 1)$, then playing T and L at stage 1 is an equilibrium in the one-shot game with $V_1 = W_1 = W_2 = V_2 = (2, 1)$ with reward $(2, 1)$. The same holds for reward $(1, 3)$.
- Case 3: x^2 uses both actions L and R with a positive probability, whereas x^1 places probability 1 on one action. This case is similar to the previous one.

Case 4: x^1 and x^2 both place probability 1 on one action. Suppose first that this yields either entry TL or entry BR . This entry is obviously also an equilibrium in the one-shot game with $V_1 = \gamma_{TL}$, $W_1 = (0, 0)$, $W_2 = (0, 0)$, $V_2 = \gamma_{BR}$, with reward $\gamma(\pi, \sigma)$. Now suppose that the entry prescribed by (x^1, x^2) is BL ; the proof is similar for entry TR . Then γ_{BL} belongs to one of the 3 sets in Theorem 12. If it belongs either to $\{(a, 1) | 1 \leq a \leq 2\}$ or to $\{(1, b) | 1 \leq b \leq 3\}$, then playing B and L at stage 1 is an equilibrium in the one-shot game with $V_1 = (1, 3)$, $W_1 = \gamma_{TR}$, $W_2 = \gamma_{BL}$, $V_2 = (2, 1)$, with reward $\gamma(\pi, \sigma)$. Finally, we prove that γ_{BL} cannot belong to the third set in Theorem 12. Indeed, in that case player 1 would receive a reward strictly less than 1. However, if player 1 deviates to action T at stage 1, then entry TL will be chosen, and player 1 will be able to guarantee reward 1. This he can do by, after entry TL , playing as follows: player 1 plays his two actions alternately, unless player 2 loses an action at some point. In that case, player 1 plays the action repeatedly which gives him at least 1. This completes the proof. \square

3.2 Games of type 2

In this section we characterize the set of equilibrium rewards in 2-restricted games of type 2. We do so by examining the following type 2 game \mathcal{H} :

$$\mathcal{H} = \begin{matrix} & L & R \\ \begin{matrix} T \\ B \end{matrix} & \begin{pmatrix} 1, 2 & 0, 0 \\ 0, 0 & 4, 3 \end{pmatrix} \end{matrix}.$$

When at stage 1 entry TL or entry BR is chosen, we find the following analogously to Theorem 3:

Theorem 13 *If in \mathcal{H} at stage 1 one of the entries TL or BR is selected, then the set of possible follow-up equilibrium rewards is $\{(1, 2), (4, 3), (\frac{5}{2}, \frac{5}{2})\}$.*

Now suppose that at stage 1 entry TR or BL is selected. The analysis is quite similar to the one of the type 1 game \mathcal{G} in the previous section. Of course, $(1, 2)$ and $(4, 3)$ are follow-up equilibrium payoffs. Also, we find follow-up equilibria $(\hat{\pi}_q, \hat{\sigma}_q)$ as in lemma 11 for \mathcal{G} , now with rewards

$$\gamma(\hat{\pi}_q, \hat{\sigma}_q) = \left(1 - \frac{3}{4^{q+1}-1}, 2 - \frac{1}{\left(\frac{3}{2}\right)^{q+1}-1} \right)$$

for all $q \in \{0, 1, 2, \dots\}$. One can check with a proof similar to the proof of lemma 6 that there are no other possibilities for follow-up equilibrium rewards, so a result similar to lemma 7 does not hold for \mathcal{H} . Hence we obtain:

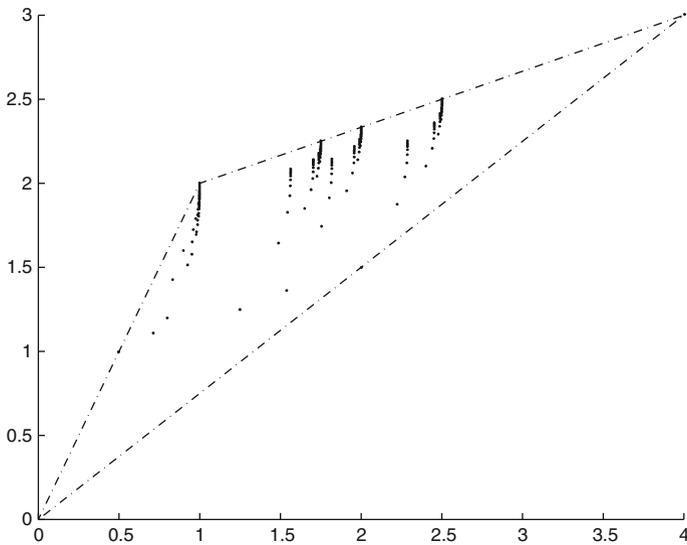


Fig. 2 Equilibrium rewards in \mathcal{H}

Theorem 14 *If in \mathcal{H} at stage 1 one of the entries TR or BL is selected, then the possible follow-up equilibrium rewards are $(1, 2)$, $(4, 3)$ and the rewards in the set*

$$\left\{ \left(1 - \frac{3}{4q+1-1}, 2 - \frac{1}{\left(\frac{3}{2}\right)^{q+1}-1} \right) \mid q \in \{0, 1, 2, \dots\} \right\}.$$

Note that in the above theorem, for $q = 0$ we obtain $(0, 0)$ as a follow-up equilibrium reward.

Based on the results of Theorems 13 and 14 we obtain the following theorem, whose proof is similar to the proof of Theorem 1.

Theorem 15 *The set of equilibrium rewards in \mathcal{H} equals the set of equilibrium rewards of the following one-shot games:*

$$\begin{pmatrix} V_1 & W_1 \\ W_2 & V_2 \end{pmatrix},$$

where $V_1, V_2 \in \left\{ (1, 2), (4, 3), \left(\frac{5}{2}, \frac{5}{2}\right) \right\}$ and

$$W_1, W_2 \in \left\{ \left(1 - \frac{3}{4q+1-1}, 2 - \frac{1}{\left(\frac{3}{2}\right)^{q+1}-1} \right) \mid q \in \{0, 1, 2, \dots\} \right\} \cup \{(1, 2), (4, 3)\}.$$

In Fig. 2 the set of equilibrium rewards in \mathcal{H} is depicted.

4 Subgame perfect equilibrium rewards in 2 × 2 coordination games

In this section we characterize the set of subgame perfect equilibria in the games \mathcal{G} and \mathcal{H} . We start with \mathcal{G} .

Lemma 16 *If in \mathcal{G} at stage 1 entry TL or respectively entry BR is selected, then the unique subgame perfect follow-up equilibrium reward is $(2, 1)$ $(1, 3)$.*

Proof We only prove the statement for action pair TL ; the arguments for BR are analogous. Let $(\hat{\pi}, \hat{\sigma})$ be a subgame perfect follow-up equilibrium in \mathcal{G} after at stage 1 entry TL is selected. In order to prove that $(\hat{\pi}, \hat{\sigma})$ induces reward $(2, 1)$, it suffices to show that if player 1 plays T at stage 2 (possibly by individually deviating at this stage), then his reward is 2. If player 1 plays T at stage 2, he unlearns action B and there are two possible scenarios:

1. If player 2 plays action L at stage 2, he unlearns action R and entry TL will be played at all further stages. This gives reward $(2, 1)$.
2. If player 2 plays action R at stage 2, then he keeps both actions alive. Since $(\hat{\pi}, \hat{\sigma})$ is subgame perfect, in this case player 2 will eventually unlearn action R and again the reward is $(2, 1)$.

In both cases player 1 receives reward 2, hence the statement. □

Lemma 17 *If in \mathcal{G} at stage 1 entry TR or entry BL is selected, then set of possible subgame perfect follow-up equilibrium rewards is*

$$\{(a, 1) \mid 1 \leq a \leq 2\} \cup \{(1, b) \mid 1 \leq b \leq 3\}.$$

Proof Let $(\hat{\pi}, \hat{\sigma})$ be a subgame perfect follow-up equilibrium in \mathcal{G} after at stage 1 one of the entries TR or BL is selected. Then $(\hat{\pi}, \hat{\sigma})$ is an equilibrium in the following stochastic game:

$(2, 1)_*$	$(0, 0)_{\rightarrow}$	$(2, 1)_*$	$(0, 0)_*$
$(0, 0)_*$	$(1, 3)_*$	$(0, 0)_{\leftarrow}$	$(1, 3)_*$
<i>state BL</i>		<i>state TR</i>	

The initial state is state BL if entry BL is chosen at stage 1, otherwise it is state TR . The interpretation of state BL is the following. If in state BL entry TR is chosen then the players receive zero payoffs and the play moves to state TR . If any other entry is chosen, then, as indicated by an asterisk, the play moves to an absorbing state in which this particular payoff will be obtained at all stages during the remainder of the play. The interpretation of state TR is similar.

Notice that player 1 can guarantee himself a reward of 1 by stationarily playing $(\frac{1}{2}, \frac{1}{2})$ in state BL and $(0, 1)$ in state TR . Similarly player 2 can guarantee himself a reward of 1 by stationarily playing $(\frac{2}{3}, \frac{1}{3})$ in state BL and $(1, 0)$ in state TR . The only follow-up equilibria in \mathcal{G} which induce a reward of at least $(1, 1)$ are, by the analysis in the previous section, the strategies as mentioned in the proof of lemma 7. Since these strategies do not make use of the $(0, 0)$ -threat, they are subgame perfect, which completes the proof. □

Based on the results of lemmas 16 and 17 we obtain following theorem, whose proof is similar to the proof of Theorem 1.

Theorem 18 *In \mathcal{G} the following rewards can be obtained by subgame perfect equilibria:*

$$\{(a, 1) \mid 1 \leq a \leq 2\} \cup \{(1, b) \mid 1 \leq b \leq 3\}.$$

For games of type 2, like \mathcal{H} , the analysis is even more straightforward. The reason is, as mentioned before, that a result similar to lemma 7 does not hold for \mathcal{H} . Hence:

Theorem 19 *If $(\hat{\pi}, \hat{\sigma})$ is a subgame perfect follow-up equilibrium in \mathcal{H} after at stage 1 entry BR is selected, then $\gamma(\hat{\pi}, \hat{\sigma}) = (4, 3)$. Furthermore if any of the other entries was selected at stage 1, then $\gamma(\hat{\pi}, \hat{\sigma}) \in \{(1, 2), (4, 3)\}$. Hence in \mathcal{H} only the rewards $(1, 2)$, $(4, 3)$ and $\frac{1}{2}(1, 2) + \frac{1}{2}(4, 3) = (\frac{5}{2}, \frac{5}{2})$ can be obtained by subgame perfect equilibria.*

5 $m \times m$ coordination games

In this section we describe a method to obtain both the set of equilibrium rewards and the set of subgame perfect equilibrium rewards in 2-restricted $m \times m$ coordination games. In Sect. 3 we saw that for 2×2 games the characteristic property of a coordination game, the $(0, 0)$ -payoffs outside the main diagonal, gave rise to a surprisingly complex set of equilibrium rewards. For bigger games however, the same characteristic ensures that the analysis does not get much more complicated. A few useful observations:

- Observation (A): After two stages each player has at most two actions left. In this subgame from stage 3 on, the action pair selected at stage 2 determines the set of follow-up (subgame perfect) equilibrium rewards. Indeed, from the actions available at stage 3 and the action pair selected at stage 2, we can deduce which actions were chosen at stage 1.
- Observation (B): Note that $(0, 0)$ is not an equilibrium reward. A player can, in expectation, get a strictly positive reward by using the following strategy: Randomize uniformly over all m actions at stages 1 and 2. Now irrespective of the other player's strategy, with probability $\frac{1}{m^2}$ both times a diagonal entry is selected. In that case the player can guarantee a strictly positive reward.
- Observation (C): If at stage 1 an off-diagonal entry is selected, then $(0, 0)$ is a follow-up equilibrium reward. Indeed, suppose that entry (i^1, i^2) is chosen at stage 1 with $i^1 \neq i^2$. Then consider the pair of follow-up strategies $(\hat{\pi}, \hat{\sigma})$ which prescribes to play as follows: At stage 2, the players should play (i^2, i^1) . If this occurs, then play the follow-up equilibrium with reward $(0, 0)$ from stage 3 on (cf. Theorems 12 and 14). On the other hand, if a player deviates

at stage 2, then his opponent should play the action chosen at stage 1 at every future stage, which gives the deviator reward 0. The pair $(\hat{\pi}, \hat{\sigma})$ is a follow-up equilibrium with reward $(0, 0)$.

5.1 Equilibrium rewards

Let Γ be a 2-restricted $m \times m$ coordination game. We will now describe a method based on backwards induction to find the set of equilibrium rewards in Γ . Observation (A) enables us to list all the essentially different game-situations after stage 2 and their corresponding sets of follow-up equilibrium rewards. Let $i = (i^1, i^2)$ denote the entry chosen at stage 1 and $j = (j^1, j^2)$ denote the entry chosen at stage 2. We use the notation $\Theta_E(i; j)$ for the follow-up equilibrium rewards after at stage 1 entry i and at stage 2 entry j are chosen. We distinguish the following game-situations:

- Game-situation (1): After stage 2 all payoffs are $(0, 0)$. In this case, $\Theta_E(i; j) = \{(0, 0)\}$.
- Game-situation (2): After stage 2 we are left with a payoff matrix

$$\begin{pmatrix} \alpha_1, \beta_1 & 0, 0 \\ 0, 0 & \alpha_2, \beta_2 \end{pmatrix},$$

where $\alpha_1, \alpha_2, \beta_1, \beta_2 > 0$. In this case, the values of the α 's and the β 's determine whether this matrix corresponds to a game of type 1 or of type 2. Recall that by observation (A) the set of follow-up equilibrium rewards depends on entry j . If j was on the main diagonal, then $\Theta_E(i; j)$ is the set described in Theorem 3 (in case of a type 1 game) or Theorem 13 (type 2 game); if j was off-diagonal, then $\Theta_E(i; j)$ is the set described in Theorem 12 (in case of a type 1 game) or Theorem 14 (type 2 game).

- Game-situation (3): After stage 2 we are left with a payoff matrix

$$\begin{pmatrix} \alpha, \beta & 0, 0 \\ 0, 0 & 0, 0 \end{pmatrix},$$

where $\alpha, \beta > 0$. In this case we have two possibilities. If entry j , which was selected at stage 2, was on the main diagonal of this matrix, then similarly to the proof of Theorem 3 we find that $\Theta_E(i; j) = \{(\alpha, \beta), \frac{1}{2}(\alpha, \beta), (0, 0)\}$. If j was off-diagonal, we now argue that $\Theta_E(i; j) = \{(0, 0), (\alpha, \beta)\}$. Clearly, both $(0, 0)$ and (α, β) can be obtained as follow-up equilibrium rewards (for $(0, 0)$, the players should alternately play the two off-diagonal entries in turn, supplemented with the $(0, 0)$ -threat). It remains to verify that there are no other follow-up equilibrium rewards. Suppose by way of contradiction that $c \cdot (\alpha, \beta)$ is a follow-up equilibrium reward where $c \in (0, 1)$. Suppose also that j is entry BL ; the proof is similar when j is entry TR . Then, at stage 3, player 1 never chooses action B , which would with certainty yield reward $(0, 0)$, and hence the only possibility is that entry TR is chosen at stage 3 with follow-up reward $c \cdot (\alpha, \beta)$. At stage 4, we find similarly that entry BL has to be chosen with probability 1, and by continuing this way we

obtain an alternating sequence on *BL* and *TR*. This however yields reward (0, 0), which is a contradiction.

- Game-situation (4): After stage 2 we are left with one of the following payoff matrices:

$$(\alpha, \beta \ 0, 0), \begin{pmatrix} \alpha, \beta \\ 0, 0 \end{pmatrix} \text{ or } (\alpha, \beta),$$

where $\alpha, \beta > 0$. Obviously here $\Theta_E(i; j) = \{(\alpha, \beta)\}$.

Now we describe the set of all follow-up equilibrium rewards after at stage 1 the action pair $i = (i^1, i^2)$ is selected. Note that at stage 2 both players still have all of their m actions available. First we need additional notations. For every player k , let V_i^k be the $m \times m$ matrix whose entry $j = (j^1, j^2)$ consists of the value of the zero-sum game which arises after at stage 1 entry i and at stage 2 entry j are played, and in which player k maximizes his own reward, whereas his opponent minimizes player k 's reward. This is player k 's punishment level in this subgame, and for its calculation we refer to [Joosten et al. \(1995\)](#).

Let S^1 and S^2 be non-empty subsets of the action sets $\{1, 2, \dots, m\}$ of the players, and let $S = S^1 \times S^2$. Let M_i^S denote an $m \times m$ matrix that is constructed as follows:

$$M_i^S(j) = \begin{cases} \text{a reward in } \Theta_E(i; j) & \text{if } j \in S \\ (V_i^1(j), V_i^2(j)) & \text{if } j \notin S. \end{cases}$$

Thus, when entry j is chosen in this auxiliary bimatrix-game M_i^S , if $j \in S$ then the players receive a follow-up equilibrium reward in the subgame after at stage 1 entry i and at stage 2 entry j are played, whereas if $j \notin S$ then they receive their punishment levels in the same subgame. We denote by $\Theta_E(M_i^S)$ the set of rewards which can be obtained by an equilibrium in M_i^S in which each player k uses every action from S^k with a positive probability and every other action with probability 0. Obviously, for some matrices M_i^S and subsets S^1 and S^2 , the set $\Theta_E(M_i^S)$ may be empty.

Let $\Theta_E(i)$ be the union of the sets $\Theta_E(M_i^S)$ for all possible matrices M_i^S and all possible subsets S^1 and S^2 of $\{1, 2, \dots, m\}$.

Note that every reward in $\Theta_E(M_i^S)$ is a follow-up equilibrium reward that can be obtained after at stage 1 action pair i is selected. Indeed, play the corresponding equilibrium in the bimatrix game M_i^S at stage 2, and subsequently, if the entry $j = (j^1, j^2)$ chosen at stage 2 is in S then continue with a follow-up equilibrium that induces the reward $M_i^S(j)$, whereas if action j^1 does not belong to S^1 then player 2 should punish player 1 by lowering player 1's reward to at most $V_i^1(j)$, and if action j^2 does not belong to S^2 then player 1 should punish player 2 by lowering player 2's reward to at most $V_i^2(j)$.

Based on this observation, one can verify that the set $\Theta_E(i)$ exactly consists of all follow-up equilibrium rewards that can be obtained after at stage 1 the action pair i is selected. Using the set $\Theta_E(i)$ for each action pair i , we can then use a similar technique for stage 1 to find the set of all equilibrium rewards in Γ .

5.2 Subgame perfect equilibrium rewards

Backwards induction can also be used to find the set of subgame perfect equilibrium rewards in an $m \times m$ game Γ . After stage 2 the play will still be in one of the four different game-situations mentioned in Sect. 5.1. We use the notation $\Theta_{SPE}(i; j)$ for the subgame-perfect follow-up equilibrium rewards after at stage 1 entry i and at stage 2 entry j are chosen. We first determine $\Theta_{SPE}(i; j)$ in each of the four game-situations:

- In game-situation (1), we obviously have $\Theta_{SPE}(i; j) = \{(0, 0)\}$.
- Now consider game-situation (2). If entry j was on the main diagonal, then $\Theta_{SPE}(i; j)$ follows from lemma 16 for a type 1 game and from Theorem 19 for a type 2 game. If j was off-diagonal, then $\Theta_{SPE}(i; j)$ follows from lemma 17 for a type 1 game and from Theorem 19 for a type 2 game.
- For game-situation (3), we find similarly to Theorem 19 that if j is entry TL then $\Theta_{SPE}(i; j) = \{(\alpha, \beta)\}$, whereas if j is any other entry then $\Theta_{SPE}(i; j) = \{(0, 0), (\alpha, \beta)\}$.
- In game-situation (4), we obviously have $\Theta_{SPE}(i; j) = \{(\alpha, \beta)\}$.

Now we describe the set of all subgame perfect follow-up equilibrium rewards after at stage 1 the action pair i is selected. Let N_i be an $m \times m$ matrix such that every entry j of N_i contains a reward from $\Theta_{SPE}(i; j)$. We denote by $\Theta(N_i)$ the set of equilibrium rewards in the bimatrix game N_i , and by $\Theta_{SPE}(i)$ the union of the sets $\Theta(N_i)$ for all possible matrices N_i .

The set $\Theta_{SPE}(i)$ exactly consists of all subgame perfect follow-up equilibrium rewards that can be obtained after at stage 1 action pair i is selected. Using the set $\Theta_{SPE}(i)$ for each action pair i , we can then use a similar technique for stage 1 to find the set of all subgame perfect equilibrium rewards in Γ .

6 Concluding remarks

General 2×2 games: In Sects. 3 and 4 we characterized the set of equilibrium and subgame perfect equilibrium rewards in all 2×2 coordination games. However, our techniques are also useful for the analysis of general 2×2 games. Unfortunately, it seems inevitable to distinguish a large number of cases and the results strongly depend on several characteristics of the payoff matrix, such as whether the players have strictly or weakly dominant actions, or are indifferent between two actions, but for instance also on the averages of the payoffs on a diagonal compared to the other payoffs. We will provide illustrative examples below for some of these cases and show that even seemingly simple cases may contain hidden difficulties and require a lengthy analysis. We note that in Borm (1987), no fewer than 15 different types of 2×2 payoff matrices are distinguished for the classification of the possible one-shot equilibria. In our setting, the averages of the payoffs on the diagonals further complicate the classification, although there might be a way to treat some cases simultaneously. To simplify the discussion, we focus on equilibrium rewards.

Consider the two games \mathcal{K}_1 and \mathcal{K}_5 where:

$$\mathcal{K}_x = \begin{matrix} & L & R \\ T & (6, 1) & (0, 0) \\ B & (x, 0) & (2, 1) \end{matrix}.$$

Notice that in the one-shot versions of both \mathcal{K}_1 and \mathcal{K}_5 there are three equilibria, namely TL , BR and a completely mixed one; this is also the case for the one-shot versions of the coordination games \mathcal{G} and \mathcal{H} that we discussed in Sects. 3 and 4. Now consider the 2-restricted versions of these games and assume that at stage 1 entry TR is selected. In \mathcal{K}_1 we find that $\frac{1}{2} \cdot (0, 0) + \frac{1}{2} \cdot (1, 0) = (\frac{1}{2}, 0)$ is not a follow-up equilibrium reward, since player 1 can get at least 1 by only playing action B . However, in \mathcal{K}_5 the payoff $\frac{1}{2} \cdot (0, 0) + \frac{1}{2} \cdot (5, 0) = (2\frac{1}{2}, 0)$ is a follow-up equilibrium reward. The corresponding follow-up strategies are as follows: Play TR at odd stages and BL at even stages, unless a player deviates. If player 1 deviates at any stage, then he unlearns an action and player 2 punishes him by playing only action R in the future. If player 2 deviates at any stage, then he too unlearns an action and player 1 punishes him by playing the action that reduces player 2’s payoff to 0 in the future.

Another factor that influences the analysis is a player having identical payoffs in different entries. Just consider the following seemingly simple game in which player 1, who has a strictly dominant action, can fully determine player 2’s payoff:

$$\begin{matrix} & L & R \\ T & (2, 1) & (5, 1) \\ B & (3, 0) & (6, 0) \end{matrix}$$

In this game, every pair in the set $[3, 5] \times [0, 1]$ is an equilibrium reward. To induce an arbitrary pair (u_1, u_2) in $[3, 5] \times [0, 1]$, we use a type of equilibrium which did not play a role in our analysis for coordination games. The players should play as follows: At stage 1, player 1 makes a lottery by choosing an action according to the probability distribution $(u_2, 1 - u_2)$, and at every later stage he repeats this action. Note that player 1 loses an action after stage 2. Player 2 simply waits by playing actions L and R at stages 1 and 2 respectively. Subsequently, player 2 makes a lottery at stage 3 depending on the remaining action of player 1, and plays the mixed action which gives player 1 expected payoff u_1 . At all stages from stage 4 onwards, player 2 keeps on repeating this action. In addition, if player 1 deviates to the other action at stage 2, then player 2 should punish him with playing action L in the future. Notice that the reward of each player is determined by the lottery of the opponent.

The above set $[3, 5] \times [0, 1]$ can be used to generate new equilibrium rewards. For instance as follows: at stage 1, each player plays the mixed action $(\frac{1}{2}, \frac{1}{2})$. If an entry on the main diagonal is chosen, then they play BR in the future (which is an equilibrium in the one-shot game), whereas if an entry on the other diagonal is chosen then they induce reward $(5, 1)$ by playing strategies as above (in this case, player 1 will make his lottery at stage 2, and player 2 should wait again by playing his two actions alternately). This induces reward $(5.5, 0.5)$. There are of course other equilibrium rewards as well, and maybe some of these can only be obtained with more complex strategies.

These examples above provided a brief illustration of the difficulties in general 2×2 games, and indicate that one may need to distinguish a large number of different cases. For equilibria in bigger games we refer to [Schoenmakers \(2004\)](#).

Discounted rewards: We will now briefly discuss what happens to the sets of equilibrium and subgame perfect equilibrium rewards when each player evaluates his sequence of payoffs by discounting, instead of taking the average reward. The bottom line is that these sets will change essentially, and will moreover contain a plethora of rewards. To illustrate the changes, we examine the set of (discounted) equilibrium rewards in the type 1 game \mathcal{G} for a fixed discount factor $\beta \in (0, 1)$. As before, the main difficulty lies in determining the possible follow-up equilibrium rewards given a certain entry is chosen at stage 1. Once this is done, we only have to investigate the resulting one-shot games as in [Theorem 1](#).

Assume first that entry TL is chosen at stage 1 (cf. [Theorem 3](#)); we obtain similar results for entry BR . In this case, we obtain many follow-up equilibrium rewards by letting the players play entries BR and TL in turn, starting with BR at stage 2, at a specified number of stages (possibly infinite), and then repeating either entry TL or entry BR at all further stages. This means in particular that, for relatively large discount factors, we obtain follow-up equilibrium rewards that are “far” from the set $\{(2, 1), (1, 3), (\frac{3}{2}, 2)\}$ as found for the average reward in [Theorem 3](#). Thus, it is not true that the discounted game approaches the average reward game when the discount factor tends to 1.

Besides, for small discount factors, we also have other types of follow-up equilibria, such as the following: at stage 2 play the unique completely mixed equilibrium in the bimatrix game

$$\begin{array}{cc} & \begin{array}{cc} L & R \end{array} \\ \begin{array}{c} T \\ B \end{array} & \begin{pmatrix} 2, 1 & \beta \cdot (2, 1) \\ \beta \cdot (2, 1) & 1, 3 \end{pmatrix} \end{array}$$

and at all further stages from stage 3 on, play entry BR if the entry at stage t was BR , and play entry TL otherwise. Note that the payoffs of this bimatrix game are exactly the corresponding follow-up rewards. In fact, the special case $\beta = 0.5$ would even allow randomization at stage 2 on the entries TR and BR .

Assume now that entry BL is chosen at stage 1 (cf. [Theorem 12](#)); we obtain similar results for entry TR . Here again, we have a large number of follow-up equilibrium rewards. Just consider the type of follow-up strategies as in [lemma 11](#). Now it will also matter, in contrast with the average reward game, when the randomizations take place, which leads to plenty of possibilities. Moreover, as part (i) of [lemma 9](#) is no longer valid for discounted rewards, part (ii) does not hold any more either, and this has the consequence that in the follow-up strategies as in [lemma 11](#) we also have to consider randomization by only one player at certain stages.

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

- Arrow KJ (1962) The economic implications of learning by doing. *Rev Econ Stud* 29:155–173
- Borm P (1987) A classification of 2×2 bimatrix games. *Cahiers du CERO* 29:69–84
- Joosten R, Peters H, Thuijsman F (1995) Unlearning by not doing: repeated games with vanishing actions. *Games Econ Behav* 9:1–7
- Mertens JF, Neyman A (1981) Stochastic games. *Int J Game Theory* 10:53–66
- Schoenmakers G (2004) The profit of skills in repeated and stochastic games. PhD dissertation, Maastricht University, The Netherlands
- Schoenmakers G, Flesch J, Thuijsman F (2002) Coordination games with vanishing actions. *Int Game Theory Rev* 4:119–126
- Shapley LS (1953) Stochastic games. *Proc Natl Acad Sci USA* 39:1095–1100
- Vieille N (2000) Two-player stochastic games I: a reduction. *Israel J Math* 119:55–91
- Vieille N (2000) Two-player stochastic games II: the case of recursive games. *Israel J Math* 119:93–126