The profit of skills in repeated and stochastic games

The profit of skills in repeated and stochastic games

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit Maastricht,
op gezag van de Rector Magnificus,
Prof. mr. G.P.M.F. Mols
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen
op woensdag 29 september 2004 om 12.00 uur

door

Gijsbertus Maria Schoenmakers

Promotor:

Prof. dr. ir. drs. O.J. Vrieze

Copromotores:

Dr. F. Thuijsman

Dr. J. Flesch

Beoordelingscommissie:

Prof. dr. ir. C.P.M. van Hoesel (voorzitter)

Dr. R. Peeters

Prof. dr. H.J.M. Peters

Prof. dr. S.H. Tijs (Universiteit van Tilburg)

The profit of skills in repeated and stochastic games

Acknowledgements

I would like to thank all the people, who helped me to write this monograph. The first name that crosses my mind is János Flesch. He invested almost all of his available time helping me with my research. For his efforts he truly deserves his supervisorship. Then there are my supervisors Koos Vrieze and Frank Thuijsman. They showed great and thorough knowledge of stochastic game theory and their discussions have always been inspiring.

I think relaxation is very important in life. Therefore I would like to thank the people who played many, many pool games with me, in particular Jeroen Kuipers and Dries Vermeulen. In this context a special thanks to Dirk Overkleeft, who showed me that there are people who sometimes not even want to quit playing when I'm getting tired of it, which is quite an achievement.

Furthermore I would like thank the members of Raes Damclub Maastricht. The atmosphere in the club is always friendly, the most important reason for me to be there.

Finally I would like to thank my family, in particular my father and my sisters, and my friends for providing me the sometimes much-needed mental as well as physical support.

Contents

Chapter 1

Introduction

1.1 Introduction and summary

Game theory describes and analyzes situations, in which several decision makers, usually called players, who may or may not have conflicting interests, interact. In repeated and stochastic games these interactions occur over and over again. For the sake of simplicity we only describe 2-player games here; the generalization to games with an arbitrary number of players is straightforward. A 2-player stochastic game can be described as follows: We have a state space S and for each state $s \in S$ the players have action sets I^s (for player 1) and J^s (for player 2). To each action pair corresponds a payoff to both players and a probability vector, the transition vector. The play of a stochastic game proceeds as follows: Play starts at stage 1 in initial state $s \in S$, where, simultaneously and independently, both players choose an action: player 1 has to choose an action $i^s \in I^s$ and player 2 has to choose an action $j^s \in J^s$. Now each player receives the payoff corresponding to the action pair (i^s, j^s) and play moves to state $s' \in S$ according to the transition vector. At stage 2 in state s' the players have to choose actions again. Again they receive the payoffs corresponding to the selected action pair and play moves to another state, where actions have to be chosen again at stage 3 and so on to infinity.

Some general assumptions in stochastic games: The game is non-cooperative, which means that the players can not make binding agreements. Furthermore the players have complete information, which means that they know the stochastic game and they have perfect recall, which means that they remember the entire history of play so far. Consequently the players can use this information, when choosing an action at the current stage.

A plan that tells a player at every decision moment which action to play given the history of play so far, is called a strategy. A strategy may prescribe to play a certain action with probability 1 as well as to randomize over the available actions. The latter is called a mixed action. If the prescribed (mixed) action depends only on the state that is currently visited, then the strategy is called stationary.

The play of the game gives rise to an infinite stream of payoffs to both players, which has to be evaluated. Such evaluations are called rewards and the aim of each

player is to maximize his own reward. The most common rewards are the β-discounted reward and the limiting average reward. We will almost exclusively deal with the latter.

We can distinguish between two types of stochastic games. In the first type of 2-player games the players have completely opposite interests; the gain of one player is the loss of the other player. These are so-called zero-sum stochastic games. For zero-sum stochastic games we assume that player 1 maximizes his reward, whereas player 2 minimizes player 1's reward (which is thus equivalent with the maximization of his own reward). It is well-known that there exists a unique reward such that for any $\varepsilon > 0$ player 1 has a strategy that guarantees this reward up to ε against any strategy of player 2, whereas player 2 has a strategy available that guarantees that player 1's reward is not more than this reward (up to ε) no matter which strategy player 1 uses. This reward is called the value of the game and the related strategies are called ε-optimal. Strategies that completely guarantee the value, 0-optimal strategies, do not necessarily exist.

Stochastic games that do not assume opposite interests for the players, are called general-sum stochastic games. Since the players in these games might, up to some extent, have matching interests the notions of value and optimality are no longer meaningful. The usual solution concept in general-sum stochastic games is the (ε-)equilibrium. Here a pair of strategies is an ε-equilibrium, if neither player can gain more than ε by making a unilateral deviation. This concept of equilibrium was introduced by Nash (1950a,1950b,1951) for bimatrix games and it is therefore known as the Nash-equilibrium. For 2-player games (ε-)equilibria always exist (cf. Vieille (2000a,2000b)). For more than 2 players however, the existence of (ε-)equilibria in general-sum stochastic games is still an open problem.

A specific type of stochastic game is a stochastic game that consists of one state only. Such a game is called a repeated game. For zero-sum repeated games it is well-known that 0-optimal strategies always exist, whereas for general-sum repeated games 0-equilibria always exist.

During the course of repeated and stochastic games the players may change their strategic behavior. The idea is that if a player figures out that he can get a higher reward by playing a different action or, on the contrary, by sticking to the same action all the time, then he will do so. This is the concept of learning in games. The models we discuss in this monograph all somehow fit in the framework of learning. The models in this monograph are on skill improvement; by repeatedly playing actions the players acquire and improve skills that, in return, are expected to yield high payoffs. Examples of skill-improvement are learning by doing, imitation learning and reinforcement learning. We shall briefly describe each of these types now.

In the learning-by-doing model by Arrow (1962), players acquire and improve skills simply by performing the same task over and over again. Arrow models the improvement in skills as follows: As a player performs a specific task more often, he is capable to do it in less time. There are different ways to model skill-improvement. The output of firms is related to labor, capital and materials and we basically have the same physical resources available now as 200 years ago. Nevertheless our standard of living is much higher now than it was 200 years ago. The field of growth theory investigates this phenomenon and states that it stems from our improved ability

to transform the same resources into products that are worth more. For example computers now are a thousand times faster than they were 20 years ago, whereas it consists of almost the same materials. For a summary on recent results in growth theory we refer to Cortright (2001) and Jovanovic (2000). The concept of learning by doing as a tool to improve skills can also be modelled as follows: the payoff corresponding to a certain action increases as it has been played more often in the recent past. So players don't do more in the same time span, they do the same things better. Similarly playing an action less often may lead to a skill-deterioration, which logically corresponds to a lower payoff. This last phenomenon is called unlearning by not doing and was introduced by Joosten, Peters and Thuijsman (1995). In chapters 2, 3, 4 and 5 of this monograph we will discuss models on skill-improvement and -deterioration.

In imitation learning (cf. e.g. Mataric (1994,1997)), the learner observes a skill being demonstrated by a teacher, then attempts to imitate that skill, and finally refines the skill through trial and error learning. The ability of imitation learning provides the opportunity to profit from knowledge of others and to acquire new skills much more quickly. Effectively, imitation learning biases a learning system towards a good solution in order to significantly reduce the search space during trial by trial learning.

Reinforcement learning (cf. e.g. Kaelbling, Littman and Moore (1996) or Sutton and Barto (1998)) is the problem faced by a player who learns strategic behavior through trial-and-error interactions with a dynamic environment. If in a certain state the player chooses one of his available actions, then he receives a (stochastic) payoff (the reinforcement) and he transfers, by some stochastic transition rule, to another state, where he has to choose an action again. A player's aim is to maximize the long-run average amount of reinforcement, where he can only find out how much reinforcement a specific action in a certain state yields by playing it a few times. For a psychological viewpoint on reinforcement learning we refer to Walker (1975).

Fictitious play can be described as the problem faced by two or more players who have no notion of the payoffs of the other players in the game. At each stage each player only observes the actions that have been played and his own payoff. If the game is zero-sum, we expect the players not to know. The discrete fictitious play process briefly consists of a (bi)matrix game that is played repeatedly, where at each stage each player selects a pure action that is a best reply against the "average" action of the other player(s). Here the average action of a player is the probability vector consisting of the frequencies by which he played his actions so far. The aim of the fictitious play processes is that it converges to an equilibrium, which is not necessarily the case. A game is said to have the fictitious play property, if every fictitious play process converges to an equilibrium. The fictitious play process was first proposed by Brown (1951) and Robinson (1951) as a tool to find optimal mixed actions in matrix games.

We now describe the setup of this monograph.

In section 1.2 we present the formal definition of a stochastic as well as a repeated game. Furthermore some results that are useful for the analysis of the models in the other chapters.

In chapter 2 we discuss the model of repeated games with bonuses. A repeated

game with bonus ξ is a zero-sum repeated game, in which player 1 can improve his action skills, where playing an action at high skill yields a payoff that is ξ higher than playing the same action at low skill. Here at stage t the action that player 1 played at stage $t-1$ is high-skill, whereas all other actions are low-skill. Within the framework of this model we investigate so-called simple strategies, strategies that prescribe to play the same mixed action at each stage, irrespective of past play. We first derive a relationship between the value of the repeated game with bonus ξ and the underlying matrix game. After that we show that in 2×2 - games player 1 always has a simple optimal strategy. Furthermore for $m \times n$ - games in which player 1 has a simple optimal strategy, we characterize the set of stationary optimal strategies for player 2 and finally this result is generalized to games with action-dependent bonuses.

In chapters 3, 4 and 5 we discuss the model of repeated games with vanishing actions. A repeated game with vanishing actions is a game, in which actions vanish from the players' action sets if they have not been played in the recent past, an extreme form of unlearning by not doing. Furthermore, once an action has vanished the player can never play it again; there is no way to regain actions that have been unlearned. Hence in this model skills can deteriorate but not improve. In chapter 3 we calculate the value for 2×2 - zero-sum games, in which one of the players is vulnerable to unlearning.

In chapter 4 we consider coordination games with vanishing actions. A coordination game is a game in which all payoffs on the matrix diagonal are positive, whereas all off-diagonal payoffs are 0. We characterize the set of equilibrium rewards in 2×2 - coordination games in which the actions of each player vanish, if they have not been played at least once at the previous 2 stages. Such a game is called a $(2,2)$-restricted game. After that we consider $(3,3)$-restricted coordination games, games in which each action of each player vanish, if they have not been played at least once at the previous 3 stages. The analysis of $(3,3)$-restricted coordination games also concerns the general $m \times m$ - sized games. With the aid of "agreements", a specific type of strategy pair, we show that all Pareto-optimal rewards are obtainable as equilibrium rewards, although a Folk-theorem is not applicable for this class of games. This is also the case for $\left(r^1, r^2\right)$-restricted games, as long as $r^1 \geq 3$ and $r^2 \geq 3$. For multiplayer $(3,3,\ldots,3)$-restricted coordination games we prove that the Folk-theorem does apply: Every feasible and individually rational rewards in an $\left(r^1, r^2\right)$-restricted game can be obtained by a pair of equilibrium strategies.

Chapter 5 deals with 2-player general-sum games with vanishing actions. Firstly we discuss for $\left(r^1, r^2\right)$-restricted 2×2 - games the possible frequencies by which action pairs can be played, such that neither player unlearns an action. The analysis concerns so-called frequency matrices, matrices in which the (i,j)-th entry shows the frequency by which the action pair (i,j) is being played. We characterize the set of obtainable frequency matrices in 2×2 - games. Then we generalize the concept of frequency matrices to $m \times n$ - games, where for some specific frequency matrices we derive a set of necessary and sufficient conditions concerning the obtainability. After that we start the search for equilibria by means of a generalized version of the agreements as defined in chapter 4. Furthermore we show that a Folk-theorem like result is not applicable for general-sum games, although it does apply for coordination games with at least 3 players.

In chapter 6 we discuss a model of fictitious play in stochastic games. Therefore we first introduce a model of fictitious play in repeated games, which is a straightforward generalization of the fictitious play model for one-shot games, using stationary strategies instead of mixed actions. Then we present 3 possible generalizations of this model to a fictitious play model in stochastic games, one of which we analyze in more detail. With the aid of an example, we show that for this model the fictitious play process does not necessarily converge to a pair of stationary equilibrium strategies.

1.2 The stochastic game model

Definition 1.2.1 *A 2-player stochastic game is determined by the following parameters:*

1. $K = \{1, 2\}$ *is the set of players;*
2. $S = \{1, 2, \ldots, z\}$ *is the set of states;*
3. $I^s = \{1, 2, \ldots, m^s\}$ *is the set of pure actions for player 1 in state* $s \in S$*;*
4. $J^s = \{1, 2, \ldots, n^s\}$ *is the set of pure actions for player 2 in state* $s \in S$*;*
5. $R^k(s, i^s, j^s)$ *is the payoff to player* $k \in K$*, if in state* $s \in S$ *the action pair* (i^s, j^s) *is played;*
6. $p(s' \mid s, i^s, j^s)$ *is the probability that play moves to state* s'*, if in state* s *the action pair* (i^s, j^s) *is played.*

The game is played at stages $1, 2, \ldots$*, where each time simultaneously and independently both players have to choose one of their available actions. The action choices are announced and then 2 things happen: Firstly each player* $k \in K$ *receives a payoff* $R^k(s, i^s, j^s)$ *and secondly with probability* $p(s' \mid s, i^s, j^s)$ *play moves to state* s'*, where actions have to be chosen again. Here obviously* $\sum_{s' \in S} p(s' \mid s, i^s, j^s) = 1$ *for all* $s \in S$*.*

Stochastic games were introduced by Shapley (1953).
For a 2-player stochastic game we define h_t to be the *history of play* up to stage t:

$$h_1 = \emptyset$$

and

$$h_t = (s_1, i_1, j_1, s_2, i_2, j_2, \ldots, s_{t-1}, i_{t-1}, j_{t-1}, s_t) \text{ for } t \geq 2. \tag{1.1}$$

At any stage t, if state $s \in S$ is currently being visited, the players are allowed to randomize over the actions in I^s and J^s, which yields a *mixed action* denoted a^s for player 1 and b^s for player 2 at stage t, and these choices may depend on h_t. Each pure action is a mixed action where something is chosen with probability 1. The set of mixed actions in state s of players 1 and 2 are denoted by A^s and B^s respectively. A sequence

$$\pi = (a_t(h_t))_{t=1}^{\infty} \tag{1.2}$$

is called a *strategy* for player 1. For player 2 a strategy σ is defined analogously. In a stochastic game player $k \in K$ evaluates the infinite stream of stage payoffs generated by the strategy pair (π, σ) and given the initial state s , i.e. the state that is visited at stage 1, by means of a *reward function* $\gamma^{ks}(\pi, \sigma)$. The most common reward functions for repeated games are the *β-discounted reward*

$$\gamma_\beta^{ks}(\pi,\sigma) = (1-\beta)\sum_{t=1}^{\infty}\beta^{t-1}E_{\pi,\sigma}^{s}\left(R_t^k\right) \tag{1.3}$$

and the *limiting average reward*

$$\gamma^{ks}(\pi,\sigma) = \liminf_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}E_{\pi,\sigma}^{s}\left(R_t^k\right), \tag{1.4}$$

which was introduced by Gillette (1957). Each player is assumed to maximize his reward. A pair of strategies (π, σ) is an *equilibrium* if for each player k and for each initial state s it holds that

$$\gamma^{1s}(\tilde{\pi},\sigma) \leq \gamma^{1s}(\pi,\sigma) \text{ for all } \tilde{\pi} \tag{1.5}$$

and

$$\gamma^{2s}(\pi,\tilde{\sigma}) \leq \gamma^{2s}(\pi,\sigma) \text{ for all } \tilde{\sigma} \tag{1.6}$$

i.e. each player is playing a *best reply* against the strategy of the other player. We will now present the formal definitions of some properties that stochastic games may have.

Definition 1.2.2 *A stochastic game has the single-controller property, if for each state $s \in S$ either*

$$p(s' \mid s, i^s, j^s) = p(s' \mid s, i^s, \hat{\jmath}^s) \text{ for all } j^s, \hat{\jmath}^s \in J^s$$

(a player-1 controlled game) or

$$p(s' \mid s, i^s, j^s) = p(s' \mid s, \hat{\imath}^s, j^s) \text{ for all } i^s, \hat{\imath}^s \in J^s$$

(a player-2 controlled game). Hence the transition probabilities in the game depend on the actions of one player only.

Definition 1.2.3 *A stochastic game has state independent transitions, if for each state $s \in S$*

$$p(s' \mid s, i, j) = p(s' \mid i, j) \text{ for all } i \in I^s \text{ and for all } j \in J^s.$$

Notice that the state independent transition property is meaningful only if $m^s = m$ and $n^s = n$ for each $s \in S$.

Definition 1.2.4 *A stochastic game is irreducible, if, irrespective of the players' strategies, each state will be visited infinitely often with probability 1.*

Some of the consequences of these properties with respect to the existence of optimal strategies or equilibria are listed below.

Theorem 1.2.5 *In irreducible zero-sum stochastic games both players have stationary optimal strategies and the value is independent of the initial state (cf. Hoffman and Karp (1966) or Thuijsman (1992)).*
In zero-sum single-controller stochastic games both players have stationary optimal strategies (cf. Filar (1981)).
In zero-sum games with state independent transitions both players have stationary optimal strategies and the value is independent of the initial state (cf. Thuijsman (1992)).
In irreducible stochastic games stationary equilibria exist (cf. Rogers (1969), Sobel (1971) or Federgruen (1978)).

For surveys on stochastic games we refer to Filar and Vrieze (1997), Thuijsman (1992) and Neyman (2004).

Definition 1.2.6 *A 2-player repeated game is a 2-player stochastic game with just one state.*

When discussing a repeated game we will suppress the state variable in the notations. A reward $\hat{\gamma} = \left(\hat{\gamma}^1, \hat{\gamma}^2\right)$ is called *individually rational* if

$$\hat{\gamma}^1 \geq \inf_{\sigma} \sup_{\pi} \gamma^1(\pi, \sigma) \tag{1.7}$$

and

$$\hat{\gamma}^2 \geq \inf_{\pi} \sup_{\sigma} \gamma^2(\pi, \sigma).$$

This means that $\hat{\gamma}^k$ is the highest reward that player k can defend against any strategy of the opponent. Furthermore a reward $\bar{\gamma} = \left(\bar{\gamma}^1, \bar{\gamma}^2\right)$ is called *feasible*, if there exists a strategy pair (π, σ) such that for each player k

$$\gamma^k(\pi, \sigma) = \bar{\gamma}^k. \tag{1.8}$$

A well-known result on repeated games with the limiting average reward (cf. e.g. Aumann (1981) or Sorin (1992)) is the Folk-theorem, arguable the most outstanding result in repeated games:

Theorem 1.2.7 *In a repeated game every feasible and individually rational reward can be obtained as an equilibrium reward.*

Chapter 2

Repeated Games With Bonuses

2.1 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.

These phenomena may be called learning by doing and unlearning by not doing respectively. From the seminal paper of Arrow (1962) 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 game-theory. However, on the unlearning side of the story, although very much 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, Peters and Thuijsman introduced a model of unlearning for infinitely repeated zero-sum games, which was generalized by Schoenmakers, Flesch and Thuijsman (2002) to non-zero-sum games. These models, called games with vanishing actions, will be discussed in chapter 4. In this chapter we discus a model that deals with learning and unlearning in infinitely repeated games in a slightly different way.

In section 2.2 the model of the (zero-sum) repeated game with bonus ξ is presented in more detail. In section 2.3 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. In section 2.4 we do the same for $2 \times n$ -games. In section 2.5 we take a look at games of arbitrary size and, under the assumption that player 1 has a simple optimal strategy, we characterize the set of stationary optimal strategies of player 2. Section 2.6 concludes. Sections 2.2-2.6 are based on Schoenmakers, Flesch, Thuijsman and Vrieze (2004).

2.2 The repeated game with bonus ξ: the model

In section 1.2 we presented the model of repeated games as a one-state stochastic game. We will now insert a skill-improvement and -deterioration component into the model of zero-sum repeated games. We do this in the following way: Suppose that player 1 played action i at stage t. Now at stage t player 1 has learned "how to handle action i" and action i becomes a high-skill action at stage t. 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 $\hat{\imath} \neq i$, then he will not receive a bonus and he will forget how to handle 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 $\hat{\imath}$ that thereby becomes high-skill, so if player 1 decides to play action $\hat{\imath}$ 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. This higher reward is expressed as the 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.
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 $\{1, \ldots, m\}$ and $\{1, \ldots, n\}$ 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

$$m_{ij} \text{ is the } (i,j)\text{-th entry of } M. \tag{2.1}$$

However, if player 1 also selected action i at the previous stage, then he receives $m_{ij} + \xi$ from player 2.

Definition 2.2.1 *The repeated game with bonus ξ corresponding to the matrix game $M \in \mathbb{R}^{m \times n}$ can be formulated a zero-sum stochastic game with finite state and action spaces with the following properties:*

1. *The set of states is $\{1, \ldots, m\}$, where state s is related to player 1's action s in the matrix game M.*

2. *In each state players 1 and 2 have action sets* $\{1,\ldots,m\}$ *and* $\{1,\ldots,n\}$ *respectively.*

3. *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 *with respect to the corresponding payoffs in* M*.*

4. *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 2.2.2 *The repeated game with bonus* ξ *corresponding to the matrix game* M *is called* M_ξ*.*

As evaluation criterion for the stream of payoffs generated by the strategy pair (π,σ), we will use the *limiting average reward*, i.e.

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

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 $s \in \{1,\ldots,m\}$ (cf. (1.4)). Player 1 maximizes $\gamma_\xi^s(\pi,\sigma)$, whereas player 2 minimizes the same reward.

Notation 2.2.3 *The unit simplex in* $\mathbb{R}^z$ *is denoted by* Δ^z*.*

Theorem 2.2.4 *Von Neumann (1928)*

Each matrix game M *has a value* $v \in \mathbb{R}$*, for which*

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

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

Notice that the transposed sign is left out. For notational purposes we will leave it out throughout this chapter.

Definition 2.2.5 *A stationary strategy* $x = \left(a^1, a^2, \ldots, a^m\right)$ *for player 1 prescribes to play the mixed action* a^s *each time state* s *is visited. Hence* x *is independent of* h_t*. A stationary strategy* $y = \left(b^1, b^2, \ldots, b^m\right)$ *for player 2 is defined analogously.*

Mertens & Neyman (1981) proved that zero-sum stochastic games, like matrix games, have a value v, which may depend on the initial state.

Definition 2.2.6 *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 Gilette (1957). For M_ξ however, not only do optimal strategies exist, but as theorem 2.2.7 shows, even stationary optimal strategies exist.

Theorem 2.2.7 *The game M_ξ has a state-independent value v_ξ and players 1 and 2 have stationary optimal strategies x^* and y^* respectively.*

Proof. Notice first that in M_0, the game without bonus, the m states are identical and the game is strategically indifferent from the ordinary repeated game. Therefore in M_0 an optimal strategy for each player is to repeatedly play a mixed action that is optimal in the underlying matrix game M and the value v_0 of M_0 equals the value v of M. According to point 4 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 (1981) or theorem 1.2.5). Furthermore Thuijsman (1992) proved that for zero-sum games with state independent transitions the value is independent of the initial state (cf. theorem 1.2.5). ■

As a consequence of theorem 2.2.7 we have:

$$\gamma_\xi^s(x^*, y^*) = v_\xi \tag{2.2}$$

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

In Sobel (1981) these type of strategies are called myopic. Notice that for the existence of simple strategies it is essential that all states have the same action sets. Notice furthermore 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

$$\gamma_\xi^s(a', y) = \gamma_\xi^{s'}(a', y) \text{ for all } s, s' \in \{1, \ldots, m\}. \tag{2.3}$$

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

Definition 2.2.9 *The carrier of a mixed action a is defined as follows: $car(a) = \{i \in \{1, \ldots, m\} \mid a_i > 0\}$. For the stationary strategy $x = \left(a^1, a^2, \ldots, a^m\right)$ let $car(x)$ denote the Cartesian product $car\left(a^1\right) \times car\left(a^2\right) \times \ldots \times car\left(a^m\right)$.*

Notation 2.2.10 *The variable s will denote an action of player 1, a state or even both at once. The line "state $s \in car(a')$" should be interpreted as the state s that is visited after player 1 plays action $s \in car(a)$, where player 1 uses the simple strategy a'. Furthermore, in this chapter 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 simple strategy a', denoting the probability to play action s.*

The main goal of sections 2.3 and 2.4 is to provide conditions, under which simple optimal strategies exist.

Notation 2.2.11 *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. But then, since for each mixed action a we have

$$\min_{j\in J}\{aMe_j+\xi\cdot a_i\}=\min_{j\in J}\{aMe_j\}+\xi\cdot a_i \text{ for all } i\in\{1,\dots,m\},$$

player 2's set of one-shot best replies against a in state s of M_ξ is the same as his set of best replies against a in M. Here e_j denotes the unit vector with the 1 at position j (cf. notation 2.5.4).

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 $s\in car(a')$ will be visited infinitely often, whereas each state outside $car(a')$ will not be visited at all, except if it happens to be the initial state. Now a stationary strategy $y=\left(b^1,b^2,\dots,b^m\right)$ is a best reply against a' if for each state $s\in car(a')$ we have: $b^s\in B(a)$. Furthermore each state $s\notin car(a')$ 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 with $car(a)=\{i_1,\dots,i_p\}\subset\{1,\dots,m\}$ we have

$$B_\xi(a')=\{(b^1,\dots,b^m)\,|\,b^s\in B(a) \text{ for all } s\in\{i_1,\dots,i_p\},\ b^s\in\Delta^n \text{ for all } s\notin\{i_1,\dots,i_p\}\} \tag{2.4}$$

Notation 2.2.12 *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 $\varphi^s_\xi(x)=\min_y\gamma^s_\xi(x,y)$. Notice that for each simple strategy a' we have that $\varphi_\xi(a')$ is independent of the initial state (cf. (2.3)). Notice furthermore that, by (2.2), for any stationary optimal strategy x^* also $\varphi_\xi(x^*)=v_\xi$ is independent of the initial state.*

Since M_0 has m identical states, which are also identical to the single state in M, we have:

$$\varphi_0(a')=\min_{b\in B}aMb.$$

Consider a strategy pair (a',y), where $y=\left(b^1,b^2,\dots,b^m\right)\in B_\xi(a')$. From (2.4) it follows that for each state $s\in car(a')$ the (mixed) action b^s is a best reply against a in M, which means that $\min_{b\in B}aMb=aMb^s$ and hence

Lemma 2.2.13 *For each* $y = \left(b^1, b^2, \ldots, b^m\right) \in B_\xi(a')$ *and for each* $s \in car\,(a')$ *we have:*

$$aMb^s = \varphi_0(a').$$

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

Given the strategy pair (a', y) if, at a certain stage, state $s \in car(a')$ 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

$$\varphi_\xi\,(a') = \sum_{s=1}^{m} a_s \cdot (\varphi_0(a') + \xi \cdot a_s) = \varphi_0(a') + \xi \cdot \sum_{s=1}^{m} a_s^2. \tag{2.5}$$

For the optimal mixed action a^* in M we have

$$v_\xi \geq \varphi_\xi\,(a^{*\prime}) = v + \xi \cdot \sum_{s=1}^{m} \left(a_s^*\right)^2, \tag{2.6}$$

since $a^{*\prime}$ might not be optimal in M_ξ. However if a' would be a simple optimal strategy, then obviously

$$v_\xi = \varphi_\xi\,(a') = \varphi_0(a') + \xi \cdot \sum_{s=1}^{m} a_s^2. \tag{2.7}$$

Suppose 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 $\hat{\imath}'$ such that $\hat{\imath} \in \arg\max_{i \in I} \min_{j \in J} m_{ij}$ player 1 can guarantee a reward of $\max_{i \in I} \min_{j \in J} m_{ij} + \xi$. Combining these observations with (2.6) we find

Lemma 2.2.14 *For each repeated game with bonus* ξ *we have:* $\max\{v + \xi \cdot \sum_{s=1}^{m} \left(a_s^*\right)^2, \max_{i \in I} \min_{j \in J} m_{ij} + \xi\} \leq v_\xi \leq v + \xi$.

In figure 2.1 v_ξ is somewhere in the dotted area. Notice that, if a^* is pure, then $\sum_{s=1}^{m} \left(a_s^*\right)^2 = 1$ and $\max_{i \in I} \min_{j \in J} m_{ij} + \xi = v$ and hence $v_\xi = v + \xi$ and $a^{*\prime}$ is optimal in M_ξ. This means that

Theorem 2.2.15 *If* a^* *is a pure optimal action of player 1 in* M*, then* $a^{*\prime}$ *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 and in the following sections we only have to consider games M_ξ for which player 1 does not have a pure optimal action in the underlying matrix game M.

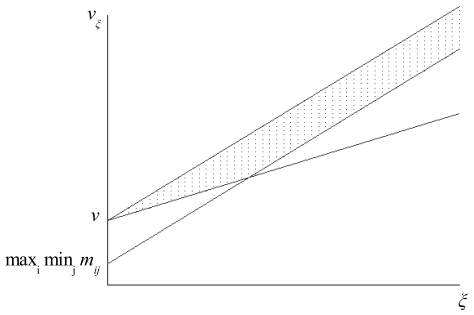


Figure 2.1: v_ξ versus ξ

Definition 2.2.16 *A cycle* $\mathbb{C}(s_1, s_2, \ldots, s_l)$ *is a sequence of consecutively visited states* $s_1, s_2, \ldots, s_l, s_{l+1}$ *such that* $s_i \neq s_1$ *for each* $i \in \{2, \ldots, l\}$ *and* $s_{l+1} = s_1$.

Recall that player 1 controls the transitions and that, if he plays a', the probability to go from state s_i to state s_j is a_{s_j}, which does not depend on s_i. Therefore, given that player 1 is playing a' and that at stage t state $s_1 \in car(a')$ is visited, the probability that, from stage t on, the cycle $\mathbb{C}(s_1, s_2, \ldots, s_l)$ will appear, is $a_{s_2} \cdot \ldots \cdot a_{s_l} \cdot a_{s_1}$, which is strictly positive, as long as $s_i \in car(a')$ for all $i \in \{1, \ldots, l\}$. For each cycle $\mathbb{C}(s_1, s_2, \ldots, s_l)$ let the pure strategy of player 1 that prescribes to play the action sequence $(s_1, s_2, \ldots, s_l)$ repeatedly, be denoted by $\mathbb{C}'(s_1, s_2, \ldots, s_l)$.
Let π be a pure strategy of player 1, prescribing to play action s_{t+1} at stage t, where s_{t+1} neither depends on player 2's action choices so far nor on the initial state s_1. Then π leads to an infinite sequence of visited states $(s_1, s_2, \ldots)$. Given this sequence the infinite sequence of cycles $(\mathbb{C}_1^\pi, \mathbb{C}_2^\pi, \ldots)$, where cycle C_l^π starts at stage t_l and ends at stage u_l is constructed as follows:
Consider state s_1 and let w be the first stage after stage 1 such that $s_w = s_1$. Then the cycle $\mathbb{C}_1^\pi$ starts at stage 1 and ends at stage $w-1$. Hence $t_1 = 1$ and $u_1 = w - 1$. If there is no stage $w \geq 2$ with $s_w = s_1$, then we skip stage 1 and we consider state s_2 instead. In that case $t_1 = 2$ and $u_1 = w - 1$, where w is the first stage after stage 2 such that $s_w = s_2$ etcetera.
For cycle $\mathbb{C}_l^\pi$ with $l \geq 2$, consider state $s_{u_{l-1}+1}$, the state that is visited at the first stage after cycle $\mathbb{C}_{l-1}^\pi$ has finished. Let w be the first stage after stage $u_{l-1} + 1$ such that $s_w = s_{u_{l-1}+1}$. Then the cycle $\mathbb{C}_l^\pi$ starts at stage $u_{l-1} + 1$ and ends at stage $w - 1$. Hence $t_l = s_{u_{l-1}+1}$ and $u_l = w - 1$. If there is no stage $w \geq u_{l-1} + 2$ with $s_w = s_{u_{l-1}+1}$, then instead of state $s_{u_{l-1}+1}$ consider state $s_{u_{l-1}+2}$ etcetera.
Since player 1 has at least one action that he plays an infinite number of times, the number of stages that is skipped in this construction is at most $m - 1$. Now consider the pure strategy $\hat{\pi}$ of player 1, leading to the sequence $(s_{t_1}, s_{t_1+1}, \ldots, s_{u_1}, s_{t_2}, s_{t_2+1}, \ldots, s_{u_2}, \ldots)$ of visited states, and a stationary optimal strategy y^* of player 2. We will now calculate an upper bound on the number of stages in which the expected payoffs of π and $\hat{\pi}$ against y^* are different. First of all notice that the number of skipped stages is bounded above by $m - 1$. Secondly, not taking the

skipped stages into account, the expected payoffs of π and $\hat{\pi}$ against y^* may differ only at those stages u_τ for which $t_{\tau+1} \neq u_\tau + 1$. These stages are exactly the stages that precede the skipped stages, and therefore this number is also bounded above by $m-1$. Consequently the total number of stages in which the expected payoff of π and $\hat{\pi}$ are unequal, is at most $2m-2$, which is finite and therefore does not influence the average reward: $\gamma_\xi(\pi, y^*) = \gamma_\xi(\hat{\pi}, y^*)$.

In lemma 2.2.17 the line "pure strategy within $car(a^{*\prime})$" should be interpreted as follows: In state s player 1 selects a pure action that is in $car(a^*)$.

Lemma 2.2.17 *For each pair $(a^{*\prime}, y^*)$ of optimal strategies in M_ξ any pure strategy within $car(a^{*\prime})$ is a best reply against y^*.*

Proof. Let $(a^{*\prime}, y^*)$ be a pair of optimal strategies. We will prove that against y^* the expected average payoff to player 1 during each cycle within $car(a^{*\prime})$ exactly equals $\varphi_\xi(a^{*\prime}) = v_\xi$.
Take an arbitrary cycle $\mathbb{C}(s_1, s_2, \ldots, s_l)$ within $car(a^{*\prime})$ and consider player 1's pure strategy $\mathbb{C}'(s_1, s_2, \ldots, s_l)$, which prescribes to play actions $(s_1, s_2, \ldots, s_l)$ repeatedly. We distinguish between 2 cases:

Case 1: $\gamma_\xi(\mathbb{C}'(s_1, s_2, \ldots, s_l), y^*) > \varphi_\xi(a^{*\prime}) = v_\xi$.

In this case y^* is not optimal, which is a contradiction.

Case 2: $\gamma_\xi(\mathbb{C}'(s_1, s_2, \ldots, s_l), y^*) < \varphi_\xi(a^{*\prime}) = v_\xi$.

Take an arbitrary stage t. Let $\mathfrak{p}$ be the probability that, given that player 1 plays $a^{*\prime}$, from stage t on, the cycle $\mathbb{C}(s_1, s_2, \ldots, s_l)$ appears. Then $\mathfrak{p} > 0$. Therefore the cycle $\mathbb{C}(s_1, s_2, \ldots, s_l)$ occurs with a strictly positive frequency and the expected average payoff in the corresponding stages is strictly lower than $\varphi_\xi(a^{*\prime})$. But then in the other stages the average payoff is strictly larger than $\varphi_\xi(a^{*\prime})$. This means that there also exists a cycle $\tilde{\mathbb{C}}$ with $\gamma_\xi(\tilde{\mathbb{C}}', y^*) > \varphi_\xi(a^{*\prime}) = v_\xi$, which, again, contradicts the optimality of y^*.

Consequently $\gamma_\xi(\mathbb{C}'(s_1, s_2, \ldots, s_l), y) = \varphi_\xi(a^{*\prime})$ for each cycle $\mathbb{C}'(s_1, s_2, \ldots, s_l)$ within $car(a^*)$ and hence each pure strategy within $car(a^*)$ is a best reply against y^*. ∎

2.3 2×2 - Games

In this section we consider repeated games with bonuses M_ξ, in which the size of the underlying matrix game M is 2×2. Theorem 2.2.15 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:

$$M = \begin{pmatrix} \mathfrak{a} & \mathfrak{b} \\ \mathfrak{c} & \mathfrak{d} \end{pmatrix} \in \mathbb{R}^{2\times 2} \tag{2.8}$$

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

$\mathfrak{a}+\xi$	$\mathfrak{b}+\xi$
1	1
$\mathfrak{c}$	$\mathfrak{d}$
2	2

state 1

$\mathfrak{a}$	$\mathfrak{b}$
1	1
$\mathfrak{c}+\xi$	$\mathfrak{d}+\xi$
2	2

state 2 (2.9)

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 states 1 and 2 the payoffs in rows 1 and 2 respectively are increased with the bonus ξ.

Notation 2.3.1 *Throughout section 2.3 in M the unique mixed optimal action for player 1 is $a^* = (a_1^*, a_2^*)$ with $a_1^* = \frac{\mathfrak{d}-\mathfrak{c}}{\mathfrak{a}-\mathfrak{b}+\mathfrak{d}-\mathfrak{c}}$ and $a_2^* = \frac{\mathfrak{a}-\mathfrak{b}}{\mathfrak{a}-\mathfrak{b}+\mathfrak{d}-\mathfrak{c}}$. Furthermore $v = \frac{\mathfrak{a}\mathfrak{d}-\mathfrak{b}\mathfrak{c}}{\mathfrak{a}-\mathfrak{b}+\mathfrak{d}-\mathfrak{c}}$ with*

$$\max\{\mathfrak{b},\mathfrak{c}\} < v < \min\{\mathfrak{a},\mathfrak{d}\}. \tag{2.10}$$

We have: $B(a^*) = \Delta^2$ and, according to (2.4), $B_\xi(a^{*\prime}) = B(a^*) \times B(a^*) = \Delta^2 \times \Delta^2$. This means that

$$\varphi_\xi(a^{*\prime}) = \gamma_\xi(a^{*\prime}, y) \text{ for all strategies } y \text{ of player 2.} \tag{2.11}$$

Theorem 2.3.2 presents a necessary and sufficient condition for $a^{*\prime}$ to be optimal in M_ξ.

Theorem 2.3.2 *The following two statements are equivalent:*

1. *$a^{*\prime}$ is optimal in M_ξ for player 1.*
2. *$\varphi_\xi(a^{*\prime}) \geq \max\{\mathfrak{b},\mathfrak{c}\} + \xi$.*

Proof. Notice first that, according to (2.11), for all (stationary) strategies y of player 2 we have: $\gamma_\xi(a^{*\prime}, y) = \varphi_\xi(a^{*\prime})$.

1 $\Rightarrow$ 2:

If $a^{*\prime}$ is optimal, then it will yield at least as much as $(1,0)'$ and $(0,1)'$, which in turn yield at least $\mathfrak{b}+\xi$ and $\mathfrak{c}+\xi$ respectively. Hence the result.

2 $\Rightarrow$ 1:

Let $y = (b^1, b^2)$ with

$$b_1^1 = \frac{v - 2\xi a_1^* a_2^* - \mathfrak{b}}{\mathfrak{a}-\mathfrak{b}},$$
$$b_2^1 = 1 - b_1^1,$$
$$b_1^2 = \frac{\mathfrak{d} + 2\xi a_1^* a_2^* - v}{\mathfrak{d}-\mathfrak{c}}$$

and

$$b_2^2 = 1 - b_1^2.$$

It can easily be shown that, if statement *2* holds, then $0 \leq b_1^1, b_1^2 \leq 1$. We show that $\gamma_\xi^s(\pi, y) = \varphi_\xi(a^{*\prime})$ 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

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

Some elementary calculations show that (cf. equation 2.6)

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

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

$$\begin{aligned}\frac{1}{2}(\mathfrak{c}b_1 + \mathfrak{d}(1-b_1)) + \frac{1}{2}(\mathfrak{a}b_2 + \mathfrak{b}(1-b_2)) &= v + \xi\cdot((a_1^*)^2 + (a_2^*)^2)\\ &= \varphi_\xi(a^{*\prime}),\end{aligned}$$

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 (1983)). This holds for a minimizing player 1 as well as for a maximizing player 1 and consequently $\gamma_\xi^s(\pi, y) = \varphi_\xi(a^{*\prime})$ for all strategies π of player 1 and each initial state s. This means that $a^{*\prime}$, guaranteeing a reward of at least $\varphi_\xi(a^{*\prime})$ to player 1, is optimal. ∎

Theorem 2.3.3 *Player 1 has a pure simple optimal strategy in M_ξ if and only if $\varphi_\xi(a^{*\prime}) \leq \max\{\mathfrak{b}, \mathfrak{c}\} + \xi$.*

Proof. The only if-part of the proof is trivial: Each of the pure stationary strategies of player 1 guarantees a reward of at most $\max\{\mathfrak{b}, \mathfrak{c}\} + \xi$, whereas $a^{*\prime}$ guarantees $\varphi_\xi(a^{*\prime})$, so if player 1 has a pure simple optimal strategy, then $\varphi_\xi(a^{*\prime}) \leq \max\{\mathfrak{b}, \mathfrak{c}\} + \xi$.

The if-part of the proof is divided into two parts: $\varphi_\xi(a^{*\prime}) = \max\{\mathfrak{b}, \mathfrak{c}\} + \xi$ and $\varphi_\xi(a^{*\prime}) < \max\{\mathfrak{b}, \mathfrak{c}\} + \xi$. Suppose first that $\varphi_\xi(a^{*\prime}) = \max\{\mathfrak{b}, \mathfrak{c}\} + \xi$. Then, according to theorem 2.3.2, the strategy $a^{*\prime}$ is optimal for player 1 with reward $\varphi_\xi(a^{*\prime}) = \max\{\mathfrak{b}, \mathfrak{c}\} + \xi$. But player 1 also has a pure simple strategy, namely either $(1,0)'$ or $(0,1)'$ that guarantees a reward of $\max\{\mathfrak{b}, \mathfrak{c}\} + \xi$, hence he has a pure simple optimal strategy.

Now suppose that $\varphi_\xi(a^{*\prime}) < \max\{\mathfrak{b}, \mathfrak{c}\} + \xi$. Then player 1 has a pure stationary strategy that guarantees a reward of at least $\max\{\mathfrak{b}, \mathfrak{c}\} + \xi > \varphi_\xi(a^{*\prime})$. Let $z^* = (z^{1*}, z^{2*})$ and $y^* = (b^{1*}, b^{2*})$ be stationary optimal strategies for players 1 and 2 respectively and suppose w.l.o.g. $z^{1*} \neq a^*$. Then in state 1 player 2 has a unique pure one-shot best reply against z^{1*}, which, according to (2.4), is the prescribed action by his stationary optimal strategy y^*. So there are two cases:

Case 1: $b_1^{1*} = 1$. Then $\gamma_\xi((1,0)', y^*) = \mathfrak{a} + \xi$, so $v_\xi \geq \mathfrak{a} + \xi$. But then $\mathfrak{a} + \xi \leq v_\xi < v_\xi + \xi$ and hence $\mathfrak{a} < v$, which contradicts (2.10). Hence case 1 does not occur.

Case 2: $b_1^{1*} = 0$. Then $\gamma_\xi((1,0)', y^*) = \mathfrak{b} + \xi$, so $v_\xi \geq \mathfrak{b} + \xi$. Now if $v_\xi = \mathfrak{b} + \xi$, then $(1,0)'$ is optimal for player 1, since the simple strategy $(1,0)'$ guarantees player 1 a reward of at least $\mathfrak{b} + \xi$. Suppose that $v_\xi > \mathfrak{b} + \xi$ and let f be the relative frequency of stages, in which player 1 plays action 1 in state 1. In each of those stages he receives a payoff of $\mathfrak{b} + \xi$. Suppose $f > 0$ and let $\hat{v}$ denote the average payoff to player 1 during the other stages. Then $v_\xi = f \cdot (\mathfrak{b} + \xi) + (1 - f) \cdot \hat{v}$. Consequently $\hat{v} > v_\xi$ and player 1 can increase his (average) reward against y^*, which contradicts the optimality of z^*. Therefore we must have $f = 0$, which means that $z_1^{1*} = 0$ or $z_1^{2*} = 0$. If $z_1^{2*} = 0$, then $(0,1)'$ is optimal for player 1, whereas if $z_1^{2*} \neq 0$ and $z_1^{1*} = 0$, then the one-shot best reply of player 2 in state 1 would be $b_1^{1*} = 1$, contradicting the fact that $b_1^{1*} = 0$. So either $(1,0)'$ or $(0,1)'$, both of which are pure simple strategies, is optimal. ■

The following corollary follows directly from theorems 2.3.2 and 2.3.3:

Corollary 2.3.4 *For each $M \in \mathbb{R}^{2\times 2}$ and $\xi \geq 0$ player 1 has a simple optimal strategy in M_ξ.*

In many 2×2 matrix games M one can easily see if player 1 in the corresponding stochastic game M_ξ has a simple optimal strategy. Firstly, if player 1 has an optimal pure action in M, then, according to theorem 2.2.15, he has an optimal simple pure strategy in M_ξ. If he has an optimal mixed action a^* in M, then theorems 2.3.5 and 2.3.6 provide respectively a necessary and a sufficient condition concerning the optimality of $a^{*\prime}$ in M_ξ. These conditions are easy in the sense that the calculations one has to make in order to check out, if they are satisfied, can easily be done without a calculator.

Theorem 2.3.5 *If $\frac{1}{2}(\mathfrak{a} + \mathfrak{d}) < \max\{\mathfrak{b}, \mathfrak{c}\} + \xi$, then $a^{*\prime}$ is not an optimal strategy for player 1 in M_ξ.*

Proof. Let y^* be a stationary optimal strategy of player 2. Suppose that $\frac{1}{2}(\mathfrak{a} + \mathfrak{d}) < \max\{\mathfrak{b}, \mathfrak{c}\} + \xi$. Suppose w.l.o.g. that $\mathfrak{b} \geq \mathfrak{c}$ and consider the pure simple strategy $(1,0)'$ and the pure stationary strategy $((0,1),(1,0))$, both of which are in $car\,(a^{*\prime})$. Then $\gamma_\xi\left((1,0)', y^*\right) \geq \mathfrak{b} + \xi = \max\{\mathfrak{b}, \mathfrak{c}\} + \xi > \frac{1}{2}(\mathfrak{a} + \mathfrak{d}) \geq \gamma_\xi^s(((0,1),(1,0)), y^*)$ for each s and hence, according to lemma 2.2.17, $a^{*\prime}$ cannot be optimal. ■

Theorem 2.3.6 *If $\mathfrak{a} \geq \mathfrak{c} + 2\xi$ and $\mathfrak{d} \geq \mathfrak{b} + 2\xi$, then $a^{*\prime}$ is an optimal strategy for player 1.*

Proof. According to theorem 2.3.2 it suffices to prove that if $\mathfrak{a} \geq \mathfrak{c} + 2\xi$ and $\mathfrak{d} \geq \mathfrak{b} + 2\xi$, then $\varphi_\xi(a^{*\prime}) \geq \max\{\mathfrak{b}, \mathfrak{c}\} + \xi$. Suppose w.l.o.g. that $\mathfrak{b} \geq \mathfrak{c}$. Furthermore suppose by means of contradiction that $\varphi_\xi(a^{*\prime}) - \xi < \max\{\mathfrak{b}, \mathfrak{c}\}$. Since

$$\varphi_\xi(a^{*\prime}) - \xi = v + \xi \cdot ((a_1^*)^2 + (a_2^*)^2) - \xi \cdot (a_1^* + a_2^*)^2 = v - 2\xi \cdot a_1^* a_2^*$$

we then have:

$$\xi > \frac{v - \mathfrak{b}}{2a_1^* a_2^*} = \frac{(\mathfrak{a} - \mathfrak{b})(\mathfrak{d} - \mathfrak{b})}{\mathfrak{a} - \mathfrak{b} + \mathfrak{d} - \mathfrak{c}} \frac{1}{2a_1^* a_2^*} = \frac{1}{2} \frac{\mathfrak{d} - \mathfrak{b}}{\mathfrak{d} - \mathfrak{c}} (\mathfrak{a} - \mathfrak{b} + \mathfrak{d} - \mathfrak{c}),$$

so

$$\begin{aligned}\mathfrak{b}+2\xi &> \mathfrak{b}+\frac{\mathfrak{d}-\mathfrak{b}}{\mathfrak{d}-\mathfrak{c}}(\mathfrak{a}-\mathfrak{b}+\mathfrak{d}-\mathfrak{c})\\ &= \frac{\mathfrak{d}(\mathfrak{d}-\mathfrak{c})+(\mathfrak{d}-\mathfrak{b})(\mathfrak{a}-\mathfrak{b})}{\mathfrak{d}-\mathfrak{c}}\\ &= \mathfrak{d}+\frac{(\mathfrak{d}-\mathfrak{b})(\mathfrak{a}-\mathfrak{b})}{\mathfrak{d}-\mathfrak{c}}\\ &> \mathfrak{d},\end{aligned}$$

which is a contradiction. ∎

The following example shows that the bounds in theorem 2.3.6 are sharp in the sense that for an arbitrarily small $\varepsilon > 0$ it might not be sufficient for $a^{*\prime}$ to be optimal, if $\mathfrak{a} = \mathfrak{c} + (2-\varepsilon)\cdot\xi$.

Example 2.1

Take $M = \begin{pmatrix} 2-\varepsilon & -2 \\ 0 & \frac{\varepsilon}{2} \end{pmatrix}$ with $a^* = (\frac{\varepsilon}{8-\varepsilon}, \frac{8}{8-\varepsilon})$ and $\xi = 1$. Then

$M_1 =$

$3-\varepsilon$ 1	-1 1
0 2	$\frac{\varepsilon}{2}$ 2

state 1

$2-\varepsilon$ 1	-2 1
1 2	$1+\frac{\varepsilon}{2}$ 2

state 2

.

Now $\gamma_1(a^{*\prime}, y) = \frac{16-4\varepsilon-\frac{5}{4}\varepsilon^2+\frac{1}{4}\varepsilon^3}{16-4\varepsilon+\frac{1}{4}\varepsilon^2} < 1 \leq \gamma_1((0,1)', y)$ for all strategies y of player 2 and hence $a^{*\prime}$ is not optimal. The fact that $a^{*\prime}$ is not optimal is actually already implied by theorems 2.3.2 and 2.3.5. □

2.4 $2 \times n$ - 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 2.2.15.

Notation 2.4.1 *Recall (cf. (2.1)) that the (i,j)-th entry of M is denoted by m_{ij}. Throughout section 2.4 we suppose without loss of generalization that in M the mixed action*

$$a^* = (a_1^*, a_2^*)$$

with

$$a_1^* = \frac{m_{2j_2} - m_{2j_1}}{m_{1j_1} - m_{1j_2} - m_{2j_1} + m_{2j_2}}$$

and

$$a_2^* = \frac{m_{1j_1} - m_{1j_2}}{m_{1j_1} - m_{1j_2} - m_{2j_1} + m_{2j_2}}$$

is optimal for player 1. Here j_1 and j_2 are 2 pure best replies against a^* in M such that

$$m_{1j_1} - m_{2j_1} = \max_{j \in B(a^*)} (m_{1j} - m_{2j}) \tag{2.12}$$

and

$$m_{1j_2} - m_{2j_2} = \min_{j \in B(a^*)} (m_{1j} - m_{2j}). \tag{2.13}$$

Furthermore

$$v = \frac{m_{1j_1} \cdot m_{2j_2} - m_{1j_2} \cdot m_{2j_1}}{m_{1j_1} - m_{1j_2} - m_{2j_1} + m_{2j_2}}.$$

Notice that $m_{1j_1} - m_{2j_1} > 0$ and $m_{1j_2} - m_{2j_2} < 0$, since otherwise there would be a dominant row and player 1 would have a pure optimal action in M.

Now consider the 2-state stochastic game M_ξ. According to (2.4), for the simple strategy $a^{*\prime}$ of player 1 we have $B_\xi(a^{*\prime}) = B(a^*) \times B(a^*)$. Take a stationary strategy $y \in B_\xi(a^{*\prime})$. Then, due to (2.6) and (2.5):

$$\varphi_\xi\left(a^{*\prime}\right) = \gamma_\xi(a^{*\prime}, y) = v + \xi \cdot ((a_1^*)^2 + (a_2^*)^2).$$

After this preliminary work we now focus on generalizing the theorems in the previous section. The first theorem in this section is the generalization of theorem 2.3.2.

Theorem 2.4.2 *The following two statements are equivalent:*

1. *$a^{*\prime}$ is optimal in M_ξ for player 1.*
2. *$\varphi_\xi\left(a^{*\prime}\right) \geq \max\{m_{1j_2}, m_{2j_1}\} + \xi$.*

Proof. To prove that statement *2* is a consequence of statement *1*, we show that, if $\varphi_\xi(a^{*\prime}) < \max\{m_{1j_2}, m_{2j_1}\} + \xi$, then for each $y \in B_\xi(a^{*\prime})$ the tuple $(a^{*\prime}, y)$ is not a pair of optimal strategies. Suppose $a^{*\prime}$ is optimal and suppose w.l.o.g. $\varphi_\xi(a^{*\prime}) < m_{2j_1} + \xi$. Take $y \in B_\xi(a^{*\prime})$. A consequence of (2.12) is that for each pure best reply j against a^* we have: $m_{2j} \geq m_{2j_1}$. But then $\gamma_\xi((0,1)', y) \geq m_{2j_1} + \xi > \varphi_\xi(a^{*\prime}) = \gamma_\xi(a^{*\prime}, y)$. Hence $a^{*\prime}$ is not a best reply against y and therefore not optimal.

The proof of *2* $\Rightarrow$ *1* is equal to the proof of theorem 2.3.2 with a few notational adjustments: $\mathfrak{a}$, $\mathfrak{b}$, $\mathfrak{c}$ and $\mathfrak{d}$ are, as in notation 2.4.1, replaced by m_{1j_1}, m_{1j_2}, m_{2j_1} and m_{2j_2} respectively and player 2's strategy $y = \left(b^1, b^2\right)$ with $b^1, b^2 \in \Delta^n$ consists of the following entries:

$$b^1_{j_1} = \frac{v - 2\xi a_1^* a_2^* - m_{1j_2}}{m_{1j_1} - m_{1j_2}}, b^1_{j_2} = 1 - b^1_{j_1},$$
$$b^2_{j_1} = \frac{m_{2j_2} + 2\xi a_1^* a_2^* - v}{m_{2j_2} - m_{2j_1}}, b^2_{j_2} = 1 - b^2_{j_1} \text{ and } b^i_j = 0 \text{ for all other } i, j.$$

■

Example 2.2 below shows that theorem 2.3.3 can not be generalized to $2 \times n$-games. Theorem 2.4.3 is the generalization of theorem 2.3.5.

Theorem 2.4.3 *If $a^{*\prime}$ is an optimal stationary strategy for player 1 in M_ξ, then $\frac{1}{2}(m_{1j_1} + m_{2j_2}) \geq \max\{m_{1j_2}, m_{2j_1}\} + \xi$.*

Proof. Identical to the proof of theorem 2.3.5. ■

The following theorem is the generalization of theorem 2.3.6.

Theorem 2.4.4 *If for actions j_1 and j_2 it holds that $m_{1j_1} \geq m_{2j_1} + 2\xi$ and $m_{2j_2} \geq m_{1j_2} + 2\xi$, then $a^{*\prime}$ is an optimal stationary strategy for player 1 in M_ξ.*

Proof. Let Υ be the set of stationary strategies of player 2, of which the carriers are within $\{j_1, j_2\}$:

$$\Upsilon = \{y = (b_1, b_2) \,|\, car(b_1) \subset \{j_1, j_2\} \text{ and } car(b_2) \subset \{j_1, j_2\}\}$$

Then, according to theorem 2.3.6, player 2 has a stationary strategy $y \in \Upsilon$ such that $\gamma_\xi^s(\pi, y) \leq \varphi_\xi(a^{*\prime})$ for all strategies π of player 1 and each initial state s. Since $a^{*\prime}$ guarantees $\varphi_\xi(a^{*\prime})$, it is optimal. ■

We continue the analysis by providing a necessary and a sufficient condition for player 1 to possess a pure simple strategy. These conditions are closely related to the necessary and sufficient conditions with respect to the optimality of $a^{*\prime}$ (theorems 2.4.3 and 2.4.4). We need some notations first. Without loss of generality we suppose that M looks as follows (notice that all waekly and strongly dominated columns are removed):

$$M = \begin{pmatrix} \check{\mathfrak{a}} & > & \ldots & > & \ldots & > & \check{\mathfrak{b}} \\ \check{\mathfrak{c}} & < & \ldots & < & \ldots & < & \check{\mathfrak{d}} \end{pmatrix}.$$

Theorem 2.4.5 *If $\frac{1}{2}(\check{\mathfrak{a}} + \check{\mathfrak{d}}) \leq \max\left\{\check{\mathfrak{b}}, \check{\mathfrak{c}}\right\} + \xi$, then player 1 has a pure simple optimal strategy in M_ξ.*

Proof. Let y be player 2's stationary strategy that prescribes to play action $\check{j}_1$ with probability 1 in state 1 and to play action $\check{j}_2$ with probability 1 in state 2. Some straightforward calculations show that

$$\gamma_\xi(x, y) \leq \max\left\{\check{\mathfrak{b}}, \check{\mathfrak{c}}\right\} + \xi$$

for all player 1's stationary strategies x. Hence player 2 has a strategy that guarantees that γ_ξ does not exceed $\max\left\{\check{\mathfrak{b}}, \check{\mathfrak{c}}\right\} + \xi$. Since player 1 has a pure simple strategy that guarantees a reward of $\max\left\{\check{\mathfrak{b}}, \check{\mathfrak{c}}\right\} + \xi$, this strategy must be optimal. ■

Theorem 2.4.6 *If $(1, 0)'$ is a pure simple optimal strategy in M_ξ for player 1, then $\check{\mathfrak{d}} \leq \check{\mathfrak{b}} + 2\xi$.*

Proof. Suppose that $\check{\mathfrak{b}} \geq \check{\mathfrak{c}}$ and suppose by means of contradiction that $\check{\mathfrak{d}} > \check{\mathfrak{b}}+2\xi$. By assumption the pure simple strategy $(1,0)'$ is optimal and $v_\xi = \check{\mathfrak{b}} + \xi$. Let $y^* = (b^{1*}, b^{2*})$ be a stationary optimal strategy for player 2. Then b^{1*} is the action that plays the utmost right column with probability 1. Now consider the pure stationary strategy $x = ((0,1),(1,0))$ for player 1. We have:

$$\gamma_\xi(x, y^*) = \frac{1}{2}\check{\mathfrak{d}} + \frac{1}{2}e_1^T M b^{2*},$$

where $e_1^T M b^{2*} \geq \check{\mathfrak{b}}$, since $e_1^T M b^{2*}$ is a convex combination of the numbers in the first row of M. Consequently:

$$\begin{aligned}\gamma_\xi(x, y^*) &\geq \frac{1}{2}\check{\mathfrak{d}} + \frac{1}{2}\check{\mathfrak{b}} \\ &> \frac{1}{2}(\check{\mathfrak{b}} + 2\xi) + \frac{1}{2}\check{\mathfrak{b}} \\ &= \check{\mathfrak{b}} + \xi\end{aligned}$$

which contradicts the optimality of $(1,0)'$. Hence $\check{\mathfrak{d}} \leq \check{\mathfrak{b}} + 2\xi$. ■

Analogously:

Theorem 2.4.7 *If $(0,1)'$ is a pure simple optimal strategy in M_ξ for player 1, then $\check{\mathfrak{a}} \leq \check{\mathfrak{c}} + 2\xi$.*

We will now present a necessary and sufficient condition for player 1 to possess a pure simple optimal strategy in M_ξ. For this purpose suppose without loss of generality that $\check{\mathfrak{b}} \geq \check{\mathfrak{c}}$ and consider the following matrix:

$$\hat{M} = \begin{pmatrix} \frac{1}{2}M_1 + 1^T(\frac{1}{2}\check{\mathfrak{d}} - \xi) \\ M_2 \end{pmatrix},$$

where M_i is the i^{th} row of M and 1^T is a row vector of the appropriate length, consisting of ones.

Theorem 2.4.8 *Given that $\check{\mathfrak{b}} \geq \check{\mathfrak{c}}$, player 1's pure simple strategy $(1,0)'$ is optimal if and only if the value of the matrix game $\hat{M}$ does not exceed $\check{\mathfrak{b}}$.*

Proof. Let $\check{\mathfrak{b}} \geq \check{\mathfrak{c}}$. Notice first that the value of the matrix game $\hat{M}$ does not exceed $\check{\mathfrak{b}}$ if and only if player 2 has a stationary optimal strategy $y^* = (b^{1*}, b^{2*})$ with b^{1*} the (pure) action that plays the utmost right column with probability 1 and b^{2*} such that

$$\hat{M} b^{2*} \leq \begin{pmatrix} \check{\mathfrak{b}} \\ \check{\mathfrak{b}} \end{pmatrix}.$$

We will first show that $(1,0)'$ is optimal for player 1 if and only if player 2 has a stationary optimal strategy $y^* = (b^{1*}, b^{2*})$, where b^{1*} is the action that plays the utmost right column with probability 1 and b^{2*} is such that the following inequalities hold:

1. $\frac{1}{2}(e_1 M b^{2*} + \check{\mathfrak{d}}) \leq \check{\mathfrak{b}} + \xi$
2. $e_2 M b^{2*} \leq \check{\mathfrak{b}}$.

Let $(1,0)'$ be optimal for player 1 and let $y^* = (b^{1*}, b^{2*})$ be a stationary optimal strategy for player 2. Since y^* is a best reply against $(1,0)'$, by 2.4 we have that b^{1*} the action that plays the utmost right column with probability 1.

"$\Rightarrow$"

Suppose by means of contradiction that inequality *1* does not hold. Then

$$\begin{aligned}\gamma_\xi\left(((0,1),(1,0)),y^*\right) &= \frac{1}{2}(e_1 M b^{2*} + \check{\mathfrak{d}}) \\ &> \check{\mathfrak{b}} + \xi \\ &= \gamma\left((1,0)', y^*\right).\end{aligned}$$

Consequently $(1,0)'$ is not a best reply against y^* and hence it can not be optimal, contradicting the assumption. If inequality 2 does not hold, then we find that

$$\begin{aligned}\gamma\left((0,1)', y^*\right) &= e_2 M b^{2*} + \xi \\ &> \check{\mathfrak{b}} + \xi \\ &= \gamma\left((1,0)', y^*\right)\end{aligned}$$

and again $(1,0)'$ can not be optimal.

"$\Leftarrow$"

Let $y^* = (b^{1*}, b^{2*})$ with b^{1*} the action that plays the utmost right column with probability 1 and b^{2*} such that $\frac{1}{2}(e_1 M b^{2*} + \check{\mathfrak{d}}) \leq \check{\mathfrak{b}} + \xi$ and $e_2 M b^{2*} \leq \check{\mathfrak{b}}$. Then $(1,0)'$ is a best reply against the optimal strategy y^* and hence it is optimal.

This completes the proof of the statement above. Writing inequalities *1* and *2* in matrix-vector notation, yields the statement of the theorem. ■

If $\check{\mathfrak{b}} \leq \check{\mathfrak{c}}$, then for the pure simple strategy $(0,1)'$ we can obtain similar results with the aid of the matrix

$$\bar{M} = \begin{pmatrix} \frac{1}{2}M_1 + 1^T(\frac{1}{2}\check{\mathfrak{d}} - \xi) \\ M_2 \end{pmatrix}.$$

This completes the analysis of necessasy and or sufficient conditions for simple optimal strategies to exist in $2 \times n$ - games. We conclude this section with two examples showing that in the $2 \times n$ - case player 1 does not necessarily possess a simple optimal strategy.

Example 2.2. Take $M = \begin{pmatrix} 3 & 0 & 1 \\ 0 & 3 & 1 \end{pmatrix}$ and let $\xi \in \langle 0, 1]$. Then the set of optimal mixed actions of player 1 in M is

$$A^* = \{a^* = (a_1^*, a_2^*) \mid \frac{1}{3} \leq a_1^* \leq \frac{2}{3}\,,\ a_1^* + a_2^* = 1\}.$$

Furthermore

$M_\xi =$

$3+\xi$ 1	ξ 1	$1+\xi$ 1
0 2	3 2	1 2

state 1

3 1	0 1	1 1
ξ 2	$3+\xi$ 2	$1+\xi$ 2

state 2

and for each $a^* \in A^*$ we have: $\varphi_\xi(a^{*\prime}) = 1 + \xi \cdot ((a_1^*)^2 + (a_2^*)^2)$, which is maximal for $a^* \in \{(\frac{1}{3}, \frac{2}{3}), (\frac{2}{3}, \frac{1}{3})\}$, guaranteeing a reward of $1 + \frac{5}{9} \cdot \xi$. It can easily be calculated that for each $a \notin A^*$ we have $\varphi_\xi(a') < 1 + \frac{5}{9} \cdot \xi$. However, the (optimal) stationary strategy $x = ((\frac{2}{3}, \frac{1}{3}), (\frac{1}{3}, \frac{2}{3}))$ guarantees a reward of $\varphi_\xi(x) = 1 + \frac{2}{3} \cdot \xi$ and hence player 1 does not have a simple optimal strategy. Example 2.2 also shows that, as long as the number of actions for player 2 is at least 3, not even for an arbitrarily small $\xi > 0$ it can be guaranteed that a simple optimal strategy for player 1 exists (cf. 2.3.4). □

Notice that in example 2.2 the stationary optimal strategy $((\frac{2}{3}, \frac{1}{3}), (\frac{1}{3}, \frac{2}{3}))$ in both states prescribes a mixed action in A^*. Example 2.3 shows that this is in general not necessarily the case.

Example 2.3 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 $(0.2, 0.8)$,

$M_1 =$

11 1	2.99 1	1.39 1
0 2	0.89 2	1.29 2

state 1

10 1	1.99 1	0.39 1
1 2	1.89 2	2.29 2

state 2

and $\varphi_\xi(a^{*\prime}) = 1.11 + 0.04 + 0.64 = 1.79$. 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 $\varphi_\xi(x) = \frac{16.51}{9} > 1.79$. □

2.5 $m \times n$ - 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. With the aid of an example we show that even in the $m \times 2$ - case player 1 does not necessarily have a simple optimal strategy.

Example 2.4

Take $M = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -\alpha & \alpha \end{pmatrix}$ with $\alpha > 2$ and $\xi = \frac{1}{\alpha}$. Then $M_\xi =$

$1+\frac{1}{\alpha}$ 1	$\frac{1}{\alpha}$ 1
0 2	1 2
$-\alpha$ 3	α 3

state 1

1 1	0 1
$\frac{1}{\alpha}$ 2	$1+\frac{1}{\alpha}$ 2
$-\alpha$ 3	α 3

1 1	0 1
0 2	1 2
$-\alpha+\frac{1}{\alpha}$ 3	$\alpha+\frac{1}{\alpha}$ 3

state 3

The optimal stationary strategy for player 1 in M_ξ is

$$x^* = ((\frac{2\alpha}{2\alpha+1}, 0, \frac{1}{2\alpha+1}), (\frac{1}{2}, \frac{1}{2}, 0), (\frac{1}{2}, \frac{1}{2}, 0))$$

providing a reward of at least

$$\varphi_\xi(x^*) = \tfrac{3+\alpha+\frac{1}{2\alpha}}{2\alpha+3} = \tfrac{8\alpha^4+32\alpha^3+30\alpha^2+10\alpha+1}{2\alpha(2\alpha+1)^2(2\alpha+3)}.$$

Notice that the simple strategies $(\frac{1}{2}, \frac{1}{2}, 0)'$ and $(\frac{2\alpha}{2\alpha+1}, 0, \frac{1}{2\alpha+1})'$ yield

$$\varphi_\xi((\tfrac{1}{2}, \tfrac{1}{2}, 0)') = \tfrac{1}{2} + \tfrac{1}{2\alpha} = \tfrac{8\alpha^4+28\alpha^3+34\alpha^2+17\alpha+3}{2\alpha(2\alpha+1)^2(2\alpha+3)}$$

and

$$\varphi_\xi((\tfrac{2\alpha}{2\alpha+1}, 0, \tfrac{1}{2\alpha+1})') = \tfrac{2\alpha^2+5\alpha+\frac{1}{\alpha}}{(2\alpha+1)^2} = \tfrac{8\alpha^4+32\alpha^3+30\alpha^2+4\alpha+6}{2\alpha(2\alpha+1)^2(2\alpha+3)}$$

respectively, both of which are smaller than $\varphi_\xi(x^*)$ for $\alpha > 2$. Other simple strategies yield rewards that are obviously lower than those of at least one of these two simple strategies. □

This example also shows that there is no number M such that in each column a difference between the payoffs of at least $M \cdot \xi$ is a sufficient condition for a simple optimal strategy to exist (cf. theorems 2.3.6 and 2.4.4).

However, there do exist sufficient conditions on M and ξ for simple optimal strategies in M_ξ to exist as will be shown in theorems 2.5.1 and 2.5.2. Theorem 2.5.1 discusses games with a small bonus, whereas in theorem 2.5.2 games with a large bonus are analyzed.

Theorem 2.5.1 *If player 1 has a unique optimal mixed action a^* in M, then there exists a number $\hat{\xi} > 0$ such that for all $\xi \in [0, \hat{\xi}]$ the simple strategy $a^{*\prime}$ is optimal in M_ξ.*

Proof. Let $a^{*\prime}$ be the unique optimal mixed action for player 1 in M. Then there exists an $\varepsilon > 0$ such that

$$\forall\, a \in A\, \exists\, b \in B : a^T M b \leq v - \varepsilon \,||\, a - a^*||_1 \,.$$

Now take

$$\hat{\xi} := \frac{1}{2}\varepsilon,$$

and let $\xi \leq \hat{\xi}$. Consider a stationary strategy $x = (a^1, a^2, \ldots, a^m)$ for player 1 and suppose without loss of generality that the set of states that are visited with a strictly positive frequency given x, is a subset of $car(a^*)$. Let $S' \subset car(a^*)$ be the set of states for which $a^s \neq a^*$. Some elementary Markov-Chain theory shows that the fraction of the stages at which a state in S' is visited, is strictly positive. Consequently we can fix

$$\delta := \sum_{s \in S'} q(s \,|\, x) \cdot ||a^s - a^*||_1 > 0,$$

where $q(s \,|\, x)$ is the fraction of stages that state s is visited given the stationary strategy x, and we have:

$$\varphi_0(x) \leq v - \varepsilon\delta.$$

Furthermore, some calculations show that the increase in the fraction of stages, in which player 1 receives the bonus, is at most 2δ. Consequently

$$\begin{aligned}
\varphi_\xi(x) &\leq v - \varepsilon\delta + \xi \cdot \left(\sum_{s \in S} (a_s^*)^2 + 2\delta \right) \\
&= v + \xi \cdot \sum_{s \in S} (a_s^*)^2 - \varepsilon\delta + 2\delta\xi \\
&= \varphi_\xi(a^{*\prime}) - \varepsilon\delta + 2\delta\xi \\
&\leq \varphi_\xi(a^{*\prime})
\end{aligned}$$

and hence $a^{*\prime}$ is optimal in M_ξ for $\xi \in [0, \hat{\xi}]$. ■

Theorem 2.5.2 discusses the case of a bonus that is large in comparison with the differences between the payoffs in M.

Theorem 2.5.2 *For each matrix game M there exists a $\check{\xi} \geq 0$ such that for all $\xi \geq \check{\xi}$ player 1 has a pure simple optimal strategy in M_ξ.*

Proof. Let

$$\bar{m} := \max_{i \in I, j \in J} m_{ij} \text{ and } \dot{m} := \min_{i \in I, j \in J} m_{ij}.$$

Now take $\check{\xi} := \bar{m} - \dot{m}$,let $\xi \geq \check{\xi}$ and consider a pure stationary strategy $y = (j^1, j^2, \ldots, j^m)$ for player 2, where $j^s \in \arg\min_{j \in J} m_{sj}$.Then obviously for all strategies π of player 1 we have:

$$\gamma(\pi, y) \leq \max_{i \in I} \min_{j \in J} m_{ij} + \xi.$$

Let i' be the pure simple strategy guaranteeing the highest reward of all pure simple strategies to player 1 in M_ξ. Then

$$\varphi_\xi(i') = \max_{i \in I} \min_{j \in J} m_{ij} + \xi$$

and hence i' is optimal. ■

Theorem 2.5.3 demonstrates that the number of simple optimal strategies in a repeated game with bonus ξ is always finite as long as $\xi > 0$. In fact, as has been shown in previous examples, simple optimal strategies might not even exist at all. Theorem 2.5.3 shows that per carrier the maximum number of simple optimal strategies that might exist, is 1.

Theorem 2.5.3 *If a'_1 and a'_2 are simple optimal strategies of player 1 in M_ξ with $car(a_1) = car(a_2)$, then $a_1 = a_2$.*

Proof. Let a'_1 and a'_2 be 2 simple optimal strategies of player 1 in M_ξ with $car(a_1) = car(a_2)$ and suppose by means of contradiction that $a_1 \neq a_2$. We know (cf. (2.7)):

$$v_\xi = \varphi_0(a'_1) + \xi \cdot \sum_{s \in I} a^2_{1s} = \varphi_0(a'_2) + \xi \cdot \sum_{s \in I} a^2_{2s}.$$

Let I_1 be the following subset of the action set $I = \{1, \ldots, m\}$ of player 1:

$$\text{Action } i \in I \text{ is in } I_1 \text{ if and only if } a_{1i} \geq a_{2i}. \tag{2.14}$$

Let I_2 consist of the other actions of player 1 and consider the stationary strategy x of player 1 that prescribes to play the mixed action a_1 whenever a state $s \in I_1$ is visited and to play a_2 otherwise. Notice that $\varphi_\xi(x)$ is state independent (cf. notation 2.2.12). Let z_1 be the probability that a state $s \in I_1$ is visited at stage $t+1$, given that at stage t a state $s \in I_2$ was visited, and let z_2 be the probability that a state $s \in I_2$ is visited at stage $t+1$, given that at stage t a state $s \in I_1$ was visited. Then the long-run frequency of visits to states in I_1 is $\frac{z_1}{z_1+z_2}$ and the long-run frequency of visits to state s is

$$\frac{z_1}{z_1 + z_2} \cdot a_{1s} + \frac{z_2}{z_1 + z_2} \cdot a_{2s}.$$

Notice that $\sum_{s \in I_1} a_{1s} = 1 - z_2$ and $\sum_{s \in I_2} a_{2s} = 1 - z_1$.

Furthermore, if state $s \in I_1$ is visited, then player 1 plays a_1 and his minimal expected stage payoff is equal to $\varphi_0(a'_1) + \xi \cdot a_{1s}$; if state $s \in I_2$ is visited, then his minimal

expected stage payoff is $\varphi_0(a_2') + \xi \cdot a_{2s}$. Hence:

$$\begin{aligned}
\varphi_\xi(x) &= \sum_{s\in I_1}(\frac{z_1}{z_1+z_2}\cdot a_{1s} + \frac{z_2}{z_1+z_2}\cdot a_{2s})\cdot(\varphi_0(a_1') + \xi\cdot a_{1s}) \\
&+ \sum_{s\in I_2}(\frac{z_1}{z_1+z_2}\cdot a_{1s} + \frac{z_2}{z_1+z_2}\cdot a_{2s})\cdot(\varphi_0(a_2') + \xi\cdot a_{2s}) \\
&= (\frac{z_1}{z_1+z_2}\cdot(1-z_2) + \frac{z_2}{z_1+z_2}\cdot z_1)\cdot\varphi_0(a_1') \\
&+ (\frac{z_1}{z_1+z_2}\cdot z_2 + \frac{z_2}{z_1+z_2}\cdot(1-z_1))\cdot\varphi_0(a_2') \\
&+ \xi\cdot\left(\sum_{s\in I_1}\frac{z_1a_{1s}^2 + z_2a_{1s}a_{2s}}{z_1+z_2} + \sum_{s\in I_2}\frac{z_1a_{1s}a_{2s} + z_2a_{2s}^2}{z_1+z_2}\right) \\
&> \frac{z_1}{z_1+z_2}\cdot\varphi_0(a_1') + \frac{z_2}{z_1+z_2}\cdot\varphi_0(a_2') \\
&+ \xi\cdot\left(\sum_{s\in I_1}\frac{z_1a_{1s}^2 + z_2a_{2s}^2}{z_1+z_2} + \sum_{s\in I_2}\frac{z_1a_{1s}^2 + z_2a_{2s}^2}{z_1+z_2}\right) \\
&= \frac{z_1}{z_1+z_2}(\varphi_0(a_1') + \xi\cdot\sum_{s\in I}a_{1s}^2) + \frac{z_2}{z_1+z_2}(\varphi_0(a_2') + \xi\cdot\sum_{s\in I}a_{2s}^2) \\
&= v_\xi
\end{aligned}$$

contradicting the optimality of a_1' and a_2'. Hence $a_1 = a_2$. ■

Notation 2.5.4 *The unit vector with the 1 at position i is denoted e_i. The length of the unit vector will be clear from the context. Furthermore e_i can also denote a pure action.*

Example 2.5 below shows that indeed it is possible that for each carrier a simple optimal strategy exists.

Example 2.5

Take $M = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $\xi = 1$. Then

$M_\xi =$

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

state 1

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

state 2

and in this game player 1 has 3 simple optimal strategies: Each of the strategies $(1,0)'$, $(\frac{1}{2},\frac{1}{2})'$ and $(0,1)'$ guarantees a reward of 1 to player 1. The (unique) stationary optimal strategy for player 2 is $((0,1),(1,0))$.

This holds in more general cases: Let M be the $m \times m$ identity matrix and let $\xi = \frac{1}{m-1}$. Then for each unit vector e_s (cf. notation 2.5.4) we have:

$$\varphi_\xi(e_s') = \varphi_\xi((\frac{1}{m},\frac{1}{m},\ldots,\frac{1}{m})') = \frac{1}{m-1}$$

and each of these simple strategies is optimal. □

The rest of this section is devoted to theorem 2.5.8 that, under the assumption that player 1 has a simple optimal strategy $a^{*\prime}$, characterizes the set of optimal stationary strategies of player 2. This theorem is a generalization of the Shapley-Snow theorem for matrix games (cf. Shapley & Snow (1950)) to the class of repeated games with bonuses.
We need some notations:

Notation 2.5.5 *Let $a^{*\prime}$ be a simple optimal strategy for player 1 in M_ξ with $car(a^*) = \{i_1, i_2, \ldots, i_p\}, i_1 < i_2 < \ldots < i_p$. Then the subvector $\bar{a}^*$ of a^* consists of the elements of a^* that are in $car(a^*)$: $\bar{a}^* = (a^*_{i_1}, a^*_{i_2}, \ldots, a^*_{i_p})$.*

Notice that

$$\sum_{i=1}^{p} \bar{a}^*_i = 1 \text{ and that } \sum_{i=1}^{p} \left(\bar{a}^*_i\right)^2 = \sum_{i=1}^{m} \left(a^*_i\right)^2. \tag{2.15}$$

Lemma 2.5.6 *Let $\check{y} = (\check{b}^1, \check{b}^2, \ldots, \check{b}^m)$ and $\hat{y} = (\hat{b}^1, \hat{b}^2, \ldots, \hat{b}^m)$ be 2 stationary optimal strategies for player 2. Then for all actions $i, j \in car(a^*)$ we have: $e_i M \check{b}^j = e_i M \hat{b}^j$.*

Proof. We start the proof by interpreting lemma 2.2.17 in terms of cycles (definition 2.2.16). Recall that $\varphi_\xi\left(a^{*\prime}\right) = v_\xi$. Consequently lemma 2.2.17 can be interpreted as follows:
For each pair of optimal strategies (a^{*1}, y^*) player 1's expected average stage payoff during each cycle within $car\left(a^*\right)$ equals v_ξ.
If player 2 plays $y^* = \left(b^{1*}, b^{2*}, \ldots, b^{m*}\right)$, then the total payoff over the stages in cycle $\mathbb{C}\left(s_1, s_2, \ldots, s_l\right)$ of length l is

$$\sum_{j=1}^{l} \left(e_{s_{j+1}} M b^{s_j *} + \xi \cdot \delta^{s_j}_{s_{j+1}}\right) \tag{2.16}$$

with $s_{l+1} = s_1$, where $\delta^{s_j}_{s_{j+1}}$ is the Kronecker delta. Since the average stage payoff over the stages in $\mathbb{C}\left(s_1, s_2, \ldots, s_l\right)$ has to equal v_ξ and given the optimality of $a^{*\prime}$, $\check{y}$ and $\hat{y}$, for each cycle $\mathbb{C}\left(s_1, s_2, \ldots, s_l\right)$ within $car(a^*)$ we have

$$v_\xi = \frac{1}{l} \sum_{j=1}^{l} \left(e_{s_{j+1}} M \check{b}^{s_j} + \xi \cdot \delta^{s_j}_{s_{j+1}}\right) = \frac{1}{l} \sum_{j=1}^{l} \left(e_{s_{j+1}} M \hat{b}^{s_j} + \xi \cdot \delta^{s_j}_{s_{j+1}}\right) \tag{2.17}$$

and hence:

$$\sum_{j=1}^{l} e_{s_{j+1}} M (\check{b}^{s_j} - \hat{b}^{s_j}) = 0 \tag{2.18}$$

and recall that $s_{l+1} = s_1$.
The number of equations in system (2.18) is infinite (notice that the length l of the cycle can be any positive integer). In particular we have the following (finite) subset of equations, corresponding to all cycles of length 1, all cycles of length 2 and a subset of the cycles of length 3:

1. $e_s M(\check{b}^s - \hat{b}^s) = 0$ for each $s \in car(a^*)$

2. $e_{s_2} M(\check{b}^{s_1} - \hat{b}^{s_1}) + e_{s_1} M(\check{b}^{s_2} - \hat{b}^{s_2}) = 0$ for each cycle (s_1, s_2) with $s_1, s_2 \in car(a^*)$, $s_1 < s_2$

3. $e_{s_1} M(\check{b}^{i_1} - \hat{b}^{i_1}) + e_{s_2} M(\check{b}^{s_1} - \hat{b}^{s_1}) + e_{i_1} M(\check{b}^{s_2} - \hat{b}^{s_2}) = 0$ for each cycle (i_1, s_1, s_2) with $s_1, s_2 \in car(a^*), i_1 < s_1 < s_2$

It can easily be proved that all equations in (2.18) hold if the ones mentioned in 1,2 and 3 do. Furthermore, since obviously $\check{y} \in B_\xi(a^{*\prime})$ and $\hat{y} \in B_\xi(a^{*\prime})$, by lemma 2.2.13 we have

$$\varphi_0(a^{*\prime}) = a^* M \check{b}^s = a^* M \hat{b}^s$$

and hence:

$$a^* M(\check{b}^s - \hat{b}^s) = 0 \text{ for each } s \in car(a^*). \tag{2.19}$$

Now let

$$\begin{aligned} \grave{d} &= \left(\grave{d}^{i_1}, \ldots, \grave{d}^{i_p}\right) \\ &= \left(\grave{d}^{i_1}_{i_1}, \grave{d}^{i_1}_{i_2}, \ldots, \grave{d}^{i_1}_{i_p}, \grave{d}^{i_2}_{i_1}, \grave{d}^{i_2}_{i_2}, \ldots, \grave{d}^{i_2}_{i_p}, \ldots, \grave{d}^{i_p}_{i_1}, \grave{d}^{i_p}_{i_2}, \ldots, \grave{d}^{i_p}_{i_p}\right) \in \mathbb{R}^{p^2} \end{aligned} \tag{2.20}$$

with $\grave{d}^{s_1}_{s_2} := e_{s_2} M(\check{b}^{s_1} - \hat{b}^{s_1})$. Then equations 1, 2, 3 and (2.19) can, using notation 2.5.5 be written in terms of $\grave{d}$ as follows:

$$\begin{cases} \grave{d}^s_s = 0 \text{ for each } s \in car(a^*) \\ \grave{d}^{s_1}_{s_2} + \grave{d}^{s_2}_{s_1} = 0 \text{ for each cycle } (s_1, s_2) \text{ with} \\ \qquad s_1, s_2 \in car(a^*),\ s_1 < s_2 \\ \grave{d}^{i_1}_{s_1} + \grave{d}^{s_1}_{s_2} + \grave{d}^{s_2}_{i_1} = 0 \text{ for each cycle } (i_1, s_1, s_2) \text{ with} \\ \qquad s_1, s_2 \in car(a^*),\ i_1 < s_1 < s_2 \\ \bar{a}^* \cdot \grave{d}^s = 0 \text{ for each } s \in car(a^*) \backslash \{i_p\}. \end{cases} \tag{2.21}$$

Notice that we left out the equation $\bar{a}^* \cdot \grave{d}^p = 0$.

Some notations: Let $H \in \mathbb{R}^{p^2 \times p^2}$ and $c \in \mathbb{R}^{p^2}$ be the matrix and vector such that $H\grave{d} = c$ is system (2.21). We will call the matrix H the *characteristic matrix for repeated games with bonus* ξ; the vector c obviously is the zero-vector. An example: For $car(a^*) = \{2, 6, 8\}$ the matrix H is the following one:

$$H = \begin{array}{c} \begin{array}{ccccccccc} \grave{d}^2_2 & \grave{d}^2_6 & \grave{d}^2_8 & \grave{d}^6_2 & \grave{d}^6_6 & \grave{d}^6_8 & \grave{d}^8_2 & \grave{d}^8_6 & \grave{d}^8_8 \end{array} \\ \left(\begin{array}{ccccccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ a^*_2 & a^*_6 & a^*_8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & a^*_2 & a^*_6 & a^*_8 & 0 & 0 & 0 \end{array}\right) \end{array}$$

and notice that $a_2^* = \bar{a}_1^*$, $a_6^* = \bar{a}_2^*$ and $a_8^* = \bar{a}_3^*$.
In theorem 2.7.1 in the appendix we prove that H is non-singular. Therefore system (2.21) of p^2 equations in p^2 variables $\grave{d}_{i_1}^{i_1}, \ldots, \grave{d}_{i_p}^{i_p}$, has a unique solution, which obviously must be the trivial one:

$$\grave{d}_{s_2}^{s_1} = 0 \text{ for all } s_1, s_2 \in car(a^*).$$

This completes the proof. ∎

We continue the analysis by showing that, if player 1 has a simple optimal strategy, then the set of stationary optimal strategies for player 2, restricted to the states in $car(a^*)$, is a Cartesian product.

Theorem 2.5.7 *Let $a^{*\prime}$ be a simple optimal strategy for player 1. The set of player 2's stationary optimal strategies restricted to the states that, given $a^{*\prime}$, are visited with a strictly positive frequency, is:*

$$Y^{a^*} = Y^{i_1} \times Y^{i_2} \times \ldots \times Y^{i_p}.$$

Proof. Since $e_i M(\lambda \cdot \check{b}^j + (1-\lambda) \cdot \hat{b}^j) = e_i M \check{b}^j$ for all $i, j \in car(a^*)$ and for all $\lambda \in [0,1]$, lemma 2.5.6 implies that any convex combination of $\check{b}^j$ and $\hat{b}^j$ in state j, in combination with $\check{b}^s$ or $\hat{b}^s$ in each state $s \in car(a^*)\backslash\{j\}$, is (part of) a stationary optimal strategy for player 2. But we can this for more than one state at once. In particular: For $\lambda_{i_1}, \lambda_{i_2}, \ldots, \lambda_{i_p} \in [0,1]$ the vector $(\lambda_{i_1} \cdot \check{b}^{i_1} + (1-\lambda_{i_1}) \cdot \hat{b}^{i_1}, \lambda_{i_2} \cdot \check{b}^{i_2} + (1-\lambda_{i_2}) \cdot \hat{b}^{i_2}, \ldots, \lambda_{i_p} \cdot \check{b}^{i_p} + (1-\lambda_{i_p}) \cdot \hat{b}^{i_p})$ is (part of) a stationary optimal strategy for player 2. This completes the proof. ∎

The set Y^{a^*} is convex and closed and therefore it is characterized by its extreme points. Since each stationary optimal strategy is a stationary best reply against $a^{*\prime}$, according to (2.4) we have $Y^{a^*} \subset B(a^*) \times \ldots \times B(a^*)$ and $Y^s \subset B(a^*)$ for each $s \in car\,(a^*)$.

Theorem 2.5.8 *Let $a^{*\prime}$ with $car(a^*) = \{i_1, \ldots, i_p\} \subset \{1, \ldots, m\}$ be a simple optimal strategy of player 1 in M_ξ and let $y^* = (b^{1*}, \ldots, b^{m*})$ with $(b^{i_1*}, \ldots, b^{i_p*}) \in Y^{a^*}$ be a stationary optimal strategy for player 2. Then the following two assertions are equivalent:*

1. *$(b^{i_1*}, \ldots, b^{i_p*})$ is an extreme point of Y^{a^*} and $\varphi_0(a^{*\prime}) \neq 0$.*
2. *For each state s corresponding to an action in $car(a^*)$ there exist a subvector $\tilde{b}^{s*}$ of b^{s*} and a non-singular square submatrix K^s of M, whose rows include $car(a^*)$ and whose columns include $car(b^{s*})$, with the following properties:*

$$K_s^s \tilde{b}^{s*} = \varphi_0(a^{*\prime}) + \xi \cdot \sum_{i=1}^{m} (a_i^*)^2 - \xi \tag{2.22}$$

$$K_{s'}^s \tilde{b}^{s*} = \varphi_0(a^{*\prime}) + \xi \cdot \sum_{i=1}^{m} (a_i^*)^2 + \xi \cdot (a_s^* - a_{s'}^*) \tag{2.23}$$

for each $s' \in car(a^*), s' \neq s$. Here $K_{s'}^s$ denotes the row corresponding to player 1's action s' of matrix K^s.

The proof of theorem 2.5.8 is split in a number of theorems all of which can be found in the appendix. To be precise: All the theorems from 2.7.3 up to and including theorem 2.7.6 together constitute the proof of theorem 2.5.8.

The following corollary provides a style of writing of the extreme points of Y^{a^*} in the Shapley-Snow fashion.

Corollary 2.5.9 *Let a' be a simple optimal strategy for player 1 in M_ξ with $car(a) = \{i_1, i_2, \ldots, i_p\}$. Then for any stationary optimal strategy $y^* = (b^{1*}, b^{2*}, \ldots, b^{m*})$ with $(b^{i_1*}, b^{i_2*}, \ldots, b^{i_p*})$ an extreme point of Y^{a^*} and for any state $s \in \{i_1, i_2, \ldots, i_p\}$ we have*

$$\tilde{b}^{s*} = (K^s)^{-1} \cdot \mathfrak{z}^s, \tag{2.24}$$

where

$$\mathfrak{z}^s = \frac{1}{\tilde{1}(K^s)^{-1}\tilde{1}} \cdot \tilde{1} + \xi \cdot \frac{\tilde{1}(K^s)^{-1}\left((K^s)^{-1}\right)^T\tilde{1}}{\left(\tilde{1}(K^s)^{-1}\tilde{1}\right)^2} \cdot \tilde{1} + \xi \cdot \frac{\tilde{1}(K^s)^{-1}e_s}{\tilde{1}(K^s)^{-1}\tilde{1}} \cdot \tilde{1} - \xi \cdot \frac{\left((K^s)^{-1}\right)^T\tilde{1}}{\tilde{1}(K^s)^{-1}\tilde{1}} - \xi \cdot e_s.$$

Here e_s is the unit vector (cf. notation 2.5.4) with the 1 at position s and $\tilde{1}$ is the vector consisting of p ones.

Proof. Using (2.52) and (2.53) the transformation of equations (2.22) and (2.23) into (2.24) is straightforward. ■

We conclude this section by an example.

Example 2.6

Take

$$M = \begin{pmatrix} 1 & 1 \\ 4 & 0 \\ 0 & 2 \end{pmatrix}$$

with optimal mixed actions $a^* = (0, \frac{1}{3}, \frac{2}{3})$ and $b^* = (\frac{1}{3}, \frac{2}{3})$ and $v = \frac{4}{3}$. We will analyze the game M_ξ for all $\xi \geq 0$. This is done in several cases:

Case 1: $\xi \in [0, \frac{3}{4}\rangle$.

It can easily be checked that in M_ξ the simple strategy $a^{*\prime} = (0, \frac{1}{3}, \frac{2}{3})'$ is optimal for player 1 and hence

$$v_\xi = \varphi_0(a^{*\prime}) + \xi \cdot \sum_{i=1}^{3} (a_i^*)^2 = v + \xi \cdot \sum_{i=1}^{3} (a_i^*)^2 = \frac{4}{3} + \frac{5}{9}\xi.$$

Given that player 1 plays the simple optimal strategy $a^{*\prime}$, theorem 2.5.8 provides us the extreme points of the set $Y^{a^*} = Y^2 \times Y^3$ of stationary optimal strategies of player 2, restricted to states 2 and 3. Notice that theorem 2.5.8 does not pronounce upon the mixed actions b^{1*} that player 2 can play in state 1, in order that $y^* = (b^{1*}, b^{2*}, b^{3*})$ with $b^{2*} \in Y^2$ and $b^{3*} \in Y^3$ is optimal. Notice furthermore that for states 2 and 3

the submatrices K^2 and K^3 both include the second and the third row of M. Since M has only 2 columns, this automatically means that in both states the only suitable submatrix K^s is $\begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}$. In state 2 we find:

$$K^2 = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix} \text{ and } K^2\tilde{b}^{2*} = \begin{pmatrix} \frac{4}{3} - \frac{4}{9}\xi \\ \frac{4}{3} + \frac{2}{9}\xi \end{pmatrix}$$

and hence

$$\tilde{b}^{2*} = b^{2*} = \begin{pmatrix} \frac{1}{3} - \frac{1}{9}\xi \\ \frac{2}{3} + \frac{1}{9}\xi \end{pmatrix},$$

which is bigger than 0, since $\xi < \frac{3}{4}$. In state 3:

$$K^3 = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix} \text{ and } K^3\tilde{b}^{3*} = \begin{pmatrix} \frac{4}{3} + \frac{8}{9}\xi \\ \frac{4}{3} - \frac{4}{9}\xi \end{pmatrix}$$

and hence

$$\tilde{b}^{3*} = b^{3*} = \begin{pmatrix} \frac{1}{3} + \frac{2}{9}\xi \\ \frac{2}{3} - \frac{2}{9}\xi \end{pmatrix},$$

which is also bigger than 0 for all $\xi < \frac{3}{4}$.

Notice that in states 2 and 3 player 2 has only one optimal mixed action, which for $\xi \neq 0$ is unequal to his optimal mixed action in M. We will now also compute the optimal mixed actions in state 1. In this example a mixed action b^{1*} in state 1 is optimal, if and only if any cycle $\mathbb{C}$ (cf. definition 2.2.16) that is not in $car(a^*)$ yields against $y^* = (b^{1*}, b^{2*}, b^{3*})$ an average payoff that is not higher than $v_\xi = \frac{4}{3} + \frac{5}{9}\xi$. Of course any cycle in $car(a^*)$ yields against y^* an average payoff of exactly v_ξ (cf. lemma 2.2.17). In the cycles below player 2 plays y^*.

The cycle (1): If player 1 plays action 1 in state 1, then his immediate payoff is $1 + \xi < \frac{4}{3} + \frac{5}{9}\xi$.

The cycle $(1, 2)$: If at stage t player 1 plays action 2 in state 1 and at stage $t + 1$ he plays action 1 in state 2, then his average payoff at these 2 stages is $2 \cdot b_1^{1*} + \frac{1}{2}$, which is not higher than v_ξ, if and only if $b_1^{1*} \leq \frac{5}{12} + \frac{5}{18}\xi$.

The cycle $(1, 3)$: If at stage t player 1 plays action 3 in state 1 and at stage $t + 1$ he plays action 1 in state 3, then his average payoff at these 2 stages is $b_2^{1*} + \frac{1}{2}$, which is not higher than v_ξ, if and only if $b_2^{1*} \leq \frac{5}{6} + \frac{5}{18}\xi$.

The cycle $(1, 2, 3)$: If at stage t player 1 plays action 2 in state 1, at stage $t + 1$ he plays action 3 in state 2 and at stage $t + 2$ he plays action 1 in state 3, then his average payoff during these 3 stages is $\frac{4}{3}b_1^{1*} + \frac{7}{9} + \frac{2}{27}\xi$, which is not higher than v_ξ, if and only if $b_1^{1*} \leq \frac{5}{12} + \frac{13}{36}\xi$.

The cycle $(1, 3, 2)$: If at stage t player 1 plays action 3 in state 1, at stage $t + 1$ he plays action 2 in state 3 and at stage $t + 2$ he plays action 1 in state 2, then his average payoff during these 3 stages is $\frac{2}{3}b_2^{1*} + \frac{7}{9} + \frac{8}{27}\xi$, which is not higher than v_ξ, if and only if $b_2^{1*} \leq \frac{5}{6} + \frac{7}{18}\xi$.

Using the fact that $b^{1*} = (b_1^{1*}, b_2^{1*}) \in \Delta^2$ we can now compute the set of mixed actions that, in combination with b^{2*} and b^{3*} as calculated above, form stationary optimal strategies for player 2 in M_ξ. For $0 \leq \xi \leq \frac{3}{5}$ we get:

$$Y^1 = \left\{ b^{1*} = (b_1^{1*}, b_2^{1*}) \in \Delta^2 \mid \tfrac{1}{6} - \tfrac{5}{18}\xi \leq b_1^{1*} \leq \tfrac{5}{12} + \tfrac{5}{18}\xi \right\}$$

and for $\frac{3}{5} < \xi < \frac{3}{4}$ we get:

$$Y^1 = \left\{ b^{1*} = (b_1^{1*}, b_2^{1*}) \in \Delta^2 \mid 0 \leq b_1^{1*} \leq \tfrac{5}{12} + \tfrac{5}{18}\xi \right\} \tag{2.25}$$

Case 2: $\xi = \frac{3}{4}$.

In this case there are 2 simple optimal strategies for player 1, namely $a^{*\prime} = (0, \frac{1}{3}, \frac{2}{3})'$ and $\check{a}' = (1, 0, 0)'$. Notice that player 1 has many other stationary optimal strategies. For example, any stationary strategy that, eventually, leads to absorption in state 1, is optimal. Player 2, on the other hand, must make sure that no cycle has an average payoff that is higher than $v_\xi = \frac{7}{4}$. Theorem 2.5.8 now tells us for all states what mixed actions for player 2 are optimal. We will separately analyze state 1 and the other states.

State 1:

State 1 is absorbing when player 1 plays $\check{a}' = (1, 0, 0)'$. The matrix K^1 includes player 1's first action. It may, however, include other rows of M as well. This leads to 4 possible submatrices of M, two of which lead to optimal actions of player 2 in state 1.

The first option is $K^1 = (1)$ corresponding to player 2's action 1. Then we get $b^{1*} = (1, 0)$, but if we consider the cycle $(1, 2)$, we see that the average payoff in this cycle, against $y = (b^{1*}, b^{2*}, b^{3*})$ with b^{2*} and b^{3*} optimal mixed actions in states 2 and 3, is $\frac{1}{2} \cdot 4 + \frac{1}{2} \cdot 1 = \frac{5}{2} > \frac{7}{4} = v_\xi$. This means that $(1, 0)$ is not an optimal action for player 2 in state 1.

The second option is $K^1 = (1)$ corresponding to player 2's action 2. Then we get $b^{1*} = (0, 1)$ and it can easily be checked that there is no cycle, of which the average payoff against $y = (b^{1*}, b^{2*}, b^{3*})$ with b^{2*} and b^{3*} optimal mixed actions in states 2 and 3 is higher than $v_\xi = \frac{7}{4}$. Hence $(0, 1)$ is indeed optimal in state 1.

The third option corresponds to player 1's actions 1 and 2 and player 2's actions 1 and 2:

$$K^1 = \begin{pmatrix} 1 & 1 \\ 4 & 0 \end{pmatrix} \text{ and } K^1 \tilde{b}^{1*} = \begin{pmatrix} 1 \\ 1 + 2\xi \end{pmatrix} = \begin{pmatrix} 1 \\ \frac{5}{2} \end{pmatrix}$$

and hence $\tilde{b}^{1*} = b^{1*} = (\frac{5}{8}, \frac{3}{8})$, which is indeed optimal in state 1.

The fourth option corresponds to player 1's actions 1 and 3 and player 2's actions 1 and 2:

$$K^1 = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} \text{ and } K^1 \tilde{b}^{1*} = \begin{pmatrix} 1 \\ \frac{5}{2} \end{pmatrix}$$

and hence $b^{1*} = (-\frac{1}{4}, \frac{5}{4})$, which is not a probability vector.

So the set of player 2's optimal mixed actions in state 1 is

$$Y^1 = \left\{ b^{1*} = (b_1^{1*}, b_2^{1*}) \in \Delta^2 \,|\, 0 \leq b_1^{1*} \leq \frac{5}{8} \right\},$$

which is exactly the same set (for $\xi = \frac{3}{4}$) we found for $\frac{3}{5} < \xi < \frac{3}{4}$ (cf. (2.25)). The intuition behind this observation: For $\frac{3}{5} < \xi < \frac{3}{4}$ this was the set of actions that made sure that there was no cycle that included state 1, in which the average payoff was bigger than $v_\xi = \frac{4}{3} + \frac{5}{9}\xi$. Since for $\xi = \frac{3}{4}$ we still have $v_\xi = \frac{4}{3} + \frac{5}{9}\xi$, for any stationary optimal strategy for player 2 there should still be no cycle that includes state 1, in which the average payoff is bigger than $v_\xi = \frac{4}{3} + \frac{5}{9}\xi$.

States 2 and 3:

In states 2 and 3 we consider player 1's simple optimal strategy $a^{*\prime} = (0, \frac{1}{3}, \frac{2}{3})'$, using actions 2 and 3. This leads, as in the case $\xi < \frac{3}{4}$, to the submatrices $K^2 = K^3 = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}$. Furthermore $b^{2*} = (\frac{1}{4}, \frac{3}{4})$ and $b^{3*} = (\frac{1}{2}, \frac{1}{2})$ are the unique optimal mixed actions for player 2 in states 2 and 3 respectively.

Case 3: $\xi > \frac{3}{4}$

For $\xi > \frac{3}{4}$ player 1 has one simple optimal strategy, namely $\breve{a}' = (1, 0, 0)'$ and applying theorem 2.5.8 leads to the same four options we found in the case $\xi = \frac{3}{4}$. The option $K^1 = (1)$ corresponding to player 2's action 2, lead to the optimal mixed action $b^{1*} = (0, 1)$ for all $\xi > \frac{3}{4}$. The option corresponding to player 1's actions 1 and 2 and player 2's actions 1 and 2 is:

$$K^1 = \begin{pmatrix} 1 & 1 \\ 4 & 0 \end{pmatrix} \text{ and } K^1 \tilde{b}^{1*} = \begin{pmatrix} 1 \\ 1 + 2\xi \end{pmatrix}$$

and hence $b^{1*} = (\frac{1}{4} + \frac{1}{2}\xi, \frac{3}{4} - \frac{1}{2}\xi)$. This is an optimal mixed action as long as $\xi \leq \frac{3}{2}$. For $\xi \geq \frac{3}{2}$ the matrix $K^1 = (1)$ corresponding to player 2's action 1, leads to the optimal mixed action $b^{1*} = (1, 0)$ and notice that against $y^* = (b^{1*}, b^{2*}, b^{3*})$ the cycle $(1, 2)$ yields an average payoff of $\frac{1}{2} \cdot 4 + \frac{1}{2} \cdot 1 = \frac{5}{2} \leq 1 + \xi = v_\xi$. The option corresponding to player 1's actions 1 and 3 and player 2's actions 1 and 2 is

$$K^1 = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} \text{ and } K^1 \tilde{b}^{1*} = \begin{pmatrix} 1 \\ 1 + 2\xi \end{pmatrix}$$

and hence $b^{1*} = (\frac{1}{2} - \xi, \frac{1}{2} + \xi)$, which is not a probability vector for $\xi > \frac{3}{4}$. Hence for $\frac{3}{4} < \xi \leq \frac{3}{2}$ we have:

$$Y^1 = \left\{ b^{1*} = (b_1^{1*}, b_2^{1*}) \in \Delta^2 \,|\, 0 \leq b_1^{1*} \leq \tfrac{1}{4} + \tfrac{1}{2}\xi \right\}$$

and for $\xi \geq \frac{3}{2}$ we have $Y^1 = \Delta^2$.

For states 2 and 3 the analysis is more complicated. In general the set of optimal mixed actions in these states depends on the choice of b^{1*} and the stationary optimal strategy space of player 2 over all states is not a Cartesian product as in theorem

2.5.7. Consequently the analysis of stationary optimal mixed actions in states 2 and 3 is beyond the scope of theorem 2.5.8 and we will omit it.

However, we will prove that the stationary optimal strategy space of player 2 over all states is in general not a Cartesian product. Take $\xi = \frac{9}{5}$ and consider the stationary strategies $\breve{y}^* = (\breve{b}^{1*}, \breve{b}^{2*}, \breve{b}^{3*}) = ((\frac{1}{3}, \frac{2}{3}), (0,1), (\frac{1}{2}, \frac{1}{2}))$ and $\bar{y}^* = (\bar{b}^{1*}, \bar{b}^{2*}, \bar{b}^{3*}) = ((\frac{1}{3}, \frac{2}{3}), (\frac{1}{4}, \frac{3}{4}), (1,0))$. Then $v_\xi = \frac{14}{5}$ and it can easily be shown that neither against $\breve{y}^*$ nor against $\bar{y}^*$ a cycle exists that provides an average payoff to player 1 that is higher than $\frac{14}{5}$. Hence $\breve{y}^*$ and $\bar{y}^*$ are both optimal. Now consider the stationary strategy $\hat{y} = ((\frac{1}{3}, \frac{2}{3}), (0,1), (1,0)) = (\breve{b}^{1*}, \breve{b}^{2*}, \bar{b}^{3*})$ and the cycle $(2,3)$. The average payoff in cycle $(2,3)$ against $\hat{y}$ is $\frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 4 = 3 > v_\xi$ and hence $\hat{y}$ is not optimal, which proves our statement. □

2.6 Generalizations of the model

The repeated game with bonus ξ model has many straightforward generalizations, the most outstanding one perhaps being the model in which the bonus is different for player 1's actions. This model is achieved by introducing numbers ξ_i that denote the bonus of action i. The game is still called M_ξ, but now $\xi = (\xi_1, \xi_2, \ldots, \xi_m)$. Many of the results that are attained in the previous sections can be generalized. Most appealing is the generalization of theorem 2.5.8:

Theorem 2.6.1 *Let $a^{*\prime}$ with $car(a^*) = \{i_1, \ldots, i_p\} \subset \{1, \ldots, m\}$ be a simple optimal strategy of player 1 in M_ξ and let $y^* = (b^{1*}, \ldots, b^{m*})$ with $(b^{i_1*}, \ldots, b^{i_p*}) \in Y^{a^*}$ be a stationary optimal strategy for player 2. Then the following two assertions are equivalent:*

1. *$(b^{i_1*}, \ldots, b^{i_p*})$ is an extreme point of Y^{a^*} and $\varphi_0(a^{*\prime}) \neq 0$.*
2. *For each state s corresponding to an action in $car(a^*)$ there exist a subvector $\tilde{b}^{s*}$ of b^{s*} and a non-singular square submatrix K^s of M, whose rows include $car(a^*)$ and whose columns include $car(b^{s*})$, with the following properties:*

$$K^s_s \tilde{b}^{s*} = \varphi_0(a^{*\prime}) + \sum_{i=1}^{m} \xi_i \left(a_i^*\right)^2 - \xi_s \tag{2.26}$$

$$K^s_{s'} \tilde{b}^{s*} = \varphi_0(a^{*\prime}) + \sum_{i=1}^{m} \xi_i \left(a_i^*\right)^2 + \xi_s a^*_s - \xi_{s'} a^*_{s'} \tag{2.27}$$

for all $s' \in \{i_1, \ldots, i_p\}, s' \neq s$.

Proof. Since the matrix game M has not changed, the non-singular submatrices K^s can be constructed as in theorems 2.7.3 and 2.7.4. If player 1 plays $a^{*\prime}$, then the average amount of bonus he receives per stage, is $\sum_{i=1}^{m} \xi_i \left(a_i^*\right)^2$. Given the optimality of $a^{*\prime}$ this means that $v_\xi = \varphi_\xi(a^{*\prime}) = \varphi_0(a^{*\prime}) + \sum_{i=1}^{m} \xi_i \left(a_i^*\right)^2$. Since lemma 2.5.6 also applies for this class of games, the statement in theorem 2.5.7 remains valid.

Now consider the proof of theorem 2.7.5 and assume without loss of generality that $car(a^*) = \{1, \ldots, p\}$. Let the vector $d \in \mathbb{R}^{p \times p}$ be defined as in (2.45). Then system (2.51), applied to the class of repeated games with bonus ξ, becomes:

$$\begin{cases} d_s^s = v_\xi - \xi_s \text{ for each } s \in \{1, \ldots, p\} \\ d_{s_2}^{s_1} + d_{s_1}^{s_2} = 2 \cdot v_\xi \text{ for each cycle } (s_1, s_2) \\ \qquad \text{with } 1 \leq s_1 < s_2 \leq p \\ d_{s_1}^1 + d_{s_2}^{s_1} + d_1^{s_2} = 3 \cdot v_\xi \text{ for each cycle } (1, s_1, s_2) \\ \qquad \text{with } 2 \leq s_1 < s_2 \leq p \\ \bar{a}^* d^s = \varphi_0(\bar{a}^{*\prime}) \text{ for each } s \in \{1, \ldots, p-1\}, \end{cases} \tag{2.28}$$

and therefore the matrix $H \in \mathbb{R}^{p^2 \times p^2}$ is, again, the same characteristic matrix as in the proof of theorem 2.7.5. This also holds for the vector $c_1 \in \mathbb{R}^{p^2}$, but $c_2 \in \mathbb{R}^{p^2}$ is different: Its first p elements are $\xi_1, \xi_2, \ldots, \xi_p$ and the last $p-1$ elements are $\sum_{i=1}^{m} \xi_i \left(a_i^*\right)^2$. The (unique) solution to the matrix-vector product $H\beta = c_2$ then is: $\beta_s^s = -\xi_s$ and $\beta_{s'}^s = \xi_s a_s^* - \xi_{s'} a_{s'}^*$ for each $s, s' \in \{1, \ldots, p\}, s' \neq s$. Consequently:

$$K_s^s \tilde{b}^{s*} = d_s^s = (\beta_1)_s^s + (\beta_2)_s^s = v_{a^*\xi} - \xi_s = \varphi_0(a^{*\prime}) + \sum_{i=1}^{m} \xi_i \left(a_i^*\right)^2 - \xi_s$$

and

$$\begin{aligned} K_{s'}^s \tilde{b}^{s*} = d_{s'}^s &= (\beta_1)_{s'}^s + (\beta_2)_{s'}^s = v_{a^*\xi} + \xi_s a_s^* - \xi_{s'} a_{s'}^* \\ &= \varphi_0(a^{*\prime}) + \sum_{i=1}^{p} \xi_i \left(a_i^*\right)^2 + \xi_s a_s^* - \xi_{s'} a_{s'}^* \text{ for all } s' \neq s, \end{aligned}$$

which is exactly (2.26) and (2.27).

Furthermore the proof of theorem 2.7.6 remains valid for this model, which completes the proof. ■

Other interesting generalizations are models, in which player 1 receives the bonus when playing action i only if he played action i at least a fixed number of times, say l^i, in the previous r^i stages or to let him, when playing action i at stage t, get an increase of $\frac{\xi \cdot l^i}{r^i}$ in payoff if, at the stages in $\left\{t - r^i, \ldots, t-1\right\}$, he played action j exactly l^i times. Furthermore all of the previous suggestions could also apply to player 2 and then also in non-zero-sum situations.

2.7 Appendix

Theorem 2.7.1 *The characteristic matrix for repeated games with bonus ξ, is non-singular.*

Proof. Let $H \in \mathbb{R}^{p^2 \times p^2}$ be the characteristic matrix for repeated games with bonus ξ and suppose without loss of generality that a^* with $car(a^*) = \{1, 2, \ldots, p\}$ is

a simple optimal strategy for player 1. Then system (2.21) is:

$$\begin{cases} d_s^s = 0 \text{ for each } s \in \{1, \dots, p\} \\ d_{s_2}^{s_1} + d_{s_1}^{s_2} = 0 \text{ for each cycle } (s_1, s_2) \text{ with } 1 \leq s_1 < s_2 \leq p \\ d_{s_1}^1 + d_{s_2}^{s_1} + d_1^{s_2} = 0 \text{ for each cycle } (1, s_1, s_2) \text{ with } 2 \leq s_1 < s_2 \leq p \\ \bar{a}^* d^s = 0 \text{ for each } s \in \{1, \dots, p-1\}. \end{cases}$$

and $Hd = 0$ is the matrix-vector notation of system (2.21) as in the proof of lemma 2.5.6 with $d = (d_1^1, d_2^1, \dots, d_p^1, d_1^2, d_2^2, \dots, d_p^2, \dots, d_1^p, d_2^p, \dots, d_p^p) \in \mathbb{R}^{p^2}$. The matrix H also appears in the proof of theorems 2.7.5 and 2.6.1. Throughout the proof we will number the columns of H as follows: The column of H that is multiplied by d_j^i is numbered ij. We will prove that H is non-singular by showing that the zero-row can only be written as a linear combination of the rows of H by assigning weight 0 to all rows.

Assume that the row corresponding to $\bar{a}^* d^i = 0$ has been assigned a weight of η_i to, let ε_i be the weight that is to be assigned to the row corresponding to the equation $d_i^i = 0$ and let ε_{ij} and ε_{1ij} be the weights that are to be assigned to the rows corresponding to the cycles (i, j) and $(1, i, j)$ respectively. In order to create the zero-row as a linear combination of the rows of H for each $i \in \{2, \dots, p-1\}$ weight 0 has to be assigned to the row corresponding to the cycle (i, p). This is the case, since the variable d_i^p appears only in this equation. Weight 0 has to be assigned to the row corresponding to $d_p^p = 1$ for the same reason. So

$$\varepsilon_p = 0$$

and

$$\varepsilon_{ip} = 0 \text{ for all } 2 \leq i \leq p-1. \tag{2.29}$$

Furthermore

$$\varepsilon_i = -\bar{a}_i^* \eta_i \text{ for all } i \leq p-1.$$

since the variable d_i^i appears only in the equations $\bar{a}^* d^i = 0$ and $d_i^i = v_\xi - \xi$.

Now we take a look at columns ip with $2 \leq i \leq p-1$. The corresponding variable d_p^i appears in the equations for the cycles (i, p) and $(1, i, p)$ and in the equation $\bar{a}^* d^i = 0$. The weight assigned to the row corresponding to the cycle (i, p) is 0 (cf. (2.29)). This means that, in order to achieve the zero-row, we must have:

$$\varepsilon_{1ip} = -\bar{a}_p^* \eta_i \text{ for all } 2 \leq i \leq p-1. \tag{2.30}$$

Now we take a look at columns ji with $2 \leq i < j \leq p-1$. The corresponding variable d_i^j appears only in the cycle (i, j) and in the equation $\bar{a}^* d^j = 0$. Therefore , in order to achieve the zero-row, we must have:

$$\varepsilon_{ij} = -\bar{a}_i^* \eta_j \text{ for all } 2 \leq i < j \leq p-1. \tag{2.31}$$

Now we take a look at columns ij with $2 \leq i < j \leq p-1$. The corresponding variable d_j^i appears in the cycle (i, j), in the cycle $(1, i, j)$ and in the equation $\bar{a}^* d^i = 0$.

The row corresponding to the cycle (i,j) and the row corresponding to the equation $\bar{a}^* d^i = 0$ have respective weights of $-\bar{a}_i^* \eta_j$, according to (2.31), and η_i been assigned to. Consequently , in order to achieve the zero-row, we must have:

$$\varepsilon_{1ij} = \bar{a}_i^* \eta_j - \bar{a}_j^* \eta_i \text{ for all } 2 \leq i < j \leq p-1. \tag{2.32}$$

Now we take a look at the columns $j1$ with $2 \leq j \leq p-1$. The corresponding variable d_1^j appears in the cycle $(1,j)$, in the cycles $(1,i,j)$ with $2 \leq i \leq j-1$ and in the equation $\bar{a}^* d^i = 0$. This means that, in order to achieve the zero-row, according to (2.32) we must have:

$$\begin{aligned} \varepsilon_{1j} &= -1 \cdot \sum_{i=2}^{j-1} \varepsilon_{1ij} - \eta_j \bar{a}_1^* = -1 \cdot \sum_{i=2}^{j-1} (\bar{a}_i^* \eta_j - \bar{a}_j^* \eta_i) - \eta_j \bar{a}_1^* \\ &= -\eta_j \cdot \sum_{i=1}^{j-1} \bar{a}_i^* + \bar{a}_j^* \cdot \sum_{i=2}^{j-1} \eta_i \text{ for all } 2 \leq j \leq p-1. \end{aligned} \tag{2.33}$$

Now we take a look at column $p1$. The corresponding variable d_1^p appears in the cycle $(1,p)$ and in the cycles $(1,i,p)$. This means that, in order to achieve the zero-row, according to (2.30) we must have:

$$\varepsilon_{1p} = -1 \cdot \sum_{i=2}^{p-1} \varepsilon_{1ip} = -1 \cdot \sum_{i=2}^{p-1} \left(-\bar{a}_p^* \eta_i\right) = \bar{a}_p^* \cdot \sum_{i=2}^{p-1} \eta_i. \tag{2.34}$$

Now we take a look at column $1p$. The corresponding variable d_p^1 appears in the cycle $(1,p)$ and in the equation $\bar{a}^* d^1 = 0$. This means that, in order to achieve the zero-row, according to (2.29) we must have:

$$\begin{aligned} \varepsilon_{1p} + \bar{a}_p^* \eta_1 = \bar{a}_p^* \cdot \sum_{i=2}^{p-1} \eta_i + \bar{a}_p^* \eta_1 = \bar{a}_p^* \cdot \sum_{i=1}^{p-1} \eta_i = 0 \\ \Leftrightarrow \sum_{i=1}^{p-1} \eta_i = 0. \end{aligned} \tag{2.35}$$

Now we take a look at the columns $1j$ with $2 \leq j \leq p-1$ (these are the only columns that we had not considered so far). The corresponding variable d_j^1 appears in the cycle $(1,j)$, in the cycles $(1,j,i)$ with $j+1 \leq i \leq p$ and in the equation $\bar{a}^* d^1 = 0$. This means that, in order to achieve the zero-row, according to (2.30), (2.32), (2.33)

and (2.35) we must have (recall that $\sum_{i=1}^{p} \bar{a}_i^* = 1$):

$$\begin{aligned}
&\varepsilon_{1j} + \sum_{i=j+1}^{p} \varepsilon_{1ji} + \bar{a}_j^* \eta_1 \\
&= -\eta_j \cdot \sum_{i=1}^{j-1} \bar{a}_i^* + \bar{a}_j^* \cdot \sum_{i=2}^{j-1} \eta_i + \sum_{i=j+1}^{p-1} \left(\bar{a}_j^* \eta_i - \bar{a}_i^* \eta_j \right) - \bar{a}_p^* \eta_j + \bar{a}_j^* \eta_1 = 0 \\
&\Leftrightarrow -\eta_j \left(1 - \bar{a}_j^*\right) + \bar{a}_j^* \cdot \sum_{\substack{i=1 \\ i \neq j}}^{p-1} \eta_i = 0 \\
&\Leftrightarrow -\eta_j + \bar{a}_j^* \cdot \sum_{i=1}^{p-1} \eta_i = 0 \\
&\Leftrightarrow \eta_j = 0.
\end{aligned}$$

So now we know that $\eta_j = 0$ for all $2 \leq j \leq p-1$. Since, according to (2.35), $\sum_{i=1}^{p-1} \eta_i = 0$, this means that also $\eta_1 = 0$, from which we can conclude, using (2.29) up to and including (2.35), that ε_i, ε_{ij} and ε_{1ij} must equal 0 for all i and j. This means that the matrix H is non-singular and therefore that systems (2.21), (2.51) and (2.28) each have exactly one solution. ■

The remainder of the appendix consists of a number of theorems that together constitute theorem 2.5.8. But firstly we have to construct the matrices K^s.

The construction of K^s

Let $a^{*\prime}$ be a simple optimal strategy for player 1 and let $y^* = (b^{1*}, \ldots, b^{m*})$ with $(b^{i_1*}, \ldots, b^{i_p*}) \in Extr(Y^{a^*})$, where $Extr$ denotes the set of extreme points, be a stationary optimal strategy for player 2. Let $\tilde{K}_1^s$ be the submatrix of M consisting of the rows of M that are in $car(a^*)$ and the columns of M that are in b^{s*} and let M_{a^*} be the submatrix of M consisting of the rows of $\tilde{K}_1^s$. Now we add to $\tilde{K}_1^s$ columns of M_{a^*} that correspond to pure actions that are one-shot best replies against a^* in M as long as those columns are linearly independent of the columns of $\tilde{K}_1^s$ and the previously adjoined ones. The remaining matrix is called $\tilde{K}_2^s$ and let $M_{b^{s*}}$ be the submatrix of M consisting of the columns of $\tilde{K}_2^s$. Now consider the subset $\tilde{I}_{b^{s*}}$ of rows of M, defined as follows:

Definition 2.7.2 *For each state $s \in car(a^*)$ the set $\tilde{I}_{b^{s*}}$ is the subset of rows of M that in state s, in combination with some pure actions in other states, can form a cycle $\mathbb{C}$ that is a (pure) best reply against y^*.*

Notice that $car(a^*) \subset \tilde{I}_{b^{s*}}$. Now we add to $\tilde{K}_2^s$ rows of $M_{b^{s*}}$ corresponding to rows in the set $\tilde{I}_{b^{s*}} \setminus car(a^*)$ in M as long as they are linearly independent of the rows in $\tilde{K}_2^s$ and the previously adjoined ones. The resulting matrix is K^s. The subvectors of a^* and b^{s*} corresponding to the rows and columns of K^s will be called $\tilde{a}^*$ and $\tilde{b}^{s*}$ respectively.

Notice that

$$\sum_i \tilde{a}_i^* = 1 \text{ and } \sum_j \tilde{b}_j^{s*} = 1$$

and that

$$\varphi_0(a^{*\prime}) = a^* M b^{s*} = \tilde{a}^* K^s \tilde{b}^{s*}. \tag{2.36}$$

Theorem 2.7.3 *If $a^{*\prime}$ is a simple optimal strategy for player 1 with $\varphi_0(a^{*\prime}) \neq 0$ and $y^* = (b^{1*}, \ldots, b^{m*})$ with $(b^{i_1*}, \ldots, b^{i_p*}) \in Extr(Y^{a^*})$ is a stationary optimal strategy for player 2, then the submatrix K^s of M as constructed above has linearly independent rows.*

Proof. Let $a^{*\prime}$ be a simple optimal strategy for player 1 with $\varphi_0(a^{*\prime}) \neq 0$ and let $y^* = (b^{1*}, \ldots, b^{m*})$ with $(b^{i_1*}, \ldots, b^{i_p*}) \in Extr(Y^{a^*})$ be a stationary optimal strategy for player 2. Suppose by means of contradiction that the, say $w \geq p$, rows of K^s are linearly dependent. We will now show that then $a^{*\prime}$ is not optimal.

Let $0 < \delta \leq \frac{1}{2} \cdot \min_{i \in car(a^*)} \{a_i^*\}$ and let $\mu \in \mathbb{R}^w$ such that $\mu K^s = 0$, $\mu_i = 0$ for all $i \notin car(a^*)$ and $||\mu||_\infty = \delta$. Such a μ exists, since K^s has linearly dependent rows and by construction the rows of K^s that are not in $car(a^*)$ can not be part of this dependency. Then

$$(\tilde{a}^{s*} + \mu)K^s = (\tilde{a}^{s*} - \mu)\, K^s = \tilde{a}^{s*} K^s. \tag{2.37}$$

Consider the following two vectors:

$$\tilde{z}_1 = \frac{\tilde{a}^{s*} + \mu}{\sum_{i=1}^{w} (\tilde{a}_i^{s*} + \mu_i)} \text{ and } \tilde{z}_2 = \frac{\tilde{a}^{s*} - \mu}{\sum_{i=1}^{w} (\tilde{a}_i^{s*} - \mu_i)}.$$

Notice that $\tilde{z}_1$ and $\tilde{z}_2$ are probability vectors with $car(\tilde{z}_1) = car(\tilde{z}_2) = car(\tilde{a}^{s*})$ and that $\tilde{a}^{s*}$ is a convex combination of $\tilde{z}_1$ and $\tilde{z}_2$. Notice furthermore that $\tilde{z}_1 \neq \tilde{a}^{s*}$ and $\tilde{z}_2 \neq \tilde{a}^{s*}$, since $\tilde{z}_1 = \tilde{z}_2 = \tilde{a}^{s*}$ only if μ and $\tilde{a}^{s*}$ are linearly dependent vectors and in that case $\tilde{a}^{s*} K^s = 0$ and hence $\varphi_0(a^{*\prime}) = a^* M b^{s*} = 0$, which contradicts our assumption. Hence $\tilde{a}^{s*}$ is a strictly convex combination of $\tilde{z}_1$ and $\tilde{z}_2$. Let $\alpha \in \langle 0, 1\rangle$ be such that $\tilde{a} = \alpha \cdot \tilde{z}_1 + (1 - \alpha) \cdot \tilde{z}_2$. Then

$$\alpha = \frac{\sum_{i=1}^{w} (\tilde{a}^{s*} + \mu_i)}{\sum_{i=1}^{w} (\tilde{a}^{s*} + \mu_i) + \sum_{i=1}^{w} (\tilde{a}^{s*} - \mu_i)} = \frac{1}{2} \cdot \sum_{i=1}^{w} (\tilde{a}^{s*} + \mu_i)\,.$$

Furthermore, due to the strict convexity of the quadratic function, we have:

$$\begin{aligned} \sum_{i=1}^{w} (\tilde{a}^{s*})^2 &= \sum_{i=1}^{w} (\alpha \cdot \tilde{z}_{1i} + (1 - \alpha) \cdot \tilde{z}_{2i})^2 \\ &< \alpha \cdot \sum_{i=1}^{w} (\tilde{z}_{1i})^2 + (1 - \alpha) \cdot \sum_{i=1}^{w} (\tilde{z}_{2i})^2\,. \end{aligned} \tag{2.38}$$

Let $z_1 \in \Delta^m$ and $z_2 \in \Delta^m$ be the mixed actions in M, whose subvectors consisting of the elements corresponding to the rows of K^s are $\tilde{z}_1$ and $\tilde{z}_2$ respectively. We compare the rewards to player 1 of the strategy pairs $(a^{*\prime}, y^*)$, (z_1', y^*) and (z_2', y^*). Since $(a^{*\prime}, y^*)$ is a pair of optimal strategies in M_ξ, according to (2.4) we have: $b^{s*} \in B(a^*)$ for each state $s \in \{i_1, \ldots, i_p\}$. We have:

$$\tilde{z}_1 K^s = \frac{(\tilde{a}^{s*} + \mu)K^s}{\sum_{i=1}^{w} (\tilde{a}^{s*} + \mu_i)} = \frac{\tilde{a}^{s*} K^s}{\sum_{i=1}^{w} (\tilde{a}^{s*} + \mu_i)} \tag{2.39}$$

and

$$\tilde{z}_2 K^s = \frac{\tilde{a}^{s*} K^s}{\sum_{i=1}^{w} (\tilde{a}^{s*} - \mu_i)}. \tag{2.40}$$

Let δ be so small that in M no action outside $B(a^*)$ is a best reply against z_1 and z_2 (it is always possible to choose a δ sufficiently small). Then from (2.39) and (2.40) we can conclude that $B(a^*) = B(z_1) = B(z_2)$. Thus $b^{s*} \in B(z_1)$ and $b^{s*} \in B(z_2)$ and therefore, according to lemma 2.2.13 for each $s \in \{i_1, \ldots, i_p\}$:

$$\varphi_0(z_1') = z_1 M b^{s*} = \tilde{z}_1 K^s \tilde{b}^{s*}$$

and

$$\varphi_0(z_2') = z_2 M b^{s*} = \tilde{z}_2 K^s \tilde{b}^{s*}.$$

But then:

$$\begin{aligned}
\varphi_0(a^{*\prime}) = a^* M b^{s*} = \tilde{a}^{s*} K^s \tilde{b}^{s*} &= \sum_{i=1}^{w} (\tilde{a}^{s*} + \mu_i) \cdot \tilde{z}_1 K^s \tilde{b}^{s*} \\
&= \sum_{i=1}^{w} (\tilde{a}^{s*} + \mu_i) \cdot z_1 M b^{s*} = \sum_{i=1}^{w} (\tilde{a}^{s*} + \mu_i) \cdot \varphi_0(z_1') \\
&= 2\alpha \cdot \varphi_0(z_1')
\end{aligned}$$

and analogously

$$\varphi_0(a^{*\prime}) = 2(1-\alpha) \cdot \varphi_0(z_2').$$

From this it easily follows that

$$\varphi_0(a^{*\prime}) = \alpha \cdot \varphi_0(z_1') + (1-\alpha) \cdot \varphi_0(z_2'). \tag{2.41}$$

Successively using (2.38) and (2.41) and noticing that

$$\sum_{i=1}^{w} (\tilde{a}_i^*)^2 = \sum_{i=1}^{m} (a_i^*)^2\,,\ \sum_{i=1}^{w} (\tilde{z}_{1i})^2 = \sum_{i=1}^{m} (z_{1i})^2 \text{ and } \sum_{i=1}^{w} (\tilde{z}_{2i})^2 = \sum_{i=1}^{m} (z_{2i})^2$$

we find:

$$\begin{aligned}\gamma_\xi(a^{*\prime}, y^*) &= \varphi_0(a^{*\prime}) + \xi \cdot \sum_{i=1}^{m} (a^{s*})^2 \\ &< \varphi_0(a^{*\prime}) + \alpha\xi \cdot \sum_{i=1}^{m} (z_{1i})^2 + (1-\alpha)\,\xi \cdot \sum_{i=1}^{m} (z_{2i})^2 \\ &= \alpha\varphi_0(z_1') + (1-\alpha)\,\varphi_0(z_2') + \alpha\xi \sum_{i=1}^{m} (z_{1i})^2 + (1-\alpha)\,\xi \sum_{i=1}^{m} (z_{2i})^2 \\ &= \alpha\cdot\left(\varphi_0(z_1') + \xi\cdot\sum_{i=1}^{m} (z_{1i})^2\right) + (1-\alpha)\cdot\left(\varphi_0(z_2') + \xi\cdot\sum_{i=1}^{m} (z_{2i})^2\right) \\ &= \alpha\cdot\gamma_\xi(z_1', y^*) + (1-\alpha)\cdot\gamma_\xi(z_2', y^*),\end{aligned}$$

which means that at least one of the two strategies z_1' and z_2' provides a higher reward against y^* than $a^{*\prime}$, thereby contradicting the optimality of $a^{*\prime}$. Hence the assumption that K^s has linearly dependent rows, is false. ■

Theorem 2.7.4 *If $a^{*\prime}$ is a simple optimal strategy for player 1 with $\varphi_0(a^{*\prime}) \neq 0$ and $y^* = (b^{1*}, \ldots, b^{m*})$ with $(b^{i_1*}, \ldots, b^{i_p*}) \in Extr(Y^{a^*})$ is a stationary optimal strategy for player 2, then the submatrix K^s of M as constructed above has linearly independent columns.*

Proof. Let $a^{*\prime}$ be a simple optimal strategy for player 1, let $y^* = (b^{1*}, \ldots, b^{m*})$ with $(b^{i_1*}, \ldots, b^{i_p*}) \in Extr(Y^{a^*})$ be a stationary optimal strategy for player 2 and let $\varphi_0(a^{*\prime}) \neq 0$. Suppose by means of contradiction that the, say $w \geq q$, columns of K^s are linearly dependent. We will show that then $(b^{i_1*}, \ldots, b^{i_p*}) \notin Extr(Y^{a^*})$.

Let $0 < \varepsilon \leq \frac{1}{2} \cdot \min_{j \in car(b^{s*})} b_j^{s*}$, let $\lambda \in \mathbb{R}^w$ such that $K^s\lambda = 0$, $\lambda_j = 0$ for all $j \notin car(b^{s*})$ and $||\lambda||_\infty = \varepsilon$. Such a λ exists, since K^s has linearly dependent columns and by construction the columns of K^s that are not in $car(a^*)$ can not be part of this dependency. Then

$$K^s(\tilde{b}^{s*} + \lambda) = K^s(\tilde{b}^{s*} - \lambda) = K^s\tilde{b}^{s*}$$

and also

$$\tilde{a}^* K^s(\tilde{b}^{s*} + \lambda) = \tilde{a}^* K^s(\tilde{b}^{s*} - \lambda) = \tilde{a}^* K^s\tilde{b}^{s*}.$$

Consider the following two vectors:

$$\tilde{b}_1 = \frac{\tilde{b}^{s*} + \lambda}{\sum_{j=1}^{w}\left(\tilde{b}_j^{s*} + \lambda_j\right)} \text{ and } \tilde{b}_2 = \frac{\tilde{b}^{s*} - \lambda}{\sum_{j=1}^{w}\left(\tilde{b}_j^{s*} - \lambda_j\right)}.$$

Notice that $\tilde{b}_1$ and $\tilde{b}_2$ are probability vectors with $car(\tilde{b}_1) = car(\tilde{b}_2) = car(\tilde{b}^{s*})$ and that $\tilde{b}^{s*}$ is a convex combination of $\tilde{b}_1$ and $\tilde{b}_2$. Notice furthermore that $\tilde{b}_1 = \tilde{b}_2 = \tilde{b}^{s*}$ if and only if λ and $\tilde{b}^{s*}$ are linearly dependent vectors. In that case we have $K^s\tilde{b}^{s*} = 0$

and hence, by (2.36), $\varphi_0(a^{*\prime}) = \tilde{a}^* K^s \tilde{b}^{s*} = 0$, contradicting one of the assumptions. This means that $\tilde{b}^{s*}$ is a strictly convex combination of $\tilde{b}_1$ and $\tilde{b}_2$. We distinguish between 2 cases.

Case 1: $\sum_{j=1}^{w} \lambda_j \neq 0$.

Let $b_1 \in \Delta^n$ and $b_2 \in \Delta^n$ be player 2's mixed actions in M, whose subvectors consisting of the elements corresponding to the columns of K^s are $\tilde{b}_1$ and $\tilde{b}_2$ respectively. Notice that

$$a^* M b_j = \tilde{a}^* K^s \tilde{b}_j,\ j \in \{1, 2\}\,. \tag{2.42}$$

We will show that in the underlying matrix game $b^{s*} \notin B(a^*)$. Suppose w.l.o.g. that $\sum_{j=1}^{w} \lambda_j > 0$. Then

$$\sum_{j=1}^{w} \left(\tilde{b}_j^{s*} + \lambda_j\right) = \sum_{j=1}^{w} \tilde{b}_j^{s*} + \sum_{j=1}^{w} \lambda_j > 1$$

and

$$\sum_{j=1}^{w} \left(\tilde{b}_j^{s*} - \lambda_j\right) < 1.$$

Recall (cf. lemma (2.2.13)) that, since $\varphi_0(a^{*\prime}) \neq 0$, we have: $a^* M b^{s*} \neq 0$. If $a^* M b^{s*} > 0$, then by (2.42)

$$\begin{aligned} a^* M b_1 = \tilde{a}^* K^s \tilde{b}_1 &= \frac{\tilde{a}^* K^s \left(\tilde{b}^{s*} + \lambda\right)}{\sum_{j=1}^{w} \left(\tilde{b}_j^{s*} + \lambda_j\right)} = \frac{\tilde{a}^* K^s \tilde{b}^{s*}}{\sum_{j=1}^{w} \left(\tilde{b}_j^{s*} + \lambda_j\right)} \\ &< \tilde{a}^* K^s \tilde{b}^{s*} = a^* M b^{s*} \end{aligned}$$

and if $a^* M b^{s*} < 0$, then

$$\begin{aligned} a^* M b_2 = \tilde{a}^* K^s \tilde{b}_2 &= \frac{\tilde{a}^* K^s \left(\tilde{b}^{s*} - \lambda\right)}{\sum_{j=1}^{w} \left(\tilde{b}_j^{s*} - \lambda_j\right)} = \frac{\tilde{a}^* K^s \tilde{b}^{s*}}{\sum_{j=1}^{w} \left(\tilde{b}_j^{s*} - \lambda_j\right)} \\ &< \tilde{a}^* K^s \tilde{b}^{s*} = a^* M b^{s*} \end{aligned}$$

and hence $b^{s*} \notin B(a^*)$.

Case 2: $\sum_{j=1}^{w} \lambda_j = 0$.

We have: $\tilde{b}_1 = \tilde{b}^{s*} + \lambda$ and $\tilde{b}_2 = \tilde{b}^{s*} - \lambda$. We will show that in state s actions b_1 and b_2, as defined in case 1, are optimal mixed actions for player 2. Let ε be so small that in M_ξ no action outside $\tilde{I}_{b^{s*}} \setminus car(a^*)$ (cf. definition 2.7.2) can, in combination with

some pure actions in other states, form a cycle $\mathbb{C}$ that is a best reply against b_1 and b_2 (it is always possible to choose ε sufficiently small). We have: $K^s\tilde{b}_1 = K^s\tilde{b}_2 = K^s\tilde{b}^{s*}$, so against the rows in K^s actions b_1 and b_2 yield the same one-shot payoff as b^{s*}. Now consider a row $i \in \tilde{I}_{b^{s*}}$ that is not a row of K^s and let λ' be the vector λ extended with zeros such that $b_1 = b^{s*} + \lambda'$ and $b_2 = b^{s*} - \lambda'$. By construction of K^s row i is linearly dependent of the rows in K^s. This means that $e_i M\lambda' = 0$ and consequently that

$$e_i M b_1 = e_i M\left(b^{s*} + \lambda'\right) = e_i M b^{s*}$$

and

$$e_i M b_2 = e_i M\left(b^{s*} - \lambda'\right) = e_i M b^{s*}$$

Now consider the two stationary strategies $y^*_+ = (b^{1*}, \ldots, b^{s-1*}, b_1, b^{s+1*}, \ldots, b^{m*})$ and $y^*_- = \left(b^{1*}, \ldots, b^{s-1*}, b_2, b^{s+1*}, \ldots, b^{m*}\right)$. Strategies y^*_+ and y^*_- are optimal for player 2 and hence $b^{s*} \notin Extr(Y^s)$.

Both cases contradict the fact that $b^{s*} \in Extr(Y^s)$ and consequently the assumption that the columns of M_{a^*} corresponding to $car(b^{s*})$ are linearly dependent, is false. ■

From theorems 2.7.3 and 2.7.4 we can conclude that K^s is a non-singular square matrix, whose rows indeed include $car(a^*)$ and whose columns include $car(b^{s*})$. Theorem 2.7.5 tells us for each state $s \in car(a^*)$ what the outcome of the matrix-vector product $K^s\tilde{b}^{s*}$ is.

Theorem 2.7.5 *For each state $s \in car\,(a^*)$, for the matrix K^s and the vector $\tilde{b}^{s*}$ as constructed on page 41 the following equations hold:*

$$K^s_s\tilde{b}^{s*} = \varphi_0(a^{*\prime}) + \xi \cdot \sum_{i=1}^{m} \left(a^*_i\right)^2 - \xi \tag{2.43}$$

$$K^s_{s'}\tilde{b}^{s*} = \varphi_0(a^{*\prime}) + \xi \cdot \sum_{i=1}^{m} \left(a^*_i\right)^2 + \xi \cdot (a^*_s - a^*_{s'}) \text{ for all } s' \in car(a^*), s' \neq s. \tag{2.44}$$

Proof. Notice that these are equations (2.22) and (2.23) in theorem 2.5.8 and that the rows of K^s outside $car(a^*)$ are not included in (2.43) and (2.44). We will prove that these equations hold. The proof is based on properties of cycles (cf. definition 2.2.16 and lemma 2.2.17) and is closely related to the proof of lemma 2.5.6. First we assume without loss of generality that $car(a^*) = \{1, \ldots, p\}$. For each state $s \in car\,(a^*)$ we introduce the vector $d^s \in \mathbb{R}^p$, which is defined as follows: d^s is the subvector of the vector $K^s\tilde{b}^{s*}$ consisting of its elements corresponding to $car(a^*)$. For $car(a^*) = \{1, \ldots, p\}$ this means that d^s consists of the first p elements of $K^s\tilde{b}^{s*}$. Furthermore

$$\begin{aligned} d &= \left(d^1, \ldots, d^p\right) \\ &= \left(d^1_1, d^1_2, \ldots, d^1_p, d^2_1, d^2_2, \ldots, d^2_p, \ldots, d^p_1, d^p_2, \ldots, d^p_p\right) \in \mathbb{R}^{p^2}. \end{aligned} \tag{2.45}$$

We will show that d^s must equal the right-hand-side of equations (2.43) and (2.44) above.

Notice that for $s, s' \in car(a^*)$ the number $K^s_{s'}\tilde{b}^{s*}$, which is equal to $d^s_{s'}$, is the immediate payoff to player 1 exclusive of the bonus, when he plays action s' in state s against y^*. Player 1 only receives the bonus when he plays action s in state s. Hence $d^s_{s'}$ is equal to the immediate payoff to player 1, when he plays action s' in state s, whenever $s' \neq s$. Furthermore $K^s_s\tilde{b}^{s*}(= d^s_s)$ is equal to player 1's immediate payoff minus the bonus, when he plays action s in state s. Now we can use lemma 2.2.17 to prove that d^s equals the right hand side of equations (2.43) and (2.44).We will do so in terms of cycles (cf. definition 2.2.16). Recall that, in terms of cycles, lemma 2.2.17 states that for each pair of optimal strategies (a^{*1}, y^*) player 1's expected average stage payoff during each cycle within $car\,(a^*)$ equals v_ξ.

Cycles of length 1: A cycle of length 1 occurs if player 1 plays action s in state s. In this particular case he receives the bonus, so his immediate payoff is

$$d^s_s + \xi = K^s_s\tilde{b}^{s*} + \xi.$$

Hence lemma 2.2.17 states that

$$K^s_s\tilde{b}^{s*} + \xi = v_\xi = \varphi_0(a^{*\prime}) + \xi \cdot \sum_{i=1}^{m} (a^*_i)^2$$

for each $s \in \{1, \ldots, p\}$, which is exactly (2.43). The consequences for d are:

$$d^s_s = K^s_s\tilde{b}^{s*} = v_\xi - \xi \text{ for all } s \in \{1, \ldots, p\}. \tag{2.46}$$

Cycles of length more than 1: Take an arbitrary cycle $\mathbb{C}\,(s_1, s_2, \ldots, s_l)$ with $l \geq 2$. The total payoff over the stages in $\mathbb{C}\,(s_1, s_2, \ldots, s_l)$ is (cf. (2.16):

$$\begin{aligned}
\sum_{j=1}^{l} \left(e_{s_{j+1}} M b^{s_j *} + \xi \cdot \delta^{s_j}_{s_{j+1}} \right) &= \sum_{j=1}^{l} \left(e_{s_{j+1}} K^{s_j} \tilde{b}^{s_j *} + \xi \cdot \delta^{s_j}_{s_{j+1}} \right) \\
&= \sum_{j=1}^{l} \left(K^{s_j}_{s_{j+1}} \tilde{b}^{s_j *} + \xi \cdot \delta^{s_j}_{s_{j+1}} \right) \\
&= \sum_{j=1}^{l} d^{s_j}_{s_{j+1}} + \xi \cdot \delta^{s_j}_{s_{j+1}}
\end{aligned}$$

with $s_{l+1} = s_1$. Applying (2.17) to these equations yields:

$$\tfrac{1}{l} \sum_{j=1}^{l} \left(d^{s_j}_{s_{j+1}} + \xi \cdot \delta^{s_j}_{s_{j+1}} \right) = v_\xi. \tag{2.47}$$

Notice that for $l = 1$ equations (2.47) are equivalent to equations (2.46). In particular we have the following subset of equations:

$$d^{s_1}_{s_2} + d^{s_2}_{s_1} = 2 \cdot v_\xi \text{ for each cycle } (s_1, s_2) \text{ with } 1 \leq s_1 < s_2 \leq p \tag{2.48}$$

and

$$d^1_{s_1} + d^{s_1}_{s_2} + d^{s_2}_1 = 3 \cdot v_\xi \text{ for each cycle } (1, s_1, s_2) \text{ with } 2 \leq s_1 < s_2 \leq p. \tag{2.49}$$

It can easily be proved that all equations in (2.47) hold, if the ones in (2.48), (2.49) and (2.46) do.

Recall that, by notation 2.5.5 we have: $\bar{a}^* = (a^*_1, a^*_2, \ldots, a^*_p)$. Because $K^s_{s'} \tilde{b}^{s*} = d^s_{s'}$ for each $s, s' \in car(a^*)$, by (2.36) we have

$$\bar{a}^* d^s = \varphi_0(a^{*\prime}) \text{ for each } s \in \{1, \ldots, p\}. \tag{2.50}$$

We now have the following equations in

$$d = \left(d^1_1, d^1_2, \ldots, d^1_p, d^2_1, d^2_2, \ldots, d^2_p, \ldots, d^p_1, d^p_2, \ldots, d^p_p\right),$$

all of which hold:

$$\begin{cases} d^s_s = v_\xi - \xi \text{ for each } s \in \{1, \ldots, p\} \\ d^{s_1}_{s_2} + d^{s_2}_{s_1} = 2 \cdot v_\xi \text{ for each cycle } (s_1, s_2) \\ \qquad \text{with } 1 \leq s_1 < s_2 \leq p \\ d^1_{s_1} + d^{s_1}_{s_2} + d^{s_2}_1 = 3 \cdot v_\xi \text{ for each cycle } (1, s_1, s_2) \\ \qquad \text{with } 2 \leq s_1 < s_2 \leq p \\ \bar{a}^* d^s = \varphi_0(a^{*\prime}) \text{ for each } s \in \{1, \ldots, p-1\}. \end{cases} \tag{2.51}$$

Notice that, except for the names of the variables, the left hand side of system (2.51) is exactly the same as the left hand side of system (2.21). Consequently, when writing system 2.51 as a matrix-vector product, again we find the characteristic matrix H for repeated games with bonus ξ, which is non-singular and therefore system 2.51 of p^2 equations in p^2 variables $d^1_1, \ldots, d^p_p$, has a unique solution.
Recall (cf. (2.7)) that

$$\varphi_0(a^{*\prime}) = v_\xi - \xi \cdot \sum_{i=1}^{m} \left(a^*_i\right)^2$$

and, by 2.15, that

$$\sum_{i=1}^{m} \left(a^*_i\right)^2 = \sum_{i=1}^{p} \left(\bar{a}^*_i\right)^2$$

Some notations: Let $c \in \mathbb{R}^{p^2}$ be vector such that $Hd = c$ is system (2.51). The vector c, being the right-hand side of this set of equations can be split into a vector c_1 that consists of constants times v_ξ and a vector c_2 that consists of terms that do not depend on v_ξ. An example: for $p = 3$ we have:

$$c = \begin{pmatrix} v_\xi - \xi \\ v_\xi - \xi \\ v_\xi - \xi \\ 2v_\xi \\ 2v_\xi \\ 2v_\xi \\ 3v_\xi \\ \varphi_0(a^{*\prime}) \\ \varphi_0(a^{*\prime}) \end{pmatrix} = \begin{pmatrix} v_\xi \\ v_\xi \\ v_\xi \\ 2v_\xi \\ 2v_\xi \\ 2v_\xi \\ 3v_\xi \\ v_\xi \\ v_\xi \end{pmatrix} + \begin{pmatrix} -\xi \\ -\xi \\ -\xi \\ 0 \\ 0 \\ 0 \\ 0 \\ -\xi \cdot \sum_{i=1}^{m} \left(a^*_i\right)^2 \\ -\xi \cdot \sum_{i=1}^{m} \left(a^*_i\right)^2 \end{pmatrix} = c_1 + c_2.$$

Let $d = \beta_1 + \beta_2$ such that $H\beta_1 = c_1$ and $H\beta_2 = c_2$. Some elementary calculations show that then

$$\beta_1 = v_\xi \cdot (1, 1, \ldots, 1)$$

and

$$(\beta_2)^s_s = -\xi \text{ and } (\beta_2)^s_{s'} = \xi \cdot (a^*_s - a^*_{s'})$$

for each $s, s' \in \{1, \ldots, p\}, s' \neq s$. But then for each $s \in \{1, \ldots, p\}$:

$$\begin{aligned} K^s_s \tilde{b}^{s*} = d^s_s &= (\beta_1)^s_s + (\beta_2)^s_s = v_\xi - \xi \\ &= \varphi_0(a^{*\prime}) + \xi \cdot \sum_{i=1}^{m} (a^*_i)^2 - \xi \end{aligned}$$

and

$$\begin{aligned} K^s_{s'} \tilde{b}^{s*} = d^s_{s'} &= (\beta_1)^s_{s'} + (\beta_2)^s_{s'} = v_\xi + \xi \cdot (a^*_s - a^*_{s'}) \\ &= \varphi_0(a^{*\prime}) + \xi \cdot \sum_{i=1}^{m} (a^*_i)^2 + \xi \cdot (a^*_s - a^*_{s'}) \text{ for all } s' \neq s,\ s' \in car(a^*), \end{aligned}$$

which is exactly (2.43) and (2.44). ■

Theorems 2.7.3, 2.7.4 and 2.7.5 together show that in theorem 2.5.8 point *1* implies point *2*. Theorem 2.7.6 states that in theorem 2.5.8 also statement *2* implies statement *1*.

Theorem 2.7.6 *If for the pair of optimal strategies $(a^{*\prime}, y^*)$ for each state $s \in car(a^*)$ there exist a subvector $\tilde{b}^{s*}$ of b^{s*} and a non-singular square submatrix K^s of M with properties (2.22) and (2.23), where the rows of K^s include $car(a^*)$ and the columns of K^s include $car(b^{s*})$, then $(b^{i_1*}, \ldots, b^{i_p*})$ is an extreme point of Y^{a^*} and $\varphi_0(a^{*\prime}) \neq 0$.*

Proof. Let $(a^{*\prime}, y^*)$ be a pair of (stationary) optimal strategies. For each state $s \in car(a^*)$ take a subvector $\tilde{b}^{s*}$ of b^{s*} and a non-singular square submatrix K^s of M such that (2.22) and (2.23) are satisfied.

We will first prove that $\varphi_0(a^{*\prime}) \neq 0$:

Recall that, according to (2.4), in M_ξ the simple strategy $a^{*\prime}$ makes player 2 indifferent between all of his (mixed and pure) actions in $B_\xi(a^{*\prime})$ and that each column of K^s is in the set $B_\xi(a^{*\prime})$. Let $\tilde{1}$ be the vector consisting of only ones. Then we have: $\tilde{a}^* K^s = \varphi_0(a^{*\prime}) \cdot \tilde{1}$, so $\tilde{a}^* = \tilde{a}^* K^s (K^s)^{-1} = \varphi_0(a^{*\prime}) \cdot \tilde{1}^T (K^s)^{-1}$ and $1 = \tilde{a}^* \tilde{1} = \varphi_0(a^{*\prime}) \cdot \tilde{1}^T (K^s)^{-1} \tilde{1}$, so

$$\varphi_0(a^{*\prime}) = \frac{1}{\tilde{1}^T (K^s)^{-1} \tilde{1}} \tag{2.52}$$

and

$$\tilde{a}^* = \frac{\tilde{1}^T (K^s)^{-1}}{\tilde{1}^T (K^s)^{-1} \tilde{1}} \tag{2.53}$$

(cf. Shapley & Snow (1950)), which means that $\varphi_0(a^{*\prime}) \neq 0$.

Now we prove that $(b^{i_1*}, \ldots, b^{i_p*})$ is an extreme point of Y^{a^*}.

According to theorems 2.7.3, 2.7.4 and 2.7.5 for each state $s \in car(a^*)$ there is a number $N(s) \in \mathbb{N}$, there exist vectors $w_1^s, w_2^s, \ldots, w_{N(s)}^s \in Extr(Y^s)$ and there exist scalars $\lambda_1, \ldots, \lambda_{N(s)}$ with $\lambda_i > 0$ and $\sum\limits_{i=1}^{N(s)} \lambda_i = 1$ such that $\tilde{b}^{s*} = \lambda_1 w_1^s + \lambda_2 w_2^s + \ldots + \lambda_{N(s)} w_{N(s)}^s$. So $\tilde{b}^{s*}$ is a convex combination of the w_i^s's and

$$car(w_i^s) \subset car(\tilde{b}^{s*}) \text{ for each } i \in \{1, \ldots, N(s)\}. \tag{2.54}$$

Since $w_i^s \in Extr(Y^s)$, theorems 2.7.3 and 2.7.4 say that there exists a non-singular submatrix K_i^s of M, where the rows of K^s include $car(a^*)$ and the columns of K^s include $car(b^{s*})$. Consider the matrix K^s. K^s is a non-singular submatrix of M, whose rows include $car(a^*)$ and whose columns include $car\,(w_i^s)$. Consequently we can take $K_i^s = K^s$. Theorem 2.7.5 says about K_i^s and w_i^s that for each state $s \in car(a^*)$ we have:

$$K_{is}^s w_i^s = K_s^s w_i^s = \varphi_0(a^{*\prime}) + \xi \cdot \sum_{i=1}^{m} (a_i^*)^2 - \xi$$

and

$$K_{is'}^s w_i^s = K_{s'}^s w_i^s = \varphi_0(a^{*\prime}) + \xi \cdot \sum_{i=1}^{m} (a_i^*)^2 + \xi \cdot (a_s^* - a_{s'}^*).$$

But then: $K^s \tilde{b}^{s*} = K^s w_i^s$. Since K^s is non-singular, this means that $\tilde{b}^{s*} = w_i^s$ for each $s \in car(a^*)$ and hence $(b^{i_1*}, \ldots, b^{i_p*}) \in Extr(Y^{a^*})$. ■

Chapter 3

Zero-sum Games With Vanishing Actions

3.1 Introduction to repeated games with vanishing actions

Unlearning is a phenomenon that regularly occurs in everyday situations. Consider the following situation: A surgeon, who has just finished medical school, learned all the operational skills just a short time ago and he is very likely to be able to perform any operation within his field of expertise extremely well. However, if during the first couple of years in practice he has never had the opportunity to perform a splenectomy, he might not feel so sure anymore about his capabilities regarding this particular operation, when he is suddenly asked to perform one.

Notice that, in this example, the surgeon does not learn anything during his period in practice; he merely performs a number of his skills. However, there is a risk that he might unlearn some of the skills that he does not need to carry out. This type of example fits in with the model that we present in this chapter: a model that exclusively deals with unlearning. The idea is that at stage 0, the start of the game, each player knows "everything", just like the surgeon right after finishing medical school, but as time goes by, some actions might be forgotten. This forgetting, or unlearning, of actions can be modelled in various ways. In this chapter we use the following model, in which the consequences of unlearning are quite severe: If player k has not played (pure) action i at any of the previous r^k stages, then he has unlearned that action, which means that he can not play it anymore. Since there is no learning embedded in the model, once a player has unlearned an action during the course of the game, he will never be able to play it again. Consequently action i practically disappears from player k's action set as soon as he has not played it for r^k consecutive stages. The assumption that at stage 0 each player knows everything, entails that at any stage $t \leq r^k$ player k does not unlearn any actions. The model as described here is called the model of *repeated games with vanishing actions* and an N-player game that fits in with this model is called an $\left(r^1, r^2, \ldots, r^N\right)$*-restricted game*.

In section 3.2 we present the formal model of repeated games with vanishing

actions. This model was introduced by Joosten, Peters and Thuijsman (1995) for zero-sum games. We will present their results as well as a new theorem in section 3.3. In chapter 4 we discuss a model on coordination games with vanishing actions that was introduced by Schoenmakers, Flesch and Thuijsman (2002). Here an N-player coordination game is a repeated game, where all players have the same number of actions and where all non-diagonal payoffs are 0. Moreover all diagonal payoffs are assumed to be strictly positive. An interesting resemblance between zero-sum games and coordination games is the fact that the payoffs of the players are very much correlated; only in coordination games this correlation is positive, whereas in zero-sum games it is negative. Since ordinary coordination games are just a special type of repeated games, the Folk-theorem applies, i.e. all feasible (and individually rational) rewards can be obtained as equilibrium rewards (cf. theorem 1.2.7). In section 4.3 a specific type of strategies, the so-called agreements, are introduced. This type of strategies will also be used in general-sum games in section 5.4, albeit in a somewhat more complex way. In section 4.4 it is exhibited that in N-player $\left(r^1, r^2, \ldots, r^N\right)$-restricted coordination games with $r^k \geq 3$ for all k the Folk-theorem applies, a result that does not hold for the general-sum case, which is the topic in chapter 5.

3.2 Repeated games with vanishing actions: The model

Definition 3.2.1 *A 2-player $\left(r^1, r^2\right)$-restricted game is determined by the following parameters:*

1. *$K = \{1, 2\}$ is the set of players;*
2. *$I = \{1, 2, \ldots, m\}$ is the initial set of pure actions for player 1;*
3. *$J = \{1, 2, \ldots, n\}$ is the initial set of pure actions for player 2;*
4. *$R^k : I \times J \to \mathbb{R}$ is the payoff function for player $k \in K$;*
5. *r^k is the level of unlearning for player $k \in K$.*

Actions disappear according to the following rule: At any stage beyond r^1, player 1's action $i \in I$ will vanish from his action set, when he has not played it at any of the previous r^1 stages. Thus the number of available actions may decrease during the course of play. Therefore we shall use I_t to denote the set of feasible or not (yet) unlearned pure actions of player 1 at stage t. At stage t player 1 is allowed to randomize over the available actions in I_t. A similar argumentation holds for player 2.

As evaluation criterion for the stream of payoffs we will again use the *limiting average reward,* i.e.

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

where $E(R_t^k)$ denotes the expected payoff to player k at stage t given that (π, σ) is being played. In the 2-player zero-sum case that is the subject of this chapter, analogously to the model in chapter 2 player 1 maximizes his limiting average reward $\gamma(\pi, \sigma)$ and player 2 minimizes the same expression.

Remark 3.2.2 *Notice that a $(1, 1, \ldots, 1)$-restriction implies that the game is essentially a one-shot game, and therefore, every equilibrium in the one-shot game is an equilibrium in the $(1, 1, \ldots, 1)$-restricted game and vice versa.*

3.3 Zero-sum games with vanishing actions

Zero-sum repeated games with vanishing actions were introduced by Joosten, Peters and Thuijsman (1995). We will briefly state their main results in lemma 3.3.2.

Notation 3.3.1 *Consider the $(m \times n)$-matrix game M. Then $I = \{1, \ldots, m\}$ and $J = \{1, \ldots, n\}$. Recall that v denotes the value of M. The value of the corresponding $\left(r^1, r^2\right)$-restricted zero-sum game is denoted by v_{r^1, r^2}.*

The matrix game M has a saddle point at (i, j) if $m_{i'j} \leq m_{ij} \leq m_{ij'}$ for all i' and for all j'.

Lemma 3.3.2 *Let $M \in \mathbb{R}^{m \times n}$ without saddle points. Then*

1. *$v_{1,1} = v$. This result is also a corollary of remark 3.2.2.*
2. *For $r^2 \geq 2$ we have: $v_{1,r^2} = \underline{v} := \max\limits_i \min\limits_j m_{ij}$ and analogously for $r^1 \geq 2$ we have: $v_{r^1,1} = \overline{v} := \min\limits_j \max\limits_i m_{ij}$.*
3. *If $m = 2$ or $n = 2$, then $v_{2,2} = \frac{1}{2}\overline{v} + \frac{1}{2}\underline{v}$.*
4. *For $r^2 \geq 3$ we have: $v_{2,r^2} = \underline{v}$ and analogously for $r^1 \geq 3$ we have: $v_{r^1,2} = \overline{v}$.*
5. *Let $M = \begin{pmatrix} \mathfrak{a} & \mathfrak{b} \\ \mathfrak{c} & \mathfrak{d} \end{pmatrix}$ and suppose without loss of generality that $\min\{\mathfrak{a}, \mathfrak{d}\} > \max\{\mathfrak{b}, \mathfrak{c}\}$. Then $v_{3,3} = median\left\{\frac{1}{4}(\mathfrak{a} + \mathfrak{b} + \mathfrak{c} + \mathfrak{d}), \underline{v}, \overline{v}\right\}$.*
6. *Algorithms are presented to calculate $v_{2,2}$ and $v_{3,3}$ for games of arbitrary size.*

We will describe the $\left(r^1, r^2\right)$-restricted zero-sum repeated game of size $m \times n$ as a special type of stochastic game. For that purpose we define a state space

$$S = \{\left(s_1^1, s_2^1, \ldots, s_m^1; s_1^2, s_2^2, \ldots, s_n^2\right) \mid \begin{array}{r} s_i^1 \in \left\{0, 1, \ldots, r^1\right\} \text{ for all } i, \\ s_j^2 \in \left\{0, 1, \ldots, r^2\right\} \text{ for all } j\} \end{array} \tag{3.1}$$

Let state $s = \left(s_1^1, s_2^1, \ldots, s_m^1; s_1^2, s_2^2, \ldots, s_n^2\right)$ be visited at stage t. Then $s_i^1 < r^1 < \infty$ is to be interpreted as follows: action $i \in \{1, \ldots, m\}$ was selected by player 1 at stage $t - s_i^1$ for the last time (notice that if player 1 plays action i at stage t, then $s_i^1 = 0$). This means, in view of the restriction r^1, that action i is still available to player 1 and hence $i \in I_t$. However, if $s_i^1 = r^1$, then action i was not selected by player 1

for at least r^1 consecutive stages and player 1 has unlearned action i, so $i \notin I_t$. The numbers s_j^2 are to be interpreted similarly. The transitions in this stochastic game are defined in the obvious way, whereas the rewards are defined analogously to the original game.

State s is defined this way merely in order to be able to see which actions are unlearned and within how many stages the not yet unlearned actions might be unlearned. When a player is not restricted, this is not a relevant issue and hence this construction of a set of states does not depend on the actions played by an unrestricted player. So if $r^1 = \infty$, i.e. player 1 is unrestricted and he never unlearns an action, then $I_t = I$ for all t and the state that is visited at stage t depends only on the actions played by player 2 at stages $1, \ldots, t-1$. Notice that if both players are unrestricted, then the state space consists of only one state and the game reduces to an ordinary zero-sum repeated game.

We will now calculate the value of 2×2 games without saddle points, in which exactly one of the two players is restricted. Let

$$M = \begin{array}{c} \\ T \\ B \end{array} \begin{array}{c} \begin{array}{cc} L & R \end{array} \\ \begin{pmatrix} \mathfrak{a} & \mathfrak{b} \\ \mathfrak{c} & \mathfrak{d} \end{pmatrix} \end{array}$$

and suppose without loss of generality that $\min\{\mathfrak{a}, \mathfrak{d}\} > \max\{\mathfrak{b}, \mathfrak{c}\}$.

The following theorem provides for the (r, ∞)-restricted and for the (∞, r)-restricted game the value as well as an optimal strategy for the restricted player.

Theorem 3.3.3 *Consider the (r, ∞)-restricted game corresponding to the matrix game M. If $v_{r,\infty} > \max\{\mathfrak{b}, \mathfrak{c}\}$, then player 1's strategy x prescribing to play $(a_1, a_2) = (\frac{\mathfrak{d}-\mathfrak{c}}{\mathfrak{a}-\mathfrak{b}-\mathfrak{c}+\mathfrak{d}}, \frac{\mathfrak{a}-\mathfrak{b}}{\mathfrak{a}-\mathfrak{b}-\mathfrak{c}+\mathfrak{d}})$ as long as none of his actions is about to vanish and to play the action that is about to vanish otherwise, is optimal. Furthermore $v_{r,\infty} = \max\{\mathfrak{b}, \mathfrak{c}, g^1\}$ where*

$$g^1 = \frac{v \cdot (\mathfrak{a} - \mathfrak{b} - \mathfrak{c} + \mathfrak{d})^r - \mathfrak{d} \cdot (\mathfrak{a} - \mathfrak{b})^r - \mathfrak{a} \cdot (\mathfrak{d} - \mathfrak{c})^r}{(\mathfrak{a} - \mathfrak{b} - \mathfrak{c} + \mathfrak{d})^r - (\mathfrak{a} - \mathfrak{b})^r - (\mathfrak{d} - \mathfrak{c})^r}.$$

In the (∞, r)-restricted game player 2 has a similar optimal strategy if $v_{\infty,r} < \min\{\mathfrak{a}, \mathfrak{d}\}$ and $v_{\infty,r} = \min\{\mathfrak{a}, \mathfrak{d}, g^2\}$ where

$$g^2 = \frac{v \cdot (\mathfrak{a} - \mathfrak{b} - \mathfrak{c} + \mathfrak{d})^r - \mathfrak{b} \cdot (\mathfrak{a} - \mathfrak{c})^r - \mathfrak{c} \cdot (\mathfrak{d} - \mathfrak{b})^r}{(\mathfrak{a} - \mathfrak{b} - \mathfrak{c} + \mathfrak{d})^r - (\mathfrak{a} - \mathfrak{c})^r - (\mathfrak{d} - \mathfrak{b})^r}.$$

Proof. Consider the stochastic game representation of the (r, ∞)-restricted game. We have:

$$S = \{(0,0), (0,1), \ldots, (0,r), (1,0), (2,0), \ldots, (r,0)\}.$$

State $(0,0)$ is the initial state, since at stage 0 player 1 "knows everything". Furthermore states $(0,r)$ and $(r,0)$ are absorbing states. If ever state $(0,r)$ is reached, then player 1 has only action T left and player 2 will play action R at each stage, leading to a reward of $\mathfrak{b}$. Since player 1 can, by unlearning action B, enforce play to

enter state $(0,r)$ and thereby to obtain a reward of $\mathfrak{b}$, we have $v_{r,\infty} \geq \mathfrak{b}$. Similarly if play absorbs in state $(r,0)$, then the reward will be $\mathfrak{c}$ and player 1 can enforce this, so $v_{r,\infty} \geq \mathfrak{c}$. Now consider the strategy x: x induces a Markov chain over the states in S, where from state $(0,l)$ with probability 1 play moves to state $(0,l+1)$ if player 1 plays action T and with probability 1 play moves to state $(1,0)$ if player 1 plays action B. Since player 1 plays action T with probability a_1, the former transition occurs with probability a_1 and the latter one with probability a_2. In state $(0,r-1)$ strategy x prescribes to play B with probability 1, since otherwise it vanishes, and play moves to state $(1,0)$ with probability 1. Furthermore from state $(l,0)$ play moves with probability 1 to state $(l+1,0)$ if player 1 plays action B and with probability 1 to state $(0,1)$ if player 1 plays action T. Since player 1 plays action B with probability 1, the former transition occurs with probability a_2 and the latter with probability a_1. In state $(r-1,0)$ the by x prescribed action is T with probability 1, since otherwise it vanishes, and play moves to state $(0,1)$ with probability 1.

Notice that this stochastic game has the single-controller property (cf. definition 1.2.2). Filar (1981) states that for single-controller stochastic games both players possess stationary optimal strategies (theorem 1.2.5). Furthermore, if player 1 plays x, then the absorbing states $(0,r)$ and $(r,0)$ are never reached and the strategy x, extended with playing actions T and B in states $(0,r)$ and $(r,0)$ respectively, is a stationary strategy. Write $x = \times_{s\in S} x^s$ with

$$x^s = \begin{cases} (a_1,a_2) \text{ for } s \in S\backslash\{(0,r-1),(0,r),(r-1,0),(r,0)\}, \\ (0,1) \text{ for } s = (0,r-1), \\ (1,0) \text{ for } s = (r-1,0), \\ (1) \text{ for } s \in \{(0,r),(r,0)\}. \end{cases}$$

From Vrieze (1987) we deduce that

1. For each state $(0,l)$ with $0 \leq l \leq r-1$ there exists a number $w^{(0,l)}$ such that

$$w^{(0,l)} + v_{r,\infty}^{(0,l)} = val \begin{pmatrix} \mathfrak{a} + w^{(0,l+1)} & \mathfrak{b} + w^{(0,l+1)} \\ \mathfrak{c} + w^{(1,0)} & \mathfrak{d} + w^{(1,0)} \end{pmatrix}. \tag{3.2}$$

2. For each state $(l,0)$ with $0 \leq l \leq r-1$ there exists a number $w^{(l,0)}$ such that

$$w^{(l,0)} + v_{r,\infty}^{(l,0)} = val \begin{pmatrix} \mathfrak{a} + w^{(0,1)} & \mathfrak{b} + w^{(0,1)} \\ \mathfrak{c} + w^{(l+1,0)} & \mathfrak{d} + w^{(l+1,0)} \end{pmatrix}. \tag{3.3}$$

3. For state $(0,r)$ we have

$$w^{(0,r)} + v_{r,\infty}^{(0,r)} = val \begin{pmatrix} \mathfrak{a} + w^{(0,r)} & \mathfrak{b} + w^{(0,r)} \end{pmatrix} = \mathfrak{b} + w^{(0,r)} \tag{3.4}$$

   and hence $v_{r,\infty}^{(0,r)} = \mathfrak{b}$.

4. For state $(r,0)$ we have

$$w^{(r,0)} + v_{r,\infty}^{(r,0)} = val \begin{pmatrix} \mathfrak{c} + w^{(r,0)} & \mathfrak{d} + w^{(r,0)} \end{pmatrix} = \mathfrak{c} + w^{(r,0)} \tag{3.5}$$

   and hence $v_{r,\infty}^{(r,0)} = \mathfrak{c}$.

5. A stationary optimal strategy for player 1 prescribes to play in state s a mixed action that is optimal in the matrix game mentioned in the equation in statement 1, 2, 3 or 4 that corresponds to state s.

Let x^* be a stationary optimal strategy for player 1. There are 3 options:

- If $v_{r,\infty} = \mathfrak{b}$, then any stationary strategy that, eventually, leads to absorption in state $(0, r)$, yields a reward of $\mathfrak{b}$ and is thereby optimal.
- If $v_{r,\infty} = \mathfrak{c}$, then any stationary strategy that, eventually, leads to absorption in state $(r, 0)$, yields a reward of $\mathfrak{c}$ and is thereby optimal.
- If $v_{r,\infty} > \max\{\mathfrak{b}, \mathfrak{c}\}$, then x^* prescribes to keep both actions available. We argue that x^* prescribes to play a completely mixed action in state $(0, r-2)$. If x^* would prescribe to play action B with probability 1, then state $(0, r-1)$ would never be visited and player 1 would have no advantage of having an r-restriction instead of an $(r-1)$-restriction. A similar argumentation can be held for state $(r-2, 0)$. Therefore x^* prescribes to play a completely mixed action in all states $(0, l)$ and $(l, 0)$ as long as $l \leq r-2$. Furthermore in state $(0, r-1)$ clearly x^* prescribes to play action B with probability 1 and in state $(r-1, 0)$ it prescribes to play action T with probability 1, since otherwise play will, eventually, absorb in one of the states $(0, r)$ and $(r, 0)$. According to (3.2) and (3.3) this means that the numbers w^s are such that in the matrix games

$$\begin{pmatrix} \mathfrak{a} + w^{(0,l+1)} & \mathfrak{b} + w^{(0,l+1)} \\ \mathfrak{c} + w^{(1,0)} & \mathfrak{d} + w^{(1,0)} \end{pmatrix} \text{ and } \begin{pmatrix} \mathfrak{a} + w^{(0,1)} & \mathfrak{b} + w^{(0,1)} \\ \mathfrak{c} + w^{(l+1,0)} & \mathfrak{d} + w^{(l+1,0)} \end{pmatrix}$$

  player 1 has a unique completely mixed optimal action as long as $l \leq r-2$. This mixed action then is $(\frac{\mathfrak{d}-\mathfrak{c}}{\mathfrak{a}-\mathfrak{b}-\mathfrak{c}+\mathfrak{d}}, \frac{\mathfrak{a}-\mathfrak{b}}{\mathfrak{a}-\mathfrak{b}-\mathfrak{c}+\mathfrak{d}}) = (a_1, a_2)$ in all of these states. Notice that this action does not depend on the w's. Furthermore in

$$\begin{pmatrix} \mathfrak{a} + w^{(0,r)} & \mathfrak{b} + w^{(0,r)} \\ \mathfrak{c} + w^{(1,0)} & \mathfrak{d} + w^{(1,0)} \end{pmatrix} \text{ and } \begin{pmatrix} \mathfrak{a} + w^{(0,1)} & \mathfrak{b} + w^{(0,1)} \\ \mathfrak{c} + w^{(r,0)} & \mathfrak{d} + w^{(r,0)} \end{pmatrix}$$

  player 1 has a pure optimal action, namely B and T respectively. Consequently, x^* prescribes to play (a_1, a_2) as long as no action is about to vanish and to play the action that is about to vanish otherwise. Hence $x^* = x$. Notice that under the stationary optimal strategy x^* there are 3 ergodic classes, one consisting of $(0, r)$ only, one consisting of $(r, 0)$ only and one consisting of the states in $S \setminus \{(0,0), (0,r), (r,0)\}$. It is well-known that for each $s, s' \in S \setminus \{(0,0), (0,r), (r,0)\}$ we have $v^s_{r,\infty} = v^{s'}_{r,\infty}$. Furthermore, since the initial state is $(0, 0)$, at stage 2 play enters state $(0, 1)$ with probability a_1 and state $(1, 0)$ with probability $a_2 = 1 - a_1$, both of which are in the third ergodic set. This means that $v^{(0,0)}_{r,\infty} = v_{r,\infty} = v^s_{r,\infty}$ for all $s \in S \setminus \{(0,0), (0,r), (r,0)\}$. But then some calculations show that in order for the equations in statements (3.2), (3.3), (3.4) and (3.5) to hold, we must have:

$$v_{r,\infty} = \frac{v \cdot (\mathfrak{a} - \mathfrak{b} - \mathfrak{c} + \mathfrak{d})^r - \mathfrak{d} \cdot (\mathfrak{a} - \mathfrak{b})^r - \mathfrak{a} \cdot (\mathfrak{d} - \mathfrak{c})^r}{(\mathfrak{a} - \mathfrak{b} - \mathfrak{c} + \mathfrak{d})^r - (\mathfrak{a} - \mathfrak{b})^r - (\mathfrak{d} - \mathfrak{c})^r} =: g^1.$$

Now when we combine the results we found for each of the three options, we obtain that $v_{r,\infty} = \max\left\{\mathfrak{b}, \mathfrak{c}, g^1\right\}$ and that, if $v_{r,\infty} = g^1 > \max\{\mathfrak{b}, \mathfrak{c}\}$, then x is an optimal strategy for player 1, which completes the proof of the (r,∞)-restricted game. The proof for the (∞, r)-restricted game is analogous. ■

We finish the chapter with an example that shows that not always $v_{r,\infty} = g^1$.

Example 3.1

Consider the following (r,∞)-restricted game:

$$\begin{pmatrix} 2^{r+1}-1 & 2^r-1 \\ 0 & 2^r \end{pmatrix}.$$

We find

$$g^1 = \frac{2^r - 3\frac{1}{2} + \left(\frac{1}{2}\right)^r}{1 - \left(\frac{1}{2}\right)^{r-1}} < 2^r - 1 = \mathfrak{b}$$

and hence for player 1 it is optimal to unlearn his second action. □

Chapter 4

Coordination Games With Vanishing Actions

4.1 Introduction

Chapter 3 provided the general model of repeated games with vanishing actions. Afterwards zero-sum repeated games with vanishing actions were analyzed. In this chapter we subject a different type of games to the vanishing actions model, namely coordination games, which are defined below. In this chapter we investigate N-player games as well as 2-player games. Consequently we need some notations:

Notation 4.1.1 *In N-player repeated games we have:*

1. $K = \{1, 2, \ldots, N\}$ *is the set of players;*
2. i^k *is a pure action for player $k \in K$;*
3. $I^k = \left\{1, 2, \ldots, m^k\right\}$ *is the set of pure actions for player $k \in K$;*
4. a^k *is a mixed action for player $k \in K$;*
5. A^k *is the set of mixed actions for player $k \in K$;*
6. $R^k : \times_{\kappa=1}^{N} I^\kappa \to \mathbb{R}$ *is the payoff function for player $k \in K$;*
7. π^k *is a strategy for player $k \in K$.*

Definition 4.1.2 *A coordination game is a game, in which each player has the same set of actions: $I^k = \{1, 2, \ldots, m\}$ for all $k \in K$. Furthermore the players have to coordinate their actions in order to receive a strictly positive payoff. If not all players coordinate their actions, then each player receives a payoff of zero. Formally:*

$$R^k\left(i^1, i^2, \ldots, i^N\right) > 0 \text{ if } i^1 = i^2 = \ldots = i^N$$

and

$$R^k\left(i^1, i^2, \ldots, i^N\right) = 0 \text{ otherwise.}$$

4.2 $(2,2)$-restricted coordination games

In this section we analyze $(2,2)$-restricted 2×2 - coordination games. We thoroughly investigate an example and characterize the set of equilibrium rewards in this example. It appears that, paradoxically, although the game in the example leaves hardly any possibilities for the players to randomize over their actions, pairs of equilibrium strategies may be very complex. However, this also means that although the number of pure strategy equilibrium rewards is only 3, with the aid of non-pure strategies uncountably many equilibrium rewards can be obtained.

Example 4.1

Consider the following $(2,2)$-restricted coordination game $\mathcal{G}$:

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

Here T stands for top, B for bottom, L for left and R for right. Analogously the entries of $\mathcal{G}$ will be called TL, TR, BL and BR. We will determine the set of equilibrium rewards in $\mathcal{G}$ with the aid of a number of lemmas:

Lemma 4.2.1 *If at stage 1 one of the entries TL or BR is selected, then during the course of play as soon as one of the entries TR or BL is selected, one of the players immediately unlearns an action.*

Proof. Suppose that, up to stage t, only entries TL and BR have been selected and that at stage $t+1$ entry TR is selected (the proof for BL is similar). Then, if TL was played at stage t, then player 1 unlearns action B at stage $t+1$, whereas if entry BR was played at stage t, then player 2 unlearns action L. ■

We consider pairs of equilibrium strategies.

Lemma 4.2.2 *With respect to any equilibrium (π, σ) in $\mathcal{G}$, if at stage 1 one of the entries TL or BR is selected, then there is a stage after which, with probability 1, entries TR and BL are never played.*

Proof. Let (π, σ) be a pair of equilibrium strategies in $\mathcal{G}$. If the action pairs TR and BL are never selected, then evidently TR and BL are played with frequency 0. Suppose that, by (π, σ), stage $t \geq 2$ is the first stage at which entry TR is selected. Then by lemma 4.2.1 one of the players unlearns an action at stage t. Suppose without loss of generality that player 2 unlearns action L at stage t. Then the 2 action pairs that are still available, are TR and BR with respective payoffs of $(0,0)$ and $(1,3)$. Strategy π, being a best reply against σ, must prescribe to, eventually, unlearn action T and in the long run action pair BR is played with frequency 1 and hence action pairs TR and BL are played with long-run frequency 0. Similar arguments are applicable, if player 1 unlearns an action or if at a certain stage entry BL is selected. ■

Lemma 4.2.2 can be interpreted as follows: If (π, σ) is a pair of equilibrium strategies in $\mathcal{G}$ and, by (π, σ), at stage 1 one of the entries TL or BR is selected, then the total number of stages in which one of the cells TR or BL is selected, is finite.

Lemma 4.2.3 *If (π, σ) is a pair of equilibrium strategies in $\mathcal{G}$ and, in accordance with (π, σ), at stage 1 entry TL is selected and at stage 2 σ prescribes to play both action L and action R with a positive probability, then only $(2,1)$ can occur as equilibrium reward. Similar statements are applicable if at stage 2 player 1's strategy π prescribes to play actions T and B each with a positive probability and if at stage 1 entry BR is selected (then with reward $(1,3)$).*

Proof. Let (π, σ) be a pair of equilibrium strategies in $\mathcal{G}$ and let, in accordance with (π, σ), at stage 1 entry TL be selected. Suppose that at stage 2 player 2's strategy σ prescribes to play both action L and action R with a strictly positive probability. If player 2 selects action L at stage 2, then he unlearns action R and by lemma 4.2.2 the long-run frequency of the action pair TL is 1 with reward $(2,1)$. Since σ is a best replay against π, player 2's expected reward, when he plays action R at stage 2, must also be 1. Suppose that player 2 plays action R at stage 2. Then, applying lemma 4.2.2, we know that there is an $\alpha \in [0,1]$ such that the expected long-run frequencies of the action pairs is α for action pair TL and $(1-\alpha)$ for BR. But then the expected reward to player 2 for playing action R at stage 2 is $\alpha \cdot 1 + (1-\alpha) \cdot 3$, which is only equal to 1 if $\alpha = 1$. Hence, since (π, σ) is an equilibrium, the expected reward if player 2 plays action R is $(2,1)$, which completes the proof. ■

Lemma 4.2.4 *If (π, σ) is a pair of equilibrium strategies in $\mathcal{G}$ and, in accordance with (π, σ), at stages $1, 2, \ldots, t$ the entries TL and BR are selected alternately, where at stage t entry TL is selected, and at stage $t+1$ σ prescribes to play both action L and action R with a positive probability, then only $(2,1)$ can occur as equilibrium reward. Similar statements are applicable if at stage $t+1$ player 1's strategy π prescribes to play actions T and B each with a positive probability and if at stage t entry BR is selected (then with reward $(1,3)$).*

Proof. Analogously to the proof of lemma 4.2.3. ■

Theorem 4.2.5 *If (π, σ) is a pair of equilibrium strategies in $\mathcal{G}$ and, in accordance with (π, σ), at stage 1 one of the entries TL or BR is selected, then only $(2,1)$, $(1,3)$ and $(\frac{3}{2}, 2)$ can occur as equilibrium rewards.*

Proof. Let (π, σ) be a pair of equilibrium strategies in $\mathcal{G}$ and let, in accordance with (π, σ), at stage 1 one of the entries TL or BR be selected. Then by lemma 4.2.2 the action pairs TR and BL are played only at a finite number of stages. Furthermore, if ever at any stage π and/or σ prescribe to play both actions with a positive probability, then by lemma 4.2.4 only $(2,1)$ or $(1,3)$ can occur as equilibrium rewards. Now if (π, σ) never prescribes to play both actions with a positive probability, then there are 3 possible scenarios:

- There are 2 consecutive stages, at which (π, σ) prescribes to play the action pair TL. Then the players unlearn actions B and R and TL will be played with long-run frequency 1, leading to a reward of $(2,1)$.
- There are 2 consecutive stages, at which (π, σ) prescribes to play the action pair BR. Then the players unlearn actions T and L and BR will be played with long-run frequency 1, leading to a reward of $(1,3)$.

- The action pairs TL and BR are played alternately. Then both TL and BR are played with long-run frequency $\frac{1}{2}$ and a reward of $\frac{1}{2} \cdot (2,1) + \frac{1}{2} \cdot (1,3) = (\frac{3}{2}, 2)$ is acquired.

This completes the proof. ■

The next part of the analysis concerns the equilibrium rewards in $\mathcal{G}$ that, from stage 2 on, can be obtained, after at stage 1 one of the entries TR or BL is selected.

Analogously to lemma 4.2.1:

Lemma 4.2.6 *If at stage 1 one of the entries TR or BL is selected, then during the course of play as soon as one of the entries TL or BR is selected, one of the players immediately unlearns an action.*

Lemma 4.2.7 *If at stage 1 one of the entries TR or BL is selected, then from stage 2 on $(0,0)$ is obtainable as an equilibrium reward.*

Proof. Let (π, σ) be a pair of equilibrium strategies in $\mathcal{G}$ and let, in accordance with (π, σ), at stage 1 one of the entries TR or BL be selected. From stage 2 on let (π, σ) prescribe to play the action pairs BL and TR in turns, starting with the one that was not selected at stage 1. Furthermore if a player deviates at stage $t \geq 2$, then by lemma 4.2.6 he immediately unlearns an action and the other player punishes him by, from state $t+1$ on, playing the action that yields the $(0,0)$-payoff. ■

Definition 4.2.8 *The retaliation strategy as used in the proof of theorem 4.2.7 is the $(0,0)$-threat.*

Remark 4.2.9 *Notice that for player 1 (2) the $(0,0)$-threat is applicable if and only if at stages $\{1, 2, \ldots, t\}$ the action pairs TR and BL are played alternately and at stage $t+1$ the strategy σ (π) prescribes to continue alternating.*

Again we focus merely on equilibrium strategies.

Lemma 4.2.10 *If (π, σ) is a pair of equilibrium strategies in $\mathcal{G}$ and, in accordance with (π, σ), the entries TR and BL are selected alternately at stages $1, 2, \ldots, t$ and entry TL at stage $t+1$, then only $(2,1)$ can occur as equilibrium reward.*

Proof. Let (π, σ) be a pair of equilibrium strategies in $\mathcal{G}$ and let, in accordance with (π, σ), the entries TR and BL be selected alternately at stages $1, 2, \ldots, t$ and let entry TL be selected at stage $t+1$. Then by lemma 4.2.6 at stage $t+1$ one of the players unlearns an action. Suppose without loss of generality that player 2 unlearns an action (action R). Then the 2 action pairs that are still available, are TL and BL with respective payoffs of $(2,1)$ and $(0,0)$. Strategy π, being a best reply against σ, must prescribe to, eventually, unlearn action B and in the long run action pair TL is played with frequency 1, yielding a reward of $(2,1)$. ■

Analogously:

Lemma 4.2.11 *If (π, σ) is a pair of equilibrium strategies in $\mathcal{G}$ and, in accordance with (π, σ), the entries TR and BL are selected alternately at stages $1, 2, \ldots, t$ and entry BR at stage $t+1$, then only $(1,3)$ can occur as equilibrium reward.*

We will now continue the analysis by showing which rewards can be obtained by a pair of equilibrium strategies given that at stage 1 one of the entries TR or BL is selected. Notice first that each equilibrium reward that can be obtained after at stage 1 entry TR is selected, can also be selected if at stage 1 entry BL is selected by, with the aid of the $(0,0)$-threat, "forcing" the players to play action pair TR at stage 2. Similarly each equilibrium reward that can be obtained after at stage 1 entry BL is selected, can also be selected if at stage 1 entry TR is selected. Let $\mathbb{E}$ be the set of equilibrium rewards that can be obtained given that at stage 1 one of the entries TR or BL is selected and let $\tilde{\gamma}(\pi,\sigma)$ denote the reward of (π,σ) from stage 2 on given that, in accordance with (π,σ), at stage 1 one of the entries TR or BL is selected.

Lemma 4.2.12 *For each $\tilde{\gamma} = \left(\tilde{\gamma}^1, \tilde{\gamma}^2\right) \in \mathbb{E}$ we have: $\tilde{\gamma} \not> (1,1)$.*

Proof. Let (π,σ) be a pair of equilibrium strategies in $\mathcal{G}$, suppose without loss of generality that, in accordance with (π,σ), at stage 1 entry BL is selected and suppose that $\tilde{\gamma}(\pi,\sigma) > (1,1)$. Then at stage 2 the players face a game G_2, in which they may expect to receive the following rewards, if they do not deviate from π and σ:

$$G_2 = \begin{pmatrix} 2,1 & a_2,b_2 \\ 0,0 & 1,3 \end{pmatrix}.$$

Here the numbers can be explained as follows: If, in accordance with (π,σ), at stage 1 entry BL and at stage 2 entry TL is selected, then player 2 has lost action R and player 1 maximizes his reward by unlearning action B, leading to the $(2,1)$ reward. If, in accordance with (π,σ), at stage 1 entry BL and at stage 2 entry BR is selected, then player 1 has lost action T and player 2 maximizes his reward by unlearning action L, leading to the $(1,3)$ reward. If at both stage 1 and stage 2 entry BL is selected, then players 1 and 2 have lost action T and R respectively and entry BL will be played at all following stages, leading to the $(0,0)$ reward. If at stage 1 entry BL and at stage 2 entry TR is selected, then each player has both actions available. In this case the tuple (a_2,b_2) is the expected reward. The reward $\tilde{\gamma}(\pi,\sigma) > (1,1)$ is the expected reward in the game G_2. Since player 1 can not get more than 1, if he plays action B, he will play action T with probability 1. Furthermore player 2 will for a similar reason play action R with probability 1 and hence the action pair TR will be selected with probability 1 and $(a_2,b_2) = \tilde{\gamma}(\pi,\sigma)$. Now we consider the situation the players face at stage 3, after having selected action pairs BL and TR at stages 1 and 2. Analogously to the situation at stage 2 this can be described by a game:

$$G_3 = \begin{pmatrix} 2,1 & 0,0 \\ a_3,b_3 & 1,3 \end{pmatrix}$$

and the expected reward in this game must be $\tilde{\gamma}(\pi,\sigma)$. This can be achieved in two ways: Firstly the players can be forced by a $(0,0)$-threat to play action pair BL, in which case we have $(a_3,b_3) = \tilde{\gamma}(\pi,\sigma)$. In this case at stage 4 the players face game G_2 again, where at stages 2 and 3 they received a payoff of $(0,0) < \tilde{\gamma}(\pi,\sigma)$. Therefore this situation can not continue infinitely; eventually the strategies must prescribe to do something else. The second possibility is that at stage 3 both players randomize. They have to do that in such a way that the opponent is indifferent between his actions, since otherwise he would not randomize. Furthermore the expected reward must be

$\tilde{\gamma}(\pi,\sigma) = (a_2, b_2)$. Some straightforward calculations show that then $a_3 = 3 - \frac{2}{a_2} > a_2$ and $b_3 = 4 - \frac{3}{b_2} > b_2$. Now if the players select action pair BL, then at stage 4 they face a game similar to G_2 but then with expected rewards that are equal to $(a_3, b_3) > \tilde{\gamma}(\pi,\sigma)$ and with expected rewards (a_4, b_4), if entry TR were to be selected. Continue the analysis to find an increasing sequence of tuples (a_t, b_t) that converges to $(2, 3)$, which is not a feasible reward. This means that, in order to receive $\tilde{\gamma}(\pi,\sigma)$ as an expected reward from stage 2 on, from some stage t on the players have to receive an infeasible expected reward after having selected BL and TR in turns so far. Of course this is impossible and hence $\tilde{\gamma}(\pi,\sigma)$ can not be obtained. ■

Lemma 4.2.13 *For each $\tilde{\gamma} = (\tilde{\gamma}^1, \tilde{\gamma}^2) \in \mathbb{E}$ we can not have that $\tilde{\gamma}^1 < 1$ and $\tilde{\gamma}^2 > 1$.*

Proof. Let (π,σ) be a pair of equilibrium strategies in $\mathcal{G}$, suppose without loss of generality that, in accordance with (π,σ), at stage 1 entry BL is selected and suppose that $\tilde{\gamma}^1(\pi,\sigma) < 1$ and $\tilde{\gamma}^2(\pi,\sigma) > 1$. Then at stage 2 the players face a game G_2, in which they may expect to receive the following rewards, if they do not deviate from π and σ:

$$G_2 = \begin{pmatrix} 2,1 & a_2, b_2 \\ 0,0 & 1,3 \end{pmatrix}$$

with expected reward $\tilde{\gamma}(\pi,\sigma)$. Notice that in G_2 action R strictly dominates action L. Consequently σ prescribes to play action R with probability 1. But then, in order to achieve an expected reward of $\tilde{\gamma}(\pi,\sigma)$ we must have: $(a_2, b_2) = \tilde{\gamma}(\pi,\sigma)$ and player 1 must by means of a $(0,0)$-threat be forced to play action T at stage 2. Now we consider the situation the players face at stage 3, after having selected action pairs BL and TR at stages 1 and 2. This situation can be described by the following game:

$$G_3 = \begin{pmatrix} 2,1 & 0,0 \\ a_3, b_3 & 1,3 \end{pmatrix}$$

and the expected reward in this game must be $\tilde{\gamma}(\pi,\sigma) = (a_2, b_2)$. Then similarly to the proof of lemma 4.2.12 we have the option to "return" to G_2 with an average payoff of 0 at stages 2 and 3. The second option, randomization, yields: $a_3 = 3 - \frac{2}{a_2} < a_2$ and $b_3 = 4 - \frac{3}{b_2} > b_2$ and if action pair BL is selected, then at stage 4 the players face a game similar to G_2 but then with expected rewards (a_3, b_3). If we continue the analysis we find a sequence of tuples (a_t, b_t), where a_t tends to $-\infty$ and b_t converges to 3. Again we come to the conclusion that $\tilde{\gamma}(\pi,\sigma)$ can not be obtained. ■

Similarly:

Lemma 4.2.14 *For each $\tilde{\gamma} = (\tilde{\gamma}^1, \tilde{\gamma}^2) \in \mathbb{E}$ we can not have that $\tilde{\gamma}^1 > 1$ and $\tilde{\gamma}^2 < 1$*

and also

Lemma 4.2.15 *For each $\tilde{\gamma} = (\tilde{\gamma}^1, \tilde{\gamma}^2) \in \mathbb{E}$ we can not have that $\tilde{\gamma}^1 < 1$ and $\tilde{\gamma}^2 = 1$*

and

Lemma 4.2.16 *For each $\tilde{\gamma} = (\tilde{\gamma}^1, \tilde{\gamma}^2) \in \mathbb{E}$ we can not have that $\tilde{\gamma}^1 = 1$ and $\tilde{\gamma}^2 < 1$.*

In the proofs of lemmas 4.2.18, 4.2.19, 4.2.21 and 4.2.22 we make use of follow-up strategies, which are defined as follows:

Definition 4.2.17 *A follow-up strategy is a substrategy that, given a specific course of the game up to and including stage t, prescribes how to play from stage $t+1$ on.*

Lemma 4.2.18 *For each $a_2 \in [1,2\rangle$ the tuple $(a_2, 1) \in \mathbb{E}$.*

Proof. We will construct a pair of equilibrium follow-up strategies, yielding a reward of $(a_t, 1)$, given that at stage 1 the entry BL is selected. Consider the following sequence of reals: $a_2 \in [1,2\rangle$ and for $t \geq 2$ let

$$a_{t+1} = \begin{cases} \frac{a_t+1}{2} \text{ if } t \text{ is even} \\ 3 - \frac{2}{a_t} \text{ if } t \text{ is odd} \end{cases}$$

Here a_t is to be interpreted as an expected future reward to player 1. Notice that $a_t \in [1,2\rangle$ for all $t \geq 2$. Suppose without loss of generality that at stage 1 entry BL is selected and consider the following pair of follow-up strategies (π, σ) in $\mathcal{G}$:

- π prescribes to play action T with probability 1 if the stage number t is even and $\left(\frac{2}{3}, \frac{1}{3}\right)$ if t is odd as long as at stages $1, 2, \ldots, t-1$ alternately action pairs BL and TR are selected. Furthermore as soon as at stage τ an entry is selected that deviates from this alternating sequence, then at all stages from $\tau+1$ on π prescribes to play the action played at stage τ.
- σ prescribes to play $(\alpha_t, 1-\alpha_t)$, where $\alpha_t = \frac{a_t - a_{t+1}}{2 - a_{t+1}} \in [0,1\rangle$ if t is even and $(\frac{1}{2}a_t, 1 - \frac{1}{2}a_t)$ if t is odd as long as at stages $1, 2, \ldots, t-1$ alternately action pairs BL and TR are selected. Furthermore as soon as at stage τ an entry is selected that deviates from this alternating sequence, then at all stages from $\tau+1$ on σ prescribes to play the action played at stage τ.

Furthermore the $(0,0)$-threat is inserted in π and σ.

Now suppose that at stages $1, 2, \ldots, t-1$ the action pairs BL and TR are selected alternately and that t is even. Then at stage $t-1$ entry BL is selected and at stage t the players face a game G_t, in which they may expect to receive the following rewards, if they do not deviate from π and σ:

$$G_t = \begin{pmatrix} 2,1 & a_{t+1},1 \\ 0,0 & 1,3 \end{pmatrix}$$

with expected reward $(a_t, 1)$. Since $a_{t+1} \geq 1$ an equilibrium strategy π must prescribe to play action T with probability 1. Furthermore this makes player 2 indifferent, allowing him to randomize. The randomization $(\frac{1}{2}a_t, 1 - \frac{1}{2}a_t)$ guarantees an expected reward of $(a_t, 1)$ in G_t and neither player 1 nor player 2 can make a profitable unilateral deviation at stage t in game $\mathcal{G}$.

Now suppose that at stages $1, 2, \ldots, t-1$ the action pairs BL and TR are selected alternately and that t is odd. Then at stage $t-1$ entry TR is selected and at stage t

the players face a game G_t, in which they may expect to receive the following rewards, if they do not deviate from π and σ:

$$G_t = \begin{pmatrix} 2,1 & 0,0 \\ a_{t+1},1 & 1,3 \end{pmatrix}$$

with expected reward $(a_t, 1)$. Given that $a_{t+1} \in [1,2\rangle$ in G_t there is one completely mixed equilibrium, where player 1 plays $(\frac{2}{3}, \frac{1}{3})$ and player 2 plays $(\frac{1}{3-a_{t+1}}, \frac{2-a_{t+1}}{3-a_{t+1}})$. For the reward, corresponding to this pair of mixed actions, to be equal to $(a_t, 1)$ we need that $a_{t+1} = 3 - \frac{2}{a_t}$. Now neither player 1 nor player 2 can make a profitable unilateral deviation at stage t in $\mathcal{G}$. ■

Similarly:

Lemma 4.2.19 *For each $b_2 \in [1,3\rangle$ the tuple $(1, b_2) \in \mathbb{E}$.*

Now we will investigate follow-up equilibrium strategies for which, given that at stage 1 one of the entries TR and BL is selected, the reward of each player is smaller than 1. Notice first that if at a certain stage exactly one player randomizes, then his reward is at least 1. Therefore we only have to consider strategy pairs by which, at each stage, either both players randomize or neither player randomizes.

Remark 4.2.20 *It is important to observe that for $(\lambda, \mu) < (1,1)$ the (unique) pair of completely mixed equilibrium strategies in the one-shot game*

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

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 pair of completely mixed equilibrium strategies in the one-shot game

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

Suppose that at stage 1 entry TR is selected. Then from remark 4.2.20 it follows that, from stage 2 on, the reward of the follow-up strategy pair $(\tilde{\pi}, \tilde{\sigma})$ that prescribes to randomize at stages 2, 3 and 4 (or until at least one of the players has unlearned an action) is the same as the reward of the follow-up strategy pair (π', σ') that prescribes to randomize at stages 4, 6 and 37 (or until at least one of the players has unlearned an action) and to keep on alternating at the stages in between. Only the number of stages, at which the strategies prescribe to randomize, is relevant for the reward. The exact same argumentation can be held if at stage 1 entry BL is selected. Furthermore we already observed that each reward that can be obtained by means of the follow-up strategy pair $(\tilde{\pi}, \tilde{\sigma})$ after at stage 1 entry TR is selected, can also be obtained by means of a follow-up strategy pair after at stage 1 entry BL is selected. This means that the reward of the follow-up strategy pair $(\tilde{\pi}, \tilde{\sigma})$, after at stage 1 entry TR is selected, is also equal to the reward of the follow-up strategy pair $(\bar{\pi}, \bar{\sigma})$ that prescribes to randomize at stages 21, 26 and 58 (or until at least one of the players has unlearned an action) after at stage 1 entry BL is selected. Therefore we only need to consider strategy pairs that prescribe to randomize at stages $\{2, 3, \ldots, q+1\}$ or until

at least one of the players has unlearned an action. Furthermore we argue that, if after stage $q+1$ each player still has both actions available, then the reward to both players must be strictly less then 1. Suppose namely that player 1 gets at least 1 after stage $q+1$. Then if, in accordance with the pair of equilibrium strategies (π, σ), at stage $t \in \{1, 2, 3, \ldots, q\}$ entry BL is selected and both players still have both of their actions available, then π will prescribe to play action T with probability 1 at stage $t+1$, which contradicts the fact that he randomizes at stage $t+1$. Since there is no randomization involved in the play from stage $q+2$ on, this means that the reward from stage $q+2$ on must be $(0,0)$, which can be obtained using the $(0,0)$-threat. Hence for $q \in \{0, 1, 2, \ldots\}$ we consider the following pair of follow-up strategies $(\hat{\pi}_q, \hat{\sigma}_q)$ after at stage 1 entry TR is selected: $(\hat{\pi}_q, \hat{\sigma}_q)$ prescribes to randomize from stage 2 on for at most a fixed number of q consecutive stages and then:

1. if after stage $q+1$ both players still have both actions alive (which happens exactly when entries TR and BL have been played in turns up to and including stage $q+1$), to play TR and BL in turns, leading to a reward of $(0,0)$ using the $(0,0)$-threat (cf. definition 4.2.8),

2. if at some stage $\tau \in \{2, 3, \ldots, q+1\}$ exactly one player unlearns an action (which happens exactly when at stage τ for the first time one of the entries TL or BR is selected), to play the action pair that is selected at that stage repeatedly or

3. if at some stage $\tau \in \{2, 3, \ldots, q+1\}$ both players unlearn an action (which happens exactly when TR and BL have been played in turns up to and including stage $\tau - 1$ and at τ the same action pair is selected as at stage $\tau - 1$), to play the only available action pair (TR or BL) repeatedly.

The pair of follow-up strategies $(\check{\pi}_q, \check{\sigma}_q)$ after at stage 1 the action pair BL is selected, is defined analogously.

Lemma 4.2.21 *After at stage 1 entry TR is selected, for each $q \in \{0, 1, 2, \ldots\}$ the pair of follow-up equilibrium strategies $(\hat{\pi}_q, \hat{\sigma}_q)$ yields a reward of $(1 - \frac{1}{2^{q+1}-1}, 1 - \frac{2}{3^{q+1}-1})$.*

Proof. If entry TR is selected at stage 1 and the follow-up strategy pair $(\hat{\pi}_q, \hat{\sigma}_q)$ is being played, then at stage 2 the players face a (sub)game leading to the same reward as the following one-shot game:

$$\hat{G}_q = \begin{pmatrix} 2,1 & 0,0 \\ \lambda_q, \mu_q & 1,3 \end{pmatrix}.$$

Here λ_q and μ_q are the rewards the players expect to receive, if they, when using the follow-up strategy pair $(\hat{\pi}_q, \hat{\sigma}_q)$ after selecting cell TR at stage 1, select cell BL at stage 2.

We will investigate the possible values λ_q and μ_q can have in case $(\hat{\pi}_q, \hat{\sigma}_q)$ is an equilibrium. Notice first that, by definition, $(\lambda_0, \mu_0) = (0,0)$. Now let entry BL be selected at stage 1 and suppose that the players use the follow-up strategy pair $(\check{\pi}_{q+1}, \check{\sigma}_{q+1})$, so from stage 2 on the players randomize for at most a fixed number of

q consecutive stages. Then at stage 2 the players face a game leading to the same reward as the following one-shot game:

$$\check{G}_{q+1} = \begin{pmatrix} 2,1 & \lambda_{q+1}, \mu_{q+1} \\ 0,0 & 1,3 \end{pmatrix}. \tag{4.1}$$

Here λ_{q+1} and μ_{q+1} are the rewards the players receive, if they, when using the follow-up strategy pair $(\check{\pi}_{q+1}, \check{\sigma}_{q+1})$ after selecting cell BL at stage 1, select cell TR at stage 2. However, if after selecting cell BL at stage 1, at stage 2 the players select entry TR, then at stage 3 they face a (sub)game leading to the same reward as the following one-shot game:

$$\hat{G}_q = \begin{pmatrix} 2,1 & 0,0 \\ \lambda_q, \mu_q & 1,3 \end{pmatrix}, \tag{4.2}$$

which is exactly the same subgame we find after stage 1, if at stage 1 TR is selected and the follow-up strategy pair $(\hat{\pi}_q, \hat{\sigma}_q)$ is to be played. The unique completely mixed equilibrium in the one-shot game $\hat{G}_q$ is $((\frac{3-\mu_q}{4-\mu_q}, \frac{1}{4-\mu_q}), (\frac{1}{3-\lambda_q}, \frac{2-\lambda_q}{3-\lambda_q}))$ with reward $(\frac{2}{3-\lambda_q}, \frac{3}{4-\mu_q})$ and hence $\lambda_{q+1} = \frac{2}{3-\lambda_q}$ and $\mu_{q+1} = \frac{3}{4-\mu_q}$. Consequently we find the following iterative set of equations (recall that $\lambda_0 = \mu_0 = 0$):

$$\begin{cases} \lambda_{q+1} = \frac{2}{3-\lambda_q} \text{ with } \lambda_0 = 0 \\ \mu_{q+1} = \frac{3}{4-\mu_q} \text{ with } \mu_0 = 0 \end{cases} \tag{4.3}$$

It can easily be shown that for equations (4.3) we have:

$$(\lambda_q, \mu_q) < (1,1) \text{ for all } q$$

and hence in each of the one-shot games (4.2) and (4.1) a completely mixed pair of equilibrium strategies exists. Furthermore from equations (4.3) it easily follows that

$$\lambda_q = 1 - \frac{1}{2^{q+1}-1} \text{ and } \mu_q = 1 - \frac{2}{3^{q+1}-1}.$$

This completes the proof. ∎

The rewards mentioned in lemma 4.2.21 can also be obtained by a pair of follow-up equilibrium strategies, if at stage 1 entry BL is selected:

Lemma 4.2.22 *After at stage 1 entry BL is selected, for each $q \in \{0,1,2,\ldots\}$ the equilibrium follow-up strategy pair $(\check{\pi}_q, \check{\sigma}_q)$ yields a reward of $(1 - \frac{1}{2^{q+1}-1}, 1 - \frac{2}{3^{q+1}-1})$.*

Theorem 4.2.23 *If (π, σ) is a pair of equilibrium strategies in $\mathcal{G}$ and, in accordance with (π, σ), at stage 1 one of the entries TL or BR is selected, then from stage 2 on, the rewards that are in the set $\mathbb{E}$ can be obtained by (π, σ), where $\mathbb{E}$ is the union of the following 3 sets:*

$$\{(a,1) \mid 1 \le a \le 2\},$$
$$\{(1,b) \mid 1 \le b \le 3\}$$

and

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

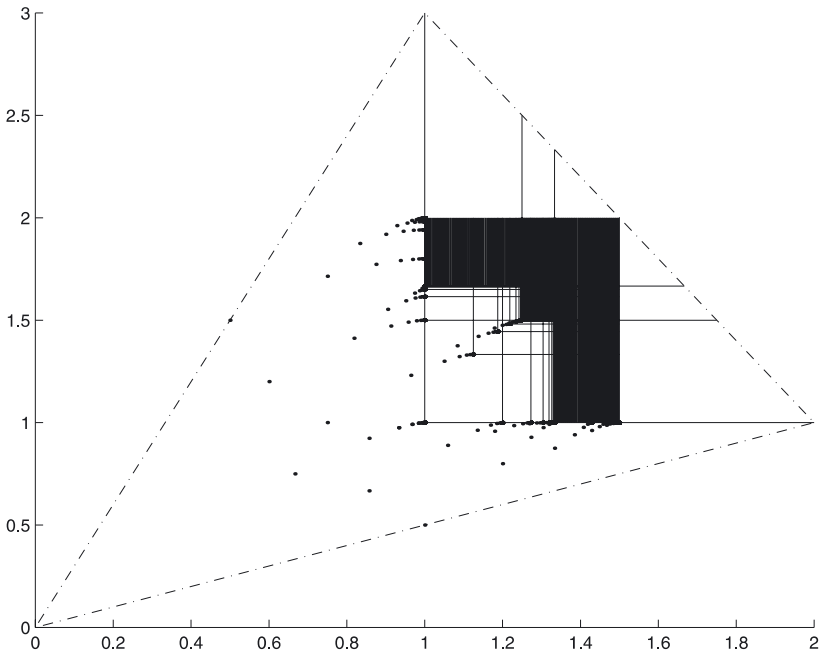


Figure 4.1: Equilibrium rewards in example 4.1.

Proof. We argued that $(0,0)$ is the only follow-up equilibrium reward, if after stage $q+1$ both players have both actions available, leading to a reward in the subgame after stage 1, that is strictly smaller than $(1,1)$. This means that, after at stage 1 one of the entries TR or BL is selected, there are no other follow-up equilibrium strategies that yield rewards that are strictly smaller than $(1,1)$. Combining this observation with lemmas 4.2.12 up to and including 4.2.22 completes the proof. ■

Theorem 4.2.24 *The set of equilibrium rewards in $\mathcal{G}$ equals the set of equilibrium rewards of the following one-shot games:*

$$\begin{pmatrix} V & W \\ W & V \end{pmatrix},$$

where $V \in \left\{(2,1),(\frac{3}{2},2),(1,3)\right\}$ and $W \in \mathbb{E}$, the set mentioned in theorem 4.2.23.

Proof. A direct consequence of theorems 4.2.5 and 4.2.23. ■

We have thus determined the set of equilibrium rewards in $\mathcal{G}$ in example 4.1. In figure 4.1 the equilibrium rewards are depicted. For other $(2,2)$-restricted 2×2 - coordination games we find similar results. In particular for each of the following $(2,2)$-restricted 2×2 - coordination games:

$$\begin{array}{cc} & \begin{array}{cc} L & R \end{array} \\ \begin{array}{c} T \\ B \end{array} & \begin{pmatrix} \mathfrak{a}_1, \mathfrak{b}_1 & 0,0 \\ 0,0 & \mathfrak{a}_2, \mathfrak{b}_2 \end{pmatrix}, \end{array}$$

where either $\mathfrak{a}_1 > \mathfrak{a}_2 > 0$ and $\mathfrak{b}_2 > \mathfrak{b}_1 > 0$ or $\mathfrak{a}_2 > \mathfrak{a}_1 > 0$ and $\mathfrak{b}_1 > \mathfrak{b}_2 > 0$, the set of equilibrium rewards is the equivalent of the set mentioned in theorem 4.2.24.

4.3 $(3,3)$-restricted coordination games

For $\left(r^1, r^2\right) \geq (3,3)$ the situation is very much different. As soon as the players each have only 2 actions available, their restrictions allow them to randomize without losing actions. This appears to be a very powerful tool for generating equilibrium rewards. In this section is we will show that every convex combination of equilibrium rewards can also be obtained as an equilibrium reward. Notice that in a coordination game each diagonal payoff, i.e. a (strictly positive) payoff on the main diagonal, is an equilibrium reward. For that reason we start by showing that any convex combination of diagonal payoffs can be obtained as an equilibrium reward. We proceed in two steps: first we show it for convex combinations of only two diagonal payoffs by means of a so-called "agreement", next we extend the result to the general case.

Lemma 4.3.1 *Consider a $(3,3)$-restricted coordination game of the form*

$$\begin{array}{cc} & \begin{array}{cc} L & R \end{array} \\ \begin{array}{c} T \\ B \end{array} & \left(\begin{array}{cc} \mathfrak{a}_1, \mathfrak{b}_1 & 0,0 \\ 0,0 & \mathfrak{a}_2, \mathfrak{b}_2 \end{array} \right) \end{array}$$

(with $\mathfrak{a}_1, \mathfrak{a}_2, \mathfrak{b}_1, \mathfrak{b}_2 > 0$). Then every convex combination of $(\mathfrak{a}_1, \mathfrak{b}_1)$ and $(\mathfrak{a}_2, \mathfrak{b}_2)$ can be obtained as an equilibrium reward.

Proof. Consider the strategy pairs $\left(\pi^T, \sigma^L\right)$ and $\left(\pi^B, \sigma^R\right)$, respectively prescribing to play the action pairs TL and BR at each stage. With respect to the threat points corresponding to these strategies it can easily be verified that:

$$\hat{\gamma}\left(\pi^T, \sigma^L\right) = (\mathfrak{a}_1, \mathfrak{b}_1) \text{ and } \hat{\gamma}\left(\pi^B, \sigma^R\right) = (\mathfrak{a}_2, \mathfrak{b}_2)$$

and that $\left(\pi^T, \sigma^L\right)$ and $\left(\pi^B, \sigma^R\right)$ are equilibria.

Step 1: Given the $(3,3)$-restriction for each $\alpha \in [\frac{1}{3}, \frac{2}{3}]$ the reward $\alpha \cdot (a_1, b_1) + (1 - \alpha) \cdot (a_2, b_2)$ is obtainable by a pair of pure equilibrium strategies $(\pi^\alpha, \sigma^\alpha)$ as follows: Let $(\pi^\alpha, \sigma^\alpha)$ prescribe to play $TL\ BR\ \lambda_1\ TL\ BR\ \lambda_2\ TL\ BR\ \lambda_3\ \ldots$, where for all i the action pairs $\lambda_i \in \{TL, BR\}$ are such that the long-run frequency of action pair TL is α. Then

$$\gamma\left(\pi^\alpha, \sigma^\alpha\right) = \alpha \cdot (\mathfrak{a}_1, \mathfrak{b}_1) + (1 - \alpha) \cdot (\mathfrak{a}_2, \mathfrak{b}_2)$$

and, since obviously neither player can make a profitable unilateral deviation at any stage, $(\pi^\alpha, \sigma^\alpha)$ is an equilibrium with reward $\alpha \cdot (a_1, b_1) + (1 - \alpha) \cdot (a_2, b_2)$.

Step 2: For any $\alpha \in [\frac{1}{6}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{5}{6}]$ the reward $\alpha \cdot (\mathfrak{a}_1, \mathfrak{b}_1) + (1 - \alpha) \cdot (\mathfrak{a}_2, \mathfrak{b}_2)$ can be obtained by a pair of equilibrium strategies.

Without loss of generality we take $\alpha \in [\frac{1}{6}, \frac{1}{3}]$. Note that $\alpha \cdot (\mathfrak{a}_1, \mathfrak{b}_1) + (1 - \alpha) \cdot (\mathfrak{a}_2, \mathfrak{b}_2)$ $= \frac{1}{2}\gamma\left(\pi^B, \sigma^R\right) + \frac{1}{2}\gamma\left(\pi^{2\alpha}, \sigma^{2\alpha}\right)$. Now consider the following pair of strategies, resulting in playing either (π^B, σ^R) or $(\pi^{2\alpha}, \sigma^{2\alpha})$, each with probability $\frac{1}{2}$, from stage 2 onwards:

Define $(\pi^\alpha, \sigma^\alpha)$ by playing $(\frac{1}{2}, \frac{1}{2})$ at stage 1, followed by playing (π^B, σ^R) if the first stage actions are the same, and followed by $(\pi^{2\alpha}, \sigma^{2\alpha})$ otherwise. Then $(\pi^\alpha, \sigma^\alpha)$ is an equilibrium as required.

Step 3: For any $\alpha \in [0,1]$ the reward $\alpha \cdot (\mathfrak{a}_1, \mathfrak{b}_1) + (1-\alpha) \cdot (\mathfrak{a}_2, \mathfrak{b}_2)$ can be obtained by a pair of non-pure strategies forming an equilibrium.

Take $\alpha \in [\frac{1}{12}, \frac{1}{6}]$ and let $(\pi^{2\alpha}, \sigma^{2\alpha})$ be an equilibrium as in the proof of step 2. We will now introduce strategies that result in playing either (π^B, σ^R) or $(\pi^{2\alpha}, \sigma^{2\alpha})$, each with probability $\frac{1}{2}$, from stage 3 onwards: Define $(\pi^\alpha, \sigma^\alpha)$ by playing $(\frac{1}{2}, \frac{1}{2})$ at stage 1, followed by playing (π^B, σ^R) if the first stage actions are the same, and followed by the alternative action at stage 2 and from stage 3 onwards start playing $(\pi^{2\alpha}, \sigma^{2\alpha})$, otherwise. When the latter is done stage 3 acts as if it were the initial stage. Then $(\pi^\alpha, \sigma^\alpha)$ is an equilibrium as required. The alternative actions at stage 2 are necessary in order not to lose any action in the selection process. Thus we have established equilibria for all rewards $\alpha \cdot (\mathfrak{a}_1, \mathfrak{b}_1) + (1-\alpha) \cdot (\mathfrak{a}_2, \mathfrak{b}_2)$ with $\alpha \in [\frac{1}{12}, \frac{1}{6}]$. By repeating this very same procedure as often as we like we obtain the statement of the lemma. ■

A pair of equilibrium strategies as in step 2 and 3 of the proof of lemma 4.3.1 is a special type of a so-called agreement, which is defined as follows:

Definition 4.3.2 *In an $\left(r^1, r^2\right)$-restricted 2×2-game, $\left(r^1, r^2\right) \geq (3,3)$, a strategy pair is an agreement, if it prescribes to play*

- $(\frac{1}{2}, \frac{1}{2})$ *at stage 1 followed by*
- *The alternative action at stage 2 followed by*
- *Playing according to a strategy pair (π^1, σ^1) if the first stage actions were the same, and playing according to a pair of strategies (π^2, σ^2), otherwise.*

An agreement is denoted by $(\pi_{\mathcal{A}}, \sigma_{\mathcal{A}})$.

Remark 4.3.3 *Notice that*

$$\gamma\left(\pi_{\mathcal{A}}, \sigma_{\mathcal{A}}\right) = \frac{1}{2}\gamma(\pi^1, \sigma^1) + \frac{1}{2}\gamma(\pi^2, \sigma^2)$$

and that the pairs (π^1, σ^1) and (π^2, σ^2) may also be agreements.

Furthermore playing "according to strategy pair $\left(\pi^1, \sigma^1\right)$" does not mean unthinkingly copying that strategy, since that might lead to the unlooked-for loss of an action. Consider for example a $(3,3)$-restricted game where π^1 prescribes to play action T at the first and the second stage. Now if the selected cell at stage 1 of the agreement is BR, then at stage 2 $(\pi_{\mathcal{A}}, \sigma_{\mathcal{A}})$ prescribes to play TL. At stage 3 player 1 plays the action that he is supposed to play at stage 1, if the strategy pair (π^1, σ^1) was used, which is action T. Now, if at stage 4 player 1 plays the action that he is supposed to play at stage 2, if (π^1, σ^1) was used, it will be action T again, but then he unlearns action B. In such a case at stage 3 we insert an extra stage, in which each player plays the action he is not supposed to play at stage 1, if (π^1, σ^1) was used and we copy (π^1, σ^1)

from stage 4 on. A slightly more complicated situation arises, if the strategy π^1 prescribes to randomize at stages 1 and 2. In this case $\pi_{\mathcal{A}}$ and $\sigma_{\mathcal{A}}$ state that after the $(\frac{1}{2},\frac{1}{2})$-randomization at stage 1 and the alternative action at stage 2, if play continues according to (π^1,σ^1), then at stage 3 the first stage of π^1 respectively σ^1 is played, at stage 4 the action that was not played at stage 3, and at stage 5 the second stage of π^1 respectively σ^1 is played. So now we insert an extra stage at stage 4, in between the 2 randomization stages, in order not to unlearn any actions. From stage 6 on we can truly copy (π^1,σ^1), without unlearning actions in an unforeseen way. Hence playing according to a strategy pair means playing the actions prescribed by the two strategies and, if necessary, insert pure actions in order not to accidentally unlearn any actions.

Notice that if (π^1,σ^1) and (π^2,σ^2) are agreements, then the agreement $(\pi_{\mathcal{A}},\sigma_{\mathcal{A}})$ playing according to each of them with probability $\frac{1}{2}$ does not need any extra insertion stages.

Another most appealing feature of the agreement is mentioned in lemma 4.3.4.

Lemma 4.3.4 *If (π^1,σ^1) and (π^2,σ^2) are equilibria, then also $(\pi_{\mathcal{A}},\sigma_{\mathcal{A}})$ is an equilibrium.*

Proof. As soon as, after stage 1, the decision to play by (π^1,σ^1) or (π^2,σ^2) has been made, neither player can make a deviation from $(\pi_{\mathcal{A}},\sigma_{\mathcal{A}})$ and benefit from it. Therefore only at stage 1 a player might be able to advantageously deviate from $(\pi_{\mathcal{A}},\sigma_{\mathcal{A}})$. He does so by making the probability of selecting an entry at the main diagonal different from $\frac{1}{2}$. Suppose that, with that aim, instead of $(\frac{1}{2},\frac{1}{2})$ player 1 decides to play $(\alpha,1-\alpha)$ for some $\alpha\in[0,1]$, a deviation from $\pi_{\mathcal{A}}$ that can not be detected by player 2. Then the probability that an entry at the main diagonal is selected is $\frac{1}{2}\alpha+\frac{1}{2}(1-\alpha)=\frac{1}{2}$ and the deviation is not profitable. A similar argument is applicable for player 2, so $(\pi_{\mathcal{A}},\sigma_{\mathcal{A}})$ is an equilibrium. ■

A direct consequence of lemma 4.3.4 in combination with remark 4.3.3 is the following theorem:

Lemma 4.3.5 *Let $\left(\pi^1,\sigma^1\right)$ and $\left(\pi^2,\sigma^2\right)$ be equilibria in a $(3,3)$-restricted coordination game. Then for each $\alpha\in[0,1]$ there exists an agreement $(\pi_{\mathcal{A}},\sigma_{\mathcal{A}})$ such that $(\pi_{\mathcal{A}},\sigma_{\mathcal{A}})$ is an equilibrium and $\gamma(\pi_{\mathcal{A}},\sigma_{\mathcal{A}})=\alpha\cdot\gamma\left(\pi^1,\sigma^1\right)+(1-\alpha)\cdot\gamma\left(\pi^2,\sigma^2\right)$.*

Proof. The agreement $\left(\pi_{\mathcal{A}}^{\frac{1}{2}},\sigma_{\mathcal{A}}^{\frac{1}{2}}\right)$ yields a reward of $\frac{1}{2}\gamma(\pi^1,\sigma^1)+\frac{1}{2}\gamma(\pi^2,\sigma^2)$. Now consider the agreements $\left(\pi_{\mathcal{A}}^{\frac{1}{4}},\sigma_{\mathcal{A}}^{\frac{1}{4}}\right)$ and $\left(\pi_{\mathcal{A}}^{\frac{3}{4}},\sigma_{\mathcal{A}}^{\frac{3}{4}}\right)$. The first one, by lemma 4.3.2, consists of playing $\left(\frac{1}{2},\frac{1}{2}\right)$ at stage 1, playing the alternative action at stage 2 and from stage 3 on playing according to $\left(\pi_{\mathcal{A}}^{\frac{1}{2}},\sigma_{\mathcal{A}}^{\frac{1}{2}}\right)$, if the selected first stage actions were the same, and according to $\left(\pi^2,\sigma^2\right)$ otherwise, thereby, indeed, yielding an equilibrium reward of $\frac{1}{2}\gamma\left(\pi_{\mathcal{A}}^{\frac{1}{2}},\sigma_{\mathcal{A}}^{\frac{1}{2}}\right)+\frac{1}{2}\gamma\left(\pi^2\sigma^2\right)=\frac{1}{4}\gamma\left(\pi^1,\sigma^1\right)+\frac{3}{4}\gamma\left(\pi^2,\sigma^2\right)$. Furthermore $\gamma\left(\pi_{\mathcal{A}}^{\frac{3}{4}},\sigma_{\mathcal{A}}^{\frac{3}{4}}\right)=\frac{3}{4}\gamma\left(\pi^1,\sigma^1\right)+\frac{1}{4}\gamma\left(\pi^2,\sigma^2\right)$ is also an equilibrium reward. But then, according to lemma 4.3.4, the agreements $\left(\pi_{\mathcal{A}}^{\frac{1}{8}},\sigma_{\mathcal{A}}^{\frac{1}{8}}\right)$, $\left(\pi_{\mathcal{A}}^{\frac{3}{8}},\sigma_{\mathcal{A}}^{\frac{3}{8}}\right)$, $\left(\pi_{\mathcal{A}}^{\frac{5}{8}},\sigma_{\mathcal{A}}^{\frac{5}{8}}\right)$ and $\left(\pi_{\mathcal{A}}^{\frac{7}{8}},\sigma_{\mathcal{A}}^{\frac{7}{8}}\right)$

also lead to equilibrium rewards. By continuing this procedure as often as we like, we find equilibrium rewards that approach $\alpha \cdot \gamma\left(\pi^1, \sigma^1\right) + (1-\alpha) \cdot \gamma\left(\pi^2, \sigma^2\right)$ arbitrarily close for any $\alpha \in [0,1]$. ■

The next theorem shows that in 2×2 -coordination games a $(3,3)$- restriction is sufficient to construct equilibrium rewards that are convex combinations of all the diagonal payoffs.

Theorem 4.3.6 *In a* $(3,3)$*-restricted* 2×2*-coordination game the set of equilibrium rewards is convex.*

Proof. Consider a reward $\check{\gamma} = \left(\check{\gamma}^1, \check{\gamma}^2\right)$ that is a convex combination of equilibrium rewards. Then there exist 3 equilibrium rewards ζ_i, ζ_j and ζ_k such that $\check{\gamma}$ is a convex combination of ζ_i, ζ_j and ζ_k. So $\check{\gamma} = \beta_1 \cdot \zeta_i + \beta_2 \cdot \zeta_j + \beta_3 \cdot \zeta_k$, where $\beta = (\beta_1, \beta_2, \beta_3) \in \Delta^3$, the unit simplex in $\mathbb{R}^3$, and we suppose without loss of generality that $\beta_1 \geq \beta_2 \geq \beta_3$. Let $\lambda = (1-2\beta_2, 2\beta_2, 0) \in \Delta^3$ and $\mu = (1-2\beta_3, 0, 2\beta_3) \in \Delta^3$, and write $\zeta = (\zeta_i, \zeta_j, \zeta_k)$. Then

$$\check{\gamma} = \left(\frac{1}{2}\lambda + \frac{1}{2}\mu\right) \cdot \zeta.$$

Since both λ and μ only put positive weight on at most two equilibrium rewards, by lemma 4.3.5 we can use agreements $\left(\pi^\lambda_{\mathcal{A}}, \sigma^\lambda_{\mathcal{A}}\right)$ and $\left(\pi^\mu_{\mathcal{A}}, \sigma^\mu_{\mathcal{A}}\right)$ to support $\lambda \cdot \zeta$ and $\mu \cdot \zeta$ as equilibrium rewards respectively. But then the agreement that results in playing either $\left(\pi^\lambda_{\mathcal{A}}, \sigma^\lambda_{\mathcal{A}}\right)$ and $\left(\pi^\mu_{\mathcal{A}}, \sigma^\mu_{\mathcal{A}}\right)$, each with probability $\frac{1}{2}$, from stage 3 onwards, is an equilibrium with reward $\check{\gamma}$. ■

Example 4.2:

Consider the following $(3,3)$-restricted coordination game:

$$\begin{array}{cc} & \begin{array}{cc} L & R \end{array} \\ \begin{array}{c} T \\ B \end{array} & \left(\begin{array}{cc} 2,1 & 0,0 \\ 0,0 & 1,1 \end{array}\right) \end{array}$$

From Joosten et al. (1995) it can be derived that for a pair of pure equilibrium strategies (π, σ) that prescribe to keep both actions available during the entire course of the game, we must have $\gamma(\pi, \sigma) \geq (\frac{3}{4}, \frac{1}{2})$. In chapter 5 the technique is presented to find that the following set of rewards is obtainable by pairs of pure strategies that prescribe to keep both actions alive: $conv\{(0,0), (\frac{1}{3}, \frac{1}{3}), (1, \frac{2}{3}), (\frac{2}{3}, \frac{1}{3}), (\frac{5}{3}, 1), (\frac{4}{3}, 1)\}$, where $conv\{.\}$ is the convex hull. Taking the intersection of this set and the set of rewards that are bigger than $(\frac{3}{4}, \frac{1}{2})$, we find the set of pure-strategy equilibrium rewards with both players keeping both actions available:

$$conv\left\{(\tfrac{3}{4}, \tfrac{1}{2}), (\tfrac{11}{12}, \tfrac{1}{2}), (\tfrac{5}{3}, 1), (\tfrac{4}{3}, 1), (\tfrac{3}{4}, \tfrac{11}{18})\right\}.$$

For pairs of strategies keeping subsets of I and J available we find only the two equilibrium rewards $(2,1)$ and $(1,1)$. Now theorem 4.3.6 tells us that each reward in the set

$$\begin{aligned} &conv\left\{(\tfrac{3}{4}, \tfrac{1}{2}), (\tfrac{11}{12}, \tfrac{1}{2}), (\tfrac{5}{3}, 1), (\tfrac{4}{3}, 1), (\tfrac{3}{4}, \tfrac{11}{18}), (2,1), (1,1)\right\} \\ &= conv\left\{(\tfrac{3}{4}, \tfrac{1}{2}), (\tfrac{11}{12}, \tfrac{1}{2}), (\tfrac{3}{4}, \tfrac{11}{18}), (2,1), (1,1)\right\} \end{aligned}$$

can be supported by an equilibrium.

It is important to notice that not all feasible and individually rational rewards can be obtained by a pair of equilibrium rewards. Suppose, by means of contradiction, that we have a pair of equilibrium strategies (π, σ) for the feasible and individually rational reward $(\frac{3}{4}, \frac{3}{4})$. Then (π, σ) must prescribe to play the action pair TL with related payoff $(2, 1)$ with long run frequency 0, because one can only obtain $(\frac{3}{4}, \frac{3}{4})$ as a convex combination of the form $\frac{3}{4}(1, 1) + \frac{1}{4}(0, 0)$. Now in case at least one player keeps both actions available throughout play, the payoff $(0, 0)$ occurs with a frequency of at least $\frac{1}{3}$. Therefore, the probability of absorption on $(0, 0)$ must be positive. The latter would contradict the fact that each player can get strictly more by keeping both actions alive. □

This completes our analysis of $(3, 3)$-restricted coordination games of size 2×2. Theorem 4.3.7 below is a generalization of lemma 4.3.1 to coordination games of arbitrary size; it shows that a $(3, 3)$-restriction suffices to make any convex combination of the diagonal payoffs obtainable as an equilibrium reward.

Theorem 4.3.7 *Consider a 2-player $(3, 3)$-restricted coordination game of size $m \times m$. Every convex combination of the diagonal payoffs can be supported by an equilibrium.*

Proof. Consider a reward $\check{\gamma} = \left(\check{\gamma}^1, \check{\gamma}^2\right)$ that is a convex combination of the diagonal payoffs. Then, because any two-dimensional polytope can be subdivided in triangles with the same set of extreme points, $\check{\gamma}$ is a convex combination of at most three diagonal payoffs D_i, D_j and D_k. Without loss of generality, suppose that $\check{\gamma} = \beta_1 \cdot D_i + \beta_2 \cdot D_j + \beta_3 \cdot D_k$, where $\beta = (\beta_1, \beta_2, \beta_3) \in \Delta^3$, and suppose that $\beta_1 \geq \beta_2 \geq \beta_3$. Let $\lambda = (\frac{3}{2}(\beta_1 + \beta_3) - \frac{1}{2}, \frac{3}{2}\beta_2, 0) \in \Delta^3$ and $\mu = (1 - 3\beta_3, 0, 3\beta_3) \in \Delta^3$, and write $D = (D_i, D_j, D_k)$. Then

$$\check{\gamma} = \left(\frac{2}{3}\lambda + \frac{1}{3}\mu\right) \cdot D.$$

Since both λ and μ only put positive weight on at most two diagonal payoffs, we can use agreements $\left(\pi^\lambda_{\mathcal{A}}, \sigma^\lambda_{\mathcal{A}}\right)$ and $(\pi^\mu_{\mathcal{A}}, \sigma^\mu_{\mathcal{A}})$ to support $\lambda \cdot D$ and $\mu \cdot D$ as equilibrium rewards respectively. We will now define strategies that result in playing either $\left(\pi^\lambda_{\mathcal{A}}, \sigma^\lambda_{\mathcal{A}}\right)$ and $(\pi^\mu_{\mathcal{A}}, \sigma^\mu_{\mathcal{A}})$ with probability $\frac{2}{3}$ and $\frac{1}{3}$ respectively, from stage 2 onwards. We can do so by having the players play actions i, j and k each with probability $\frac{1}{3}$ at stage 1, followed by playing $\left(\pi^\lambda_{\mathcal{A}}, \sigma^\lambda_{\mathcal{A}}\right)$ if the first stage actions are the same, and followed by $(\pi^\mu_{\mathcal{A}}, \sigma^\mu_{\mathcal{A}})$ otherwise. ■

Theorems 4.3.6 and 4.3.7 can easily be generalized to games with a milder restriction: $\left(r^1, r^2\right) \geq (3, 3)$. Then we obtain the following results:

Theorem 4.3.8 *In an $\left(r^1, r^2\right)$-restricted 2×2-coordination game with $\left(r^1, r^2\right) \geq (3, 3)$ every convex combination of equilibrium rewards can be obtained as an equilibrium reward.*

and

Theorem 4.3.9 *In an $\left(r^1, r^2\right)$-restricted $m \times m$ - coordination game with $\left(r^1, r^2\right) \geq (3,3)$ every convex combination of the diagonal payoffs can be supported by an equilibrium.*

4.4 N-player $(3,3,\ldots,3)$-restricted coordination games

Although not true for the 2-player case (cf. example 4.2), we now show that for the N-player case with $N \geq 3$ the set of equilibrium rewards in $(3,3,\ldots,3)$-restricted coordination games equals the set of feasible individually rational rewards. The proof uses lemmas 4.4.1 and 4.4.2 below.

Lemma 4.4.1 *Let $p \geq 2$. For any $z \in \Delta^p$ there exist $y_1, y_2, ..., y_{p-1} \in \Delta^p$ such that $z = \frac{1}{p-1} \sum_{i=1}^{p-1} y_i$ and for any y_i at most two coordinates are non-zero.*

Proof. We prove the result by induction.

If $p = 2$, then z has at most two non-zero coordinates, so we can take $y_1 = z$. Suppose the result is true for p. We now show that it is also true for $p+1$. Take $z = (z_1, z_2, \ldots, z_{p+1}) \in \Delta^{p+1}$ and suppose, without loss of generality, that $z_1 \geq z_2 \geq \cdots \geq z_p \geq z_{p+1}$. Notice that $pz_1 + z_{p+1} \geq z_1 + z_2 + \cdots z_{p+1} = 1$, so $z_1 - \frac{1}{p} + z_{p+1} \geq 0$. Then by letting $w := \left(\frac{p}{p-1}(z_1 - \frac{1}{p} + z_{p+1}, z_2, \ldots, z_p)\right) \in \Delta^p$ and $y_p := (1 - pz_{p+1}, 0, \ldots, 0, pz_{p+1}) \in \Delta^{p+1}$ we have

$$z = (z_1, z_2, \ldots, z_p, z_{p+1}) = \frac{p-1}{p} \cdot (w, 0) + \frac{1}{p} \cdot y_p.$$

By induction there are $w_1, w_2, \ldots, w_{p-1} \in \Delta^p$, such that $w = \frac{1}{p-1} \sum_{q=1}^{p-1} w_q$, while each w_q has at most two nonzero coordinates. Then, by letting $y_q = (w_q, 0)$ for $q = 1, 2, \ldots, p$, we get

$$z = \frac{1}{p} \sum_{q=1}^{p} y_q$$

which completes the proof. ■

Lemma 4.4.2 *In any N-player, $N \geq 3$, $(3,3,\ldots,3)$-restricted coordination game $(0,0,\ldots,0)$ can be obtained as an equilibrium reward.*

Proof. Recall that each player has at least 2 actions. Let player 1 play $(1,2,1,2,\ldots)$ and let player 2 play $(2,1,2,1,\ldots)$, and let all other players play $(1,1,1,1,\ldots)$, while in addition if player 2 does not play according to this plan, then player 1 continues by playing action 2 exclusively and similarly, if player 1 does not play according to this plan, then player 2 continues by playing action 2 exclusively. Then, clearly, the rewards are 0 and none of the players has a profitable deviation. ■

We are now ready to prove a Folk-theorem for $(3,3,\ldots,3)$-restricted coordination games:

Theorem 4.4.3 *In any N-player, $N \geq 3$, $(3,3,\ldots,3)$-restricted coordination game all feasible (and individually rational) rewards can be obtained as limiting average equilibrium rewards.*

Proof. The proof proceeds in a number of steps. In step 1 we prove that any convex combination of two payoffs can be obtained as an equilibrium reward by an agreement. Thus we obtain a skeleton of equilibrium rewards. In step 2 we fill the space inside the skeleton.

Step 1: Any convex combination of two diagonal payoffs can be obtained in a similar way as this was done for the 2-player case in Lemma 4.3.1. Now take a convex combination of 0 and a diagonal payoff, let's say D_1 corresponding to action 1 for all players. So we can write this convex combination as αD_1. Then, similarly to the 2-player case we can get αD_1 by focussing on entries $(1,1,\ldots,1)$ giving D_1 and $(2,2,1,1,\ldots,1)$ giving 0. Almost the same strategies can be used as in the 2-player case; the only difference is that in case $N = 3$ players 1 and 2 should prevent player 3 from playing action 2 along with them; this can be done by using the threat that players 1 and 2 will play actions 1 and 2 respectively from that moment onwards. As far as the randomizations are concerned, it is only players 1 and 2 who may need to randomize and they only randomize on actions 1 and 2. Thus every convex combination of two payoffs can be obtained by an agreement.

Step 2: If all players have m actions then there are at most $m+1$ different payoffs. Let u be an arbitrary convex combination of these $m+1$ payoffs $D_0, D_1, \ldots, D_m$, where $D_0 = 0 \in \mathbb{R}^N$. So $u = \sum_{l=0}^{m} \alpha_l D_l$, where $\alpha \in \Delta^{m+1}$. By Lemma 4.4.1 there are $\beta_1, \beta_2, \ldots, \beta_m \in \Delta^{m+1}$ such that each β_j has at most 2 non-zero coordinates and $\alpha = \frac{1}{m}\sum_{j=1}^{m} \beta_j$. Then, each β_j corresponds to an agreement, since for each of them at most 2 payoffs are involved. Moreover

$$u = \sum_{l=0}^{m} \alpha_l \cdot D_l = \frac{1}{m}\sum_{j=1}^{m}\sum_{l=0}^{m} \beta_j(l) \cdot D_l = \frac{1}{m}\sum_{j=1}^{m} \gamma(\beta_j)$$

where $\gamma(\beta_j)$ is the reward corresponding to β_j. At step 1 suppose players 1 and 2 both play $(\frac{1}{m}, \frac{1}{m}, \ldots, \frac{1}{m})$, while all other players play action 1. If $(a, b, 1, 1, \ldots, 1)$ is the entry selected at stage 1, the from stage 2 onwards the players play the agreement that corresponds to β_j, where $j = (a+b) \bmod m + 1$. This yields an equilibrium with reward u which completes the proof. ∎

As in the previous section the result presented in theorem 4.4.3 can easily be generalized to games with a milder restriction:

Theorem 4.4.4 *In any N-player, $N \geq 3$, $\left(r^1, r^2, \ldots, r^N\right)$-restricted coordination game with $\left(r^1, r^2, \ldots, r^N\right) \geq (3,3,\ldots,3)$ all feasible (and individually rational) rewards can be obtained as limiting average equilibrium rewards.*

Chapter 5

General-sum Games With Vanishing Actions

5.1 Introduction

In this chapter we analyze games with vanishing actions without any additional structure in the payoffs, general-sum games with vanishing actions. In these games pure strategies play an important role, since if a player uses a pure strategy, then we can pinpoint in advance the stages, at which he unlearns one (or more) of his actions. Furthermore any deviation from a pure strategy can immediately be detected by (the) other player(s) and hence every sin brings its punishment with it. One more attractive property of pure strategies is that in ordinary repeated games every feasible reward can be obtained by a pair of pure strategies. Before turning our attention to the actual games, we first take a look at the frequencies, by which the players have to play their different actions in order not to unlearn them, and at pairs of pure strategies that guarantee that certain actions will remain alive during the entire course of the game. For pure strategy pairs whose long-run joint action frequencies converge, we analyze these frequencies in section 5.2 by means of frequency matrices. A frequency matrix is a matrix whose (i, j)-th entry consists of the long-run frequency of the related action pair. In section 5.3 we apply the results of section 5.2 to find pure-strategy equilibrium rewards in general-sum games with vanishing actions. In section 5.4 we use a generalized version of the agreement (cf. definition (4.3.2)) to obtain convex combinations of pure-strategy equilibria as equilibrium rewards. This chapter is based on Schoenmakers, Joosten, Peters & Thuijsman (forthcoming). In Joosten, Brenner & Witt (2002) several models making use of this type of strategies, are discussed.

5.2 Frequency matrices

This entire section merely deals with so-called jointly-convergent strategies and their corresponding frequency matrices, which are defined as follows:

Definition 5.2.1 *A pair of pure strategies (π, σ) is jointly-convergent if there exists a nonnegative $(m \times n)$-matrix F, such that*

$$F_{ij} = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \#\left\{t \,|\, \pi_t = i, \sigma_t = j\right\}.$$

Hence π and σ are geared to one another in such a way that in the long run action pair (i, j) is played with frequency F_{ij}. A pair of jointly-convergent strategies is denoted by (π_c, σ_c) and the matrix $F(\pi_c, \sigma_c)$ is called the frequency matrix corresponding to the pair of jointly convergent strategies (π_c, σ_c).

Notice that

$$F_{ij} \geq 0 \text{ for all } i, j \text{ and } \sum_{i,j} F_{ij} = 1.$$

Clearly a pair of pure strategies generally is not jointly-convergent, because the action frequencies do not necessarily converge. We will now characterize the set $\mathbb{F}^{r^1,r^2}$ of frequency matrices that can be obtained, if the players make use of a pair of jointly-convergent strategies in an $\left(r^1, r^2\right)$-restricted game. Notice first that if π_c prescribes to keep action $i \in I$ alive, then i must be played at least once in each set of r^1 consecutive stages (since otherwise it vanishes) and therefore its frequency must be at least $\frac{1}{r^1}$. Furthermore if π_c prescribes to unlearn action i, then in the long run its frequency will always converge to 0. A similar argument holds for σ_c. In terms of F this means that:

$$\sum_{j\in J} F_{ij}(\pi_c, \sigma_c) \geq \frac{1}{r^1} \text{ or } \sum_{j\in J} F_{ij}(\pi_c, \sigma_c) = 0 \text{ for each } i \in I \tag{5.1}$$

and

$$\sum_{i\in I} F_{ij}(\pi_c, \sigma_c) \geq \frac{1}{r^2} \text{ or } \sum_{i\in I} F_{ij}(\pi_c, \sigma_c) = 0 \text{ for each } j \in J. \tag{5.2}$$

Let $\mathbb{F}^{r^1,r^2}_{I',J'}$ be the set of obtainable frequency matrices in an $\left(r^1, r^2\right)$-restricted game, in which players 1 and 2 keep the actions in $I' \subset I$ and $J' \subset J$ respectively alive, whereas the other actions will be unlearned during the course of play. Then from (5.1) and (5.2) it follows that necessary conditions for a frequency matrix F to be in $\mathbb{F}^{r^1,r^2}_{I',J'}$ are:

$$\sum_{j\in J} F_{ij} \geq \frac{1}{r^1} \text{ for each } i \in I', \ \sum_{j\in J} F_{ij} = 0 \text{ for each } i \notin I' \tag{5.3}$$

and

$$\sum_{i\in I} F_{ij} \geq \frac{1}{r^2} \text{ for each } j \in J' \text{ and } \sum_{i\in I} F_{ij} = 0 \text{ for each } j \notin J'. \tag{5.4}$$

The set $\mathbb{F}^{r^1,r^2}$ of frequency matrices that can be obtained, if the players make use of a pair of jointly-convergent strategies in an $\left(r^1, r^2\right)$-restricted game, is the union of all sets $\mathbb{F}^{r^1,r^2}_{I',J'}$:

$$\mathbb{F}^{r^1,r^2} = \bigcup_{I' \subset I, J' \subset J} \mathbb{F}^{r^1,r^2}_{I',J'}.$$

We will characterize $\mathbb{F}^{r^1,r^2}$ by characterizing all sets $\mathbb{F}^{r^1,r^2}_{I',J'}$. This is necessary, since conditions (5.3) and (5.4) are not always sufficient for a frequency matrix F to belong to $\mathbb{F}^{r^1,r^2}$ as will become clear later (cf. theorem 5.2.8).

5.2.1 Frequency matrices in 2×2 - games with $(r^1, r^2) \geq (3, 3)$ and $\gcd\left\{r^1, r^2\right\} \geq 2$

The analysis of frequency matrices is a rather complex procedure. Therefore we will start by analyzing the frequency matrices of 2×2-games with an $\left(r^1, r^2\right) \geq (3, 3)$-restriction. The analysis is split into 2 parts:

1. $\gcd\left\{r^1, r^2\right\} \geq 2$ (gcd = greatest common divisor),

2. $\gcd\left\{r^1, r^2\right\} = 1$.

In this section we analyze frequency matrices of games with $\gcd\left\{r^1, r^2\right\} \geq 2$; frequency matrices of games with $\gcd\left\{r^1, r^2\right\} = 1$ will be discussed in section 5.2.2.

We start with an example.

Example 5.1:

Consider a $(4, 6)$-restricted 2×2-game and the frequency matrix

$$F^1 = \begin{array}{cc} & \begin{array}{cc} L & R \end{array} \\ \begin{array}{c} T \\ B \end{array} & \frac{1}{12}\begin{pmatrix} 7 & 2 \\ 3 & 0 \end{pmatrix} \end{array}$$

Notice that $F^1_{BL} + F^1_{BR} = \frac{1}{4} = \frac{1}{r^1}$ and $F^1_{TR} + F^1_{BR} = \frac{1}{6} = \frac{1}{r^2}$. The following pair of jointly-convergent strategies exactly leads to the frequency matrix F:

$\pi_c = TTTBTTTBTTTB$ repeatedly

and

$\sigma_c = RLLLLLRLLLLL$ repeatedly.

Notice that indeed during each period of 12 consecutive stages the action pair TL is selected 7 times, TR is selected 2 times and BL 3 times, where player 1's low-frequency action B is played only at some stage numbers that are divisible by 2 and player 2's low-frequency action R is only selected at some stages that are not

divisible by 2. Notice that it is impossible for player 1 to play action T with a higher frequency without losing action B and that player 2 can not play action L with a higher frequency without unlearning action R. □

Consider the frequency matrix

$$F^1 = \begin{array}{cc} & \\ T \\ B \end{array} \begin{array}{c} \begin{array}{ccc} L & & R \end{array} \\ \frac{1}{r^1 r^2} \begin{pmatrix} r^1 r^2 - r^1 - r^2 & & r^1 \\ r^2 & & 0 \end{pmatrix} \end{array}$$

and suppose without loss of generality that the greatest common divisor of r^1 and r^2 is $\mathfrak{g} \geq 2$. We will show that $F^1 \in \mathbb{F}_{I,J}^{r^1,r^2}$. Notice first that F^1 satisfies conditions (5.3) and (5.4). Since r^1 and r^2 are both divisible by $\mathfrak{g}$, we can easily construct a jointly-convergent strategy pair (π_c, σ_c) that leads to the frequency matrix F^1: Let π_c prescribe to play action B only at stages $\mathfrak{g}$, $\mathfrak{g} + r^1$, $\mathfrak{g} + 2r^1$, $\mathfrak{g} + 3r^1, \ldots$ and let σ_c prescribe to play action R only at stages 1, $1 + r^2$, $1 + 2r^2$, $1 + 3r^2, \ldots$ Then (π_c, σ_c) is jointly-convergent with $F(\pi_c, \sigma_c) = F^1$.

We will now prove that

Theorem 5.2.2 *F^1 is an extreme point of $\mathbb{F}_{I,J}^{r^1,r^2}$.*

Proof. We have $F^1_{BL} + F^1_{BR} = \frac{1}{r^1}$ and $F^1_{TR} + F^1_{BR} = \frac{1}{r^2}$ and $F^1 \in \mathbb{F}_{I,J}^{r^1,r^2}$. Take $\alpha \in (0,1)$ and $\tilde{F}, \hat{F} \in \mathbb{F}_{I,J}^{r^1,r^2}$ such that $F^1 = \alpha\tilde{F} + (1-\alpha)\hat{F}$. Then

$$\begin{aligned} \tilde{F}_{BL} + \tilde{F}_{BR} &= \frac{1}{r^1} \\ \tilde{F}_{TR} + \tilde{F}_{BR} &= \frac{1}{r^2} \\ \hat{F}_{BL} + \hat{F}_{BR} &= \frac{1}{r^1} \\ \hat{F}_{TR} + \hat{F}_{BR} &= \frac{1}{r^2} \end{aligned}$$

and

$$\tilde{F}_{BR} = \hat{F}_{BR} = 0.$$

But then $\tilde{F}_{BL} = \hat{F}_{BL} = \frac{1}{r^1} = F^1_{BL}$ and $\tilde{F}_{TR} = \hat{F}_{TR} = \frac{1}{r^2} = F^1_{TR}$ and hence $\tilde{F} = \hat{F} = F^1$. ■

Theorem 5.2.4 below characterizes the set of frequency matrices in 2×2 - games, if both players keep both of their actions alive. In that case we have $I' = I$ and $J' = J$ with $|I| = |J| = 2$.

Theorem 5.2.3 *If* $\gcd\{r^1, r^2\} \geq 2$*, then all of the following frequency matrices are extreme points of* $\mathbb{F}_{I,J}^{r^1,r^2}$*:*

$$F^1 = \frac{1}{r^1 r^2}\begin{pmatrix} r^1r^2 - r^1 - r^2 & r^1 \\ r^2 & 0 \end{pmatrix}, F^2 = \frac{1}{r^1 r^2}\begin{pmatrix} r^1 & r^1r^2 - r^1 - r^2 \\ 0 & r^2 \end{pmatrix},$$
$$F^3 = \frac{1}{r^1 r^2}\begin{pmatrix} r^2 & 0 \\ r^1r^2 - r^1 - r^2 & r^1 \end{pmatrix}, F^4 = \frac{1}{r^1 r^2}\begin{pmatrix} 0 & r^2 \\ r^1 & r^1r^2 - r^1 - r^2 \end{pmatrix},$$
$$F^5 = \frac{1}{\min\{r^1, r^2\}}\begin{pmatrix} \min\{r^1, r^2\} - 1 & 0 \\ 0 & 1 \end{pmatrix},$$
$$F^6 = \frac{1}{\min\{r^1, r^2\}}\begin{pmatrix} 0 & \min\{r^1, r^2\} - 1 \\ 1 & 0 \end{pmatrix},$$
$$F^7 = \frac{1}{\min\{r^1, r^2\}}\begin{pmatrix} 0 & 1 \\ \min\{r^1, r^2\} - 1 & 0 \end{pmatrix}$$

and

$$F^8 = \frac{1}{\min\{r^1, r^2\}}\begin{pmatrix} 1 & 0 \\ 0 & \min\{r^1, r^2\} - 1 \end{pmatrix}.$$

Proof. For each of the mentioned matrices we can prove that they are extreme points of $\mathbb{F}_{I,J}^{r^1,r^2}$ in a fashion similar to the proof of theorem 5.2.2. ■

Example 5.1 (continued)

The extreme points of $\mathbb{F}_{I,J}^{4,6}$ are:

$$F^1 = \frac{1}{12}\begin{pmatrix} 7 & 2 \\ 3 & 0 \end{pmatrix}, F^2 = \frac{1}{12}\begin{pmatrix} 2 & 7 \\ 0 & 3 \end{pmatrix}, \tag{5.5}$$
$$F^3 = \frac{1}{12}\begin{pmatrix} 3 & 0 \\ 7 & 2 \end{pmatrix}, F^4 = \frac{1}{12}\begin{pmatrix} 0 & 3 \\ 2 & 7 \end{pmatrix}, \tag{5.6}$$
$$F^5 = \frac{1}{4}\begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}, F^6 = \frac{1}{4}\begin{pmatrix} 0 & 3 \\ 1 & 0 \end{pmatrix}, \tag{5.7}$$
$$F^7 = \frac{1}{4}\begin{pmatrix} 0 & 1 \\ 3 & 0 \end{pmatrix}, F^8 = \frac{1}{4}\begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}. \tag{5.8}$$

Notice that in matrices F^5, F^6, F^7 and F^8 only two cells are selected with positive frequency. In that case the tightest restriction (in this case r^1) determines the minimal frequency of the low-frequency action: For $F^8 = \frac{1}{4}\begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}$ we have that since player 1 has to play action T once every 4 stages ($r^1 = 4$), player 2 has to play action L once every 4 stages as well. The extreme points of the other subsets of $\mathbb{F}^{4,6}$ can

easily be calculated. They are:

$$\begin{aligned}
&\text{for } \mathbb{F}^{4,6}_{I,\{L\}} : \frac{1}{4}\begin{pmatrix} 3 & 0 \\ 1 & 0 \end{pmatrix} \text{ and } \frac{1}{4}\begin{pmatrix} 1 & 0 \\ 3 & 0 \end{pmatrix}, \\
&\text{for } \mathbb{F}^{4,6}_{I,\{R\}} : \frac{1}{4}\begin{pmatrix} 0 & 3 \\ 0 & 1 \end{pmatrix} \text{ and } \frac{1}{4}\begin{pmatrix} 0 & 1 \\ 0 & 3 \end{pmatrix}, \\
&\text{for } \mathbb{F}^{4,6}_{\{T\},J} : \frac{1}{6}\begin{pmatrix} 1 & 5 \\ 0 & 0 \end{pmatrix} \text{ and } \frac{1}{6}\begin{pmatrix} 5 & 1 \\ 0 & 0 \end{pmatrix}, \\
&\text{for } \mathbb{F}^{4,6}_{\{B\},J} : \frac{1}{6}\begin{pmatrix} 0 & 0 \\ 1 & 5 \end{pmatrix} \text{ and } \frac{1}{6}\begin{pmatrix} 0 & 0 \\ 5 & 1 \end{pmatrix}, \\
&\text{for } \mathbb{F}^{4,6}_{\{T\},\{L\}} : \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \text{ for } \mathbb{F}^{4,6}_{\{T\},\{R\}} : \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \\
&\text{for } \mathbb{F}^{4,6}_{\{B\},\{L\}} : \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \text{ and for } \mathbb{F}^{4,6}_{\{B\},\{R\}} : \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.
\end{aligned} \tag{5.9}$$

□

Switching from one frequency matrix to another

The next part of the analysis deals with frequency matrices F' that are convex combinations of the extreme frequency matrices. In order to obtain these matrices we use jointly-convergent strategies (π_c, σ_c) that prescribe to play according to an extreme frequency matrix $\tilde{F}$ for a while, then switch to another extreme frequency matrix $\hat{F}$, play according to $\hat{F}$ for a while and then switch to the next extreme frequency matrix etc. In order that (π_c, σ_c) exactly leads to the frequency matrix F', the following conditions have to be taken into consideration. Firstly, no action may be unlearned when switching from one extreme frequency matrix to another and secondly the limit frequency of "switching stages" should be 0. Here a switching stage is a stage at which (π_c, σ_c) does not prescribe to play according to any of the extreme frequency matrices, but merely prescribes an action pair in order that neither player loses an action.

Example 5.1 (continued)

Take matrices F^2 and F^4 in (5.5) respectively (5.6). If there is a switch from F^2 to F^4, then player 2 keeps playing action L with frequency $\frac{1}{r^2}$, but for player 1 the low-frequency action changes from B to T and π_c prescribes to play action T at those stages that player 2 does not play L. For example (π_c, σ_c) could prescribe to play $TL\ TR\ TR\ BR\ TR\ TR\ TL\ BR\ TR\ TR\ TR\ BR$ repeatedly during the F^2-stages and $BL\ TR\ BR\ BR\ BR\ TR\ BL\ BR\ BR\ TR\ BR\ BR$ repeatedly during the F^4-stages. In this case switching between F^2 and F^4 never leads to the loss of an action of any player. Furthermore, there is no switching stage needed.

Now consider matrices F^1 and F^5. For each player the low-frequency action in F^1 is the same one as the low-frequency action in F^5. In that case (π_c, σ_c) could prescribe to play $TL\ TL\ TL\ BL\ TR\ TL\ TL\ BL\ TL\ TL\ TR\ BL$ repeatedly during the F^1-stages and $TL\ TL\ TL\ BR$ repeatedly during the F^5-stages and, again, neither player loses an action and no switching stages are needed. □

It is, however, not always possible to make switches between extreme frequency matrices without using switching stages. Consider for example the frequency matrices F^1 and F^5 in theorem 5.2.3 with $r^1 = r^2 = r$:

$$F^1 = \frac{1}{r}\begin{pmatrix} r-2 & 1 \\ 1 & 0 \end{pmatrix} \text{ and } F^5 = \frac{1}{r}\begin{pmatrix} r-1 & 0 \\ 0 & 1 \end{pmatrix}.$$

For F^5 to be played repeatedly, in each set of r consecutive stages there must be exactly one stage in which both player 1 and player 2 play their low-frequency action: $TL\ TL \ldots TL\ BR$ repeatedly. To play F^1 repeatedly, in each set of r consecutive stages there must be exactly one stage, in which player 1 plays his low-frequency action and there must be exactly one stage, in which player 2 plays his low-frequency action and these stages may not be the same one: for example $TL\ TL \ldots TL\ TR\ BL$ repeatedly. Now consider the frequency matrix $F = \alpha F^1 + (1-\alpha) F^5$ with $\alpha \in (0,1)$. In order to obtain F as a result of a jointly-convergent pair of strategies (π_c, σ_c), we have to make an infinite number of switches from F^5 to F^1 and back to F^5. Suppose (π_c, σ_c) prescribes $TL\ TL \ldots TL\ BR$ (leading to F^5) for a number of times, then follows $TL\ TL \ldots TL\ TR\ BL$ (according to F^1) for a while, such that the ratio between the number of F^5-stages and the number of F^1-stages is $\frac{\alpha}{1-\alpha}$. Here an F^5-stage is a stage at which (π_c, σ_c) prescribes to play according to F^5. A sequence of such stages will be called an F^5-sequence. In order to switch back to F^5 after the first F^1-sequence (π_c, σ_c) can not prescribe the action sequence $TL\ TL \ldots TL\ BR$ again, since player 2 would then play action L for r consecutive stages and thereby unlearn action R. Therefore the action pair BR has to move forward (at least) one position in the sequence and the second time the players play according to F^5 the prescribed action sequence is $TL\ TL \ldots TL\ BR\ TL$. But then, after the next switch, when (π_c, σ_c) again prescribes to play according to F^1, the TR and the BL have to be moved forward one position in the sequence as well (or the BL has to be moved forward 2 positions, which makes no difference for the analysis). This means that the second time the players play according to F^1, they repeat the following sequence: $TL\ TL \ldots TL\ TR\ BL\ TL$. But then, after the next switch to F^5 the BR has to be moved forward one position in the sequence again. This pattern continues until the BR action pair is first in the F^5-sequence: $BR\ TL\ TL \ldots TL$. Now it is impossible to make a switch to the F^1-sequence again, since at the first stage of this F^1-sequence both players have to play their low-frequency action in order not to unlearn it and this action pair, namely BR, is never to be played in an F^1-sequence. This means that at the end of this F^5-sequence one extra BR has to be inserted in order to make sure that both players keep both actions alive and the last $r+1$ stages of this "F^5-sequence" are actually played according to $\frac{1}{r+1}\begin{pmatrix} r-1 & 0 \\ 0 & 2 \end{pmatrix}$. The next F^1-sequence then can be chosen equal to the first one: $TL\ TL \ldots TL\ TR\ BL$. From here on the reasoning starts all over again and eventually an infinite number of extra BR's has to be inserted.

Now suppose for simplicity's sake that $\alpha \in \mathbb{Q}$. Then $\alpha = \frac{\alpha_1}{\alpha_5}$ for some $\alpha_1, \alpha_5 \in \{1, 2, 3, \ldots\}$ and consider the following strategy pair $(\hat{\pi}_c, \hat{\sigma}_c)$: at stages $1, 2, \ldots, \alpha_1 \cdot r$ the prescribed actions by $(\hat{\pi}_c, \hat{\sigma}_c)$ are $TL\ TL \ldots TL\ TR\ BL$ repeatedly, an F^1-sequence of length $\alpha_1 \cdot r$. Then at stages $\alpha_1 \cdot r + 1, \alpha_1 \cdot r + 2, \ldots, (\alpha_1 + \alpha_5) \cdot r$ the prescribed actions are $TL\ TL \ldots BR\ TL$ repeatedly, an F^5-sequence of length $\alpha_5 \cdot r$.

Now a switching stage is inserted, at which $(\hat{\pi}_c, \hat{\sigma}_c)$ prescribes to play action pair BR. Now at the next $2\alpha_1 \cdot r$ stages $(\hat{\pi}_c, \hat{\sigma}_c)$ prescribes an F^1-sequence again and at the subsequent $2\alpha_5 \cdot r$ stages $(\hat{\pi}_c, \hat{\sigma}_c)$ prescribes an F^5-sequence again. Now, again, a switching stage is inserted, at which $(\hat{\pi}_c, \hat{\sigma}_c)$ prescribes to play action pair BR. At the next $3\alpha_1 \cdot r$ stages thereafter $(\hat{\pi}_c, \hat{\sigma}_c)$ prescribes an F^1-sequence again, at the subsequent $3\alpha_5 \cdot r$ stages $(\hat{\pi}_c, \hat{\sigma}_c)$ prescribes an F^5-sequence again and then a switching stage is inserted once again, at which the action pair BR is selected. The strategy pair $(\hat{\pi}_c, \hat{\sigma}_c)$ keeps on prescribing to alternate between F^1-sequences, F^5-sequences and switching stages, where each time the length of the F^1-sequences and the F^5-sequences increases by $\alpha_1 \cdot r$ and $\alpha_5 \cdot r$ respectively. Notice that the frequency of switching stages converges to 0 and that the pair $(\hat{\pi}_c, \hat{\sigma}_c)$ is jointly-convergent with frequency matrix $F(\hat{\pi}_c, \hat{\sigma}_c) = F$.

Example 5.1 (continued)

We will show how to obtain the frequency matrix

$$F = \begin{pmatrix} \frac{5}{12} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{12} \end{pmatrix}$$

as a result of jointly-convergent strategies in a $(3,3)$-restricted game. Notice that

$$F = \frac{3}{4} \cdot \begin{pmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 \end{pmatrix} + \frac{1}{4} \cdot \begin{pmatrix} \frac{2}{3} & 0 \\ 0 & \frac{1}{3} \end{pmatrix} = \frac{3}{4}F^1 + \frac{1}{4}F^5,$$

where F^1 and F^5 are as in theorem 5.2.4. Consider the following part of the jointly-convergent strategy pair (π_c, σ_c): (π_c, σ_c) prescribes to play $TL\ TR\ BL$ and to repeat this sequence $3n+2$ times. After that (π_c, σ_c) prescribes to play $TL\ BR\ TL$ and to repeat this sequence n times. After that (π_c, σ_c) prescribes to play $TL\ BR$ once and then the sequence $TL\ TR\ BL$ again for another $3n+2$ times etc. Then neither player loses an action and the frequency matrix corresponding to these $12n+8$ stages of (π_c, σ_c) is

$$\hat{F}_n = \frac{1}{12n+8} \begin{pmatrix} 5n+3 & 3n+2 \\ 3n+2 & n+1 \end{pmatrix}.$$

The jointly-convergent strategy pair (π_c, σ_c) prescribes to first play according to $\hat{F}_1$, then $\hat{F}_2$, then $\hat{F}_3$ etcetera, all in the fashion described above. Notice that neither player loses an action and that indeed this strategy pair is jointly-convergent. Let F'_n denote the joint frequency matrix after the stages in $\hat{F}_1, \hat{F}_2, \ldots, \hat{F}_n$. Some calculations show that

$$F'_n = \frac{1}{6n^2+14n} \begin{pmatrix} \frac{5}{2}n^2 + \frac{11}{2}n & \frac{3}{2}n^2 + \frac{7}{2}n \\ \frac{3}{2}n^2 + \frac{7}{2}n & \frac{1}{2}n^2 + \frac{3}{2}n \end{pmatrix}.$$

Furthermore the frequency matrix corresponding (π_c, σ_c) is exactly the limit for n tending to infinity of F'_n or

$$F(\pi_c, \sigma_c) = \lim_{n\to\infty} F'_n = \begin{pmatrix} \frac{5}{12} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{12} \end{pmatrix} = F.$$

For any pair of extreme frequency matrices F^i and F^j as mentioned in theorem 5.2.3 we can, in a similar fashion, show that $\alpha F^i + (1-\alpha)F^j$ is obtainable by means of a pair of jointly-convergent strategies. But then any convex combination of $F^1, F^2, \ldots, F^8$ can be obtained by means of a jointly-convergent strategy pair.

Theorem 5.2.4 *For any* $\left(r^1, r^2\right)$*-restricted* 2×2*-game with* $\gcd\left\{r^1, r^2\right\} \geq 2$ *we have:*

$$\mathbb{F}^{r^1,r^2} = \bigcup_{I'\subset I, J'\subset J} \mathbb{F}^{r^1,r^2}_{I',J'}$$

where $\mathbb{F}^{r^1,r^2}_{I,J}$ *is the convex hull of the frequency matrices* $F^1, F^2, \ldots, F^8$ *as mentioned in theorem 5.2.3.*

Proof. Since every convex combination of $F^1, F^2, \ldots, F^8$ can be obtained by a pair of jointly-convergent strategies, it is sufficient to show that every frequency matrix in $\mathbb{F}^{r^1,r^2}_{I,J}$ is a convex combination of matrices $F^1, F^2, \ldots, F^8$.

Consider a frequency matrix

$$F = \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix}$$

such that $F \in \mathbb{F}^{r^1,r^2}_{I,J}$. F satisfies conditions 5.3 and 5.4 and hence

$$f_{11} + f_{12} \geq \frac{1}{r^1}$$
$$f_{21} + f_{22} \geq \frac{1}{r^1}$$
$$f_{11} + f_{21} \geq \frac{1}{r^2}$$

and

$$f_{12} + f_{22} \geq \frac{1}{r^2}.$$

Suppose without loss of generality that $r^1 \leq r^2$ and that $\frac{f_{11}}{f_{22}} = \min\{\frac{f_{11}}{f_{22}}, \frac{f_{22}}{f_{11}}, \frac{f_{12}}{f_{21}}, \frac{f_{21}}{f_{12}}\}$. We split the analysis in two cases:

Case 1: $\frac{f_{11}}{f_{22}} \geq \frac{1}{r^1-1}$.

In this case F can straightforwardly be constructed as a convex combination of F^5, F^6, F^7 and F^8.

Case 2: $\frac{f_{11}}{f_{22}} < \frac{1}{r^1-1}$.

Notice that F is a linear combination of $F^1, F^2, \ldots, F^8$. Let

$$F = \sum_{j=1}^{8} \alpha^k \cdot F^k.$$

Then, since

$$\sum_{i,j} F^k_{ij} = 1 \text{ for all } k \in \{1, 2, \ldots, 8\}$$

we have that

$$\sum_{j=1}^{8} \alpha^k = 1.$$

Now take $\alpha^1 = \alpha^2 = \alpha^3 = \alpha^5 = 0$ and $\alpha^8 = f_{11} \cdot r^1$, which, by assumption, is in $[0, 1)$. Take furthermore

$$\alpha^4 = \frac{\big(f_{22} - f_{11} \cdot (r^1 - 1)\big) \cdot r^1 r^2}{r^1 r^2 - r^1 - r^2},$$

$$\alpha^6 = (r^1 - 1 - \tfrac{1}{r^1 - 1}) \cdot r^1 \cdot \Big(f_{12} + f_{11} - \tfrac{1}{r^1 - 1} \cdot (f_{21} + f_{22})\Big)$$

and

$$\alpha^7 = 1 - \alpha^4 - \alpha^6 - \alpha^8.$$

Some straightforward calculations show that in that case we have that $\alpha^k \geq 0$ for all $k \in \{1, 2, \ldots, 8\}$ and hence, since F is a linear combination of $F^1, F^2, \ldots, F^8$, it must even be a convex combination of $F^1, F^2, \ldots, F^8$. ■

Theorem 5.2.4 states that for $\big(r^1, r^2\big)$-restricted 2×2-games with $\gcd\big\{r^1, r^2\big\} \geq 2$ the closure of the set of obtainable frequency matrices is exactly characterized by (5.1) and (5.2).

5.2.2 Frequency matrices in 2×2 - games with $(r^1, r^2) \geq (3, 3)$ and $\gcd\{r^1, r^2\} = 1$

For 2×2-games with an $\big(r^1, r^2\big)$-restriction with $\gcd\big\{r^1, r^2\big\} = 1$, the analysis is more complicated, as the following example shows:

Example 5.2

Consider a 2×2-game with a $(3, 4)$-restriction and the following frequency matrix:

$$F = \begin{array}{c} \\ T \\ B \end{array} \begin{array}{c} \begin{array}{cc} L & R \end{array} \\ \frac{1}{12} \begin{pmatrix} 5 & 3 \\ 4 & 0 \end{pmatrix} \end{array}$$

Notice that F satisfies conditions (5.1) and (5.2). Neither player is to lose an action. It is impossible to arrange 5 TL-stages, 3 TR-stages and 4 BL-stages in a block of 12 consecutive stages, such that neither player loses an action when repeatedly playing this block. This is the case, since player 1 has to play the action sequence $T\ T\ B$ repeatedly and player 2 has to repeat playing $L\ L\ L\ R$ and the prescribed sequence of joint actions must therefore be something like $TL\ TL\ BL\ TR\ TL\ BL\ TL\ TR\ BL\ TL\ TL\ BR$ repeatedly. But then the action pair BR is played with a strictly positive frequency (namely once every 12 stages) and hence the matrix F can never be obtained as frequency matrix of a pair of jointly-convergent strategies. This phenomenon occurs, if the strategies of both players prescribe to play one of the 2 actions with a frequency that is just slightly above or equal to $\frac{1}{r^k}$. □

We will now characterize the set $\mathbb{F}_{I,J}^{r^1, r^2}$, again by finding its extreme points. Therefore we use the following theorem by Euclid:

Theorem 5.2.5 *Euclid*
For every pair of integers α and β there exist two integers ζ and η such that $\zeta\cdot\alpha+\eta\cdot\beta = \gcd\{\alpha,\beta\}$.

From this theorem we deduce the following corollary:

Corollary 5.2.6 *For every pair of positive integers α and β there exist two positive integers ζ and η such that $\zeta\cdot\alpha-\eta\cdot\beta = \gcd\{\alpha,\beta\}$.*

Proof. Let α and β be two strictly positive integers. By theorem 5.2.5 we can take integers $\tilde{\zeta}$ and $\tilde{\eta}$ such that $\tilde{\zeta}\cdot\alpha+\tilde{\eta}\cdot\beta = \gcd\{\alpha,\beta\}$. Since $\min\{\alpha,\beta\} \geq \gcd\{\alpha,\beta\}$ we either have $\tilde{\zeta}\geq 0$ and $\tilde{\eta}\leq 0$ or $\tilde{\zeta}\leq 0$ and $\tilde{\eta}\geq 0$. If $\tilde{\zeta}\geq 0$ and $\tilde{\eta}\leq 0$, then we can take $\zeta = \tilde{\zeta}\geq 0$ and $\eta = -\tilde{\eta}\geq 0$ and then $\zeta\cdot\alpha-\eta\cdot\beta = \gcd\{\alpha,\beta\}$. If $\tilde{\zeta}\leq 0$ and $\tilde{\eta}\geq 0$, then we can take

$$\zeta = -\tfrac{\alpha\cdot\tilde{\zeta}}{\gcd\{\alpha,\beta\}} + \tilde{\zeta} + 1 \geq 0$$

and

$$\eta = \tilde{\eta}\cdot\left(\tfrac{\alpha}{\gcd\{\alpha,\beta\}} - 1\right) \geq 0$$

and, again, $\zeta\cdot\alpha-\eta\cdot\beta = \gcd\{\alpha,\beta\}$. ■

Now consider the following pair of strategies: π_c prescribes to play $B\ T\ T\ldots T$ repeatedly, where the number of T's is r^1-1 and σ_c prescribes to play the sequence $L\ L\ldots L\ R$ repeatedly, where the number of L's is r^2-1 until this would lead to an action pair of BR. In that case σ_c prescribes to play R one stage earlier. It is important to observe that against any strategy of player 1 that prescribes to play action R with frequency $\frac{1}{r^1}$, player 2's strategy σ_c is the one that puts the lowest frequency on action R, under the condition that R should not be unlearned and the action pair BR should be played with frequency 0. Let ω be the smallest number in $\mathbb{N} = \{1,2,3,\ldots\}$ such that $\omega\cdot r^2-1$ is divisible by r^1. Since $\gcd\left\{r^1,r^2\right\} = 1$ by corollary 5.2.6 this number exists and clearly $\omega\in\left\{1,2,\ldots,r^2-1\right\}$. Then σ_c prescribes to play the following sequence consisting of ω subsequences: $L\ L\ldots L\ R\quad\ldots\ldots\quad L\ L\ldots L\ R\quad L\ L\ldots L\ R$ repeatedly, where the number of L's is r^2-1 in the first $\omega-1$ subsequences and the number of L's is r^2-2 in the last subsequence. This means that the long-run frequency of action R is $\frac{\omega}{\omega r^2-1}$. The (jointly-convergent) strategy pair (π_c,σ_c) leads to the following frequency matrix:

$$F^1 = \frac{1}{\omega r^2-1}\begin{pmatrix} \frac{(r^1-1)(\omega r^2-1)}{r^1} - \omega & \omega \\ \frac{\omega r^2-1}{r^1} & 0 \end{pmatrix}.$$

Theorem 5.2.7 *F^1 is an extreme point of $\mathbb{F}_{I,J}^{r^1,r^2}$.*

Proof. Notice that π_c prescribes to play action B with frequency $\frac{1}{r^1}$. Notice furthermore that, given π_c, any strategy of player 2 that prescribes to play action R with a long-run frequency that is lower than $\frac{\omega}{\omega r^2-1}$, leads to either the unlearning of action R or to a strictly positive frequency of stages in which the action pair

BR is being played. This means that for player 2, using σ_c, to be able to put a lower frequency on action R without losing it and without getting a strictly positive frequency of the action pair BR, player 1 has to play another strategy that also prescribes to play action B with frequency $\frac{1}{r^1}$. This means playing the strategy $\hat{\pi}$ prescribing the sequence $T\ T\ldots T\ B\ T\ T\ldots T$ of length r^1 repeatedly, where the B is not at position 1. But then, since $\gcd\{r^1,r^2\}=1$, there is a stage $\hat{t}\in\{1,\ldots,r^1r^2\}$, for which we have: $(\sigma_c)_{\hat{t}}=R$ and $\hat{\pi}_{\hat{t}+1}=B$. Consequently for each stage t we have: $(\pi_c,\sigma_c)_t=(\hat{\pi},\sigma_c)_{t+\hat{t}}$ and hence the long-run frequency matrix corresponding to the strategy pair $(\hat{\pi},\sigma_c)$ is also F^1. Therefore the lowest frequency player 2 can put on action R, such that he does not unlearn it and the action pair BR is played with frequency 0 is $\frac{\omega}{\omega r^2-1}$ and F^1 is an extreme point of $\mathbb{F}_{I,J}^{r^1,r^2}$. ∎

Let ϑ be the smallest number in $\mathbb{N}$ such that $\vartheta\cdot r^1-1$ is divisible by r^2, let σ_c prescribe to play the sequence $R\ L\ L\ldots L$ repeatedly, where the number of L's is r^2-1 and let π_c prescribe to play the following sequence consisting of ϑ subsequences: $T\ T\ldots T\ B$ $\ldots\ldots\ldots$ $T\ T\ldots T\ B$ $T\ T\ldots T\ B$ repeatedly, where the number of T's is r^1-1 in the first $\vartheta-1$ subsequence and r^1-2 in the last subsequence. Then, analogously to the argumentation above we have that $\vartheta\in\{1,2,\ldots,r^1-1\}$, that (π_c,σ_c) leads to the frequency matrix

$$F^2=\frac{1}{\vartheta r^1-1}\begin{pmatrix}\frac{(r^2-1)(\vartheta r^1-1)}{r^2}-\vartheta & \frac{\vartheta r^1-1}{r^2}\\ \vartheta & 0\end{pmatrix}$$

and that F^2 is an extreme point of $\mathbb{F}_{I,J}^{r^1,r^2}$. In a similar fashion the other extreme points of $\mathbb{F}_{I,J}^{r^1,r^2}$ can be found leading to:

Theorem 5.2.8 *For any (r^1,r^2)-restricted 2×2-game with $\gcd\{r^1,r^2\}=1$ we have:*

$$\mathbb{F}^{r^1,r^2}=\bigcup_{I'\subset I,J'\subset J}\mathbb{F}_{I',J'}^{r^1,r^2}$$

where $\mathbb{F}_{I,J}^{r^1,r^2}$ is the convex hull of the following frequency matrices:

$$F^1=\frac{1}{\omega r^2-1}\begin{pmatrix}\frac{(r^1-1)(\omega r^2-1)}{r^1}-\omega & \omega\\ \frac{\omega r^2-1}{r^1} & 0\end{pmatrix},$$

$$F^2=\frac{1}{\vartheta r^1-1}\begin{pmatrix}\frac{(r^2-1)(\vartheta r^1-1)}{r^2}-\vartheta & \frac{\vartheta r^1-1}{r^2}\\ \vartheta & 0\end{pmatrix},$$

$$F^3=\frac{1}{\omega r^2-1}\begin{pmatrix}\omega & \frac{(r^1-1)(\omega r^2-1)}{r^1}-\omega\\ 0 & \frac{\omega r^2-1}{r^1}\end{pmatrix},$$

$$F^4=\frac{1}{\vartheta r^1-1}\begin{pmatrix}\frac{\vartheta r^1-1}{r^2} & \frac{(r^2-1)(\vartheta r^1-1)}{r^2}-\vartheta\\ 0 & \vartheta\end{pmatrix},$$

$$F^5=\frac{1}{\omega r^2-1}\begin{pmatrix}\frac{\omega r^2-1}{r^1} & 0\\ \frac{(r^1-1)(\omega r^2-1)}{r^1}-\omega & \omega\end{pmatrix},$$

$$F^6 = \frac{1}{\vartheta r^1 - 1}\begin{pmatrix} \vartheta & 0 \\ \frac{(r^2-1)(\vartheta r^1-1)}{r^2} - \vartheta & \frac{\vartheta r^1-1}{r^2} \end{pmatrix},$$

$$F^7 = \frac{1}{\omega r^2 - 1}\begin{pmatrix} 0 & \frac{\omega r^2-1}{r^1} \\ \omega & \frac{(r^1-1)(\omega r^2-1)}{r^1} - \omega \end{pmatrix},$$

$$F^8 = \frac{1}{\vartheta r^1 - 1}\begin{pmatrix} 0 & \vartheta \\ \frac{\vartheta r^1-1}{r^2} & \frac{(r^2-1)(\vartheta r^1-1)}{r^2} - \vartheta \end{pmatrix},$$

$$F^9 = \frac{1}{\min\{r^1, r^2\}}\begin{pmatrix} \min\{r^1, r^2\} - 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$F^{10} = \frac{1}{\min\{r^1, r^2\}}\begin{pmatrix} 0 & \min\{r^1, r^2\} - 1 \\ 1 & 0 \end{pmatrix},$$

$$F^{11} = \frac{1}{\min\{r^1, r^2\}}\begin{pmatrix} 0 & 1 \\ \min\{r^1, r^2\} - 1 & 0 \end{pmatrix}$$

and

$$F^{12} = \frac{1}{\min\{r^1, r^2\}}\begin{pmatrix} 1 & 0 \\ 0 & \min\{r^1, r^2\} - 1 \end{pmatrix}.$$

Proof. For each of the 12 matrices mentioned in the theorem, it can, similarly to the proof of theorem 5.2.2, be proved that they are extreme points of the set $\mathbb{F}_{I,J}^{r^1,r^2}$. Similarly to the proof of theorem 5.2.4 it can be shown that no other frequency matrices are obtainable. ■

Together theorems 5.2.4 and 5.2.8 cover all possibilities concerning the sets $\mathbb{F}_{I,J}^{r^1,r^2}$ in 2×2 - games. The sets $\mathbb{F}_{I',J'}^{r^1,r^2}$ corresponding to jointly-convergent strategies (π_c, σ_c), leading to the loss of an action for at least one player, are independent of $\gcd\{r^1, r^2\}$, since at least one of the players has only 1 action left. These sets are all similar to the ones mentioned in (5.9).

5.2.3 Frequency matrices in $m \times n$ - games

For $m \times n$ - games the figuring out if a frequency matrix is obtainable by jointly-convergent strategies, is very complex. Therefore we will not make a complete characterization of the set of obtainable frequency matrices in $m \times n$-games. However, we have established some results, the most straightforward one being theorem 5.2.9.

Theorem 5.2.9 *Consider an (r^1, r^2)-restricted game and let $|I'| = |J'| = 2$. If $\gcd\{r^1, r^2\} \geq 2$, then the set $F_{I',J'}^{r^1,r^2}$ coincides with the set we found in theorem 5.2.4, whereas if $\gcd\{r^1, r^2\} = 1$, then the set $F_{I',J'}^{r^1,r^2}$ coincides with set we found in theorem 5.2.8.*

Proof. Use the same pair of jointly-convergent strategies (π_c, σ_c) but then applied to the action sets I' and J'. ■

From now on we consider subsets I' and J' with $\max\{|\,I'\,|, |\,J'\,|\} \geq 3$. In a $(3,3)$-restricted game there is no way to switch from

$$F^1 = \frac{1}{3}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ to } F^2 = \frac{1}{3}\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \tag{5.10}$$

However, in the slightly milder $(3,4)$-restricted game this switch can be made as the following joint action sequence shows (here the actions of players 1 and 2 are named $T\ M\ B$ and $L\ M\ R$ respectively):

$$TL\ MM\ BR\ TM\ ML\ BM\ TR\ ML\ BM.$$

Notice that the first 3 stages of this action sequence are played according to F^1 and the last 3 according to F^2. Therefore, when considering frequency matrices and corresponding pairs of jointly-convergent strategies (π_c, σ_c) keeping the actions in I' and J' respectively alive, we distinguish between 2 cases:

1. $|\,I'\,| = r^1$ and $|\,J'\,| = r^2$
2. $|\,I'\,| \leq r^1 - 1$ and $|\,J'\,| \leq r^2$ or $|\,I'\,| \leq r^1$ and $|\,J'\,| \leq r^2 - 1$.

For frequency matrices $\tilde{F}$ and $\hat{F}$ belonging to the first category, it is impossible to switch from $\tilde{F}$ to $\hat{F}$ and the set $\mathbb{F}^{r^1,r^2}_{I',J'}$ consists of a finite number of distinct frequency matrices. The following theorem, stated without proof, gives the exact number.

Theorem 5.2.10 *In an $\left(r^1, r^2\right)$-restricted $m \times n$-game with $m \geq r^1$ and $n \geq r^2$ for each pair I' and J' with $|\,I'\,| = r^1$ and $|\,J'\,| = r^2$ the total number of frequency matrices in $\mathbb{F}^{r^1,r^2}_{I',J'}$ is equal to*

$$\frac{r^1!}{(\frac{r^1}{\kappa})!} \cdot \frac{r^2!}{(\frac{r^2}{\kappa})!} \cdot \frac{1}{\kappa}.$$

Here $\kappa := \gcd\left\{r^1, r^2\right\}$.

In particular if $\gcd\left\{r^1, r^2\right\} = 1$, then the only frequency matrix in $\mathbb{F}^{r^1,r^2}_{I',J'}$ is the one with the number $\frac{1}{r^1 r^2}$ in each entry and if $\gcd\left\{r^1, r^2\right\} = \kappa$, then each of the frequency matrices in $\mathbb{F}^{r^1,r^2}_{I',J'}$ consist of $\frac{r^1 r^2}{\kappa}$ entries with the number $\frac{\kappa}{r^1 r^2}$ in it and $r^1 r^2 - \frac{r^1 r^2}{\kappa}$ entries containing a 0.

Theorem 5.2.11 *In an $\left(r^1, r^2\right)$-restricted game for each pair of subsets I' and J' with $|\,I'\,| \leq r^1 - 1$ and $|\,J'\,| \leq r^2$ or $|\,I'\,| \leq r^1$ and $|\,J'\,| \leq r^2 - 1$, the set $\mathbb{F}^{r^1,r^2}_{I',J'}$ is convex.*

Proof. Notice that the statement is equivalent to the statement that for any two frequency matrices $F, F' \in \mathbb{F}^{r^1,r^2}_{I',J'}$ it is possible to switch from F to F' and back. If $|\,I'\,| \leq r^1 - 1$ and $|\,J'\,| \leq r^2 - 1$, then this statement is obviously true, since after they stopped playing $F \in \mathbb{F}^{r^1,r^2}_{I',J'}$ both players can alter their action sequence, in order to start F' without losing an action. Now suppose w.l.o.g. that $|\,I'\,| = r^1$ and $|\,J'\,| \leq r^2 - 1$ and take $F, F' \in \mathbb{F}^{r^1,r^2}_{I',J'}$. Then there exist pure strategies π, σ and σ' such that (π, σ) is jointly-convergent with $F(\pi, \sigma) = F$ and (π, σ') is jointly-convergent with $F(\pi, \sigma') = F'$. Since player 2 can still change his action sequence, a strategy $\tilde{\sigma}$ exists that plays according to σ for a number of stages and then, via a switching period in which the action sequence of player 2 is altered, without losing an action starts playing according to σ', keeps playing that way for a while and then switches back to σ etcetera. ■

From now on we analyze for (r^1, r^2)-restricted $m \times n$-games frequency matrices $F \in \mathbb{F}^{r^1,r^2}_{I',J'}$ with $|I'| = m' \in \{2, \ldots, m\}$ and $|J'| = n' \in \{2, \ldots, n\}$ of the following form:

$$F = \frac{1}{r^1 r^2} \begin{pmatrix} \kappa & r^1 & \ldots & r^1 & 0 & \ldots & 0 \\ r^2 & & & & & & \\ \vdots & & & & & & \\ r^2 & & & 0 & & & \\ 0 & & & & & & \\ \vdots & & & & & & \\ 0 & & & & & & \end{pmatrix}. \tag{5.11}$$

Here the number of r^1's in the first row is $n' - 1$, the number of r^2's in the first column is $m' - 1$ and $\kappa \geq 0$ is the number that makes the entries of F add up to 1:

$$\kappa = r^1 r^2 - (n' - 1) \cdot r^1 - (m' - 1) \cdot r^2,$$

which means that for F to be a frequency matrix we must have:

$$(n' - 1) \cdot r^1 + (m' - 1) \cdot r^2 \leq r^1 r^2. \tag{5.12}$$

Furthermore for F to satisfy conditions (5.3) and (5.4) also the sum of the elements in the first row must be at least $\frac{1}{r^1}$ or

$$\kappa + (n' - 1) \cdot r^1 \geq r^2. \tag{5.13}$$

Similarly for the elements in the first column we find:

$$\kappa + (m' - 1) \cdot r^2 \geq r^1. \tag{5.14}$$

Definition 5.2.12 *A pure action of player k is a low-frequency action, if its prescribed frequency is $\frac{1}{r^k}$.*

Theorem 5.2.13 *Consider a frequency matrix F of the form (5.11) that satisfies conditions (5.13) and (5.14). Then F is obtainable by a jointly-convergent strategy pair (π_c, σ_c) if and only if the following inequality holds:*

$$\left\lceil \frac{(m' - 1) \cdot \gcd\{r^1, r^2\}}{r^1} \right\rceil + \left\lceil \frac{(n' - 1) \cdot \gcd\{r^1, r^2\}}{r^2} \right\rceil \leq \gcd\{r^1, r^2\}. \tag{5.15}$$

The proof of theorem 5.2.13 can be found in the appendix.

Theorem 5.2.14 *If the frequency matrix F, mentioned in (5.11), is obtainable, then it is an extreme point of $\mathbb{F}^{r^1 r^2}_{I',J'}$ with $I' = \{1, 2, \ldots, m'\}$ and $J' = \{1, 2, \ldots, n'\}$.*

Proof. Suppose that $F \in \mathbb{F}^{r^1 r^2}_{I',J'}$ and consider frequency matrices F^1 and F^2 in $\mathbb{F}^{r^1 r^2}_{I',J'}$ such that $F = \alpha \cdot F^1 + (1-\alpha) \cdot F^2$ for some $\alpha \in \langle 0, 1\rangle$. Then $F^1_{1j} = F^2_{1j} = \frac{1}{r^2} = F_{1j}$ for all $j \in \{2, 3, \ldots, n'\}$. Furthermore $F^1_{i1} = F^2_{i1} = \frac{1}{r^1} = F_{i1}$ for all $i \in \{2, 3, \ldots, m'\}$ and obviously for each action pair (i, j) for which $F_{ij} = 0$, we must also have: $F^1_{ij} = F^2_{ij} = 0$. The only elements of the $m \times n$ -matrices F^1 and F^2 that are not fixed by these restrictions, F^1_{11} and F^2_{11}. Since $\sum_{i=1}^{m}\sum_{j=1}^{n} F^1_{ij} = \sum_{i=1}^{m}\sum_{j=1}^{n} F^2_{ij} = 1$ we have: $F^1_{11} = F^2_{11} = 1 - (m'-1)\cdot\frac{1}{r^1} - (n'-1)\cdot\frac{1}{r^2} = \frac{\kappa}{r^1 r^2} = F_{11}$. Therefore $F^1 = F^2 = F$ and hence F is an extreme point of $\mathbb{F}^{r^1 r^2}_{I',J'}$. ∎

Some elementary calculations show that the statement

$$\left\lceil \frac{(m'-1)\cdot \gcd\left\{r^1, r^2\right\}}{r^1} \right\rceil + \left\lceil \frac{(n'-1)\cdot \gcd\left\{r^1, r^2\right\}}{r^2} \right\rceil \leq \gcd\left\{r^1, r^2\right\}. \tag{5.16}$$

as in theorem 5.2.13 is equivalent with the statement

$$r^2 - \left\lceil \frac{(m'-1)\cdot \gcd\left\{r^1, r^2\right\}}{r^1} \right\rceil \cdot \frac{r^2}{\gcd\left\{r^1, r^2\right\}} \geq n' - 1 \tag{5.17}$$

and also with the statement

$$r^1 - \left\lceil \frac{(n'-1)\cdot \gcd\left\{r^1, r^2\right\}}{r^2} \right\rceil \cdot \frac{r^1}{\gcd\left\{r^1, r^2\right\}} \geq m' - 1. \tag{5.18}$$

These inequalities have an intuitive explanation. The strategy σ_c prescribes to repeat playing the action sequence played at stages $\left\{1, 2, \ldots, r^2\right\}$. Consider the set of $m'-1$ stages in $\left\{1, 2, \ldots, r^1\right\}$ at which π_c prescribes to play a low-frequency action. If this set is the set $\left\{\hat{t}_1, \hat{t}_2, \ldots, \hat{t}_{m'-1}\right\}$ as constructed in the appendix (page 109), then the number of stages in $\left\{1, 2, \ldots, r^2\right\}$ at which σ_c can not prescribe to play a low-frequency action, is equal to $\left\lceil \frac{(m'-1)\cdot\gcd\left\{r^1, r^2\right\}}{r^1} \right\rceil \cdot \frac{r^2}{\gcd\left\{r^1, r^2\right\}}$. The number of stages in $\left\{1, 2, \ldots, r^2\right\}$ at which σ_c has to prescribe to play a low-frequency action, is equal to $n'-1$ and the total number of stages in $\left\{1, 2, \ldots, r^2\right\}$ is obviously equal to r^2. Consequently for F to be obtainable as a frequency matrix, a necessary condition is that inequality (5.17) holds. The fact that this inequality is also sufficient, is due to the fact that it is equivalent with inequality (5.18), the player-1-equivalent of (5.17).

These types of inequalities we will use for the analysis of general frequency matrices. Suppose without loss of generality that $r^1 \geq r^2$ and consider the following frequency matrix:

$$F=\frac{1}{r^1 r^2}\begin{pmatrix}
\kappa_1\, r^1\, r^1 \dots r^1 & & & & & & & \\
r^2 & & & & & & & \\
r^2 & & & & & & & \\
\vdots & & & & & & & \\
r^2 & & & & & & & \\
 & \kappa_2\, r^1\, r^1 \dots r^1 & & 0 & & & & \\
 & r^2 & & & & & & \\
 & r^2 & & & & & & \\
 & \vdots & & & & & & \\
 & r^2 & & & & & & \\
 & & \ddots & & & & & \\
 & & & \kappa_{\mathfrak{k}}\, r^1\, r^1 \dots r^1 & & & & \\
 & & & r^2 & & & & \\
 & & & r^2 & & & & \\
 & & & \vdots & & & & \\
 & 0 & & r^2 & & & & \\
 & & & & r^1 & & & \\
 & & & & & \ddots & & \\
 & & & & & & r^1 & \\
 & & & & & & 0 & \\
 & & & & & & & \ddots \\
 & & & & & & & 0
\end{pmatrix} \tag{5.19}$$

where κ_i is the smallest nonnegative integer such that

$$\kappa_i + n^i r^1 \geq r^2 \text{ for all } i = 2, 3, \dots, \mathfrak{k} \tag{5.20}$$

and

$$\kappa_i + m^i r^2 \geq r^1 \text{ for all } i = 2, 3, \dots, \mathfrak{k} \tag{5.21}$$

and where κ_1 contains the rest of the weight such that the sum of the entries in F add up to 1. Notice that, since $r^1 \geq r^2$, inequality (5.20) automatically holds for $\kappa_i = 0$. Now in (5.19) let m_i be the number of r^2's that is below the cell containing the κ_i and, similarly, let n_i denote the number of r^1's that is to the right of the cell containing the κ_i. Furthermore let $n_{\mathfrak{k}+1}$ be the number of r^1's in the lower right part of the matrix. Then player 1 has $\sum_{i=1}^{\mathfrak{k}} m_i$ low-frequency actions and player 2 has $\sum_{i=1}^{\mathfrak{k}+1} n_i$ low-frequency actions and we obtain the following generalization of inequalities (5.17) and (5.18):

$$r^2 - \left\lceil \frac{\sum_{i=1}^{\mathfrak{k}} m_i \cdot \gcd\left\{r^1, r^2\right\}}{r^1} \right\rceil \cdot \frac{r^2}{\gcd\left\{r^1, r^2\right\}} \geq \sum_{i=1}^{\mathfrak{k}+1} n_i \tag{5.22}$$

and

$$r^1 - \left\lceil \frac{\sum_{i=1}^{\mathfrak{k}+1} n_i \cdot \gcd\{r^1, r^2\}}{r^2} \right\rceil \cdot \frac{r^1}{\gcd\{r^1, r^2\}} \geq \sum_{i=1}^{\mathfrak{k}} m_i, \tag{5.23}$$

both of which are equivalent with

$$\left\lceil \frac{\sum_{i=1}^{\mathfrak{k}} m_i \cdot \gcd\{r^1, r^2\}}{r^1} \right\rceil + \left\lceil \frac{\sum_{i=1}^{\mathfrak{k}+1} n_i \cdot \gcd\{r^1, r^2\}}{r^2} \right\rceil \leq \gcd\{r^1, r^2\} \tag{5.24}$$

Now theorem 5.2.15 below is the generalization of theorem 5.2.13 to frequency matrices of the form (5.19).

Theorem 5.2.15 *Consider a frequency matrix F of the form (5.19) that satisfies conditions (5.20) and (5.21) and let $r^1 \geq r^2$. Then F is obtainable by a jointly-convergent strategy pair (π_c, σ_c) if and only if inequality (5.24) holds.*

Proof. Similar to the proof of theorem 5.2.13. ∎

Notice that if $r^1 = r^2$, then although the number of low-frequency actions of player 1 increases from $\sum_{i=1}^{\mathfrak{k}} m_i$ to $\sum_{i=1}^{\mathfrak{k}} m_i + n_{\mathfrak{k}+1}$, the statement in theorem 5.2.15 still holds. At any corresponding stage each player plays a low-frequency action, but such a stage still only counts for one. The generalization of theorem 5.2.14 is:

Theorem 5.2.16 *If the frequency matrix F, mentioned in (5.19), is obtainable, then it is an extreme point of $\mathbb{F}^{r^1 r^2}_{I',J'}$ with $I' = \left\{1, 2, \ldots, \mathfrak{k} + \sum_{i=1}^{\mathfrak{k}} m_i + n_{\mathfrak{k}+1}\right\}$ and $J' = \left\{1, 2, \ldots, \mathfrak{k} + \sum_{i=1}^{\mathfrak{k}+1} n_i\right\}$.*

Proof. Analogously to the proof of theorem 5.2.14. ∎

5.2.4 Frequency matrices in N-player games

In this subsection we analyze the obtainability of frequency matrices in N-player restricted games. For N-player games as well as notation 4.1.1 we use following notations:

Notation 5.2.17 *For N-player games:*

1. *$i = \left(i^1, i^2, \ldots, i^N\right)$ is a joint pure action (for all players);*
2. *$I = I^1 \times I^2 \times \ldots \times I^N$ is the set of joint pure actions;*
3. *$i^{-k} = \left(i^1, \ldots, i^{k-1}, i^{k+1}, \ldots, i^N\right)$ is a joint pure action for all players except player k;*

4. $I^{-k} = I^1 \times \ldots \times I^{k-1} \times I^{k+1} \times \ldots \times I^N$ *is the set of joint pure actions for all players except player* k.

As long as each player has at least 2 actions at the start of the game, the frequency matrices are N-dimensional. For example in a $(3,3,3)$-restricted $2 \times 2 \times 2$-game we can have the following frequency matrix:

$$F = \begin{array}{cc} & \begin{array}{cc} L & R \end{array} \\ \begin{array}{c} T \\ B \end{array} & \frac{1}{3}\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \end{array} \qquad \begin{array}{cc} & \begin{array}{cc} L & R \end{array} \\ \begin{array}{c} T \\ B \end{array} & \frac{1}{3}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \end{array}$$

$$N(ear) \qquad\qquad F(ar)$$

and F obviously is obtainable by jointly-convergent strategies.

Let $\pi_c = \left(\pi_c^1, \pi_c^2, \ldots, \pi_c^N\right)$ denote a set of jointly-convergent strategies. If π_c^k prescribes to keep action $i^k \in I^k$ alive, then the frequency of action i^k must be at least $\frac{1}{r^k}$, whereas if π_c^k prescribes to unlearn action i^k, then in the long run its frequency will converge to 0. Let $I^{k\prime} \subset I^k$ denote the action set that is kept available in the long run by player k and let F_i denote the long-run frequency of the joint pure action i. Then for each $k \in K$:

$$\sum_{i^{-k} \in I^{-k}} F_{i^k, i^{-k}}(\pi_c) \geq \frac{1}{r^k} \text{ for each } i^k \in I^{k\prime} \tag{5.25}$$

and

$$\sum_{i^{-k} \in I^{-k}} F_{i^k, i^{-k}}(\pi_c) = 0 \text{ for each } i^k \notin I^{k\prime}. \tag{5.26}$$

This is the N-player equivalent of (5.1) and (5.2).

Let

$$I' = I^{1\prime} \times I^{2\prime} \times \ldots \times I^{N\prime}$$

denote the joint pure action set that is kept alive by all players, let

$$r = \left(r^1, r^2, \ldots, r^N\right)$$

denote the joint restriction of the game and let, analogously to the 2-player case, $\mathbb{F}_{I'}^r$ be the set of frequency matrices that, while keeping the joint action set I' available, are obtainable by jointly-convergent strategies in an $r = \left(r^1, r^2, \ldots, r^N\right)$-restricted game. We now present a few results in $2 \times 2 \times \ldots \times 2$ - games. Notice first that $I^k = \{1, 2\}$ for each $k \in K$. We obviously have:

Theorem 5.2.18 *Consider an* r*-restricted* $2 \times 2 \times \ldots \times 2$ *game with* $\hat{r} := \min\{r^1, r^2, \ldots, r^N\}$ *such that* $\hat{r} \geq 2$. *Then each frequency matrix* F *with*

$$F_{(i^1, i^2, \ldots, i^N)} = \frac{\hat{r} - 1}{\hat{r}},$$

$$F_{(3-i^1,3-i^2,\ldots,3-i^N)} = \frac{1}{\hat{r}}$$

and

$F_i = 0$ *for all other* $i \in I$

is an extreme point of $\mathbb{F}_I^r$.

From now on we concentrate on N-player frequency matrices of the following form (below the 3-player version is shown):

$$F(3) = \begin{array}{c} \\ T \\ B \end{array} \begin{array}{c} \begin{array}{cc} L & R \end{array} \\ \begin{pmatrix} \kappa & \frac{1}{r^2} \\ \frac{1}{r^1} & 0 \end{pmatrix} \end{array} \qquad \begin{array}{c} \\ T \\ B \end{array} \begin{array}{c} \begin{array}{cc} L & R \end{array} \\ \begin{pmatrix} \frac{1}{r^3} & 0 \\ 0 & 0 \end{pmatrix} \end{array}$$

$$N(ear) \qquad\qquad F(ar)$$

where $\kappa = 1 - \frac{1}{r^1} - \frac{1}{r^2} - \frac{1}{r^3}$. If $F(N)$ is obtainable, then in a fashion similar to the proof of theorem 5.2.2 it can be shown that it is an extreme point of $\mathbb{F}_I^r$. Some staightforward statements concerning $F(N)$:

Lemma 5.2.19 *In an* $\left(r^1, r^2, \ldots, r^N\right)$*-restricted* $2 \times 2 \times \ldots \times 2$ *game* $F(N)$ *is obtainable if and only if the stages in* $\left\{1, 2, \ldots, \text{lcm}\left\{r^1, r^2, \ldots, r^N\right\}\right\}$ *can be partitioned in* N *parts such that each player plays his low-frequency action in a different part. (Here* lcm *is the lowest common multiple.)*

Theorem 5.2.20 *Consider an* $\left(r^1, r^2, \ldots, r^N\right)$*-restricted* $2 \times 2 \times \ldots \times 2$ *game. If*

$$\gcd\left\{r^1, r^2, \ldots, r^N\right\} =: \Omega \geq N,$$

then $F(N)$ *is obtainable.*

Proof. Let π_c be a set of jointly-convergent strategies such that:

1. for each $k \in K$: π_c^k prescribes to play the low-frequency action only at (some of the) stages t for which $t - k$ is divisible by Ω,
2. for each $k \in K$: π_c^k prescribes to play the low-frequency action with frequency $\frac{1}{r^k}$.

Then $F(\pi_c) = F(N)$. ∎

Theorem 5.2.21 *Consider an* $\left(r^1, r^2, \ldots, r^N\right)$*-restricted* $2 \times 2 \times \ldots \times 2$ *game. If for some* $k_1, k_2 \in K$*:*

$$\gcd\left\{r^{k_1}, r^{k_2}\right\} = 1,$$

then $F(N)$ *is not obtainable.*

Proof. Similar to example 5.2 at some stage players k_1 and k_2 each have to play their low-frequency actions. ∎

Theorem 5.2.22 *Consider an* $\left(r^1, r^2, r^3\right)$*-restricted* $2\times2\times2$ *game.* $F(3)$ *is obtainable if and only if for each* $k_1, k_2 \in K$*:*

$$\gcd\left\{r^{k_1}, r^{k_2}\right\} \geq 2$$

and for at least one pair $\kappa_1, \kappa_2 \in K$*,* $\kappa_1 \neq \kappa_2$*:*

$$\gcd\left\{r^{\kappa_1}, r^{\kappa_2}\right\} > 2.$$

Proof. The "⇒"-part of the statement is an immediate consequence of lemma 5.2.19 and theorem 5.2.21.
To prove the "⇐"-part of the statement let

$$\tilde{r}^{ij} := \gcd\left\{r^i, r^j\right\}$$

and suppose without loss of generality that

$$\tilde{r}^{12} \leq \tilde{r}^{13} \leq \tilde{r}^{23}.$$

Notice that r^1 is divisible by $\operatorname{lcm}\left\{\tilde{r}^{12}, \tilde{r}^{13}\right\}$ etcetera. Let π_c be a set of jointly-convergent strategies such that

1. π_c^1 prescribes to play a low-frequency action only at (some of the) stages that are divisible by $\operatorname{lcm}\left\{\tilde{r}^{12}, \tilde{r}^{13}\right\}$,
2. π_c^2 prescribes to play a low-frequency action only at some stages τ for which $\tau - 1$ is divisible by $\operatorname{lcm}\left\{\tilde{r}^{12}, \tilde{r}^{23}\right\}$,
3. π_c^3 prescribes to play a low-frequency action only at some stages τ for which $\tau + 1$ is divisible by $\operatorname{lcm}\left\{\tilde{r}^{13}, \tilde{r}^{23}\right\}$,
4. π_c^1 prescribes to play action B with frequency $\frac{1}{r^1}$,
5. π_c^2 prescribes to play action R with frequency $\frac{1}{r^2}$,
6. π_c^3 prescribes to play action F with frequency $\frac{1}{r^3}$.

Then by assumption $\operatorname{lcm}\left\{\tilde{r}^{13}, \tilde{r}^{23}\right\} \geq 3$ and the low-frequency actions of the different players are not played at the same stages. Hence $F(\pi_c) = F(3)$, which completes the proof. ■

5.3 Pure strategy equilibria

In the previous sections we figured out which frequency matrices can be obtained by pairs of jointly-convergent strategies. In this section we use this information to find out which rewards can be obtained by jointly-convergent and other pure strategies. We concentrate on 2-player games solely. However, the concepts of threat points (cf. definition 5.3.1) and agreements (cf. definition 5.4.2) that are discussed in the following subsections, can in a straightforward fashion be generalized to N-player games.

For a pair of jointly-convergent strategies (π_c, σ_c) with corresponding frequency matrix $F(\pi_c, \sigma_c)$ for $k \in \{1, 2\}$ we have:

$$\gamma^k(\pi_c, \sigma_c) = \sum_{i \in I} \sum_{j \in J} F_{ij}(\pi_c, \sigma_c) \cdot R^k(i, j). \tag{5.27}$$

We now continue with the search for equilibria that make use of jointly-convergent strategies, in which threats are implemented. In normal repeated games many of the equilibrium rewards, mentioned in the Folk-theorem, exist merely by means of threats. This is also the case in restricted games, but where ordinary repeated games contain a threat point, which is not subject to changes during the course of play, for restricted games such a fixed threat point does not exist. The following example clarifies this statement:

Example 5.3

Consider the following $(4, 5)$-restricted prisoner's dilemma:

$$\begin{array}{c} \\ T \\ B \end{array} \begin{array}{c} \begin{array}{cc} L & R \end{array} \\ \left(\begin{array}{cc} 5,5 & 0,6 \\ 6,0 & 1,1 \end{array} \right) \end{array}$$

The jointly-convergent strategy pair (π_c, σ_c) that prescribes to play the action pair BR at each stage, clearly is an equilibrium with reward $(1, 1)$. In the unrestricted repeated game $(5, 5)$ also is an equilibrium reward corresponding to a strategy pair $(\check{\pi}, \check{\sigma})$, where $\check{\pi}$ prescribes to play T as long as player 2 played action L at all previous stages and to play action B otherwise and, similarly, $\check{\sigma}$ prescribes to play L as long as player 1 played action T at all previous stages and to play action R otherwise. The pair $(\check{\pi}, \check{\sigma})$ is jointly-convergent with implemented threats. In the $(4, 5)$-restricted game the same strategy pair does not lead to an equilibrium reward, since at stage 5 player 1 has unlearned action B and thereby the game is reduced to

$$\begin{array}{c} \\ T \end{array} \begin{array}{c} \begin{array}{cc} L & R \end{array} \\ \left(\begin{array}{cc} 5,5 & 0,6 \end{array} \right) \end{array}$$

In this subgame $(5, 5)$ is not an equilibrium reward: Playing action R at each stage provides a higher reward to player 2 than playing action L at each stage and player 1 is unable to punish player 2 for this deviation. □

This example shows that the Folk-theorem, as stated in theorem 1.2.7 for repeated games, is not valid for repeated games with vanishing actions. This is caused by the fact that the threat point in the restricted game changes in time and where at stages 1, 2, and 3 the threat point is $(1, 1)$, suddenly at stage 4, if player 1 decides to play action T for the fourth time in a row and thereby to unlearn action B, the threat point changes to $(0, 6)$ and suddenly at stage 5 player 2 can guarantee himself a payoff of 6 and player 1 can no longer guarantee himself a reward of more than 0. If, however, player 1 plays a strategy that keeps only action B alive, then player 2 does not have a strategy, that reduces player 1's reward to an amount strictly smaller than 1. This means that the threat point in a restricted game depends on the strategies.

Consider the pure strategy pair (π, σ), including the following threats: If at stage t player 1 deviates from π by selecting an action $i \neq \pi_t$, then from stage $t+1$ onwards σ prescribes to play according to the strategy $\sigma'_t(i)$ that, given the deviating action i, from stage $t+1$ on minimizes player 1's maximum reward. Let $\hat{\pi}_t(i)$ denote player 1's maximizing strategy, if he selects action i at stage t. Then player 1's reward by deviating from π at stage t is at most

$$\hat{\gamma}^1_t(\pi, \sigma) = \inf_{\sigma'_t(i)} \sup_{i \in I_t} \gamma^1(\hat{\pi}_t(i), \sigma'_t(i)).$$

Analogously if at stage t player 2 deviates from σ by selecting an action j that was supposed to be played with probability 0 according to σ_t, then from stage $t+1$ onwards π prescribes to play according to the strategy $\pi'_t(j)$ that, given the deviating action j, from stage $t+1$ on minimizes player 2's maximum reward. Let $\hat{\sigma}_t(j)$ denote player 2's maximizing strategy, if he selects action j at stage t. Then player 2's reward by deviating at stage t is at most

$$\hat{\gamma}^2_t(\pi, \sigma) = \inf_{\pi'_t(j)} \sup_{j \in I_t} \gamma^2(\pi'_t(j), \hat{\sigma}_t(j)).$$

Definition 5.3.1 *The threat point corresponding to the strategy pair (π, σ) is*

$$\left(\hat{\gamma}^1(\pi, \sigma), \hat{\gamma}^2(\pi, \sigma)\right) := \left(\sup_{t \in \mathbb{N}} \hat{\gamma}^1_t(\pi, \sigma), \sup_{t \in \mathbb{N}} \hat{\gamma}^2_t(\pi, \sigma)\right).$$

and (π, σ) is an equilibrium if

$$\gamma^1(\pi, \sigma) \geq \hat{\gamma}^1(\pi, \sigma) \text{ and } \gamma^2(\pi, \sigma) \geq \hat{\gamma}^2(\pi, \sigma).$$

Example 5.3 (continued)

We will now calculate the set of equilibrium rewards of pairs of jointly-convergent strategies in the $(4,5)$-restricted prisoner's dilemma discussed in example 5.3. We have:

$$\begin{array}{cc} & \begin{array}{cc} L & R \end{array} \\ \begin{array}{c} T \\ B \end{array} & \left(\begin{array}{cc} 5,5 & 0,6 \\ 6,0 & 1,1 \end{array}\right) \end{array}$$

First consider the strategy pair $(\check{\pi}, \check{\sigma})$ adjusted to the $(4,5)$-restriction: Player 1 will only start playing action B at each stage, if he has not unlearned it yet and for player 2 a similar argument holds with respect to action R. Then $\gamma(\check{\pi}, \check{\sigma}) = (5,5)$. Furthermore $\hat{\gamma}^1(\check{\pi}, \check{\sigma}) = (1,1,1,1,5,5,5,\ldots)$, since from stage 5 on player 1 can no longer deviate from $\check{\pi}$, and $\hat{\gamma}^2(\check{\pi}, \check{\sigma}) = (1,1,1,6,6,5,5,5,5,\ldots)$, since a deviation from $\check{\sigma}$ at stages 4 or 5 can not be punished by player 1, because he unlearns action B at stage 4 by playing T for the fourth consecutive time. The 5's from stage 6 on mean that if player 2 has not deviated from $\check{\sigma}$ at one of the first 5 stages, then he unlearns action R and his payoff against $\check{\pi}$ will be 5 at all remaining stages. This means that $\gamma^2(\check{\pi}, \check{\sigma}) = 5 < \hat{\gamma}^2(\check{\pi}, \check{\sigma})$ and $(\check{\pi}, \check{\sigma})$ is not an equilibrium.

In a prisoner's dilemma the threat point of the pure strategy pair (π, σ) can easily be found: As long as, after stage t, player 1 has action B alive and player 2 has action R

alive, we have: $\hat{\gamma}_t = (\hat{\gamma}_t^1, \hat{\gamma}_t^2) = (1,1)$. If after stage t player 1 has only action T left, whereas player 2 has both actions, then $\hat{\gamma}_t(\pi,\sigma) = (0,6)$. If after stage t player 2 has only action L, whereas player 1 still has both actions available, then $\hat{\gamma}_t(\pi,\sigma) = (6,0)$. Furthermore, if $I_t^1 = \{T\}$ and $I_t^2 = \{L\}$, then $\hat{\gamma}_t(\pi,\sigma) = (5,5)$ and if according to (π,σ) at stage t player 1 unlearns action B and player 2 unlearns action R, then $\hat{\gamma}_t(\pi,\sigma) = (6,6)$.

Suppose that π and σ prescribe to keep only actions T and L respectively available. Then $\gamma(\pi,\sigma) = (5,5)$ and there is a stage t, at which either $\hat{\gamma}_t^1(\pi,\sigma) = 6$ or $\hat{\gamma}_t^2(\pi,\sigma) = 6$. Consequently (π,σ) is not an equilibrium (and hence $(5,5)$ is not an equilibrium reward).

Now suppose that π prescribes to keep both actions alive, whereas σ prescribes to keep only action L available. Then from some stage t on we have: $\hat{\gamma}_t^1(\pi,\sigma) = 6$, whereas $\gamma^1(\pi,\sigma) \leq 5\frac{4}{5}$ and hence player 1 will deviate and unlearn action T. Therefore (π,σ) is not an equilibrium.

Now suppose that π prescribes to keep both actions alive, whereas σ prescribes to keep only action R available. Then we have: $\hat{\gamma}^1(\pi,\sigma) = 1$ and $\gamma^1(\pi,\sigma) \leq \frac{4}{5}$. Hence player 1 will deviate and unlearn action T and (π,σ) is not an equilibrium.

For strategy pairs, of which π prescribes to unlearn an action and σ prescribes to keep both actions available, similar arguments show that (π,σ) is not an equilibrium.

Now suppose that π prescribes to keep only action B alive, whereas σ prescribes to keep only action R available. Then we have: $\hat{\gamma}^1(\pi,\sigma) = \hat{\gamma}^2(\pi,\sigma) = 1$ and $\gamma(\pi,\sigma) = (1,1)$ and hence (π,σ) is an equilibrium.

Now the most difficult case: Suppose that π and σ each prescribe to keep both actions alive. Then $\hat{\gamma}^1(\pi,\sigma) = \hat{\gamma}^2(\pi,\sigma) = 1$. Now assume that the pair (π,σ) is jointly-convergent. Then $F(\pi,\sigma) \in \mathbb{F}_{I,J}^{4,5}$ and, using theorem 5.2.8 and equation (5.27), we find:

$$\gamma(\pi,\sigma) \in conv\left\{(4,4), (\tfrac{64}{15}, \tfrac{58}{15}), (\tfrac{3}{2}, \tfrac{9}{2}), (\tfrac{19}{15}, \tfrac{67}{15}), (\tfrac{9}{2}, \tfrac{3}{2}), (\tfrac{71}{15}, \tfrac{23}{15}), (2,2), (\tfrac{26}{15}, \tfrac{32}{15})\right\}.$$

All of these rewards are bigger than $(1,1)$ and hence they are all equilibrium rewards. In figure 5.1 the set of feasible rewards by means of jointly-convergent strategies is depicted on the left; the set of jointly-convergent equilibrium rewards is shown on the right. □

Let $\gamma_c(F)$ be the reward corresponding to a pair of jointly-convergent strategies (π_c, σ_c) with corresponding frequency matrix F:

$$\gamma_c^k(F) = \sum_{i \in I} \sum_{j \in J} F_{ij} \cdot R^k(i,j)$$

for $k \in \{1,2\}$. Furthermore let

$$\gamma_c\left(\mathbb{F}_{I',J'}^{r^1,r^2}\right) = \left\{\gamma_c(F) \mid F \in \mathbb{F}_{I',J'}^{r^1,r^2}\right\}$$

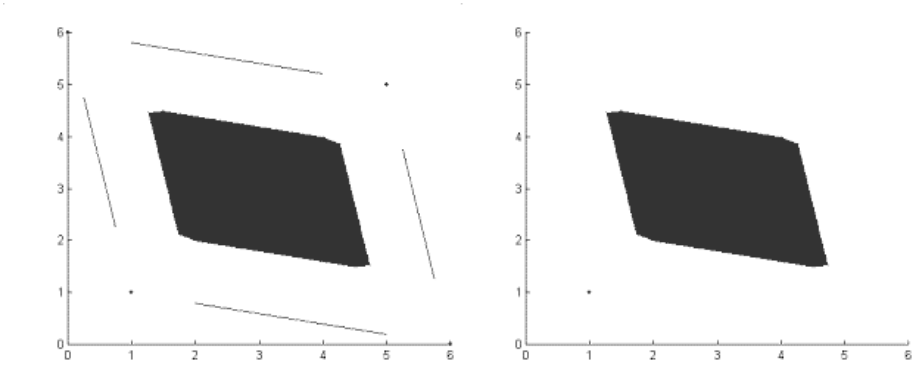


Figure 5.1: Left: Feasible rewards by jointly-convergent strategies. Right: Equilibrium rewards by jointly-convergent strategies

and

$$\gamma_c\left(\mathbb{F}^{r^1,r^2}\right) = \bigcup_{I'\subset I, J'\subset J} \gamma_c\left(\mathbb{F}^{r^1,r^2}_{I',J'}\right).$$

Then $\gamma_c\left(\mathbb{F}^{r^1,r^2}\right)$ is the set of rewards that can be obtained by a pair of jointly-convergent strategies (π_c, σ_c). Using the convexity of $\mathbb{F}^{r^1,r^2}_{I',J'}$ we can conclude that $\gamma_c\left(\mathbb{F}^{r^1,r^2}_{I',J'}\right)$ is convex.
The rest of this section deals with the rewards of pure but not jointly-convergent strategies. Here we need the concept of the average stage payoff, which is defined as follows:

Definition 5.3.2 *The average stage payoff of a pair of pure strategies (π, σ) after stage τ is*

$$\frac{1}{\tau}\sum_{t=1}^{\tau} R_t(\pi, \sigma).$$

Notice that for a pair of jointly-convergent strategies the average stage payoff converges to the limiting average reward and that for a pair of pure but not jointly-convergent strategies the average stage payoff does not converge to a unique reward.

Lemma 5.3.3 *In an $\left(r^1, r^2\right)$-restricted game, for any pair of pure strategies (π, σ) prescribing to keep only the actions in I' and J' available, every limit point of the average stage payoff of (π, σ) is in $\gamma_c\left(\mathbb{F}^{r^1,r^2}_{I',J'}\right)$.*

Proof. Consider a pair of pure strategies (π, σ) prescribing to keep only the actions in I' and J' alive and suppose that from stage τ on only the actions in I' and J' are available. Suppose first that $r^1 = r^2 = r$ and consider the r consecutive stages

$\tau, \tau+1, \ldots, \tau+r-1$. Let $i_\tau, i_{\tau+1}, \ldots, i_{\tau+r-1}$ and $j_\tau, j_{\tau+1}, \ldots, j_{\tau+r-1}$ denote the (pure) actions prescribed by π and σ respectively during these r stages and let $\tilde{F}^1$ be the frequency matrix corresponding to $(i_t, j_t)_{t=\tau}^{\tau+r-1}$. Notice that during the stages $\tau+r, \tau+r+1, \ldots, \tau+2r-1$ the players can play the same actions $i_\tau, i_{\tau+1}, \ldots, i_{\tau+r-1}$ and $j_\tau, j_{\tau+1}, \ldots, j_{\tau+r-1}$ again without losing an action and they can continue doing so. This means that $\tilde{F}^1 \in \mathbb{F}_{I',J'}^{r,r}$ and $\gamma_c(\tilde{F}^1) \in \gamma_c\left(\mathbb{F}_{I',J'}^{r,r}\right)$. Now consider the r consecutive stages $\tau+r, \tau+r+1, \ldots, \tau+2r-1$. Let $i_{\tau+r}, i_{\tau+r+1}, \ldots, i_{\tau+2r-1}$ and $j_{\tau+r}, j_{\tau+r+1}, \ldots, j_{\tau+2r-1}$ denote the (pure) actions prescribed by π and σ respectively during these r stages and let $\tilde{F}^2$ be the frequency matrix corresponding to $(i_t, j_t)_{t=\tau+r}^{\tau+2r-1}$. Then during the stages $\tau+2r, \tau+2r+1, \ldots, \tau+3r-1$ the players can play the same actions $i_{\tau+r}, i_{\tau+r+1}, \ldots, i_{\tau+2r-1}$ and $j_{\tau+r}, j_{\tau+r+1}, \ldots, j_{\tau+2r-1}$ again without losing an action and they can continue doing so. Consequently $\tilde{F}^2 \in \mathbb{F}_{I',J'}^{r,r}$. The strategy pair (π, σ) generates a sequence $(\tilde{F}^i)_{i\in\mathbb{N}}$ with $\tilde{F}^i \in \mathbb{F}_{I',J'}^{r,r}$ for all i and therefore, due to the convexity of $\gamma_c\left(\mathbb{F}_{I',J'}^{r,r}\right)$, the average stage payoff from stage τ on is in $\gamma_c\left(\mathbb{F}_{I',J'}^{r,r}\right)$. Consequently, taking the finiteness of τ into account, the average stage payoff of (π, σ) converges to $\gamma_c\left(\mathbb{F}_{I',J'}^{r,r}\right)$.

Now suppose that $r^1 \neq r^2$ and consider the $r^1 \cdot r^2$ consecutive stages $\tau, \tau+1, \ldots, \tau + r^1 \cdot r^2 - 1$. It can easily be shown that the frequency matrix $\tilde{F}^1$ corresponding to the actions $(i_t, j_t)_{t=\tau}^{\tau+r^1r^2-1}$ is in the set $\mathbb{F}_{I',J'}^{r^1,r^2}$. This is also the case for the frequency matrix $\tilde{F}^2$ corresponding to the actions $(i_t, j_t)_{t=\tau+r^1r^2}^{\tau+2r^1r^2-1}$ etc. Again this leads to the conclusion that the average stage payoff of (π, σ) converges to $\gamma_c\left(\mathbb{F}_{I',J'}^{r^1,r^2}\right)$. ∎

For player $k \in \{1, 2\}$ let $l_{I',J'}^{r^1,r^2}(k)$ denote the lowest reward that player k can get, if a pair of jointly-convergent strategies keeping action sets I' and J' alive, is played. Furthermore let $l_{I',J'}^{r^1,r^2} = \left(l_{I',J'}^{r^1,r^2}(1), l_{I',J'}^{r^1,r^2}(2)\right)$, let $\Gamma_{I',J'}^{r^1,r^2}$ be the set of rewards that can be obtained, if the players make use of pure strategies that keep the actions in I' and J' available, and let $\Gamma^{r^1,r^2} = \bigcup_{I'\subset I, J'\subset J} \Gamma_{I',J'}^{r^1,r^2}$.

Theorem 5.3.4 *For an (r^1, r^2)-restricted 2×2-game with $r^1 \neq r^2$ and $(r^1, r^2) \geq (3, 3)$, we have:*

$$\Gamma_{I',J'}^{r^1,r^2} = conv\left\{\gamma_c\left(\mathbb{F}_{I',J'}^{r^1,r^2}\right), l_{I',J'}^{r^1,r^2}\right\} \text{ for each } I' \subset I, J' \subset J.$$

Proof. We prove that

$$\Gamma_{I',J'}^{r^1,r^2} \subset conv\left\{\gamma_c\left(\mathbb{F}_{I',J'}^{r^1,r^2}\right), l_{I',J'}^{r^1,r^2}\right\}$$

and

$$\Gamma_{I',J'}^{r^1,r^2} \supset conv\left\{\gamma_c\left(\mathbb{F}_{I',J'}^{r^1,r^2}\right), l_{I',J'}^{r^1,r^2}\right\}.$$

To prove the $\subset$-statement notice that for each $\tilde{\gamma} \in \Gamma_{I',J'}^{r^1,r^2}$ we have that $\tilde{\gamma} \geq l_{I',J'}^{r^1,r^2}$ and that, according to lemma 5.3.3, the average stage payoff converges to $\gamma_c\left(\mathbb{F}_{I',J'}^{r^1,r^2}\right)$. Then automatically $\tilde{\gamma} \in conv\left\{\gamma_c\left(\mathbb{F}_{I',J'}^{r^1,r^2}\right), l_{I',J'}^{r^1,r^2}\right\}$.

The proof of the $\supset$-statement:

Consider a reward

$$\breve{\gamma} = \left(\breve{\gamma}^1, \breve{\gamma}^2\right) \in conv\left\{\gamma_c\left(\mathbb{F}_{I',J'}^{r^1,r^2}\right), l_{I',J'}^{r^1,r^2}\right\}$$

If $\breve{\gamma} \in \gamma_c\left(\mathbb{F}_{I',J'}^{r^1,r^2}\right)$, then a pair of jointly-convergent strategies (π_c, σ_c) exists with $\gamma\left(\pi_c, \sigma_c\right) = \breve{\gamma}$.
If $\breve{\gamma} \notin \gamma_c\left(\mathbb{F}_{I',J'}^{r^1,r^2}\right)$, then there are pairs of jointly-convergent strategies $\left(\pi_c^1, \sigma_c^1\right)$ and $\left(\pi_c^2, \sigma_c^2\right)$ keeping action sets I' and J' available, such that $\gamma^1\left(\pi_c^1, \sigma_c^1\right) = \breve{\gamma}^1$ and $\gamma^2\left(\pi_c^2, \sigma_c^2\right) = \breve{\gamma}^2$. Notice that $\gamma^2\left(\pi_c^1, \sigma_c^1\right) > \breve{\gamma}^2$ and $\gamma^1\left(\pi_c^2, \sigma_c^2\right) > \breve{\gamma}^1$. Now consider the following pair of pure strategies (π, σ):

Strategies π and σ prescribe to play according to $\left(\pi_c^1, \sigma_c^1\right)$ during the first 100 stages. Thereafter they prescribe to make a switch and play according to $\left(\pi_c^2, \sigma_c^2\right)$. Notice that, although it might take a stage or 2, this switch can always be made without any player losing an action. After the switch is made, (π, σ) keeps prescribing to play according to $\left(\pi_c^2, \sigma_c^2\right)$, until the average stage payoff approaches $\gamma\left(\pi_c^2, \sigma_c^2\right)$ up to a small number $\varepsilon \in (0, 1)$. Suppose this happens at stage t_1. Then

$$\left|\left|\frac{1}{t_1}\sum_{t=1}^{t_1} R_t\left(\pi, \sigma\right) - \gamma\left(\pi_c^2, \sigma_c^2\right)\right|\right| < \varepsilon.$$

At stage $t_1 + 1$ Now (π, σ) prescribes to switch back to $\left(\pi_c^1, \sigma_c^1\right)$ and (π, σ) keeps prescribing to play according to $\left(\pi_c^1, \sigma_c^1\right)$ until stage t_2, the stage at which the average stage payoff approaches $\gamma\left(\pi_c^2, \sigma_c^2\right)$ up to ε^2:

$$\left|\left|\frac{1}{t_2}\sum_{t=1}^{t_2} R_t\left(\pi, \sigma\right) - \gamma\left(\pi_c^1, \sigma_c^1\right)\right|\right| < \varepsilon^2.$$

Now (π, σ) prescribes to switch back to $\left(\pi_c^2, \sigma_c^2\right)$ and (π, σ) keeps prescribing to play according to $\left(\pi_c^2, \sigma_c^2\right)$ until stage t_3, the stage at which the average stage payoff approaches $\gamma\left(\pi_c^3, \sigma_c^3\right)$ up to ε^3:

$$\left|\left|\frac{1}{t_3}\sum_{t=1}^{t_3} R_t\left(\pi, \sigma\right) - \gamma\left(\pi_c^2, \sigma_c^2\right)\right|\right| < \varepsilon^3.$$

Now (π, σ) prescribes to switch back to $\left(\pi_c^1, \sigma_c^1\right)$ etc.

We have:

$$\gamma^1(\pi, \sigma) = \liminf_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} R_t^1\left(\pi, \sigma\right) = \breve{\gamma}^1 \text{ and}$$

$$\gamma^2(\pi, \sigma) = \liminf_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} R_t^2\left(\pi, \sigma\right) = \breve{\gamma}^2$$

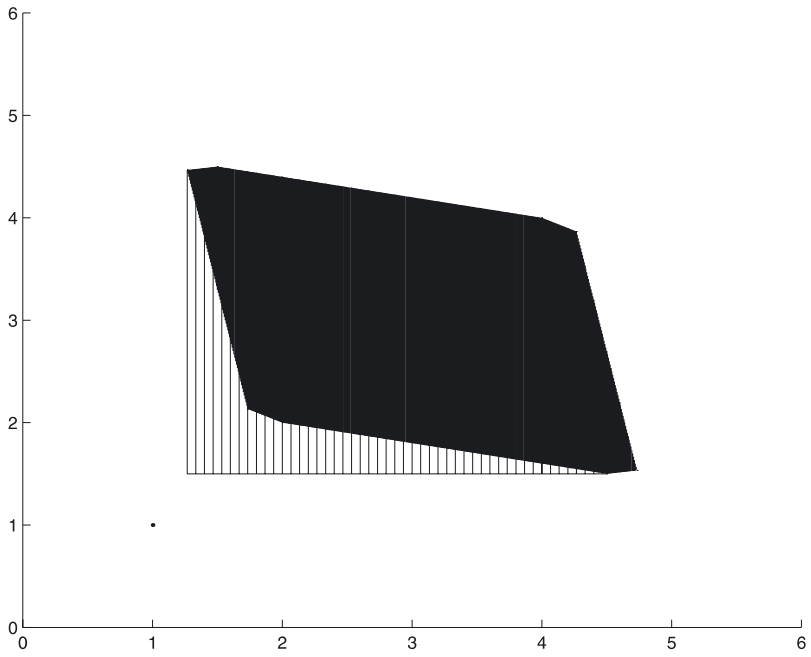


Figure 5.2: Equilibrium rewards by pure strategies

and hence $\breve{\gamma} \in \Gamma^{r^1,r^2}_{I',J'}$. ■

Example 5.3 (continued):

We have $l^{4,5}_{I,J} = (\frac{19}{15}, \frac{3}{2})$ and hence

$$\Gamma^{4,5}_{I,J} = conv\left\{(4,4), (\tfrac{3}{2}, \tfrac{9}{2}), (\tfrac{19}{15}, \tfrac{67}{15}), (\tfrac{9}{2}, \tfrac{3}{2}), (\tfrac{71}{15}, \tfrac{23}{15}), (\tfrac{19}{15}, \tfrac{3}{2})\right\}.$$

Furthermore $l^{4,5}_{\{B\},\{R\}} = (1,1)$ and $\Gamma^{4,5}_{\{B\},\{R\}} = \{(1,1)\}$ and the set of pure-strategy equilibrium rewards is

$$\begin{aligned}\Gamma^{4,5} &= \bigcup_{I' \subset I, J' \subset J} \Gamma^{4,5}_{I',J'} \\ &= conv\left\{(4,4), (\tfrac{3}{2}, \tfrac{9}{2}), (\tfrac{19}{15}, \tfrac{67}{15}), (\tfrac{9}{2}, \tfrac{3}{2}), (\tfrac{71}{15}, \tfrac{23}{15}), (\tfrac{19}{15}, \tfrac{3}{2})\right\} \cup \{(1,1)\}.\end{aligned}$$

The shaded area in figure 5.2 is the set of rewards that can be obtained by general pure strategies, but not by jointly-convergent strategies. □

5.4 Non-pure equilibria

Section 5.3 considered pure strategies, prescribing pure actions at each stage. In this section we consider strategies that allow the players to use mixed actions at certain

stages. We will call these strategies non-pure strategies and we will show that, with the aid of non-pure strategies, the players can obtain convex combinations of pure strategy equilibrium rewards as equilibrium rewards. For that purpose we make use of a generalized version of the agreement (cf. definition 4.3.2) that can be described as follows: At stage 1 the strategy pair (π, σ) prescribes to play mixed actions. This means that a probability distribution arises over the cells that can be selected at stage 1. Now for each cell c_A that can be selected, π and σ prescribe to play, from stage 2 on, the pure strategy equilibrium (π_A, σ_A). Then, if the randomization at stage 1 is done properly, (π, σ) is an equilibrium with as reward a convex combination of the rewards of the pure strategy equilibria (π_A, σ_A). Of course the randomization is not necessarily restricted to stage 1; non-pure strategies may prescribe to randomize at a large, even infinite, number of stages.

The first result in this section, theorem 5.4.1, is a generalization of theorem 4.3.8, the main result obtained in section 4.3. After that we consider games of size $m \times n$.

Theorem 5.4.1 *In an $\left(r^1, r^2\right)$-restricted 2×2-game with $\left(r^1, r^2\right) \geq (3,3)$, every convex combination of equilibrium rewards can be obtained as an equilibrium reward.*

Proof. Identical to the proof of theorem 4.3.6. ■

Example 5.3 (continued):

The set of equilibrium rewards in the game of example 5.3 is

$$conv\left\{(4,4), (\tfrac{3}{2}, \tfrac{9}{2}), (\tfrac{19}{15}, \tfrac{67}{15}), (\tfrac{71}{15}, \tfrac{23}{15}), (1,1)\right\}$$

as depicted in figure 5.3. □

Now we define agreements for sets of actions in $m \times n$-games. Let $I' = \{i_1, i_2, \ldots, i_{q^1}\}$ with $q^1 \leq r^1 - 1$ and $J' = \left\{j_1, j_2, \ldots, j_{q^2}\right\}$ with $q^2 \leq r^2 - 1$ be subsets of $I = \{1, \ldots, m\}$ and $J = \{1, \ldots, n\}$.

Definition 5.4.2 *In an $\left(r^1, r^2\right)$-restricted 2×2-game with $\left(r^1, r^2\right) \geq (3,3)$, for (I', J') an agreement $\left(\pi_{\mathcal{A}}(I', J'), \sigma_{\mathcal{A}}(I', J')\right)$ is defined as follows:*

- *At stage 1*

$$\begin{cases} \pi_{\mathcal{A}}(I', J') \text{ prescribes to play actions } i_1 \text{ and } i_2 \text{ each with probability } \frac{1}{2} \\ \sigma_{\mathcal{A}}(I', J') \text{ prescribes to play actions } j_1 \text{ and } j_2 \text{ each with probability } \frac{1}{2} \end{cases}$$

- *At stages $t \in \left\{2, 3, \ldots, \operatorname{lcm}\left\{q^1, q^2\right\} + 1\right\}$*

$$\begin{cases} \pi_{\mathcal{A}}(I', J') \text{ prescribes to play action } i(t) \\ \sigma_{\mathcal{A}}(I', J') \text{ prescribes to play action } j(t) \end{cases}$$

  where $i(t) = i_{(t-2) \bmod q^1 + 1}$ and $j(t) = j_{(t-2) \bmod q^2 + 1}$.
  Here lcm *stands for lowest common multiple.*

- *From stage $\operatorname{lcm}\left\{q^1, q^2\right\} + 2$ on $\left(\pi_{\mathcal{A}}(I', J'), \sigma_{\mathcal{A}}(I', J')\right)$ prescribes to play according to a strategy pair (π^1, σ^1) if the first stage actions were the same, and playing according to a pair of strategies (π^2, σ^2) otherwise.*
  Here the strategy pairs (π^1, σ^1) and (π^2, σ^2) are such that the action sets that might be kept available by strategies π^1 and π^2 are subsets of I' and the action sets that might be kept available by strategies σ^1 and σ^2 are subsets of J'.

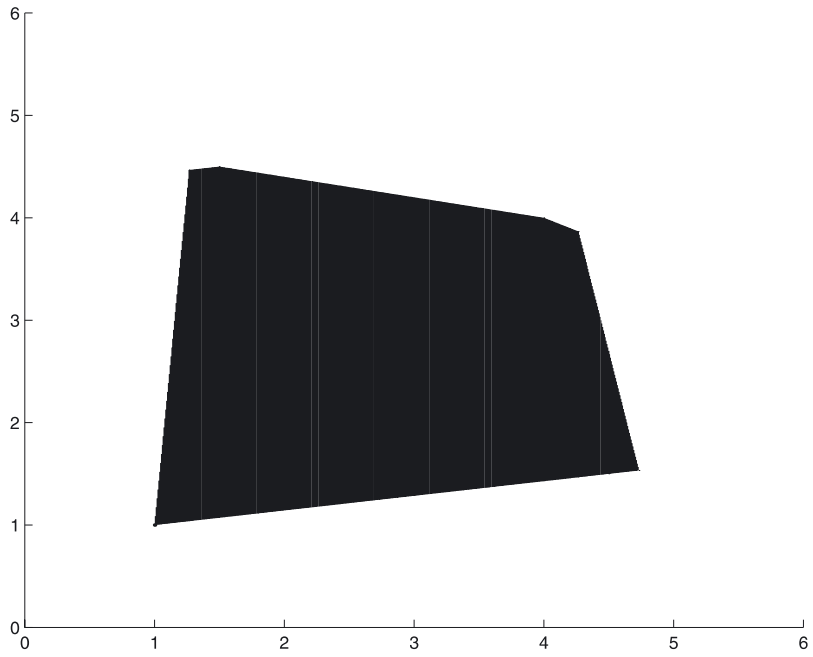


Figure 5.3: Equilibrium rewards by non-pure strategies

Remark 5.4.3 *Notice that*

$$\gamma\left(\pi_{\mathcal{A}}\left(I',J'\right),\sigma_{\mathcal{A}}\left(I',J'\right)\right)=\frac{1}{2}\gamma(\pi^1,\sigma^1)+\frac{1}{2}\gamma(\pi^2,\sigma^2)$$

and that the pairs (π^1,σ^1) *and* (π^2,σ^2) *may also be agreements (cf. remark 4.3.3).*

Notice that in an agreement for (I',J') the numbers of actions in I' and J' both are strictly less than the restrictions of the players. This has to do with the fact that randomizations at certain stages may not lead to the loss of actions. We have now made all the necessary preparations to present the generalization of theorem 4.3.6 to general-sum games of size $m\times n$ with an $\left(r^1,r^2\right)\geq(m+1,n+1)$-restriction. These games have a mild restriction in comparison with their size: both players can randomize at several stages and still keep all of their action alive. The proof is, using agreements as defined above for the complete action sets $I=\{1,2,\ldots,m\}$ and $J=\{1,2,\ldots,n\}$, identical to the proof of theorem 4.3.6 and therefore omitted.

Theorem 5.4.4 *Consider an* $m\times n$*-game with an* $\left(r^1,r^2\right)$*-restriction. If* $r^1\geq m+1$ *and* $r^2\geq n+1$*, then every convex combination of equilibrium rewards can be obtained as an equilibrium reward.*

For games, in which the number of actions is at least as large as the restriction, the analysis is much more complicated as example 5.4 shows.

Example 5.4

Consider the following $(3,3)$-restricted game:

$$\begin{array}{c} \\ T \\ C \\ B \end{array} \begin{array}{c} \begin{array}{ccc} L & M & R \end{array} \\ \left(\begin{array}{ccc} 1,2 & 3,4 & 0,0 \\ 5,3 & 2,1 & 0,0 \\ 0,0 & 0,0 & 0,0 \end{array} \right) \end{array}$$

In this game $(1,2)$ is an equilibrium reward corresponding a pair of strategies (π,σ), of which first σ prescribes to unlearn action M and then π prescribes to unlearn action C, with the obvious threats, if the other player deviates. After this π and σ prescribe to unlearn actions B and R and the reward is $(1,2)$. In a similar fashion $(2,1)$ can be supported by an equilibrium. Now consider the reward $\alpha \cdot (1,2) + (1-\alpha) \cdot (2,1)$ for some $\alpha \in (0,1)$. We will show that the agreement that yields $(1,2)$ with probability α and $(2,1)$ with probability $1-\alpha$ is not an equilibrium.

According to definition 5.4.2 we have: $\{T,C\} \subset I'$ and $\{L,M\} \subset J'$. Furthermore the sizes of the subsets I' and J' can be at most $2 = r^k - 1$ and hence $I' = \{T,C\}$ and $J' = \{L,M\}$. At the first stage $\pi_{\mathcal{A}}(I',J')$ prescribes to play actions T and C each with probability $\frac{1}{2}$ and $\sigma_{\mathcal{A}}(I',J')$ prescribes to play actions L and M each with probability $\frac{1}{2}$. At stages 2 and 3 the prescribed action pairs are TL and CM respectively. Now after stage 3 players 1 and 2 have unlearned actions B and R respectively. In the subgame that they face now $(1,2)$ and $(2,1)$ are no longer equilibrium rewards; actually (cf. theorem 3.3.2) the threat point in the subgame is $(\frac{11}{4}, \frac{5}{2})$ and the agreement that yields $(1,2)$ with probability α and $(2,1)$ with probability $1-\alpha$ is not an equilibrium.

Now consider the same game with a $(4,4)$-restriction. Then by theorem 5.4.4 every convex combination of the payoffs can be supported by an equilibrium, so in particular $\alpha \cdot (1,2) + (1-\alpha) \cdot (2,1)$ is an equilibrium reward. The agreement that yields $(1,2)$ with probability α and $(2,1)$ with probability $1-\alpha$ must keep the complete action sets I and J available though, whereas the union of the action sets that might be kept alive by π^1 and π^2 is only $\{T,C\}$ and the union of the action sets that might be kept alive by σ^1 and σ^2 is only $\{L,M\}$. □

Example 5.4 shows that a statement like theorem 5.4.4 can not be sustained. However, it is clear that, once the sizes of the action sets of the players are reduced so far that they are smaller than their restrictions, then theorem 5.4.4 does apply. Hence we can prove a weaker statement. Let

$$\mathcal{Z}(\zeta) = \{(\pi,\sigma) \mid (\pi,\sigma) \text{ is a pure equilibrium and } \gamma(\pi,\sigma) = \zeta\}.$$

In words: $\mathcal{Z}$ maps each pair of equilibrium rewards to the set of pure strategies that induce these rewards. Lemma 5.4.5 states that, if the players are not offered an opportunity to make a profitable deviation before this reduction is completed, then convex combinations of 2 equilibrium rewards making use of (subsets of) the reduced sets of actions, can also be obtained as equilibrium rewards. Theorem 5.4.6 uses lemma 5.4.5 to show that, under certain similar conditions, convex combinations of 3 pure equilibrium rewards can be supported by equilibria.

Lemma 5.4.5 *If two pairs of pure strategies* $\left(\pi^1, \sigma^1\right) \in \mathcal{Z}\left(\zeta^1\right)$ *and* $\left(\pi^2, \sigma^2\right) \in \mathcal{Z}\left(\zeta^2\right)$ *exist such that for some stage* $\tau \in \mathbb{N} : (\pi_t^1, \sigma_t^1) = (\pi_t^2, \sigma_t^2)$ *for all* $t \in \{1, 2, \ldots, \tau\}$ *and* $| I'_\tau | \leq r^1 - 1$ *and* $| J'_\tau | \leq r^2 - 1$ *for* $\left(\pi^1, \sigma^1\right)$*, then for each* $\alpha \in [0, 1]$ *there exists a pair of strategies* $(\bar{\pi}, \bar{\sigma})$ *such that* $(\bar{\pi}, \bar{\sigma})$ *is an equilibrium with* $\gamma(\bar{\pi}, \bar{\sigma}) = \alpha\zeta^1 + (1 - \alpha)\zeta^2$.

Proof. Consider the agreement $(\pi_{\mathcal{A}}(I', J'), \sigma_{\mathcal{A}}(I', J'))$ for (I'_τ, J'_τ) that plays $\left(\pi^1, \sigma^1\right)$ with probability α and $\left(\pi^2, \sigma^2\right)$ with probability $1 - \alpha$ and define $(\bar{\pi}, \bar{\sigma})$ as follows:

$$(\bar{\pi}_t, \bar{\sigma}_t) = \begin{cases} (\pi_t^1, \sigma_t^1) \text{ for } t \leq \tau \text{ followed by} \\ \text{a switching period (of length } T) \\ (\pi_{\mathcal{A}}(I', J'), \sigma_{\mathcal{A}}(I', J'))_{t-\tau-T} \text{ for } t > \tau + T, \end{cases}$$

where the switch from (π^1, σ^1) to $(\pi_{\mathcal{A}}(I', J'), \sigma_{\mathcal{A}}(I', J'))$ is done in such a way that neither player loses an action. By theorem 5.4.4 neither player can make a profitable deviation from $(\bar{\pi}, \bar{\sigma})$ at any stage beyond stage τ. Notice furthermore that if player 1 deviates at stage $t \leq \tau$, then player 2 can always reduce player 1's reward to $\min\left\{\zeta_1^1, \zeta_1^2\right\} \leq \alpha\zeta_1^1 + (1 - \alpha)\zeta_1^2$. Hence $(\bar{\pi}, \bar{\sigma})$ is an equilibrium with reward $\alpha\zeta^1 + (1 - \alpha)\zeta^2$. ∎

Theorem 5.4.6 *In an* $\left(r^1, r^2\right)$*-restricted game let* ζ^1 *as well as* ζ^2 *and* ζ^3 *be pairs of pure equilibrium rewards. If for each tuple* $(i, j) \in \{(1, 2), (1, 3), (2, 3)\}$*, there exist pairs of pure strategies* $\left(\pi^i, \sigma^i\right) \in \mathcal{Z}\left(\zeta^i\right)$ *and* $\left(\pi^j, \sigma^j\right) \in \mathcal{Z}\left(\zeta^j\right)$ *such that for some stage* $\tau \in \mathbb{N}$ *we have:*

- $\left(\pi^i, \sigma^i\right) = \left(\pi^j, \sigma^j\right)$ *for all* $t \leq \tau$ *and*
- $\left| I_\tau^l \right| \leq r^1 - 1$ *and* $\left| J_\tau^l \right| \leq r^2 - 1$, $l = i, j$

then every convex combination of ζ^1, ζ^2 *and* ζ^3 *can be obtained as an equilibrium reward.*

Proof. Consider, analogously to the proof of theorem 4.3.7, a reward $\check{\gamma} = \beta_1 \cdot \zeta^1 + \beta_2 \cdot \zeta^2 + \beta_3 \cdot \zeta^3$, where $\beta = (\beta_1, \beta_2, \beta_3) \in \Delta^3$ and assume without loss of generality that $\beta_1 \geq \beta_2 \geq \beta_3$. Then $\lambda = (\frac{3}{2}(\beta_1 + \beta_3) - \frac{1}{2}, \frac{3}{2}\beta_2, 0) \in \Delta^3$ and $\mu = (1 - 3\beta_3, 0, 3\beta_3) \in \Delta^3$. Write $\zeta = (\zeta^1, \zeta^2, \zeta^3)$. Then

$$\check{\gamma} = \left(\frac{2}{3}\lambda + \frac{1}{3}\mu\right) \cdot \zeta.$$

Since both λ and μ only put positive weight on at most two of the rewards, by lemma 5.4.5 there exist strategies $\left(\pi^\lambda, \sigma^\lambda\right)$ and (π^μ, σ^μ) to support $\lambda \cdot \zeta$ and $\mu \cdot \zeta$ as equilibrium rewards respectively. Now we define the pair of strategies $(\bar{\pi}, \bar{\sigma})$ as follows:

$$(\bar{\pi}_t, \bar{\sigma}_t) = \begin{cases} ((\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0, 0, \ldots, 0), (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0, 0, \ldots, 0)) \text{ at stage 1 followed by} \\ (\pi^\mu, \sigma^\mu) \text{ if the first stage actions were the same} \\ \left(\pi^\lambda, \sigma^\lambda\right) \text{ otherwise.} \end{cases}$$

Then from stage 2 on $\left(\pi^{\lambda}, \sigma^{\lambda}\right)$ will be played with probability $\frac{2}{3}$ and $(\pi^{\mu}, \sigma^{\mu})$ with probability $\frac{1}{3}$. Notice that at stage 1 an alternative randomization over actions 1, 2 and 3 by one of the players still leads to a probability of $\frac{1}{3}$ to select either entry $(1,1)$, $(2,2)$ or $(3,3)$. This means that a profitable deviation from $(\bar{\pi}, \bar{\sigma})$ does not exist and $(\bar{\pi}, \bar{\sigma})$ is an equilibrium with reward $\check{\gamma}$. ■

5.5 Appendix

In the appendix we prove theorem 5.2.13.

Theorem 5.5.1 *(Theorem 5.2.13).*
Consider a frequency matrix F of the form (5.11) that satisfies conditions (5.13) and (5.14). Then F is obtainable by a jointly-convergent strategy pair (π_c, σ_c) if and only if the following inequality holds:

$$\left\lceil \frac{(m'-1)\cdot \gcd\left\{r^1, r^2\right\}}{r^1} \right\rceil + \left\lceil \frac{(n'-1)\cdot \gcd\left\{r^1, r^2\right\}}{r^2} \right\rceil \leq \gcd\left\{r^1, r^2\right\}. \qquad (5.28)$$

Proof. We will first prove the "⇐"-part of the statement in the theorem. For that purpose suppose that inequality (5.28) holds. We will show that in that case player 1 can play his low-frequency actions (all actions except action 1) at such stages that for player 2 it is possible to play his low-frequency actions only at stages at which player 1 does not play a low-frequency action. Notice first that, in order to obtain F as a frequency matrix, for each stage $t \in \left\{1, 2, \ldots, r^2\right\}$ we have that at all stages in $\left\{t + \kappa \cdot r^2 \,|\, \kappa \in \{0, 1, 2, \ldots\}\right\}$ player 2 has to play the same action. Consequently player 2 has to repeat playing the action sequence he plays at stages $1, 2, \ldots, r^2$. During stages $t \in \{1, 2, \ldots, r^2\}$ player 2 has to play a low-frequency action exactly $n'-1$ times. Now we divide the set $\left\{1, 2, \ldots, r^2\right\}$ into $\frac{r^2}{\gcd\{r^1, r^2\}}$ parts: $\left\{1, 2, \ldots, \gcd\left\{r^1, r^2\right\}\right\}$, $\left\{\gcd\left\{r^1, r^2\right\}+1, \gcd\left\{r^1, r^2\right\}+2, \ldots, 2\cdot\gcd\left\{r^1, r^2\right\}\right\}$, $\ldots$, $\left\{r^2 - \gcd\left\{r^1, r^2\right\}+1, r^2 - \gcd\left\{r^1, r^2\right\}+2, \ldots, r^2\right\}$. Since during stages $1, 2, \ldots, r^1$ and also during each subsequent set of r^1 consecutive stages, player 1 plays a low-frequency action exactly $m'-1$ times, on average he will play a low-frequency action $\frac{m'-1}{r^1} \cdot \gcd\left\{r^1, r^2\right\}$ times during each of the parts

$$\mathfrak{T}_{l+1} := \left\{l\cdot\gcd\left\{r^1, r^2\right\}+1, l\cdot\gcd\left\{r^1, r^2\right\}+2, \ldots, (l+1)\cdot\gcd\left\{r^1, r^2\right\}\right\}. \qquad (5.29)$$

Let $\mathfrak{r}^1 = \frac{r^1}{\gcd\{r^1, r^2\}}$ and $\mathfrak{r}^2 = \frac{r^2}{\gcd\{r^1, r^2\}}$ and consider the following set of $m'-1$ integers in $\left\{1, 2, \ldots, r^1\right\}$ at which player 1 plays his low-frequency actions: Take $\hat{t}_1 \in \mathfrak{T}_1$ and then take $\hat{t}_2 = \hat{t}_1 + \gcd\left\{r^1, r^2\right\} \in \mathfrak{T}_2$, $\hat{t}_3 = \hat{t}_1 + 2\cdot\gcd\left\{r^1, r^2\right\} \in \mathfrak{T}_3$ etc. up to $\hat{t}_{\mathfrak{r}^1} = \hat{t}_1 + \left(\mathfrak{r}^1 - 1\right)\cdot\gcd\left\{r^1, r^2\right\} \in \mathfrak{T}_{\mathfrak{r}^1}$ or, if $m'-1 \leq \mathfrak{r}^1 - 1$, up to $\hat{t}_{m'-1} = \hat{t}_1 + (m'-2)\cdot\gcd\left\{r^1, r^2\right\} \in \mathfrak{T}_{m'-1}$. Now if $m'-1 \geq \mathfrak{r}^1+1$, then take $\hat{t}_{\mathfrak{r}^1+1} \in \mathfrak{T}_1 \setminus \left\{\hat{t}_1\right\}$ and then $\hat{t}_{\mathfrak{r}^1+2} = \hat{t}_{\mathfrak{r}^1+1} + \gcd\left\{r^1, r^2\right\} \in \mathfrak{T}_2$, $\hat{t}_{\mathfrak{r}^1+3} = \hat{t}_{\mathfrak{r}^1+1} + 2\cdot\gcd\left\{r^1, r^2\right\} \in \mathfrak{T}_3$ etc. up to $\hat{t}_{2\mathfrak{r}^1} = \hat{t}_{\mathfrak{r}^1+1} + \left(\mathfrak{r}^1 - 1\right)\cdot\gcd\left\{r^1, r^2\right\} \in \mathfrak{T}_{\mathfrak{r}^1}$ or, if $m'-1 \leq 2\mathfrak{r}^1 - 1$, up to $\hat{t}_{m'-1} = \hat{t}_{\mathfrak{r}^1+1} + \left(m' - \mathfrak{r}^1 - 2\right)\cdot\gcd\left\{r^1, r^2\right\} \in \mathfrak{T}_{m'-1-\mathfrak{r}^1}$. If $m'-1 \geq 2\mathfrak{r}^1+1$, then we continue this procedure by taking $\hat{t}_{2\mathfrak{r}^1+1} \in \mathfrak{T}_1 \setminus \left\{\hat{t}_1, \hat{t}_{\mathfrak{r}^1+1}\right\}$ etcetera until

we have a set of $m'-1$ integers. Now consider the following set of $n'-1$ integers in $\{1,2,\ldots,r^2\}$ at which player 2 plays his low-frequency actions: Take $\hat{\tau}_1 \in \mathfrak{T}_1$ and then take $\hat{\tau}_2 = \hat{\tau}_1 + \gcd\{r^1,r^2\} \in \mathfrak{T}_2$, $\hat{\tau}_3 = \hat{\tau}_1 + 2\cdot\gcd\{r^1,r^2\} \in \mathfrak{T}_3$ etc. up to $\hat{\tau}_{\mathfrak{k}^2} = \hat{\tau}_1 + (\mathfrak{k}^2-1)\cdot\gcd\{r^1,r^2\} \in \mathfrak{T}_{\mathfrak{k}^2}$ or, if $n'-1 \leq \mathfrak{k}^2-1$, up to $\hat{\tau}_{n'-1} = \hat{\tau}_1 + (n'-2)\cdot\gcd\{r^1,r^2\} \in \mathfrak{T}_{n'-1}$. Now if $n'-1 \geq \mathfrak{k}^2+1$, then take $\hat{\tau}_{\mathfrak{k}^2+1} \in \mathfrak{T}_1 \setminus \{\hat{\tau}_1\}$ and then $\hat{\tau}_{\mathfrak{k}^2+2} = \hat{\tau}_{\mathfrak{k}^2+1} + \gcd\{r^1,r^2\} \in \mathfrak{T}_2$, $\hat{\tau}_{\mathfrak{k}^2+3} = \hat{\tau}_{\mathfrak{k}^2+1} + 2\cdot\gcd\{r^1,r^2\} \in \mathfrak{T}_3$ etc. up to $\hat{\tau}_{2\mathfrak{k}^2} = \hat{\tau}_{\mathfrak{k}^2+1} + (\mathfrak{k}^2-1)\cdot\gcd\{r^1,r^2\} \in \mathfrak{T}_{\mathfrak{k}^2}$ or, if $n'-1 \leq 2\mathfrak{k}^2-1$, up to $\hat{\tau}_{n'-1} = \hat{\tau}_{\mathfrak{k}^2+1} + (n'-\mathfrak{k}^2-2)\cdot\gcd\{r^1,r^2\} \in \mathfrak{T}_{n'-1-\mathfrak{k}^2}$. If $n'-1 \geq 2\mathfrak{k}^2+1$, then we continue this procedure by taking $\hat{\tau}_{2\mathfrak{k}^2+1} \in \mathfrak{T}_1 \setminus \{\hat{\tau}_1, \hat{\tau}_{\mathfrak{k}^2+1}\}$ etcetera until we have a set of $n'-1$ integers.

Notice that for each part $\mathfrak{T}_l$ in (5.29) the subset of stages at which player 1 plays a low-frequency action, is also a subset of

$$\left\{\hat{t}_l, \hat{t}_{\mathfrak{k}^1+l}, \hat{t}_{2\mathfrak{k}^1+l}, \ldots, \hat{t}_{\left(\left\lceil\frac{(m'-1)}{\mathfrak{k}^1}\right\rceil-1\right)\cdot\mathfrak{k}^1+l}\right\} \subset \mathfrak{T}_l. \tag{5.30}$$

and hence that for each part $\mathfrak{T}_l$ the number of stages at which player 1 plays a low-frequency action, does not exceed $\left\lceil\frac{(m'-1)}{\mathfrak{k}^1}\right\rceil$. Furthermore, if player 1 were to repeat playing the action sequence he plays at stages $\{1,2,\ldots,r^1\}$, then (5.30) also holds for $l > \mathfrak{k}^1$. Similarly for player 2 for each $l \in \{1,2,\ldots\}$ we have:

$$\left\{\hat{\tau}_l, \hat{\tau}_{\mathfrak{k}^1+l}, \hat{\tau}_{2\mathfrak{k}^1+l}, \ldots, \hat{\tau}_{\left(\left\lceil\frac{(n'-1)}{\mathfrak{k}^2}\right\rceil-1\right)\cdot\mathfrak{k}^2+l}\right\} \subset \mathfrak{T}_l, \tag{5.31}$$

whose size is bounded above by $\left\lceil\frac{(n'-1)}{\mathfrak{k}^2}\right\rceil$.

Now, given that inequality (5.28) holds, let the sets

$$\mathcal{S}_1 = \left\{\hat{t}_1, \hat{t}_{\mathfrak{k}^1+1}, \hat{t}_{2\mathfrak{k}^1+1}, \ldots, \hat{t}_{\left(\left\lceil\frac{(m'-1)}{\mathfrak{k}^1}\right\rceil-1\right)\cdot\mathfrak{k}^1+1}\right\} \subset \mathfrak{T}_1$$

and

$$\mathcal{S}_2 = \left\{\hat{\tau}_1, \hat{\tau}_{\mathfrak{k}^1+1}, \hat{\tau}_{2\mathfrak{k}^1+1}, \ldots, \hat{\tau}_{\left(\left\lceil\frac{(n'-1)}{\mathfrak{k}^2}\right\rceil-1\right)\cdot\mathfrak{k}^2+1}\right\} \subset \mathfrak{T}_1$$

be constructed in such a way that

$$\mathcal{S}_1 \cap \mathcal{S}_2 = \emptyset$$

and consider the jointly-convergent strategy pair (π_c, σ_c), where π_c prescribes to play the low-frequency actions (in a specific order) at stages $\hat{t}_\kappa, \kappa \in \{1,2,\ldots\}$ and σ_c prescribes to play the low-frequency actions at stages $\hat{\tau}_\kappa, \kappa \in \{1,2,\ldots\}$. Then $F(\pi_c,\sigma_c) = F$, which completes the proof of the "$\Leftarrow$"-statement in the theorem.

We continue the analysis by proving the "$\Rightarrow$"-part of the statement. We will do so by means of contradiction. Therefore we assume that inequality (5.28) does not hold, or:

$$\left\lceil\frac{(m'-1)\cdot\gcd\{r^1,r^2\}}{r^1}\right\rceil + \left\lceil\frac{(n'-1)\cdot\gcd\{r^1,r^2\}}{r^2}\right\rceil > \gcd\{r^1,r^2\}. \tag{5.32}$$

Now let (π_c, σ_c) be a pair of jointly-convergent strategies, where π_c prescribes to play a low-frequency action at $m'-1$ out of every r^1 consecutive stages and σ_c prescribes to play a low-frequency action at $n'-1$ out of every r^2 consecutive stages. Notice that there is an $\tilde{l} \in \{1, 2, \ldots, \mathfrak{r}^1\}$, such that the number of stages in $\mathfrak{T}_{\tilde{l}}$, at which π_c prescribes to play a low-frequency action, is at least equal to $\left\lceil \frac{(m'-1)\cdot \gcd\{r^1, r^2\}}{r^1} \right\rceil$. Similarly there is an $\hat{l} \in \{1, 2, \ldots, \mathfrak{r}^2\}$, such that the number of stages in $\mathfrak{T}_{\hat{l}}$, at which σ_c prescribes to play a low-frequency action, is at least equal to $\left\lceil \frac{(n'-1)\cdot \gcd\{r^1, r^2\}}{r^2} \right\rceil$. But then, by (5.32) and taking into consideration that

$$\left| \mathfrak{T}_{\tilde{l}} \right| = \left| \mathfrak{T}_{\hat{l}} \right| = \gcd\{r^1, r^2\},$$

there are 2 stages $\tilde{t} \in \mathfrak{T}_{\tilde{l}}$ and $\hat{t} \in \mathfrak{T}_{\hat{l}}$ with

$$\tilde{t} = \hat{t} + \kappa \cdot \gcd\{r^1, r^2\}$$

for some integer κ, such that π_c prescribes to play a low-frequency action at stage $\tilde{t}$ and σ_c prescribes to play a low-frequency action at stage $\hat{t}$. But then π_c prescribes to play a low-frequency action at each stage in

$$\{\tilde{t} + \lambda \cdot r^1 \mid \lambda \in \{0, 1, 2, \ldots\}\}$$

and σ_c prescribes to play a low-frequency action at each stage in

$$\{\hat{t} + \mu \cdot r^2 \mid \mu \in \{0, 1, 2, \ldots\}\}.$$

Applying Euclid's theorem (cf. theorem 5.2.5 and corollary 5.2.6) yields that there is a stage $t \in \{1, 2, \ldots, \operatorname{lcm}\{r^1, r^2\}\}$ such that at stage t the strategies π_c and σ_c each prescribe to play a low-frequency action at stage t and at each stage in

$$\{t + \nu \cdot \operatorname{lcm}\{r^1, r^2\} \mid \nu \in \{0, 1, 2, \ldots\}\}$$

strategies π_c and σ_c each prescribe to play a low-frequency action. This means that the frequency matrix corresponding to (π_c, σ_c) can not be F. ■

Chapter 6

Fictitious Play in Stochastic Games

6.1 Introduction

In game theory many tools have been designed in order to construct optimal strategies in zero-sum games or equilibria in general-sum games. One of these tools is fictitious play. The (discrete) fictitious play process can roughly be described as follows: The game is played repeatedly and at each stage each player selects a (pure) action that is a best reply against the "average" action of the other player. Here at stage t the average action of a player is the probability vector consisting of the frequencies by which he played his actions at stages $1, 2, \ldots, t-1$. A game is said to have the fictitious play property if every fictitious play process converges to an equilibrium.

Fictitious play processes were introduced by Brown (1951) and Robinson (1951), who proved the fictitious play property for two-player zero-sum games. Miyasawa (1961) proved the fictitious play property for generic 2×2-games. A geometric proof for this class of games is provided by Metrick and Polak (1994). Convergence was also shown by Monderer and Shapley (1996) for N-player games in which all players have the same number of actions and identical payoff functions. Shapley (1964), however, provided an example of a 2-player repeated game where each player has 3 actions and where the fictitious play process does not converge.

There also exists a different type of fictitious play, the so-called continuous fictitious play. We will not discuss this type of fictitious play here. For recent results on continuous fictitious play processes we refer to Krishna and Sjöström (1998) and Sela (2000).

Notice that during the fictitious play process the (bi)matrix game is played an infinite number of times. Therefore, regarding the fictitious play process as a way to play the game, the players actually play a repeated game instead of a one-shot game and fictitious play can be considered a strategy pair in this repeated game. Furthermore, if the frequencies of the fictitious play process converge, then the limit distribution of these frequencies can be considered a stationary strategy in the repeated game as well as a mixed action in the (bi)matrix game (notice that in repeated games repeatedly playing the same pair of mixed actions forming an equilibrium in the corresponding

one-shot game, always is an equilibrium in stationary strategies). Consequently the fictitious play process as described above can be regarded as a fictitious play process for repeated games as well as for one-shot games. In section 6.2 we will formalize this idea and present a generalization of the fictitious play process for repeated games to a fictitious play process for stochastic games. In section 6.3 we present a specific 2-player 2-state stochastic game in which each player has 2 actions in each state and in section 6.4 we prove that this game does not have the fictitious play property. Sections 6.2-6.4 are based on Schoenmakers, Flesch & Thuijsman (2001). In section 6.5 we present some other fictitious play processes for stochastic games, one of which is a model introduced by Vrieze and Tijs (1982) for β-discounted games.

6.2 Fictitious play in repeated and stochastic games

In order to obtain a clear description of the fictitious play process for $m \times n$ - repeated games, we introduce some notations first. Let $i_\tau^* \in I$ denote the pure action that is selected by player 1 at stage τ of the fictitious play process. Furthermore let f_t denote the action frequencies of the pure actions of player 1 up to and including stage t of the fictitious play process, i.e.

$$f_t = \frac{1}{t} \sum_{\tau=1}^{t} i_\tau^* \in \Delta^m.$$

For player 2 $j_t^* \in J$ and $g_t \in \Delta^n$ are defined analogously. Notice that f_t and g_t can be seen as mixed actions for players 1 and 2 respectively. Now let $x(f_t)$ be the stationary strategy for player 1 that prescribes to play the mixed action f_t at each stage and let $y(g_t)$ be the stationary strategy for player 2 that prescribes to play the mixed action g_t at each stage. Then the fictitious play process for 2-player repeated games can be defined as follows:

Definition 6.2.1 *The fictitious play process for 2-player repeated games is a sequence $((i_t^*, j_t^*))_{t=1}^{\infty}$, with $(i_t^*, j_t^*) \in I \times J$ for all t, recursively defined as follows: $i_1^* = j_1^* = 1$ and hence $f_1 = i_1^*$ and $g_1 = j_1^*$. Furthermore at stage $t \geq 2$ we consider the stationary strategies $x(f_{t-1})$ and $y(g_{t-1})$ and we take $i_t^* \in I$ and $j_t^* \in J$ such that $x(i_t^*)$ and $y(j_t^*)$ are pure stationary best replies against $y(g_{t-1})$ and $x(f_{t-1})$ respectively. Thereafter we update f and g as follows:*

$$f_t = \frac{t-1}{t} \cdot f_{t-1} + \frac{1}{t} \cdot i_t^* \in \Delta^m$$

and

$$g_t = \frac{t-1}{t} \cdot g_{t-1} + \frac{1}{t} \cdot j_t^* \in \Delta^n.$$

Obviously these formulas are equivalent with

$$f_t = \frac{1}{t} \sum_{\tau=1}^{t} i_\tau^* \in \Delta^m$$

and

$$g_t = \frac{1}{t}\sum_{\tau=1}^{t} j_\tau^* \in \Delta^n.$$

The fictitious play process is said to converge if $(f_t, g_t)_{t=1}^{\infty}$ converges. A game has the fictitious play property if every fictitious play process converges to the set of stationary equilibrium strategies.

We now define a generalization of the above fictitious play process for repeated games to a fictitious play process for stochastic games. Suppose that the stochastic game has state space $S = \{1, 2, \ldots, z\}$ and that players 1 and 2 have m^s respectively n^s actions in state $s \in S$. We suppose that during the fictitious play process the players select joint pure actions at each stage. Let

$$i_\tau^* = \left(i_\tau^1, i_\tau^2, \ldots, i_\tau^z\right) \in \Delta^{m^1} \times \Delta^{m^2} \times \ldots \times \Delta^{m^z}$$

denote the joint pure action that is selected by player 1 at stage τ of the fictitious play process. Furthermore let f_t^s denote the action frequencies of the pure actions of player 1 in state s up to and including stage t of the fictitious play process, i.e.

$$f_t^s = \frac{1}{t}\sum_{\tau=1}^{t} i_\tau^s \in \Delta^{m^s}$$

and let

$$f_t = (f_t^1, f_t^2, \ldots, f_t^z).$$

For player 2 the (joint) actions $j_\tau \in \Delta^{n^1} \times \Delta^{n^2} \times \ldots \times \Delta^{n^z}$, $g_t^s \in \Delta^{n^s}$ and $g_t \in \Delta^{n^1} \times \Delta^{n^2} \times \ldots \times \Delta^{n^z}$ are defined analogously. Then f_t and g_t can be seen as joint mixed actions for players 1 and 2 respectively. Now let $x(f_t)$ be the stationary strategy for player 1 that prescribes to play the mixed action f_t^s each time that state s is visited, and let $y(g_t)$ be the stationary strategy for player 2 that prescribes to play the mixed action g_t^s at each visit to state s. Then the fictitious play process for 2-player stochastic games can, analogously to definition (6.2.1), be defined as follows:

Definition 6.2.2 *The fictitious play process for 2-player stochastic games is a sequence $((i_t^*, j_t^*))_{t=1}^{\infty}$, with i_t^* and j_t^* joint pure actions for players 1 and 2 respectively for all t, recursively defined as follows: i_1^* and j_1^* play action 1 in each state and $f_1 = i_1^*$ and $g_1 = j_1^*$. Furthermore at stage $t \geq 2$ we consider the stationary strategies $x(f_{t-1})$ and $y(g_{t-1})$ and we take i_t^* and j_t^* such that $x(i_t^*)$ and $y(j_t^*)$ are pure stationary best replies against $y(g_{t-1})$ and $x(f_{t-1})$ respectively. Thereafter we update f and g as follows:*

$$f_t = \frac{t-1}{t} \cdot f_{t-1} + \frac{1}{t} \cdot i_t^* \in \Delta^{m^1} \times \Delta^{m^2} \times \ldots \times \Delta^{m^z} \tag{6.1}$$

and

$$g_t = \frac{t-1}{t} \cdot g_{t-1} + \frac{1}{t} \cdot j_t^* \in g_t \in \Delta^{n^1} \times \Delta^{n^2} \times \ldots \times \Delta^{n^z}. \tag{6.2}$$

Observe the equivalence between formulas (6.1) and $f_t = \frac{1}{t}\sum_{\tau=1}^{t} i_\tau^* \in \Delta^{m^1} \times \Delta^{m^2} \times \ldots \times \Delta^{m^z}$ and between (6.2) and $g_t = \frac{1}{t}\sum_{\tau=1}^{t} j_\tau^* \in \Delta^n$.

We would like to emphasize that this definition of a fictitious play process for stochastic games does not correspond to any play of the game itself. Nevertheless, for one-state stochastic games the extensions coincide with the original fictitious play process for repeated games.

The fictitious play process is said to converge if $(f_t, g_t)_{t=1}^{\infty}$ converges. A game has the fictitious play property if every fictitious play process converges to the set of stationary equilibrium strategies. Notice that stationary equilibria do not necessarily exist in stochastic games (cf. e.g. the Big Match by Gilette (1957)). In the next sections we examine a $2 \times 2 \times 2$ stochastic game. Here $2 \times 2 \times 2$ stands for 2 players, 2 states and 2 actions for each player in each state. Moreover the stochastic game in the example is an irreducible single-controller stochastic game with state independent transitions (cf. definitions 1.2.4, 1.2.2 and 1.2.3). It is well known for irreducible stochastic games and for single controller stochastic games that stationary equilibria always exist (cf. Rogers (1969), Sobel (1971), Filar (1981) and theorem 1.2.5). We show however, that the fictitious play process in this example does not converge. This means that

Theorem 6.2.3 *The fictitious play process for* $2 \times 2 \times 2$ *games does not necessarily converge.*

6.3 The example

Consider the following $2 \times 2 \times 2$ stochastic game:

$2,1$ $(\frac{9}{10},\frac{1}{10})$	$4,0$ $(\frac{9}{10},\frac{1}{10})$	$0,1$ $(\frac{9}{10},\frac{1}{10})$	$2,0$ $(\frac{9}{10},\frac{1}{10})$
$0,0$ $(\frac{1}{10},\frac{9}{10})$	$7,1$ $(\frac{1}{10},\frac{9}{10})$	$2,0$ $(\frac{1}{10},\frac{9}{10})$	$4,1$ $(\frac{1}{10},\frac{9}{10})$
state 1		*state* 2	

Notice that the transition probabilities in this game depend only on the action of player 1 and they are independent of the state. Furthermore the game is irreducible, which means that irrespective of the players' strategies both states will be visited infinitely often with probability 1 and the limiting average rewards of the game do not depend on the starting state.

We now show that this game has a unique statioanry equilibrium $(x(f^*), y(g^*))$ where $f^* = ((\frac{1}{2},\frac{1}{2}),(\frac{1}{2},\frac{1}{2}))$ and $g^* = ((\frac{1}{5},\frac{4}{5}),(\frac{9}{20},\frac{11}{20}))$. Suppose that player 1 plays the stationary strategy

$$x = ((a_1, 1-a_1),(a_2, 1-a_2))$$

and that player 2 plays

$$y = ((b_1, 1-b_1), (b_2, 1-b_2)).$$

Given these strategies the invariant distribution over the states, i.e. the proportions of time that the states are being visited, is given by

$$\left(\frac{\frac{1}{10} + \frac{4}{5}a_2}{1 - \frac{4}{5}a_1 + \frac{4}{5}a_2}, \frac{\frac{9}{10} - \frac{4}{5}a_1}{1 - \frac{4}{5}a_1 + \frac{4}{5}a_2} \right)$$

and, using the expected payoffs in each of these states, it follows that

$$\gamma_1(x,y) = \frac{\frac{1}{10}+\frac{4}{5}a_2}{1-\frac{4}{5}a_1+\frac{4}{5}a_2} \cdot (5a_1b_1 - 3a_1 - 7b_1 + 7) + \frac{\frac{9}{10}-\frac{4}{5}a_1}{1-\frac{4}{5}a_1+\frac{4}{5}a_2} \cdot (4 - 2a_2 - 2b_2)$$
$$\gamma_2(x,y) = \frac{\frac{1}{10}+\frac{4}{5}a_2}{1-\frac{4}{5}a_1+\frac{4}{5}a_2} \cdot (1 - a_1 - b_1 + 2a_1b_1) + \frac{\frac{9}{10}-\frac{4}{5}a_1}{1-\frac{4}{5}a_1+\frac{4}{5}a_2} \cdot (1 - a_2 - b_2 + 2a_2b_2)$$

It is straightforward to verify that there are no equilibria in which at least one player uses a pure stationary strategy. To see that $(x(f^*), y(g^*))$ is an equilibrium observe that, if $h_1 = ((1,0),(1,0))$, $h_2 = ((1,0),(0,1))$, $h_3 = ((0,1),(1,0))$ and $h_4 = ((0,1),(0,1))$, then

$$\begin{aligned}
\gamma_1(x(h_1), y(g^*)) &= \gamma_1(x(h_2), y(g^*)) = \gamma_1(x(h_3), y(g^*)) \\
&= \gamma_1(x(h_4), y(g^*)) = \gamma_1(x(f^*), y(g^*)) = 3.35 \\
\gamma_2(x(f^*), y(h_1)) &= \gamma_2(x(f^*), y(h_2)) = \gamma_2(x(f^*), y(h_3)) \\
&= \gamma_2(x(f^*), y(h_4)) = \gamma_2(x(f^*), y(g^*)) = 0.5
\end{aligned}$$

and, therefore, by Hordijk, Vrieze and Wanrooij (1983), $(x(f^*), y(g^*))$ is an equilibrium. Uniqueness of this equilibrium follows straightforwardly from the best reply structure, which is examined in more detail in the next section.

Theorem 6.3.1 *The fictitious play process in the* $2 \times 2 \times 2$ *stochastic game above does not converge.*

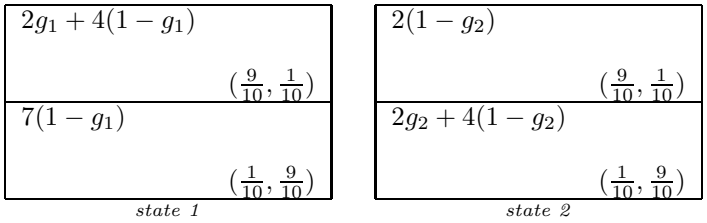
The proof of this theorem, which is based on an analysis of the best reply structure in the stationary strategy spaces, is given in the next section. The key of the proof is the observation of a cyclic pattern in the fictitious play process for the example presented.

6.4 The proof

We examine the best reply structure for stationary strategies in the example. We start with player 1. Take a fixed stationary strategy $y(g)$ with $g = ((g_1, 1-g_1), (g_2, 1-g_2))$ of player 2.

Notation 6.4.1 *During section 6.4 instead of* $g = ((g_1, 1-g_1), (g_2, 1-g_2))$ *we write* $g = (g_1, g_2)$ *with* $g_1, g_2 \in [0,1]$ *where* g_s *is the probability or the frequency of action 1 in state* s*. For joint pure actions we will also use* (j_1, j_2) *instead of* (g_1, g_2)*. Similarly for player 1 we have* $f = (f_1, f_2)$ *with* $f_1, f_2 \in [0,1]$ *and* (i_1, i_2) *for joint pure actions.*

Then player 1 faces the following Markov Decision Problem (MDP):



Let $v_{(i_1,i_2)}$ denote player 1's limiting average reward in the above MDP, when he plays the pure stationary strategy $x(i_1, i_2)$. Notice that $i_1, i_2 \in \{0, 1\}$ and that $x(i_1, i_2)$ is a best reply against $y(g)$ if and only if $v_{(i_1,i_2)}$ is maximal.

We can calculate $v_{(1,1)}$ as follows. Suppose player 1 plays $x(1, 1)$, then state 1 will, in expectation, be visited 9 stages out of 10 and

$$\begin{aligned} v_{(1,1)} &= 0.9 \cdot (2g_1 + 4(1-g_1)) + 0.1 \cdot 2(1-g_2) \\ &= 3.8 - 1.8g_1 - 0.2g_2. \end{aligned}$$

The other values are:

$$\begin{aligned} v_{(1,0)} &= 4 - g_1 - g_2 \\ v_{(0,1)} &= 4.5 - 3.5g_1 - g_2 \\ v_{(0,0)} &= 4.3 - 0.7g_1 - 1.8g_2. \end{aligned}$$

Now we calculate the values of g_1 and g_2 for which player 1 is indifferent between some of his pure stationary strategies:

$$v_{(1,1)} = v_{(1,0)} \iff 3.8 - 1.8g_1 - 0.2g_2 = 4 - g_1 - g_2,$$

hence

$$v_{(1,1)} = v_{(1,0)} \iff g_2 = g_1 + \frac{1}{4}.$$

Analogously

$$\begin{aligned} v_{(1,1)} = v_{(0,1)} &\iff g_2 = -\tfrac{17}{8}g_1 + \tfrac{7}{8} \\ v_{(1,0)} = v_{(0,0)} &\iff g_2 = \tfrac{3}{8}g_1 + \tfrac{3}{8} \\ v_{(0,1)} = v_{(0,0)} &\iff g_2 = \tfrac{7}{2}g_1 - \tfrac{1}{4}. \end{aligned}$$

From these equations we deduce the left part of figure 6.1, showing the best replies of player 1 against $y(g)$. The lines in this figure correspond with the equations above. The lines divide the square into four regions. If (g_1, g_2) is in one of the regions, then the pure stationary strategy mentioned in the region is the pure best reply for player

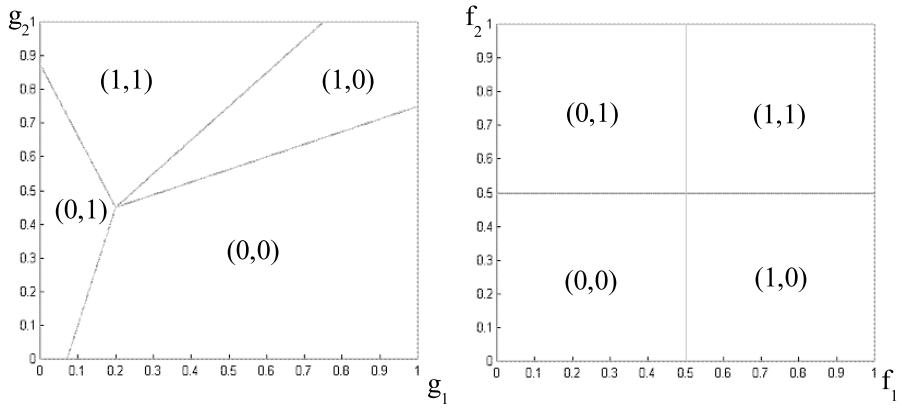


Figure 6.1: Best reply structure for stationary strategies

1 against $y(g)$. The common point of these regions corresponds to the equilibrium strategy $y(g^*)$.

Since player 2 can only maximize his one-shot payoff we can easily deduce the right part of figure 6.1 showing the best replies of player 2 against an arbitrary stationary strategy $x(f)$ of player 1. The two relevant indifference lines are $f_1 = \frac{1}{2}$ and $f_2 = \frac{1}{2}$.

Notice that figure 6.1 also indicates that $x(1,1)$ and $x(0,0)$ as well as $y(1,1)$ and $y(0,0)$ can only be best replies simultaneously at the equilibrium point. The same holds for $x(1,0)$ and $x(0,1)$ and for $y(1,0)$ and $y(0,1)$. From figure 6.1 it is clear that for each player there is a unique stationary strategy against which all pure strategies of the opponent are best replies. This implies the uniqueness of the stationary equilibrium $(x(f^*), y(g^*))$.

In general the best reply structure in stochastic games is non-linear. The single-controller condition in our example guarantees the linearity. We will now derive some properties on how the fictitious play process evolves. This will be done in terms of so-called runs:

Definition 6.4.2 *A run* $[(i_1, i_2), (j_1, j_2)]$ *is a part* $((i^*_{t_1}, j^*_{t_1}), (i^*_{t_1+1}, j^*_{t_1+1}), \ldots, (i^*_{t_2}, j^*_{t_2}))$ *of the sequence* $(i^*_t, j^*_t)_{t=1}^{\infty}$ *such that* $(i^*_\tau, j^*_\tau) = ((i_1, i_2), (j_1, j_2))$ *for all* $\tau \in \{t_1, \ldots, t_2\}$, *whereas equality fails for* $\tau = t_1 - 1$ *and for* $\tau = t_2 + 1$.

The next lemma shows how the different runs follow each other.

Lemma 6.4.3 *The following runs will succeed each other cyclically: First* $[(1,1), (1,1)]$, *then* $[(1,0), (1,1)]$, *then* $[(1,0), (1,0)]$, *then* $[(0,0), (1,0)]$, *then* $[(0,0), (0,0)]$, *then* $[(0,1), (0,0)]$, *then* $[(0,1), (0,1)]$, *then* $[(1,1), (0,1)]$ *and then we return to* $[(1,1), (1,1)]$ *and start a new cycle.*

Proof. The proof is based on the fact that if we are in run $[(i_1, i_2), (j_1, j_2)]$ at stage t, then the action frequencies will change in the following way:

$$f(t) = \frac{t-1}{t} \cdot f(t-1) + \frac{1}{t} \cdot (i_1, i_2)$$
$$g(t) = \frac{t-1}{t} \cdot g(t-1) + \frac{1}{t} \cdot (j_1, j_2).$$

So, as t increases, $f(t)$ and $g(t)$ move along a straight line in the direction of the corner points (i_1, i_2) respectively. (j_1, j_2).

Recall that the fictitious play process starts with a $[(1,1),(1,1)]$-run, hence both $f(1)$ and $g(1)$ are $(1,1)$. So at stage 2 a $[(1,0),(1,1)]$-run is started, hence f moves in the direction of $(1,0)$ and g stays at $(1,1)$. At a certain stage in the right part of figure 6.1 the line $f_2 = \frac{1}{2}$ will be crossed and g starts moving towards $(1,0)$, causing a $[(1,0),(1,0)]$-run to start. During this run both f and g move towards $(1,0)$. But then at a certain stage in the left part of figure 1 the line between the $(1,0)$-part and the $(0,0)$-part will be crossed and the $(0,0)$-part will be entered, which causes a $[(0,0),(1,0)]$-run to start. Analogous reasonings can be held to prove the occurrence of the other switches of run types. ■

We will prove the nonconvergence of the fictitious play process by defining other processes on the left part of figure 6.1, called trajectories. We will show that these trajectories do not converge to the equilibrium point and that the fictitious play process follows lines that run even further away from the equilibrium point than the trajectories do.

Definition 6.4.4 *Consider figure 6.2. A trajectory τ is a set of four connected line segments in $[0,1]^2$ that satisfies the following conditions:*

1. *A trajectory starts and ends at line segment a, which corresponds to the equation $g_2 = \frac{3}{8}g_1 + \frac{3}{8}$, where $g_1 \in [\frac{1}{5}, 1]$. The starting point of a trajectory τ is called $s(\tau)$ and the end point $e(\tau)$.*

2. *In areas A, B, C and D the trajectory moves in the direction of the respective corner points $(0,0)$, $(0,1)$, $(1,1)$ and $(1,0)$.*

A trajectory τ is called an orbit if $s(\tau) = e(\tau)$. An orbit $\overline{\tau}$, with $s(\overline{\tau}) = e(\overline{\tau}) = \psi$ is stable if for some small $\delta > 0$ the following contraction property holds: for all trajectories $\tau \neq \overline{\tau}$, if $||s(\tau) - \psi|| < \delta$, then $||e(\tau) - \psi|| < ||s(\tau) - \psi||$.

Lemma 6.4.5 *There are precisely 2 orbits, a stable one with starting point $(\frac{15}{19}, \frac{51}{76})$ and a non-stable one being the equilibrium point $(\frac{1}{5}, \frac{9}{20})$.*

Proof. Finding orbits boils down to finding fixed points of a function h that assigns the finishing value $e(\tau)$ to the starting value $s(\tau)$ for each trajectory t.

For an arbitrary trajectory we have $s(\tau) = (\frac{1}{5} + \varepsilon, \frac{9}{20} + \frac{3}{8}\varepsilon)$ with $\varepsilon \in [0, \frac{4}{5}]$, which is at line segment a in figure 6.2, corresponding to the equation $g_2 = \frac{3}{8}g_1 + \frac{3}{8}$. The trajectory enters area A and moves in the direction of $(0,0)$. As long as the trajectory

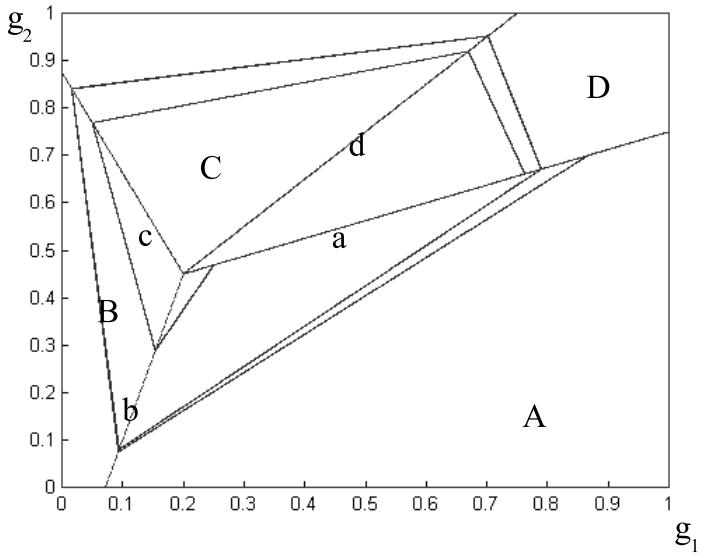


Figure 6.2: Areas and trajectories

is in area A it moves on the line $g_2 = \frac{\frac{9}{20}+\frac{3}{8}\varepsilon}{\frac{1}{5}+\varepsilon} g_1$. The trajectory leaves area A at line segment b, which corresponds to the equation $g_2 = \frac{7}{2}g_1 - \frac{1}{4}$. So at that moment we have

$$g_1 = \frac{\frac{1}{5}+\varepsilon}{1+12\frac{1}{2}\varepsilon} \quad \text{and} \quad g_2 = \frac{\frac{9}{20}+\frac{3}{8}\varepsilon}{1+12\frac{1}{2}\varepsilon}$$

and the trajectory enters area B. As long as the trajectory is in area B it moves on the line $1 - g_2 = \frac{\frac{11}{20}+12\frac{1}{8}\varepsilon}{\frac{1}{5}+\varepsilon} g_1$. The trajectory leaves area B and enters area C at line segment c, corresponding to the equation $g_2 = -\frac{17}{8}g_1 + \frac{7}{8}$, so at that moment we have

$$g_1 = \frac{\frac{1}{5}+\varepsilon}{1+80\varepsilon} \quad \text{and} \quad g_2 = \frac{\frac{9}{20}+67\frac{7}{8}\varepsilon}{1+80\varepsilon}.$$

As long as the trajectory is in area C it moves on the line $1-g_2 = \frac{\frac{11}{20}+12\frac{1}{8}\varepsilon}{\frac{4}{5}+79\varepsilon}(1-g_1)$. The trajectory leaves area C and enters area D at line segment d, which has $g_2 = g_1 + \frac{1}{4}$ as its equation, meaning that at that moment we have

$$g_1 = \frac{\frac{1}{5}+188\frac{1}{2}\varepsilon}{1+267\frac{1}{2}\varepsilon} \quad \text{and} \quad g_2 = \frac{\frac{9}{20}+255\frac{3}{8}\varepsilon}{1+267\frac{1}{2}\varepsilon}.$$

As long as the trajectory is in area D it moves on the line $g_2 = \frac{\frac{9}{20}+255\frac{3}{8}\varepsilon}{\frac{4}{5}+79\varepsilon}(1-g_1)$. At the end of the trajectory we are back on line segment a, so at that moment

$$g_1 = \frac{\frac{1}{5}+301\varepsilon}{1+380\varepsilon} \quad \text{and} \quad g_2 = \frac{\frac{9}{20}+255\frac{3}{8}\varepsilon}{1+380\varepsilon}.$$

Hence the function h is as follows:

$$h(\frac{1}{5}+\varepsilon, \frac{9}{20}+\frac{3}{8}\varepsilon) = (\frac{\frac{1}{5}+301\varepsilon}{1+380\varepsilon}, \frac{\frac{9}{20}+255\frac{3}{8}\varepsilon}{1+380\varepsilon}).$$

We have $h(\frac{1}{5}+\varepsilon, \frac{9}{20}+\frac{3}{8}\varepsilon) = (\frac{1}{5}+\varepsilon, \frac{9}{20}+\frac{3}{8}\varepsilon)$ if and only if $\varepsilon = 0$ or $\varepsilon = \frac{56}{95}$. Therefore there are precisely 2 orbits with starting points $(\frac{1}{5}, \frac{9}{20})$, which is the equilibrium point, and $(\frac{15}{19}, \frac{51}{76})$.
For all $\varepsilon \in (0, \frac{56}{95})$ we have that $(\frac{15}{19}, \frac{51}{76}) > h(\frac{1}{5}+\varepsilon, \frac{9}{20}+\frac{3}{8}\varepsilon) > (\frac{1}{5}+\varepsilon, \frac{9}{20}+\frac{3}{8}\varepsilon)$ and for all $\varepsilon \in (\frac{56}{95}, \frac{4}{5}]$ we have that $(\frac{15}{19}, \frac{51}{76}) < h(\frac{1}{5}+\varepsilon, \frac{9}{20}+\frac{3}{8}\varepsilon) < (\frac{1}{5}+\varepsilon, \frac{9}{20}+\frac{3}{8}\varepsilon)$. Hence the orbit starting at the point $(\frac{15}{19}, \frac{51}{76})$ is stable and the equilibrium point by itself is a non-stable orbit. ■

Now we need a few notations and definitions. Let e^* be the equilibrium point: $e^* = (\frac{1}{5}, \frac{9}{20})$. In view of lemma 6.4.5 there is a stable orbit $\overline{\tau}^*$ with starting point $(\frac{15}{19}, \frac{51}{76})$. For each region $X \in \{A, B, C, D\}$ let τ_X be the point where the orbit $\overline{\tau}^*$ enters region X. For each $x, y \in [0,1]^2$ let $l[x, y]$ denote the line segment starting at x and finishing at y and let T^* be the area in $[0,1]^2$ that is enclosed by $\overline{\tau}^*$. Then T^* is a compact and convex subset of $[0,1]^2$ with boundary $\overline{\tau}^*$ and extreme points τ_A, τ_B, τ_C and τ_D.

Lemma 6.4.6 *g_t is outside of T^* for each stage t.*

Proof. Since $g_t = (1,1)$ is outside of T^*, it is sufficient to show that if g_t is outside of T^*, then also g_{t+1} is outside of T^*. Suppose that g_t is outside of T^*. Suppose also that $g_t \in A$. For the other areas similar proofs can be given. Notice that by lemma 6.4.3, if the fictitious play process is in area A, then the current run can only be $[(0,0),(1,0)]$ or $[(0,0),(0,0)]$. Consequently either

$$g_{t+1} = \tfrac{t}{t+1} \cdot g_t + \tfrac{1}{t+1} \cdot (0,0) \quad \text{or} \quad g_{t+1} = \tfrac{t}{t+1} \cdot g_t + \tfrac{1}{t+1} \cdot (1,0),$$

so g can only move towards $(0,0)$ or $(1,0)$. In the latter case g_{t+1} is clearly outside of T^*, while in the former case $g_{t+1} \in l[(0,0), g_t]$.
Since $l[(0,0), g_t]$ and $l[(0,0), \tau_A]$ intersect only in $(0,0)$, where $l[(0,0), \tau_A] \supset l[\tau_A, \tau_B]$, the line segments $l[(0,0), g_t]$ and $l[\tau_A, \tau_B]$ do not intersect and hence g_{t+1} is outside of T^*. ■

Proof of theorem 6.3.1 According to lemma 6.4.3 the different runs follow each other cyclically. This means that if the fictitious play process converges, then it must converge to the unique common point of the areas in figure 6.2, which is the equilibrium point. However, according to lemma 6.4.6 the fictitious play process is always outside of the region T^*, which includes the equilibrium point and is bounded by the stable orbit $\overline{\tau}^*$. Therefore it cannot converge to a single point at all. □

6.5 Other models on fictitious play in stochastic games

We have analyzed one specific model of fictitious play in stochastic games. However, the model we used is not the only possible generalization of the fictitious play process to stochastic games. In this section we will mention 2 other models.

Model 2 (Update only in the state that is visited):

Suppose that the stochastic game has state space $S = \{1, 2, \ldots, z\}$ and that players 1 and 2 have m^s respectively n^s actions in state $s \in S$. In this model the fictitious play process proceeds as follows: Play can start in any state and as long as not all states have been visited at least once, each player selects action 1 at each stage, irrespective of the play so far. During the fictitious play process the players select pure actions at each stage. Suppose that from stage τ' on each state has been visited at least once and let s_τ denote the state that is visited at stage τ. Let $i^{s_\tau} \in I^{s_\tau}$ denote the pure action that is selected by player 1 at stage τ of the fictitious play process, for stage $t \geq \tau'$ let f_t^s denote the action frequencies of the pure actions in state s up to and including stage t of the fictitious play process, i.e.

$$f_t^s = \frac{1}{\#\{\tau \leq t \,|\, s_\tau = s\}} \sum_{\tau \leq t \,|\, s_\tau = s} i^{s_\tau} \in \Delta^{m^s} \tag{6.3}$$

and let

$$f_t = (f_t^1, f_t^2, \ldots, f_t^z).$$

For player 2 the (joint) actions $j_\tau \in \Delta^{n^1} \times \Delta^{n^2} \times \ldots \times \Delta^{n^z}$, $g_t^s \in \Delta^{n^s}$ and $g_t \in \Delta^{n^1} \times \Delta^{n^2} \times \ldots \times \Delta^{n^z}$ are defined analogously. Then f_t and g_t can be seen as joint mixed actions for players 1 and 2 respectively. Now let $x(f_t)$ be the stationary strategy for player 1 that prescribes to play the mixed action f_t^s each time that state s is visited, and let $y(g_t)$ be the stationary strategy for player 2 that prescribes to play the mixed action g_t^s at each visit to state s. This model for the fictitious play process for 2-player stochastic games can, as a generalization of the fictitious play process for repeated games (cf. definition 6.2.1), be defined as follows:

Definition 6.5.1 *The fictitious play process for 2-player stochastic games is a sequence $((i_t^{s_t}, j_t^{s_t}))_{t=1}^{\infty}$, with $i_t^{s_t}$ and $j_t^{s_t}$ pure actions for players 1 and 2 respectively for all t, recursively defined as follows: For each stage $\tau \leq \tau'$ we have: $i_\tau^{s_\tau} = j_\tau^{s_\tau} = 1$ irrespective of the state that is visited and $f_\tau^s = (1, 0, \ldots, 0) \in \Delta^{m^s}$ and $g_\tau^s = (1, 0, \ldots, 0) \in \Delta^{n^s}$. Furthermore at stage $t \geq \tau + 1$ we consider the stationary strategies $x(f_{t-1})$ and $y(g_{t-1})$ and we take i_t and j_t such that $x(i_t)$ and $y(j_t)$ are pure stationary best replies against $y(g_{t-1})$ and $x(f_{t-1})$ respectively. Thereafter we update f and g as follows:*

$$f_t^{s_t} = \frac{\#\{\tau \leq t \,|\, s_\tau = s_t\} - 1}{\#\{\tau \leq t \,|\, s_\tau = s_t\}} \cdot f_{t-1}^{s_t} + \frac{1}{\#\{\tau \leq t \,|\, s_\tau = s_t\}} \cdot i_t^{s_t} \in \Delta^{m^{s_t}}, \tag{6.4}$$

$$f_t^s = f_{t-1}^s \text{ for all } s \neq s_t,$$

$$g_t^{s_t} = \frac{\#\{\tau \leq t \,|\, s_\tau = s_t\} - 1}{\#\{\tau \leq t \,|\, s_\tau = s_t\}} \cdot g_{t-1}^{s_t} + \frac{1}{\#\{\tau \leq t \,|\, s_\tau = s_t\}} \cdot j_t^{s_t} \in \Delta^{n^{s_t}} \tag{6.5}$$

and

$$g_t^s = g_{t-1}^s \text{ for all } s \neq s_t.$$

Notice that formulas (6.3) and (6.4) are equivalent.

This model does, in contrast with the one we analyzed, correspond to a play of the game. A drawback of this model is that there is no certainty with respect to the number of times the several states have been visited, which complicates the analysis. We advance the following conjecture with respect to this model:

Conjecture 6.5.2 *The fictitious play process as specified in definition 6.5.1 does not converge in the* $2 \times 2 \times 2$ *- example discussed in this chapter.*

Model 3: (Auxiliary matrix games)

There also is another model on fictitious play in stochastic games, which is rather different from the one we analyzed. We will not analyze this model here, but for the sake of completeness we will mention it. In this model the fictitious play process, in common with the model we analyzed, updates in all states simultaneously. However, the method used for the updating procedure is different. Suppose that the stochastic game has state space $S = \{1, 2, \ldots, z\}$ and that players 1 and 2 have m^s respectively n^s actions in state $s \in S$. We suppose that during the fictitious play process the players select joint pure actions at each stage. Let

$$i_\tau = \left(i_\tau^1, i_\tau^2, \ldots, i_\tau^z\right) \in \Delta^{m^1} \times \Delta^{m^2} \times \ldots \times \Delta^{m^z}$$

denote the joint pure action that is selected by player 1 at stage τ of the fictitious play process. Furthermore let f_t^s denote the action frequencies of the pure actions of player 1 in state s up to and including stage t of the fictitious play process, i.e.

$$f_t^s = \frac{1}{t} \sum_{\tau=1}^{t} i_\tau^s \in \Delta^{m^s}$$

and let

$$f_t = (f_t^1, f_t^2, \ldots, f_t^z).$$

For player 2 the (joint) actions $j_\tau \in \Delta^{n^1} \times \Delta^{n^2} \times \ldots \times \Delta^{n^z}$, $g_t^s \in \Delta^{n^s}$ and $g_t \in \Delta^{n^1} \times \Delta^{n^2} \times \ldots \times \Delta^{n^z}$ are defined analogously. Then f_t and g_t can be seen as joint mixed actions for players 1 and 2 respectively. In this model at each stage τ and for each state $s \in S$ the players play a best reply against the average action of the other player in an auxiliary matrix game $M^s(V_\tau)$, where $M^s(V_\tau)$ denotes the $(m^s \times n^s)$ -matrix, whose (i, j)-th entry contains the following number:

$$R^s(i, j) + \sum_{s' \in S} \Pr\{s' \mid s, i, j\} \cdot V_\tau^s.$$

We are now ready to present the definition of the fictitious play process for this model.

Definition 6.5.3 *The fictitious play process is recursively defined as follows: Choose* $\mathfrak{x}_1^s \in \mathbb{R}^{n^s}$ *and* $\mathfrak{y}_1^s \in \mathbb{R}^{m^s}$ *such that*

$$\min\{\mathfrak{x}_1^s\} = \max\{\mathfrak{y}_1^s\} \text{ and } \min\{\mathfrak{y}_1^s\} \geq v^s \text{ for each } s \in S.$$

Furthermore take $V_1^s := \max\{\mathfrak{y}_1^s\}$ *for each* $s \in S$ *and take* $f_1^s = (1, 0, \ldots, 0) \in \Delta^{m^s}$ *and* $g_1^s = (1, 0, \ldots, 0) \in \Delta^{n^s}$.
Let $\tau \geq 2$. *Take for each* $s \in S$ *a pure action* $i_\tau^s \in I^s$ *such that*

$$(\mathfrak{y}_{\tau-1}^s)_{i_\tau^s} = \max\left\{\mathfrak{y}_{\tau-1}^s\right\}$$

and a pure action $j_\tau^s \in J^s$ *such that*

$$(\mathfrak{x}_{\tau-1}^s)_{i_\tau^s} = \max\left\{\mathfrak{x}_{\tau-1}^s\right\}.$$

Now for each $s \in S$ *make the following updates:*

1. $V_\tau^s := \min\left\{\frac{1}{\tau} \cdot \max\left\{\mathfrak{y}_{\tau-1}^s\right\}, V_{\tau-1}^s\right\}$
2. $\mathfrak{x}_\tau^s := \mathfrak{x}_{\tau-1}^s + (i_\tau^s)^T M^s(V_\tau)$
3. $\mathfrak{y}_\tau^s := \mathfrak{y}_{\tau-1}^s + M^s(V_\tau) j_\tau^s$
4. $f_\tau^s := \frac{\tau-1}{\tau} \cdot f_{\tau-1}^s + \frac{1}{\tau} \cdot i_\tau^s$
5. $g_\tau^s := \frac{\tau-1}{\tau} \cdot g_{\tau-1}^s + \frac{1}{\tau} \cdot j_\tau^s$.

This model is introduced by Vrieze and Tijs (1982) for β-discounted zero-sum stochastic games (cf. definition 1.3) and uses the fact that stationary β-discounted optimal strategies always exist. Vrieze and Tijs prove that this fictitious play process, being the sequence $(f_\tau, g_\tau)_{\tau=1}^\infty$, converges to the set of stationary optimal strategy pairs and that furthermore the sequence $(V_\tau^s)_{\tau=1}^\infty$ converges to the value v of the stochastic game.

References

Arrow, K.J. (1962): The economic implications of learning by doing. Review of economic studies 29, 155-173.

Aumann, R.J. (1981): Survey of repeated games. In: Essays in Game Theory and Mathematical Economics in honor of Oskar Morgenstern. Bibliographisches Institut Mannheim, pp. 11-42.

Brown, G.W. (1951): Iterative solution of games by fictitious play. In: Koopmans, T.C. (Ed.), Activity Analysis of Production and Allocation, New York: Wiley, pp. 374-376.

Cortright, J. (2001): New Growth Theory, Technology and Learning: A Practitioner's Guide. Reviews of Economic Development Literature and Practice: No. 4.

Federgruen, A. (1978): On n-person stochastic games with denumerable state space. Advances in Applied Probability 10, 452-471.

Filar, J.A. (1981): Ordered field property for stochastic games when the player who controls transitions changes from state to state. Journal of Optimization Theory and Applications 34, 503-515.

Filar, J.A. & Vrieze, O.J. (1997): Competitive Markov Decision Processes. Springer-Verlag, New York.

Gillette, D. (1957): Stochastic games with zero stop probabilities. In: Dresher, M., Tucker, A.W., Wolfe, P. (Eds.), Contributions to the Theory of Games III, Annals of Mathematical Studies, vol. 39. Princeton University Press, pp. 179-187.

Hoffman, A.J. & Karp, R.M. (1966): On nonterminating stochastic games. Management Science 12, 359-370.

Hordijk, A., Vrieze, O.J. & Wanrooij, G.L. (1983): Semo-Markov strategies in stochastic games. International Journal of Game Theory 12, 81-89.

Joosten, R., Brenner, B., Witt, U., 2002. Games with frequency-dependent stage payoffs. International Journal of Game Theory 31, 609-620.

Joosten, R., Peters, H. & Thuijsman, F. (1995): Unlearning by not doing: repeated games with vanishing actions. Games and Economic Behavior 9, 1-7.

Jovanovic, B. (2000): Growth theory. NBER Working Paper 7468, National Bureau of Economic Research, Inc.

Kaelbling, L.P., Littman, M.L. & Moore, A.W. (1996): Reinforcement Learning: A Survey. Internet address: http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume4/kaelbling96a-html/rl-survey.html

Krishna, V. & Sjöström, T. (1998): On the convergence of fictitious play. Mathematics of Operations Research 23, 479-511.

Mataric, M.J. (1995): Issues and approaches in the design of collective autonomous agents. Robotics and Autonomous Systems, 16(2-4), 321-331.

Mataric, M.J. (1997): Learning social behavior. Robotics and Autonomous Systems 20, 191-204.

Mertens, J.F. & Neyman, A. (1981): Stochastic games. International Journal of Game Theory 10, 53-66.

Metrick, A. & Polak, B. (1994): Fictitious play in 2×2 games: a geometric proof of convergence. Economic Theory 4, 923-933.

Miyasawa, K. (1961): On the convergence of the learning process in 2×2 non-zero-sum two-person game. Research Memorandum No. 33, Economic Research Program, Princeton University.

Monderer, D. & Shapley, L.S. (1996): Fictitious play property for games with identical interests. Journal of Economic Theory 68, 258-265.

Nash, J. (1950a): Non-cooperative games. Ph.D. dissertation. Princeton University.

Nash, J. (1950b): Equilibrium points in n-person games. Proceedings of the National Academy of Sciences (USA) 36, 48-49.

Nash, J. (1951): Non-cooperative games. Annals of Mathematics 54, 284-295.

Neumann, von, J. (1928): Zur Theorie der Gesellschaftsspiele. Mathematische Annalen 100, 295-320.

Robinson, J. (1951): An iterative method of solving a game. Annals of Mathematics 54, 296-301.

Rogers, P.D. (1969): Non-zerosum stochastic games. Ph.D. thesis, Report ORC 69-8, Operations Research Center, University of California, Berkeley.

Schoenmakers, G., Flesch, J. & Thuijsman, F. (2001): Fictitious play in stochastic games. Working paper.

Schoenmakers, G., Flesch, J. & Thuijsman, F. (2002): Coordination games with vanishing actions. International Game Theory Review 4, 119-126.

Schoenmakers, G., Flesch, J., Thuijsman, F. & Vrieze, O.J. (2004). Repeated games with bonuses. Working paper.

Schoenmakers, G., Joosten, R., Peters, H. & Thuijsman, F. (forthcoming). Equilibria in games with vanishing actions.

Sela, A. (2000): Fictitious play in 2×3-games. Games and economic behavior 31, 152-162.

Shapley, L.S. (1953): Stochastic games. Proceedings of the National Academy of Sciences U.S.A. 39, 1095-1100.

Shapley, L.S. (1964): Some topics in two-person games. In: Dresher, L., Shapley, S., Tucker, A.W. (Eds.), Advances in Game Theory, Princeton University Press, pp. 1-28.

Shapley, L.S. & Snow, R.N. (1950): Basic solutions of discrete games. Annals of Mathematics Studies 24, 27-35.

Sobel, M.J. (1971): Noncooperative stochastic games. Annals of Mathematical Statistics 42, 1930-1935.

Sobel, M.J. (1981): Myopic solutions of Markov decision processes and stochastic games. Operations Research 29, 995-1009.

Sorin, S. (1992): Repeated games with complete information. In: R.J. Aumann, S. Hart eds, Handbook of Game Theory with Economic Applications, Volume 1. Elsevier Science Publishers, North-Holland, pp. 72-107.

Sutton, R.S. & Barto, A.G. (1998): Reinforcement Learning: An Introduction. Cambridge: MIT Press.

Thuijsman, F. (1992): Optimality and equilibria in stochastic games. CWI-Tract 82, Centre for Mathematics and Computer Science, Amsterdam.

Thuijsman, F. & Vrieze, O.J. (1993): Stationary ε-optimal strategies in stochastic games. OR Spektrum 15, 9-15.

Vieille, N. (2000a): Two-player stochastic games I: a reduction. Israel Journal of Mathematics 119, 55-91.

Vieille, N. (2000b): Two-player stochastic games II: the case of recursive games. Israel Journal of Mathematics 119, 93-126.

Vrieze, O.J. (1987): Stochastic game with finite state and action spaces. CWI-Tract 33, Centre for Mathematics and Computer Science, Amsterdam.

Vrieze, O.J. & Tijs, S.H. (1982): Fictitious play applied to sequences of games and discounted stochastic games. International Journal of Game Theory 11, 71-85.

Walker, S. (1975): Learning and reinforcement. London: Methuen.

Author index

Subject index

Symbol index

Greek

Latin

Samenvatting

Speltheorie beschrijft en analyseert situaties, waarin verschillende beslissingnemers, gewoonlijk spelers genoemd, verkeren, die al dan niet tegenstrijdige belangen hebben. De interacties tussen de spelers kunnen op een enkel tijdstip plaatshebben, zoals bijvoorbeeld bij een gesloten-bod veiling. Bij zo'n veiling doen alle spelers simultaan en onafhankelijk van elkaar een bod, waarna de hoogste bieder het goed ontvangt. Interacties tussen spelers kunnen echter ook herhaaldelijk voorkomen, hetgeen typisch het geval is in herhaalde en stochastische spelen. Een 2-speler stochastisch spel kan als volgt beschreven worden: We hebben een toestandsruimte en in iedere toestand hebben beide spelers een verzameling acties ter beschikking. Aan ieder actiepaar is zowel een uitbetaling voor beide spelers als een kansvector, de zogenaamde overgangskansenvector, verbonden. Het verloop van een stochastisch spel geschiedt als volgt: Het spel begint op tijdstip 1 in een begintoestand, alwaar, simultaan en onafhankelijk, beide spelers een actie moeten selecteren. Nu krijgen beide spelers hun uitbetaling, behorende bij het geselecteerde actiepaar. Bovendien verhuist het spel naar een andere toestand, bepaald door de overgangskansenvector, alwaar de spelers opnieuw acties moeten kiezen. Opnieuw krijgen zij de uitbetalingen corresponderend met het geselecteerde actiepaar en het spel verhuist naar een andere toestand, waar opnieuw acties gekozen dienen te worden. Dit proces duurt voort tot in het oneindige.

Het spelverloop leidt tot een eindeloze stroom uitbetalingen aan de spelers, die geëvalueerd dienen te worden. Een evaluatiecriterium worden een opbrengst genoemd en het doel van iedere speler is de eigen opbrengst zo groot mogelijk te krijgen. Hierbij geldt dat er geen bindende afspraken gemaakt kunnen maken. Verder kennen de spelers de uitbetalingen behorend bij ieder actiepaar in iedere toestand en onthouden ze de gehele geschiedenis van het spel tot het huidige tijdstip. Derhalve kunnen de spelers deze informatie gebruiken bij het kiezen van een actie op het huidige tijdstip. Een plan dat een speler voorschrijft welke actie te spelen, gegeven de geschiedenis van het spel tot het huidige tijdstip, wordt een strategie genoemd. Als de voorgeschreven actie alleen afhangt van de toestand die momenteel bezocht wordt, dan heet de strategie stationair. Een speciaal soort stochastisch spel is een stochastisch spel dat slechts uit een toestand bestaat. Zo'n spel wordt een herhaald spel genoemd.

Tijdens het verloop van herhaalde en stochastische spelen kunnen de spelers hun strategisch gedrag veranderen. Het idee hierachter is dat als een speler er achter komt dat hij een hogere opbrengst kan krijgen door een andere actie te spelen of door juist constant dezelfde actie de blijven spelen, dan zal hij niet nalaten dit te doen. Dit is het concept van leren in spelen. De modellen die in dit proefschrift behandeld worden, gaan over het verkrijgen, vergroten en ook het verliezen van vaardigheden en vallen

derhalve binnen het raamwerk van leren. In onze modellen verkrijgen en vergroten de spelers hun vaardigheden door acties herhaaldelijk te spelen, hetgeen leidt tot hogere uitbetalingen. Tevens leidt het niet herhaaldelijk spelen van acties tot een verlies aan vaardigheden en derhalve tot lagere uitbetalingen (zoals in het model in hoofdstuk 2) of zelfs tot het verleren van de betreffende acties (zoals in de modellen in hoofdstukken 3, 4 en 5).

In hoofdstuk 2 wordt het model van herhaalde spelen met bonussen behandeld. Zo'n spel is een herhaald spel, waarin speler 1 zijn vaardigheden kan vergroten door dezelfde actie meerdere malen achtereen te spelen. Om precies te zijn, speler 1 krijgt een bonus, als hij de actie speelt die hij op het vorige tijdstip ook gespeeld heeft. Speelt hij een andere actie, dan krijgt hij geen bonus.

In hoofdstukken 3, 4 en 5 wordt het model van herhaalde spelen met verdwijnende acties behandeld. Het idee hier is dat, wanneer een speler een bepaalde actie een tijd lang niet gespeeld heeft, hij deze actie verleert en derhalve niet meer in staat is deze actie te spelen. Vanaf het moment dat een actie verleerd is, kan deze nooit meer gebruikt worden door de betreffende speler; in dit model kunnen de spelers alleen vaardigheden verliezen, niet verkrijgen. In hoofdstuk 3 wordt het nul-sommodel behandeld, het model waarin de winst van de ene speler automatisch het verlies van de andere speler is. Hoofdstuk 4 gaat over coördinatiespelen, spelen, waarin de spelers hun acties moeten coördineren om een goede opbrengst te krijgen. In hoofdstuk 5 tenslotte worden spelen geanalyseerd, waarin geen aannames gemaakt worden met betrekking tot de structuur van de uitbetalingen.

In hoofdstuk 6 wordt een model van fictitious play in stochastische spelen behandeld. Fictitious play kan als volgt omschreven worden: de spelers kennen de uitbetalingen aan de andere speler niet, maar ze nemen wel waar welke acties de andere speler neemt. Op ieder tijdstip beschouwen beide spelers de reeks acties van de andere speler tot op het huidige moment en bepalen ze diens "gemiddelde" actie. Vervolgens bepalen ze welke van hun eigen acties tegen deze gemiddelde actie de hoogste opbrengst oplevert. Dit is de actie die uitgevoerd wordt.

Curriculum Vitae

Gijs Schoenmakers werd geboren op 15 oktober 1974 te Schijndel. Van 1987 tot 1993 volgde hij middelbaar onderwijs aan Gymnasium Bernrode te Heeswijk-Dinther, alwaar hij het VWO-diploma behaalde. Vervolgens begon hij aan een studie econometrie aan de Katholieke Universiteit Brabant. In 1998 legde hij het doctoraalexamen af in de richting besliskunde. Van december 1998 tot juni 2004 was hij werkzaam als AIO bij de capaciteitsgroep wiskunde van de Universiteit Maastricht, alwaar hij onderzoek verrichtte binnen het gebied van de niet-coöperatieve speltheorie. De resultaten van dit onderzoek zijn weergegeven in dit proefschrift.