Stochastic games on a product state space

János Flesch, Gijs Schoenmakers, Koos Vrieze

Stochastic Games on a Product State Space

RM/07/010

JEL code: C73

Maastricht research school of **E**conomics
of **TE**chnology and **OR**ganizations

Universiteit Maastricht
Faculty of Economics and Business Administration
P.O. Box 616
NL - 6200 MD Maastricht

phone : ++31 43 388 3830
fax : ++31 43 388 4873

Stochastic Games on a Product State Space

János Flesch[*], Gijs Schoenmakers, Koos Vrieze[†]

March 27, 2007

Abstract

We examine product-games, which are n-player stochastic games satisfying: (1) the state space is a product $S^1 \times \cdots \times S^n$; (2) the action space of any player i only depends of the i-th coordinate of the state; (3) the transition probability of moving from $s^i \in S^i$ to $t^i \in S^i$, on the i-th coordinate S^i of the state space, only depends on the action chosen by player i. So, as far as the actions and the transitions are concerned, every player i can play on the i-th coordinate of the product-game without interference of the other players. No condition is imposed on the payoff structure of the game.

We focus on product-games with an aperiodic transition structure, for which we present an approach based on so-called communicating states. For the general n-player case, we establish the existence of 0-equilibria, which makes product-games one of the first classes within n-player stochastic games with such a result. In addition, for the special case of two-player zero-sum games of this type, we show that both players have stationary 0-optimal strategies. Both proofs are constructive by nature.

Keywords: Noncooperative Games, Stochastic Games, Markov Decision Problems, equilibria.

1 Introduction

Stochastic games and product-games. An n-player *stochastic game* is given by (1) a set of players $N = \{1, \ldots, n\}$, (2) a nonempty and finite set of states S, (3) for

[*] Address: University of Maastricht, Department of Quantitative Economics, P.O.Box 616, 6200 MD Maastricht, The Netherlands

[†] Address of Gijs Schoenmakers & Koos Vrieze: University of Maastricht, Department of Mathematics, P.O.Box 616, 6200 MD Maastricht, The Netherlands

each state $s \in S$, a nonempty and finite set of actions A^i_s for each player i, (4) for each state $s \in S$ and each joint action $a_s \in \times_{i \in N} A^i_s$, a payoff $r^i_s(a_s) \in \mathbb{R}$ to each player i, (5) for each state $s \in S$ and each joint action $a_s \in \times_{i \in N} A^i_s$, a transition probability distribution $p_{sa_s} = (p_{sa_s}(t))_{t \in S}$.

The game is to be played at stages in $\mathbb{N}$ in the following way. Play starts at stage 1 in an initial state, say in state $s_1 \in S$. In s_1, each player $i \in N$ is to choose an action a^i_1 from his action set $A^i_{s_1}$. These choices have to be made independently. The chosen joint action $a_1 = (a^1_1, \dots, a^n_1)$ induces an immediate payoff $r^i_{s_1}(a_1)$ to each player i. Next, play moves to a new state according to the transition probability distribution $p_{s_1 a_1}$, say to state $s_2 \in S$. At stage 2, a new action $a^i_2 \in A^i_{s_2}$ is to be chosen by each player i in state s_2. Then, given action combination $a_2 = (a^1_2, \dots, a^n_2)$, player i receives payoff $r^i_{s_2}(a_2)$ and the play moves to some state s_3 according to the transition probability distribution $p_{s_2 a_2}$, and so on. We assume complete information (i.e. the players know all the data of the stochastic game), full monitoring (i.e. the players observe the present state and the actions chosen by all the players), and perfect recall (i.e. the players remember all previous states and actions).

A *Markov transition structure* Γ^i for player $i \in N$ is given by (1) a nonempty and finite state space S^i; (2) a nonempty and finite action set $A^i_{s^i}$ for each state $s^i \in S^i$; (3) a transition probability distribution $p^i_{s^i a^i_{s^i}}$ over the state space S^i for each state $s^i \in S^i$ and for each action $a^i_{s^i} \in A^i_{s^i}$. Note that, if we also assigned a payoff in every state to every action, then we would obtain the well-known model of Markov decision problems for player i.

We will now consider a special type of n-player stochastic games in which the transition structure is derived by taking the product of these n Markov transition structures. For the sake of simplicity, we will call such a game a product-game. A *product-game* G, associated to the Markov transition structures $\Gamma^1, \Gamma^2, \dots, \Gamma^n$, is an n-player stochastic game for which (1) the set of players is $N = \{1, \dots, n\}$; (2) the state space is $S = S^1 \times \cdots \times S^n$; (3) the action set for each player $i \in N$ in each state $s = (s^1, \dots, s^n) \in S$ is $A^i_s = A^i_{s^i}$; (4) the transition probability distribution p_{sa_s}, for each state $s = (s^1, \dots, s^n) \in S$ and for each joint action $a_s = (a^1_s, \dots, a^n_s) \in \times_{i \in N} A^i_s$, is

$$p_{sa_s}(\bar{s}) = \prod_{i \in N} p^i_{s^i a^i_s}(\bar{s}^i)$$

for state $\bar{s} = (\bar{s}^1, \dots, \bar{s}^n) \in S$. Note that there is no condition imposed on the payoff structure.

Observe that (1) the action space of player i only depends on the i-th coordinate of the state, (2) the i-th coordinate of the transitions from any state s only depend on the i-th coordinate s^i of the state and on the action a^i_s chosen by player i, i.e. for any $\bar{s}^i \in S^i$ we have

$$p_{sa_s}(S^1, \ldots, S^{i-1}, \bar{s}^i, S^{i+1}, \ldots, S^n) = p^i_{s^i a^i_s}(\bar{s}^i).$$

Therefore, as far as the actions and the transitions are concerned, player i can play on the i-th coordinate of the game G without the interference of the other players. As a consequence, play of the product game G can be viewed as simultaneous play of the n Markov transition structures $\Gamma^1, \ldots, \Gamma^n$, which are linked by payoff functions $r^1, \ldots, r^n$ that may depend on all n current states as well as on all n actions chosen by the players.

Product-games have been introduced in Altman et al. [2005], although in a somewhat different fashion. They only examined two-player games in which the sum of the payoffs is always equal to zero (zero-sum games), and dropped the assumption of full monitoring by letting each player only observe his own coordinate of the present state and only the action chosen by himself. As a result, both players have to make choices without noticing anything about the other player's behavior. They showed that a linear programming formulation is sufficient to solve these games, i.e. to find the value and stationary optimal strategies (cf. the definitions below).

Note that the class of product-games, as defined in our paper, differs essentially from other known classes of n-player stochastic games. Stochastic games with a single controller, i.e. when one player controls the transitions, however, fall into the class of product-games. Indeed, a stochastic game which is controlled by player i can be seen as a product-game in which S^j is a singleton for all players $j \neq i$. Finally, we wish to mention the class of stochastic games with additive transitions (AT-games, cf. Flesch et al. [2007]), i.e. when the transitions are additively decomposable into player-dependent components, in contrast with a product decomposition. Not surprisingly, the structure of product-games and AT-games differ essentially, and product-games require new ideas and an entirely different approach.

From now on, we will consequently use the upper-index for the player and the lower-index for the state. Whenever one of them is omitted, we will then mean a vector in the case of quantities and a product in the case of sets, for all possible players or states respectively. For example, A^i denotes $\times_{s \in S} A^i_s$. Finally, we denote the set of opponents of any player i by $-i := N - \{i\}$. Then, $-i$ in the upper-index will mean a vector or product for all players $j \neq i$. For example, S^{-i} denotes $\times_{j \in N - \{i\}} S^j$.

Strategies. A mixed action x^i_s for player i in state $s \in S$ is a probability distribution on A^i_s. The set of mixed actions for player i in state s is denoted by X^i_s. A mixed action is called completely mixed, if it assigns a positive probability to each available action. A (history dependent) strategy π^i for player i is a decision rule that prescribes a mixed action $\pi^i_s(h) \in X^i_s$ in the present state s depending on the past history h of play (i.e. the sequence of all past states and all past actions chosen by the players). We use the notation Π^i for the set of strategies for player i. A strategy π^i for player i is called pure if π^i prescribes, for every state and every possible history, one specific action to be played with probability 1. Given a strategy π^i for player i and a history h, the strategy π^i conditional on h, denoted by $\pi^i[h]$, is the strategy which prescribes a mixed action $\pi^i_s[h](h')$ in any present state s for any history h' as if h had happened before h', i.e. $\pi^i_s[h](h') = \pi^i_s(h \oplus h')$, where $h \oplus h'$ is the history consisting of h concatenated by h'. In fact, $\pi^i[h]$ is just the continuation strategy of π^i after history h.

If the mixed actions prescribed by a strategy only depend on the present state then the strategy is called stationary. Thus, the stationary strategy space for player i is $X^i = \times_{s \in S} X^i_s$. We use the notation x^i for stationary strategies for player i, while x^i_s refers to the corresponding mixed action for player i in state s. Note that the set of pure stationary strategies for player i is simply $A^i = \times_{s \in S} A^i_s$.

A joint stationary strategy $x = (x^i)_{i \in N}$ induces a Markov-chain on the state space S with transition matrix $P(x)$, where entry $(s, \bar{s})$ of $P(x)$ gives the transition probability $p_{sx_s}(\bar{s})$ for moving from state s to state $\bar{s}$ when the joint mixed action x_s is played in state s. With respect to this Markov-chain, we can speak of transient and recurrent states. A state is called recurrent if, when starting there, play will eventually return with probability 1; otherwise the state is called transient. If play is in a recurrent state, then this state will be visited infinitely often with probability 1, while transient states can only be visited finitely many times, with probability 1. We can group the recurrent states into minimal closed sets, into so-called ergodic sets. An ergodic set is a collection F of recurrent states with the property that, when starting in any of the states in F, all states in F will be visited infinitely often and the play will remain in F forever with probability 1.

Let

$$Q(x) := \lim_{M \to \infty} \frac{1}{M} \sum_{m=1}^{M} P^m(x); \tag{1}$$

the limit is known to exist (cf. Doob [1953], theorem 2.1, page 175). Entry $(s, \bar{s})$ of the stochastic matrix $Q(x)$, denoted by $q_{sx}(\bar{s})$, is the expected frequency of stages for which the process is in state $\bar{s}$ when starting in s. The matrix $Q(x)$ has the well known properties (cf. Doob [1953]) that

$$Q(x) = Q(x)\,P(x) = P(x)\,Q(x) = Q^2(x). \tag{2}$$

Note that $Q(\cdot)$ is in general not continuous on the set X of joint stationary strategies. Indeed, if x_m converges to x but the probabilities on certain actions vanish in the limit, then the ergodic structure of the induced Markov chains may change drastically in the limit.

Rewards. For a joint strategy $\pi = (\pi^i)_{i\in N}$ and initial state $s \in S$, the sequences of payoffs are evaluated by the (expected) average reward, which is given for player i by

$$\gamma_s^i(\pi) := \liminf_{M\to\infty} \mathbb{E}_{s\pi}\left(\frac{1}{M}\sum_{m=1}^{M} R_m^i\right) = \liminf_{M\to\infty} \frac{1}{M}\sum_{m=1}^{M} \mathbb{E}_{s\pi}\left(R_m^i\right),$$

where R_m^i is the random variable for the payoff for player i at stage m, and where $\mathbb{E}_{s\pi}$ stands for expectation with respect to the initial state s and the joint strategy π.

With regard to a joint stationary strategy $x = (x^i)_{i\in N}$, we obtain more explicit formulas for the average reward. Let $r_s^i(x_s)$ denote the expected immediate payoff for player i in state s if the joint mixed action x_s is played. By definition, for the average reward of every player i we have

$$\gamma^i(x) = Q(x)\,r^i(x), \tag{3}$$

hence by (2) we also obtain

$$\gamma^i(x) = P(x)\,\gamma^i(x) \tag{4}$$

$$\gamma^i(x) = Q(x)\,r^i(x) = Q^2(x)\,r^i(x) = Q(x)\,\gamma^i(x). \tag{5}$$

Note that, as $Q(\cdot)$ is not necessarily continuous on the set X of joint stationary strategies, the same holds for the average reward γ^i of any player i. This possible discontinuity causes the main difficulties in the analysis of stochastic games with the average reward.

Nevertheless, every player i has a stationary best reply against any fixed joint stationary strategy of his opponents (cf. Hordijk et al. [1983]), i.e. for any $x^{-i} \in X^{-i}$ there exists an $x^i \in X^i$ such that $\gamma_s^i(x^i, x^{-i}) \geq \gamma_s^i(\pi^i, x^{-i})$ for all initial states $s \in S$ and for all strategies $\pi^i \in \Pi^i$.

For any player $i \in N$ and initial state $s \in S$, let

$$v_s^i := \inf_{\pi^{-i}\in\Pi^{-i}} \sup_{\pi^i\in\Pi^i} \gamma_s^i(\pi^i, \pi^{-i}). \tag{6}$$

Here v_s^i is called the minmax-level for player i in state s. Intuitively, this is the highest reward that player i can defend against any strategies of the other players if the initial state is s. Note that, against different joint strategies of players $-i$, player i may have to use different strategies to defend his minmax-level (as changing the order of infimum and supremum may yield a lower reward). It is known that the minmax-level of any player i satisfies

$$v_s^i = \min_{x_s^{-i} \in X_s^{-i}} \max_{x_s^i \in X_s^i} \sum_{t \in S} p_{s,(x_s^i, x_s^{-i})}(t)\, v_t^i, \tag{7}$$

which is an easy consequence of the definition of v_s^i and equality (4). Furthermore, by Thuijsman & Vrieze [1991] (their proof is given for only two players but directly extends to the n-player case), there always exists an initial state s in the set $\{t \in S|\, v_t^i = \min_{t' \in S} v_{t'}^i\}$ for which players $-i$ have a joint stationary strategy x^{-i} such that $\gamma_s^i(\pi^i, x^{-i}) \leq v_s^i$ for all strategies π^i for player i. In other words, the infimum in expression (6) is attained for state s at stationary strategies.

Equilibria. A joint strategy $\pi = (\pi^i)_{i \in N}$ is called a (Nash) ε-equilibrium for initial state $s \in S$, for some $\varepsilon \geq 0$, if

$$\gamma_s^i\left(\sigma^i, \pi^{-i}\right) \leq \gamma_s^i\left(\pi\right) + \varepsilon \qquad \forall \sigma^i \in \Pi^i,\ \forall i \in N,$$

which means that no player can gain more than ε by a unilateral deviation. Equivalently, for each player i, strategy π^i is an ε-best reply for initial state s against π^{-i}. If π is an ε-equilibrium for all initial states, then we call π an ε-equilibrium. It is clear from the definition of the minmax-level v that if π is an ε-equilibrium then $\gamma_s^i(\pi) \geq v_s^i - \varepsilon$ for each player i and each initial state $s \in S$.

Regarding general stochastic games, the famous game called the Big Match, which was introduced by Gillette [1957] and solved by Blackwell & Ferguson [1968], and the game in Sorin [1986] demonstrated that 0-equilibria do not necessarily exist with respect to the average reward. They made it clear, moreover, that history dependent strategies are indispensable for establishing ε-equilibria, for $\varepsilon > 0$.

For two-player stochastic games, Vieille [2000-a,b] managed to establish the existence of ε-equilibria, for all $\varepsilon > 0$. However, only little is known about n-player stochastic games, and it is unresolved whether they always possess ε-equilibria, for all $\varepsilon > 0$. This is probably the most challenging open problem in the field of stochastic games these days.

For the class of n-player aperiodic product-games, we will answer this question in the affirmative by proving the existence of 0-equilibria (cf. **Main Theorem 1**).

Here aperiodicity refers to an aperiodic transition structure, and will be given a precise definition later. Our proof is constructive by nature. The approach we present relies on so-called communicating states. The notion of communicating states we use is borrowed from the literature of Markov decision problems. We call two states of the game communicating if from either state, the players can move to the other state in finite time with probability 1 by choosing appropriate joint actions. We could call these states weakly communicating in order to emphasise that all players are needed to move between the states. The applicability of this form of weak communication is limited for general stochastic games for the simple reason that there is in general no guarantee that it is in all players' interest to follow the path between such communicating states. Nevertheless, for the class of product-games, due to their specific transition structure, this weak communication plays a fundamental role in the analysis, as we will demonstrate below.

Zero-sum games and optimality. In the development of stochastic games, a special role has been played by the class of zero-sum stochastic games, which are two-player stochastic games for which $r_s^2(a_s) = -r_s^1(a_s)$ (meaning that the sum of the payoffs is zero), for each state s and for each joint action a_s. In these games the two players have completely opposite interests. Mertens & Neyman [1981] showed that for such games $v^2 = -v^1$. Here $v := v^1$ is called the value of the game. They also showed that, if instead of using liminf one uses limsup in the definition of the average reward, one would find precisely the same value v. Thus, in a zero-sum game, player 1 wants to maximize his own reward, while at the same time player 2 tries to minimize player 1's reward. For simplicity, let $\gamma = \gamma^1$. A strategy π^1 for player 1 is called ε-optimal for initial state $s \in S$, for some $\varepsilon \geq 0$, if $\gamma_s(\pi^1, \pi^2) \geq v_s - \varepsilon$ for any strategy π^2 of player 2, while a strategy π^2 for player 2 is called ε-optimal for initial state $s \in S$ if $\gamma_s(\pi^1, \pi^2) \leq v_s + \varepsilon$ for any strategy π^1 of player 1. If π^1 or π^2 is ε-optimal for all initial states, then we call π^1 or π^2 an ε-optimal strategy. For simplicity, 0-optimal strategies are briefly called optimal. Mertens and Neyman [1981] proved that both players have ε-optimal strategies for any $\varepsilon > 0$, even though history dependent strategies are necessary for ε-optimality.

For the class of aperiodic zero-sum product-games, we will provide a proof that both players have stationary 0-optimal strategies (cf. **Main Theorem 2**). In addition, we analyse the structure of the value of these games.

The structure of the article. In section 2, we will present the main results and a detailed outline of the proofs, together with illustrative examples. The formal proofs are given in section 3. Finally, section 4 concludes with a short discussion on the case of periodic product-games.

2 The main results and a detailed outline of the proofs

2.1 The main results

For the class of product-games, we present the following result concerning existence of equilibria.

Main Theorem 1. *There exists a* 0*-equilibrium in every aperiodic* n*-player product-game.*

Aperiodicity refers to an aperiodic transition structure, and we will give a precise definition in section 2.2. In addition, for the special case of two-player zero-sum product-games, we show the existence of stationary solutions.

Main Theorem 2. *In two-player aperiodic zero-sum product-games, both players have a stationary 0-optimal strategy.*

As Main Theorem 2 will follow without much difficulty (cf. the end of section 3.1.3) from our extensive study of the minmax-levels in general n-player product-games, we will focus here on the proof of Main Theorem 1.

Now we provide a detailed outline of the proof of Main Theorem 1; the formal proof is given in section 3. The proof of Main Theorem 1 is constructive by nature. After some preliminary concepts and results in section 2.2, the first main step is to analyse the minmax-levels of the players in depth in section 2.3. Given the structural properties we achieve, we finally discuss the construction of 0-equilibria in section 2.4.

2.2 Preliminary concepts and results

Some of the contents of this section is very similar to the decomposition presented in Ross and Varadarajan [1991] for Markov decision problems (i.e. stochastic games with only one player).

Classification of states. First, we analyse the Markov transition structure Γ^i of each player i separately. We distinguish between two basic types of states in the state space S^i of Γ^i, based on the possibilities that player i has at his disposal to move between states.

A state $s^i \in S^i$ belongs to type 1 if it has the properties that (1) regardless the action of player i in state s^i, play leaves s^i with a positive probability, and (2) after leaving s^i through any action, the probability that player i ever comes back to s^i is strictly less than 1, regardless his strategy. Let $S^{\diamond i}$ denote the set of states of type 1 for player i. Hence, player i can only be in $S^{\diamond i}$ at finitely many stages, with probability 1.

On the other hand, a state $s^i \in S^i$ belongs to type 2 if it has the property that either (1) player i has an action in state s^i which keeps play in s^i with probability 1, or (2) player i has an action in state s^i such that, given play leaves s^i through this action, player i is able to come back to state s^i, possibly in a number of moves, with probability 1. Hence, given player i is in a state of type 2, player i can visit this state infinitely often, if he wishes so.

It is clear that each state in S^i belongs to precisely one type, and that there is always at least one state belonging to type 2.

Maximal communicating sets. Two states s^i_1 and s^i_2 of type 2 are said to communicate with each other, if, starting in state s^i_1, player i is able to go to state s^i_2 with probability 1, possibly in a number of moves, and vice versa. This relationship of communication is an equivalence relation (reflexive, symmetric and transitive) on the set of states of type 2. As such, it induces equivalence classes, which we call maximal communicating sets. So, by definition, two states of type 2 belong to the same maximal communicating set if and only if they communicate with each other.

Therefore, every maximal communicating set E^i has the properties that (1) player i can go from any state in E^i to any other state in E^i, possibly in a number of moves, without leaving E^i with probability 1 and (2) if player i decides to leave E^i, the probability that he ever comes back to E^i is strictly less than 1, regardless his strategy. The latter observation further implies that (3) the total number of times during the whole play that player i switches from a maximal communicating set to another one is finite with probability 1, regardless the initial state and player i's strategy; (4) there is always at least one amongst the maximal communicating sets which player i is unable to leave, i.e. there are no transitions to states outside; (5) regardless the initial state and player i's strategy, player i eventually settles, with probability 1, in one of his maximal communicating sets E^i, i.e. after finitely many stages, player i remains forever in E^i (it is possible that player i would be able to leave E^i with a different strategy).

Let $E^i_{k^i}$, where $k^i \in K^i$, denote the maximal communicating sets for player i.

Within the index-set K^i, we distinguish $K^{*i} \subset K^i$ for those maximal communicating sets which player i is not able to leave. In view of observation (4), K^{*i} is always nonempty. Further, let $K := \times_{i=1}^n K^i$. For any $k = (k^1, \ldots, k^n) \in K$, the product $E_k := \times_{i=1}^n E^i_{k^i}$ of the maximal communicating sets $E^i_{k^i}$, with $i = 1, \ldots, n$, is called a joint maximal communicating set.

In every state s^i of the communicating set $E^i_{k^i}$, for every $k^i \in K^i$, let $\bar{A}^i_{s^i}$ denote the set of those actions $a^i_{s^i} \in A^i_{s^i}$ which keep play in $E^i_{k^i}$ with probability 1. The sets $\bar{A}^i_{s^i}$ are clearly nonempty. We denote the mixed actions of player i on $\bar{A}^i_{s^i}$ by $\bar{X}^i_{s^i}$. For every state $s = (s^1, \ldots, s^n) \in S$, we also let $\bar{A}^i_s := \bar{A}^i_{s^i}$ and $\bar{X}^i_s := \bar{X}^i_{s^i}$.

Aperiodicity. A maximal cummunicating set $E^i_{k^i}$ of player i is called aperiodic, if there exists a number m such that, for any initial state in $E^i_{k^i}$, if player i plays a strategy that only uses completely mixed actions on $\bar{A}^i_{s^i}$ for all $s^i \in E^i_{k^i}$, then the probability that play at stage m is in state s^i is positive for all $s^i \in E^i_{k^i}$. Of course, this property is independent of the particular choice of this strategy of player i and remains valid for all stages larger than m. The notion of aperiodicity captures the idea that player i can be anywhere in $E^i_{k^i}$ with positive probability, after a certain finite number of moves.

For instance, if player i can move from any state $s^i \in E^i_{k^i}$ to every state $t^i \in E^i_{k^i}$ (thus including $t^i = s^i$) in one single move with a positive probability through an action in $\bar{A}^i_{s^i}$, then $E^i_{k^i}$ is obviously aperiodic. On the other hand, a trivial example of a periodic maximal communicating set is $E^i_{k^i} = \{1, \ldots, z\}$, with $z \geq 2$, when the transitions yield a cycle, i.e. player i's only choice in state $s^i < z$ is to move to state $s^i + 1$, and in state z to move to state 1.

We will call a product-game aperiodic if all maximal communicating sets, for all players, are aperiodic. From now on, we will only consider aperiodic product-games, with the exception of section 4.

Restricted games. Take an arbitrary aperiodic product-game and some $k = (k^1, \ldots, k^n) \in K$. By restricting the state space to $E_k \subset S$, and the action set of each player i in any state $s \in E_k$ to $\bar{A}^i_s$, we obtain a restricted game $\bar{G}_k$. Obviously, $\bar{G}_k$ is an aperiodic product-game itself, and the underlying Markov transition structure of any player i is obtained from Γ^i by restricting player i's state space to $E^i_{k^i}$, and by restricting player i's action set in any state $s^i \in E^i_{k^i}$ to $\bar{A}^i_{s^i}$.

These restricted games play a key role in the analysis of product-games, which is due to the following observation. As is pointed out above, regardless the initial state and the strategies of the players, each player i eventually settles in one of his maximal

$2,-2$ $\to (1,1)$	$0,0$ $\to (1,2)$
$0,0$ $\to (2,1)$	$0,0$ $\to (2,2)$

state $(1,1)$

$0,0$ $\to (1,2)$
$1,1$ $\to (2,2)$

state $(1,2)$

$3,-1$ $\to (1,1)$	$0,0$ $\to (1,2)$
$0,0$ $\to (3,1)$	$0,0$ $\to (3,2)$

state $(2,1)$

$-2,0$ $\to (1,2)$
$0,0$ $\to (3,2)$

state $(2,2)$

$0,0$ $\to (3,1)$	$0,0$ $\to (3,2)$

state $(3,1)$

$1,-1$ $\to (3,2)$

state $(3,2)$

Figure 1: Game of Example 1

communicating sets $E^i_{k^i}$. This yields a joint maximal communicating set $E_k = \times_{i=1}^n E^i_{k^i}$, which the players will never leave. Since actions outside $\bar{A}^i_s$, for any player i and in any state $s \in E_k$, would leave E_k with a positive probability, this means that such actions will be taken only finitely many times, with probability 1. Hence, with probability 1, play will eventually *settle* in a restricted game $\bar{G}_k$. The study of these restricted games is therefore of great importance.

Example 1. As an illustration, consider the product-game with two players given in figure 1. This is a game with six states. In each state, the actions of player 1 are represented by the rows, while the actions of player 2 by the columns. So each cell of each state corresponds to a pair of actions. In each cell, the two payoffs to the respective players are given in the upper-left corner, while the next state is indicated in the bottom-right corner. In this game all the transitions are pure, i.e. each transition probability distribution assigns probability 1 to a certain state.

The underlying Markov transition structure for player 1 is given by state space $S^1 = \{1, 2, 3\}$, action sets

$$A^1_1 = A^1_2 = \{1, 2\}, \quad A^1_3 = \{1\},$$

and transitions

$$p^1_{11} = (1, 0, 0),\ \ p^1_{12} = (0, 1, 0),\ \ p^1_{21} = (1, 0, 0),\ \ p^1_{22} = (0, 0, 1),\ \ p^1_{31} = (0, 0, 1).$$

So in state 1, player 1 can either stay or leave for state 2, from state 2 he can either go to state 1 or to state 3, while state 3 is absorbing. Regarding the classification of the states in S^1, both $E^1_I := \{1, 2\}$ and $E^1_{II} := \{3\}$ are maximal communicating sets. Moreover, they are both aperiodic. Since player 1 can leave E^1_I but not state 3, we have $K^1 = \{I, II\}$ and $K^{*1} = \{II\}$. As for the actions which keep play in these maximal communicating sets, we obtain $\bar{A}^1_1 = \{1, 2\}$, $\bar{A}^1_2 = \{1\}$, $\bar{A}^1_3 = \{1\}$.

The underlying Markov transition structure for player 2 is given by state space $S^2 = \{1, 2\}$, action sets $A^2_1 = \{1, 2\}$, $A^2_2 = \{1\}$, and transitions $p^2_{11} = (1, 0)$, $p^2_{12} = (0, 1)$, $p^2_{21} = (0, 1)$. Further, both $E^2_I := \{1\}$ and $E^2_{II} := \{2\}$ are aperiodic maximal communicating sets, with $K^2 = \{I, II\}$ and $K^{*2} = \{II\}$, and $\bar{A}^2_1 = \bar{A}^2_2 = \{1\}$.

As all maximal communicating sets are aperiodic, we may conclude that the game is aperiodic as well. Finally, we have $K = \{I, II\}^2$, which yields four joint maximal communicating sets and four corresponding restricted games. For example, $E_{(I,I)} = \{1, 2\} \times \{1\}$, and the corresponding restricted game $\bar{G}_{(I,I)}$ consists of cells $(1, 1)$ and $(2, 1)$ in state $(1, 1)$ and cell $(1, 1)$ in state $(2, 1)$.

2.3 The structure of the minmax-levels

We refer to section 3.1 for the formal discussion. Recall that the rewards corresponding to a 0-equilibrium are always individually rational, i.e. the equilibrium reward for each player i from any initial state s is at least his minmax-level v^i_s. It is therefore essential, for the construction of 0-equilibria, to learn more about the minmax-levels of the players in these product-games.

The analysis of the minmax-levels is split into three sub-steps. In section 2.3.1, we study the minmax-levels of the players in the restricted-games. Then, in section 2.3.2., we introduce the notion of simple product-games and explore the structure of their minmax-levels. Finally, in section 2.3.3, by combining the first two sub-steps, we are able to demonstrate the most essential structural properties of the minmax-levels in general product-games.

2.3.1 The minmax-level $\bar{v}_k$ of a restricted game $\bar{G}_k$

Consider a restricted game $\bar{G}_k$ corresponding to the joint maximal communicating set E_k, for some $k = (k^1, \ldots, k^n) \in K$. Let $\bar{v}^i_{k,s}$ denote each player i's minmax-level in $\bar{G}_k$

for initial state $s \in E_k$. By using the aperiodicity of each $E^i_{k^i}$ and that each player i can go from any state in $E^i_{k^i}$ to any other one in $E^i_{k^i}$, we will be able to show that any player i's minmax-level $\bar{v}^i_{k,s}$ in $\bar{G}_k$ is constant on the whole state space E_k of $\bar{G}_k$ (cf. lemma 1). This means that the players are indifferent between the states in E_k, as far as their minmax-levels in $\bar{G}_k$ are concerned. It will also follow that for any player i, players $-i$ have a joint stationary strategy which guarantees within $\bar{G}_k$ that player i's reward from any initial state $s \in E_k$ is at most his minmax-level $\bar{v}^i_{k,s}$. In other words, the infimum in expression (6) is attained at joint stationary strategies, for all restricted games. This will become important later, as our ultimate goal is the existence of 0-equilibria, which do not allow even small positive error terms.

As an illustration, we now revisit the game in example 1. Take first the restricted game $\bar{G}_{(I,I)}$, consisting of cells $(1,1)$ and $(2,1)$ in state $(1,1)$ and cell $(1,1)$ in state $(2,1)$. Let us examine player 1's minmax-level $\bar{v}^1_{(I,I)}$ in $\bar{G}_{(I,I)}$. In $\bar{G}_{(I,I)}$, it is only player 1 who has a choice and only in state $(1,1)$. By choosing the first action, he receives payoff 2, while by playing the second one he receives payoff 0 and subsequently payoff 3 in state $(2,1)$ before returning to state $(1,1)$. As the second action gives payoff $3/2$ on average, we may conclude that he cannot do better than to keep on choosing action 1 in state $(1,1)$. Hence, for both initial states in $\bar{G}_{(I,I)}$, player 1's minmax-level $\bar{v}^1_{(I,I)}$ is 2, whereas, for similar reasons, player 2's minmax-level $\bar{v}^2_{(I,I)}$ is -2. Thus, both minmax-levels are constant on the state space $E_{(I,I)} = \{(1,1),(2,1)\}$ of $\bar{G}_{(I,I)}$.

Now consider the restricted game $\bar{G}_{(I,II)}$, consisting of the whole state $(1,2)$ and the upper cell in state $(2,2)$. By using similar arguments, player 1's minmax-level $\bar{v}^1_{(I,II)}$ is 0, and player 2's minmax-level $\bar{v}^2_{(I,II)}$ is also 0 for both initial states in $\bar{G}_{(I,II)}$.

Finally, the restricted games $\bar{G}_{(II,I)}$ and $\bar{G}_{(II,II)}$ are both trivial, i.e. they consist of one single state and one action for both players. In $\bar{G}_{(II,I)}$, both minmax-levels $\bar{v}^1_{(II,I)}$ and $\bar{v}^2_{(II,I)}$ are equal to 0, whereas in $\bar{G}_{(II,II)}$, player 1's minmax-level $\bar{v}^1_{(II,II)}$ equals 1 and player 2's minmax-level $\bar{v}^2_{(II,II)}$ equals -1.

2.3.2 The minmax-levels in simple product-games

We now examine a special class of product-games. We call a product-game G simple if it holds for all restricted games $\bar{G}_k$ that, for all players i, all payoffs to player i within $\bar{G}_k$ are equal. This way, all restricted games are trivial.

For the minmax-levels of the players in simple product-games, we will derive several results. These results will be illustrated throughout this section by the following example.

$2,-2$ $\to (1,1)$	$0,0$ $\to (1,2)$
$2,-2$ $\to (2,1)$	$0,0$ $\to (2,2)$

state $(1,1)$

$0,0$ $\to (1,2)$
$0,0$ $\to (2,2)$

state $(1,2)$

$2,-2$ $\to (1,1)$	$0,0$ $\to (1,2)$
$0,0$ $\to (3,1)$	$0,0$ $\to (3,2)$

state $(2,1)$

$0,0$ $\to (1,2)$
$0,0$ $\to (3,2)$

state $(2,2)$

$0,0$ $\to (3,1)$	$0,0$ $\to (3,2)$

state $(3,1)$

$1,-1$ $\to (3,2)$

state $(3,2)$

Figure 2: Game of Example 2

Example 2: Consider the simple product-game G with two players given in figure 2. The underlying Markov transition structures are identical to those in example 1. Hence, this game is aperiodic as well. Actually, this game is obtained from the game in example 1 by replacing all payoffs for player 1 by 2 and for player 2 by -2 in the restricted game $\bar{G}_{(I,I)}$, and all payoffs for either player by 0 in the restricted game $\bar{G}_{(I,II)}$. Hence, the only possible pair of payoffs is $(2,-2)$ in $G_{(I,I)}$, and $(0,0)$ in $\bar{G}_{(I,II)}$. Finally, for restricted game $\bar{G}_{(II,I)}$, the only possible pair of payoffs remained $(0,0)$, while in $\bar{G}_{(II,II)}$, it remained $(1,-1)$. So, the game is simple, indeed. In fact, this is a zero-sum game, but we will not pay much attention to this aspect.

Let us examine the players' minmax-levels in G. For player 1, we will argue that

$$v^1_{(1,1)} = v^1_{(1,2)} = v^1_{(2,1)} = v^1_{(2,2)} = v^1_{(3,2)} = 1, \quad v^1_{(3,1)} = 0.$$

Player 1's minmax-level is clearly 0 for initial state $(3,1)$, in view of player 2's first action. Now consider an arbitrary other initial state $s \in S - \{(3,1)\}$. By moving to his second state, player 2 can always make sure that player 1's reward is at most 1. On the other hand, player 1 can guarantee reward 1 for state s by the pure stationary strategy x^1 defined as

$$x^1_{(1,1)} = (1,0),\ x^1_{(1,2)} = (0,1)\,,\ x^1_{(2,1)} = (1,0),\ x^1_{(2,2)} = (0,1),\ x^1_{(3,1)} = x^1_{(3,2)} = (1).$$

Hence, player 1's minmax-level equals 1 for all $s \in S - \{(3,1)\}$, indeed. We similarly find that $v^2_s = -v^1_s$ for all $s \in S$.

Given this example, we would like now to explain and to illustrate the most important results that we will achieve regarding the minmax-levels of simple aperiodic product-games. The explanation below of each of these results is given for an arbitrary simple aperiodic product-game G, which is then followed by an illustration with the help of example 2.

Observation A (cf. lemma 2). In any state $s \in S$, even if any player i had a "solitary move", i.e. he could play an action while every other player j remains in the same state s^j, he cannot improve on his minmax-level v^i in expectation. Similarly, players $-i$ cannot decrease player i's minmax-level v^i in expectation by executing a solitary move. (This important result heavily relies on the aperiodicity of the product-game, and would fail in general, cf. section 4.) Consider in example 2, for instance, state $(2,1)$ and a solitary move for player 1. Now given player 2 stays in state 1, player 1's first action yields state $(1,1)$, while the second one state $(3,1)$. As $v^1_{(2,1)} = v^1_{(1,1)} = 1$ and $v^1_{(3,1)} = 0$, player 1 is indeed unable to improve on his minmax-level by such a solitary move.

Observation B (cf. lemma 4). On any joint maximal communicating set E_k, each player i's minmax-level v^i is constant. In example 2, for instance, on both states of $E_{(I,I)}$, player 1's minmax-level is 1.

Observation C (cf. lemma 5). If player i is in a state of type 2, then the actions of the corresponding restricted game provide the best possible transitions with respect to the expected minmax-level v^i, regardless the actions of the opponents. Somewhat similarly, if players $-i$ all play actions in a restricted game, then player i's minmax-level cannot increase in expectation. In example 2, consider, for instance, player 1 in state $(2,1)$. If player 1 plays his first action (the action of the restricted game $\bar{G}_{(I,I)}$) then his minmax-level will remain 1, regardless the action chosen by player 2. Therefore, action 2 can never be better than action 1 for player 1, with respect to player 1's expected minmax-level after transition.

Observation D (cf. lemma 6). In any restricted game $\bar{G}_k$, if any player i's unique reward in $\bar{G}_k$ is strictly less than his minmax-level v^i on E_k (which is constant on E_k, cf.

observation B above), then player i is able to leave $\bar{G}_k$ (or actually the set of states E_k) in a satisfactory way. More precisely, player i has a state $s^i \in E^i_{k^i}$ and a corresponding pure "exit" action $a^i_{s^i}$ such that by playing $a^i_{s^i}$ in any state $t \in E_k$ with $t^i = s^i$, play leaves E_k with a positive probability and at the same time player i's minmax-level cannot decrease in expectation, regardless the actions of the opponents. As player i can move to s^i from any state in $E^i_{k^i}$, he is always able on his own to make play leave such an "unfavorable" E_k. Similarly, if player i's unique reward in $\bar{G}_k$ is strictly larger than his minmax-level v^i on E_k, then players $-i$ are able to leave $\bar{G}_k$ in an analogous manner. As an illustration, in example 2, consider the restricted game $\bar{G}_{(I,II)}$, in which player 1's unique reward is 0 while his minmax-level v^1 is 1. Indeed, player 1 can leave $\bar{G}_{(I,II)}$ by moving to state $(2,2)$ (or actually to state $2 \in E^1_{(I,II)}$) and playing his second action there. Note that by playing this action, player 1's minmax-level v^1 remains unchanged. Similarly, player 2 is unsatisfied with the restricted game $\bar{G}_{(I,I)}$, as his unique reward is -2, which is strictly less than his minmax-level on $E_{(I,I)}$, which equals -1. Notice that player 2 can leave $\bar{G}_{(I,I)}$ by playing his second action, and by doing so, regardless whether play is in state $(1,1)$ or in state $(2,1)$, and regardless the action chosen by player 1, the minmax-level v^2 of player 2 cannot decrease (as -1 is his lowest minmax-level in the whole game).

2.3.3 The minmax-levels in general product-games

This section is devoted to the analysis of the minmax-levels of the players in the context of general aperiodic product-games. Take an arbitrary aperiodic product-game G. As we know from section 2.3.1, the minmax-level $\bar{v}^i_k$ of each player i in any restricted game $\bar{G}_k$ is constant on the whole state space E_k of $\bar{G}_k$. Let $\widetilde{G}$ denote the simple aperiodic product-game which is derived from G by replacing each player i's payoffs in any restricted game $\bar{G}_k$ by his minmax-level $\bar{v}^i_k$. Let w^i_s denote player i's minmax-level in $\widetilde{G}$ from initial state s. For an illustration, we refer to the game in example 2 (which is now game $\widetilde{G}$ with minmax-levels w), which is obtained exactly by this very procedure from the game in example 1 (which is now game G with minmax-levels v). Recall for this example that $w^1_{(3,1)} = w^2_{(3,1)} = 0$ while $w^1_s = 1$ and $w^2_s = -1$ for all $s \in S - \{(3,1)\}$.

The transformation above of G into $\widetilde{G}$ is of course very natural, and we will be able to show in general that the minmax-levels of the players remain unchanged under this transformation (cf. lemma 7), i.e. $w^i_s = v^i_s$ for all players i and all initial states $s \in S$.

Let us explain in detail why $w^i_s \geq v^i_s$ holds in general. For this it is sufficient to show that players $-i$ have a joint stationary strategy x^{-i} which guarantees in the original game G that player i's (expected) reward is not more than w_s^i for any initial state s. In our illustrative game in example 1, for $i = 2$, one can take the stationary strategy $y^1 (= x^{-2})$ for player 1 defined as

$$y^1_{(1,1)} = (1,0),\ y^1_{(1,2)} = (0,1)\ ,\ y^1_{(2,1)} = (1,0),\ y^1_{(2,2)} = (0,1),\ y^1_{(3,1)} = y^1_{(3,2)} = (1),$$

which guarantees in G that player 2's reward is not more than w_s^2 for all initial states $s \in S$. Now we turn back to the general case, but we will indicate between brackets the corresponding events in this example.

As is pointed out in observation B in section 2.3.2, w_s^i is a constant w_k^i on each E_k. Thus, we obtained two constants $\bar{v}_k^i$ and w_k^i for any player i in any joint maximal communicating set E_k. Now, players $-i$ should use a joint stationary strategy x^{-i} which prescribes to play, roughly speaking, as follows:

1. In any joint maximal communicating set E_k in which $\bar{v}_k^i \le w_k^i$, players $-i$ should play a joint stationary strategy in $\bar{G}_k$ which guarantees in $\bar{G}_k$ that player i's reward is not more than $\bar{v}_k^i$. Such a joint stationary strategy exists, as is discussed in section 2.3.1. (In our example, this happens with y^1 in $E_{(I,I)}$, $E_{(II,I)}$ and $E_{(II,II)}$.)

2. In any joint maximal communicating set E_k in which $\bar{v}_k^i > w_k^i$, players $-i$ should leave E_k, as is discussed in observation D (with respect to the minmax-level w^i of the game $\tilde{G}$) in section 2.3.2 above. This can be done in a stationary way by moving to the joint states where exit can take place and then playing the joint "exit" actions. (In our example, this happens with y^1 in $E_{(I,II)}$.)

3. In states in which at least one player is in a state of type 1, players $-i$ should play joint mixed actions which take care that the value of w^i cannot increase in expectation after transition. Such joint actions obviously exist, as w^i is the minmax-level of game $\tilde{G}$. (In our example, there are no such states.)

We will now argue that x^{-i} guarantees in G that player i cannot receive a reward higher than w_s^i for any initial state s, as desired. Take an arbitrary stationary strategy x^i for player i and an arbitrary initial state s. Consider the joint stationary strategy (x^i, x^{-i}). First notice that the value of w^i cannot increase in expectation during play. For case 3 it is immediate. On the other hand, in cases 1 and 2, players $-i$ always use actions of the corresponding restricted game or they leave E_k with a joint "exit" action. And indeed, in both cases, as is discussed in observations C and D (with respect to the minmax-level w^i of the game $\tilde{G}$) in section 2.3.2, w^i cannot increase in expectation.

As we know, with respect to (x^i, x^{-i}) and initial state s, play eventually settles, with probability 1, in a restricted game. Let ξ denote the random variable for the index of this restricted game (so play settles in restricted game $\bar{G}_\xi$). Since w^i cannot increase in expectation during play, it follows that w^i_ξ is then at most w^i_s in expectation, i.e. $\mathbb{E}_{s,(x^i,x^{-i})}(w^i_\xi) \leq w^i_s$. Because E_ξ can only fall under case 1, and not case 2 due to the "exit" actions, we have for player i's reward (in G) that

$$\gamma^i_s(x^i, x^{-i}) \leq \mathbb{E}_{s,(x^i,x^{-i})}(\bar{v}^i_\xi) \leq \mathbb{E}_{s,(x^i,x^{-i})}(w^i_\xi) \leq w^i_s.$$

As x^i was arbitrary and player i has a stationary best reply to x^{-i}, we conclude that x^{-i} guarantees in the original game G that player i's reward is not more than w^i_s for any initial state s. This implies $w^i_s \geq v^i_s$, as desired.

One can similarly show that $w^i_s \leq v^i_s$, yielding $w^i_s = v^i_s$ for all players i and all initial states $s \in S$. This has important consequences.

First, x^{-i} thus guarantees in the original game G that player i's reward is not more than v^i_s for any initial state s. This implies, in the context of two-player aperiodic zero-sum product-games, that the stationary strategy x^{-1} of player 2 (as $-1 = \{2\}$) guarantees that player 1's reward is not more than v^1_s for any initial state s. Hence, x^{-1} is 0-optimal for player 2. One similarly finds that x^{-2} is 0-optimal for player 1. Thus, both players have stationary 0-optimal strategies (cf. the end of section 3.1.3), which proves Main Theorem 2.

Second, the structural properties (i.e. observations A,B,C and D) that we achieved in section 2.3.2 for the minmax-levels of simple product-games are now applicable to all product-games (cf. corollary 8). With this knowledge on the minmax-levels, we are now sufficiently prepared to tackle the problem of the existence of 0-equilibria.

2.4 The construction of 0-equilibria in general product-games

Take an arbitrary aperiodic product-game G. In this section, we will show that there exists a 0-equilibrium in G, as is claimed by Main Theorem 1. The construction will make extensive use of the results we obtained for the minmax-levels of the players.

The first step again is to examine the existence of equilibria in the restricted games. We will show for any restricted game $\bar{G}_k$ (cf. lemma 10) that there exists a 0-equilibrium σ_k in $\bar{G}_k$ such that the corresponding rewards are independent of the initial state and all the continuation rewards remain unchanged with probability 1 during the whole play. More precisely, if σ_k induces reward $z^i_k \in \mathbb{R}$ for some player i, then $\bar{\gamma}^i_s(\sigma_k[h]) = z^i_k$ holds for every state $s \in E_k$ and for every history h with a positive probability of occurrence with respect to σ_k. Here $\bar{\gamma}^i$ denotes player i's average reward in the restricted game $\bar{G}_k$. So if no player deviates, every player i's future expectations remain z^i_k during the whole play. This will guarantee that no player will change his mind and decides to leave E_k just because a certain history took place. Since σ_k is a 0-equilibrium in $\bar{G}_k$, we have $z_k^i \geq \bar{v}_k^i$ for all players i, where $\bar{v}_k^i$ is the minmax-level of player i in $\bar{G}_k$ ($\bar{v}_k^i$ is constant on E_k, as we know from section 2.3.1).

The idea of the proof that such a 0-equilibrium σ_k exists in $\bar{G}_k$ is simple. Notice first that, as the state space E_k of $\bar{G}_k$ is a joint maximal communicating set and each $E_{k^i}^i$ is aperiodic, the players can move from any state in E_k to any other one in E_k, possibly in a number of steps, if they wish so. Thus, the set of feasible rewards (i.e. the rewards that can be obtained by some joint strategy) is the same from any initial state in E_k. Moreover, we also know that each player i's minmax-level $\bar{v}_k^i$ in $\bar{G}_k$ is constant on E_k. Hence, this game situation in $\bar{G}_k$ is fairly similar to an ordinary repeated game, and the construction of such a σ_k is then a simple task by applying ideas and arguments taken from the well-known Folk-theorem for repeated games.

Hence, we may fix a 0-equilibrium σ_k with some reward z_k in every restricted game $\bar{G}_k$. Fix further, for every player i, a joint stationary strategy y^{-i} for players $-i$ which guarantees in the original game G that player i's reward is not more than v_s^i for any initial state s. Such joint stationary strategies exist, as is discussed in the first conclusion at the end of section 2.3.3 (where we used the notation x^{-i} for such a joint stationary strategy).

We are now ready to discuss the proof of Main Theorem 1, which claimed the existence of a 0-equilibrium η in G. The proof is constructive by nature. The main body of the proof is to construct a joint strategy π with important properties, amongst others that:

property (1): the rewards for π are individually rational, i.e. $\gamma_s^i(\pi) \geq v_s^i$ for all initial states $s \in S$ and for all players i;

property (2): no player i has an incentive to deviate from π^i inside the support of π^i, i.e. by redistributing the probabilities over the actions to which π^i would assign a positive probability (such deviations are difficult to detect, as player i still chooses actions which have positive probability according to π^i).

The joint strategy η will then, roughly speaking, prescribe to play as follows: the players play the joint strategy π as long as no player i deviates from π^i by playing an action on which π^i puts probability zero. If player i deviates in such a way, then from the next state, say state s, players $-i$ switch to the joint strategy y^{-i} and play it for the rest of the time. By doing so, they push down player i's reward to a level of at most v_s^i. As π induces individually rational rewards, y^{-i} acts as a threat strategy, which forces player i to follow the prescriptions of π^i. The use of such threat strategies for the construction of equilibria is standard in the theory of stochastic games.

Now the remaining task is to construct π. For simplicity, suppose that there are no states of type 1. As we know, states of type 1 have a transient nature, and while those states will cause no fundamental difficulties in the formal proofs, they do involve some technicalities (we would have to define two additional auxiliary games). So, by this assumption, S splits up into joint maximal communicating sets. Now π prescribes for the players to play as follows. On each joint maximal communicating set E_k, we compare the rewards z_k that the players could obtain as an equilibrium reward inside $\bar{G}_k$ with the minmax-levels of the players in the original game G, which are some constant v_k on E_k according to observation B in section 2.3.2 (and the final conclusion of section 2.3.3 as well). Now, if play enters an ergodic set E_k in which $z_k^i \geq v_k^i$ for all players i, then $\bar{G}_k$ is a "satisfactory" restricted game and π prescribes to switch to σ_k and collect reward z_k. On the other hand, if play enters an ergodic set E_k in which $z_k^i < v_k^i$ for some player i while $z_k^j \geq v_k^j$ for all players $j < i$, then player i is unsatisfied with $\bar{G}_k$, and π lets accordingly player i leave $\bar{G}_k$ as is given in observation D in section 2.3.2 (cf. the final conclusion of section 2.3.3 as well). As long as player i has not made the exit yet, players $-i$ will simply play a joint stationary strategy in the restricted game $\bar{G}_k$ which guarantees that player i cannot receive more than $\bar{v}_k^i$ inside $\bar{G}_k$. Such a strategy exists as is mentioned in section 2.3.1. As $\bar{v}_k^i \leq z_k^i < v_k^i$, by doing so, players $-i$ force player i to eventually leave $\bar{G}_k$.

Notice that π only prescribes actions within the restricted games, except for the exit actions. Therefore, it will follow from observations C and D in section 2.3.2 (and the final conclusion of section 2.3.3) that the joint strategy π satisfies properties (1) and (2) above, as desired. Given π, the construction of the 0-equilibrium η is complete. This concludes the outline of the proof of Main Theorem 1.

We wish to add that it remains unclear whether 0-equilibria always exist within the class of stationary strategies. This question is already challenging in the situation when each player i's state space S^i is just one maximal communicating set (precisely the situation we have in a restricted game), meaning that S consists of one joint maximal communicating set. While there are indications that stationary equilibria may exist, for example that all minmax-levels are constant on the whole state space S, it is still not evident how one should get a grip on the problem.

Finally, let us revisit example 1. As we know, the minmax-levels of this game coincide with the minmax-levels of the game in example 2, hence

$$v^1_{(1,1)} = v^1_{(1,2)} = v^1_{(2,1)} = v^1_{(2,2)} = v^1_{(3,2)} = 1, \qquad v^1_{(3,1)} = 0,$$

while $v^2 = -v^1$. Now consider the pure stationary strategy x^1 for player 1 defined as

$$x^1_{(1,1)} = (1,0),\ x^1_{(1,2)} = (0,1)\,,\ x^1_{(2,1)} = (1,0),\ x^1_{(2,2)} = (0,1),\ x^1_{(3,1)} = x^1_{(3,2)} = (1),$$

and the pure stationary strategy x^2 for player 2 given as

$$x^2_{(1,1)} = x^2_{(2,1)} = (0,1), \quad x^2_{(3,1)} = (1,0), \quad x^2_{(1,2)} = x^2_{(2,2)} = x^2_{(3,2)} = (1).$$

This pair (x^1, x^2) actually could play the role of π in this example. Indeed, each of the joint maximal communicating sets $E_{(II,I)}$ and $E_{(II,II)}$ is "satisfactory" to the players, and trivially, (x^1, x^2) lets the players play a 0-equilibrium in each of the restricted games $\bar{G}_{(II,I)}$ and $\bar{G}_{(II,II)}$. On the other hand, as we know, $\bar{G}_{(I,I)}$ is unsatisfactory to player 2 and $\bar{G}_{(I,II)}$ is unsatisfactory to player 1, and x^2 leaves $E_{(I,I)}$ while x^1 leaves $E_{(I,II)}$ accordingly. Notice that we need no threat strategies here, so (x^1, x^2) is a 0-equilibrium.

3 The formal proof of Main Theorems 1 and 2

In this section, we provide a formal proof for Main Theorems 1 and 2. We will focus on Main Theorem 1, as Main Theorem 2 will follow (cf. the end of section 3.1.3) along the way without major additional difficulties.

For the main ideas and the intuition behind the proofs, we also refer to the discussion in section 2. Recall the classification of states for each player's Markov transition structure from section 2.2.

3.1 The structure of the minmax-levels

3.1.1 The minmax-levels of the restricted games

Let $k = (k^1, \ldots, k^n) \in K$. As in section 2.3.1, by restricting the state space to $E_k = E^1_{k^1} \times \cdots \times E^n_{k^n}$ and the action set of each player i to $\bar{A}^i_s$ in all states $s \in E_k$, we obtain a restricted product-game, which we denote by $\bar{G}_k$. Let $\bar{v}^i_{k,s}$ denote the minmax-level of player i in $\bar{G}_k$ for initial state $s \in E_k$.

Lemma 1 *Let G be an arbitrary aperiodic product-game and consider the restricted game $\bar{G}_k$, for any $k = (k^1, \ldots, k^n) \in K$, and an arbitrary player i. Then, the minmax-level $\bar{v}^i_k$ of any player i in $\bar{G}_k$ is constant, i.e. $\bar{v}^i_{k,s} = \bar{v}^i_{k,t}$ for all states $s, t \in E_k$. Moreover, in $\bar{G}_k$, players $-i$ have a joint stationary strategy x^{-i} which guarantees that* *player i's reward from any initial state $s \in E_k$ is at most his minmax-level $\bar{v}^i_{k,s}$, i.e. for all strategies π^i for player i in $\bar{G}_k$ we have*

$$\bar{\gamma}^i_s(\pi^i, x^{-i}) \leq \bar{v}^i_{k,s},$$

where $\bar{\gamma}$ denotes the average reward for the game $\bar{G}_k$.

Proof. Consider such a restricted game $\bar{G}_k$ and a player i. Let $\alpha^i := \min_{t \in E_k} \bar{v}^i_{k,t}$. As is mentioned in the introduction, by applying Thuijsman & Vrieze [1991] for the game $\bar{G}_k$, there exists a state $s' \in \{t \in E_k | \bar{v}^i_{k,t} = \alpha^i\}$ for which players $-i$ have a joint stationary strategy x^{-i} such that for all strategies π^i for player i in $\bar{G}_k$ we have

$$\bar{\gamma}^i_{s'}(\pi^i, x^{-i}) \leq \bar{v}^i_{k,s'} = \alpha^i.$$

Let Z denote the set of all those states $s \in \{t \in E_k | \bar{v}^i_{k,t} = \alpha^i\}$ for which this x^{-i} satisfies for all strategies π^i for player i in $\bar{G}_k$ that

$$\bar{\gamma}^i_s(\pi^i, x^{-i}) \leq \alpha^i.$$

Let x^i be a completely mixed stationary strategy in $\bar{G}_k$ for player i. For the joint stationary strategy (x^i, x^{-i}), take an arbitrary ergodic set $F \subset E_k$ which is reached from some initial state $s \in Z$ with a positive probability. Then, by the definition of x^{-i}, we have $F \subset Z$. Due to the aperiodicity of $E^i_{k^i}$ and the definition of x^i, it holds that if $u \in F$ then $(t^i, u^{-i}) \in F$ for all states $t^i \in E^i_{k^i}$. Thus, the ergodic set F must be of the form $F = \widetilde{F} \times E^i_{k^i}$ for some non-empty

$$\widetilde{F} \subset E^{-i}_{k^{-i}} = \times_{j \in N - \{i\}} E^j_{k^j}.$$

Define a joint stationary strategy y^{-i} for players $-i$ in $\bar{G}_k$ as follows: let $y^{-i}_t = x^{-i}_t$ for all $t \in F$ and let y^{-i}_t be an arbitrary completely mixed action on $\bar{A}^{-i}_t$ for all $t \in (E_k - F)$. Now, y^{-i} satisfies the following two properties, regardless the initial state and player i's strategy in $\bar{G}_k$.

Property 1: play eventually visits F. This follows from the observation that players $-i$ eventually visit $\widetilde{F}$, due to the aperiodicity of $E^j_{k^j}$ for all $j \neq i$ and the choice of y^{-i} outside F.

Property 2: once play reaches F, it will never leave it. This is so because F was closed with respect to (x^i, x^{-i}) and y^{-i} equals x^{-i} on F.

In view of

$$F \subset Z \subset \{t \in E_k | \bar{v}^i_{k,t} = \alpha^i\},$$

property 1 implies that the minmax-level $\bar{v}^i_k$ of player i in $\bar{G}_k$ equals the constant α^i, so the first part of the lemma follows.

We finally show that y^{-i} satisfies the second part of the lemma. Let y^i be a stationary best reply of player i against y^{-i} in $\bar{G}_k$, and consider the joint stationary strategy (y^i, y^{-i}). Suppose U is an ergodic set with respect to (y^i, y^{-i}). Then, by properties 1 and 2, we have $U \subset F$, and hence $U \subset Z$ as well. Because y^{-i} equals x^{-i} on F, it follows for all $u \in U$ that

$$\bar{\gamma}^i_u(y^i, y^{-i}) \leq \alpha^i.$$

Since play eventually reaches an ergodic set, we conclude

$$\bar{\gamma}^i_u(y^i, y^{-i}) \leq \alpha^i = \bar{v}^i_{k,u}$$

for all initial states $u \in E_k$. For y^i is a best reply to y^{-i}, we proved that y^{-i} satisfies the second part of the lemma. ■

We remark that if $k^i \in K^{*i}$ for all players i, then for all initial states in E_k, the restricted game $\bar{G}_k$ is strategically equivalent to the original game G, and therefore $\bar{v}^i_{k,s} = v^i_s$ for all players i and for all states $s \in E_k$. In view of the previous lemma, minmax-level v^i of any player i in the original game G is also constant on such an E_k.

3.1.2 The minmax-levels in simple product-games

We call a product-game G simple if it holds within any restricted game $\bar{G}_k$ for any player i that, all payoffs to player i are equal, i.e. for any $k \in K$ and for any player i, we have $r^i_s(a_s) = r^i_s(b_s)$ for any state $s \in E_k$ and for any joint actions $a_s, b_s \in \bar{A}_s$. Hence, in simple product-games, all restricted games are trivial.

The following lemma deals with "solitary moves" of the players, as is described in property A in section 2.3.2.

Lemma 2 *Let G be a simple aperiodic product-game. Take an arbitrary player i and a state $s = (s^1, \ldots, s^n) \in S$.*

(1) For any action $a^i_s \in A^i_s$ of player i, it holds that

$$\sum_{t^i \in S^i} p^i_{s^i a^i_s}(t^i)\, v^i_{(t^i, s^{-i})} \leq v^i_s.$$

(2) For any joint action $a^{-i}_s \in A^{-i}_s$ of players $-i$, it holds that

$$\sum_{t^{-i} \in S^{-i}} p^{-i}_{s^{-i} a^{-i}_s}(t^{-i})\, v^i_{(t^{-i}, s^i)} \geq v^{-i}_s.$$

Proof. We only show part (1) of the lemma; part (2) can be proven similarly. Take an arbitrary state $s = (s^1, \ldots, s^n) \in S$, a player i and an action a^i_s of player i in state s. Let $\varepsilon > 0$. The idea of the proof is as follows. We construct a joint strategy (σ^i, π^{-i}) for initial state s and another joint strategy (π^i, σ^{-i}) for all initial states of the form (t^i, s^{-i}), with $t^i \in S^i$; for other initial states the joint strategies are arbitrary. These joint strategies will have the properties:

Property (a): with respect to (σ^i, π^{-i}) and initial state s, player i's reward is at most his minmax-level up to ε, i.e. $\gamma^i_s(\sigma^i, \pi^{-i}) \leq v^i_s + \varepsilon$.

Property (b): with respect to (π^i, σ^{-i}) and any initial state of the form (t^i, s^{-i}), player i's reward is at least his minmax-level up to ε, i.e. $\gamma^i_{(t^i,s^{-i})}(\pi^i, \sigma^{-i}) \geq v^i_{(t^i,s^{-i})} - \varepsilon$.

Property (c): player i's expected reward is the same with respect to the following two ways of playing from initial state s: (i) according to (σ^i, π^{-i}) and (ii) player i first executes the solitary move a^i_s in state s, by which play moves to a state of the form (t^i, s^{-i}), and subsequently from state (t^i, s^{-i}) the players start playing (π^i, σ^{-i}). Formally,

$$\gamma^i_s(\sigma^i, \pi^{-i}) = \sum_{t^i \in S^i} p^i_{s^i a^i_{s^i}}(t^i) \cdot \gamma^i_{(t^i,s^{-i})}(\pi^i, \sigma^{-i}).$$

Properties (a) and (b) will follow immediately from the definitions of the strategies, cf. step 1 below. On the other hand, property (c) will be implied by the observation, cf. step 2 below, that play will settle in any restricted game $\bar{G}_k$ with equal probabilities with respect to both ways of playing as mentioned in property (c). At this point, it is essential that the game is simple and therefore all payoffs within any $\bar{G}_k$ are identical.

It follows from properties (a), (b) and (c) that

$$\begin{aligned} v^i_s + \varepsilon &\geq \gamma^i_s(\sigma^i, \pi^{-i}) \qquad (8) \\ &= \sum_{t^i \in S^i} p^i_{s^i a^i_{s^i}}(t^i) \cdot \gamma^i_{(t^i,s^{-i})}(\pi^i, \sigma^{-i}) \\ &\geq \sum_{t^i \in S^i} p^i_{s^i a^i_{s^i}}(t^i) \cdot \Big(v^i_{(t^i,s^{-i})} - \varepsilon\Big) \\ &= \sum_{t^i \in S^i} p^i_{s^i a^i_{s^i}}(t^i) \cdot v^i_{(t^i,s^{-i})} - \varepsilon. \end{aligned}$$

As $\varepsilon > 0$ was arbitrary, the proof will then be complete.

<u>Step 1.</u> *The construction of two joint strategies:* (σ^i, π^{-i}) *for initial state* s *and* (π^i, σ^{-i}) *for all initial states of the form* (t^i, s^{-i}). Before the construction of the strategies, we define two maps ϕ and ψ, both of which will "transform" possible histories

of play. More precisely, ϕ will transform histories with initial state s into histories with initial states of the form (t^i, s^{-i}), while ψ will do it the other way around.

Let s_m denote the state that play visited at stage m and let a_m denote the joint action played by the players in state s_m at stage m. The history up to stage m is then the sequence $h_m = (s_1, a_1; s_2, a_2; \ldots; s_m, a_m)$. For initial state $s_1 = s$, let $\phi(h_m)$ denote the sequence of states and joint actions

$$\phi(h_m) := ((s_2^i, s_1^{-i}), (a_2^i, a_1^{-i}); (s_3^i, s_2^{-i}), (a_3^i, a_2^{-i}); \ldots; (s_m^i, s_{m-1}^{-i}), (a_m^i, a_{m-1}^{-i})),$$

which is derived from h_m by simply letting player i one step ahead of players $-i$. Note that, as the action space of a player j in a product-game only depends on the j-th coordinate of the state, the sequence $\phi(h_m)$ could arise as a possible history up to stage $m-1$ (as it consists of $m-1$ states and $m-1$ corresponding joint actions) with initial state $(s_2^i, s^{-i}) = (s_2^i, s_1^{-i})$. For infinite histories, ϕ is defined similarly.

For an initial state $s_1 = (t^i, s^{-i})$, for some $t^i \in S^i$, we also define the transformation ψ for h_m by

$$\psi(h_m) := (s, (a_s^i, a_1^{-i}); (s_1^i, s_2^{-i}), (a_1^i, a_2^{-i}); \ldots; (s_{m-1}^i, s_m^{-i}), (a_{m-1}^i, a_m^{-i})),$$

where s and a_s^i are the state and action that we fixed initially. (The notation ψ^{s,a_s^i} would be more precise to indicate the dependence of ψ on s and a_s^i, but since we only consider one s and one a_s^i througout this proof, we omit the upper index here.) Note that $\psi(h_m)$ could arise as a possible history up to stage m (as it consists of m states and m corresponding joint actions) with initial state s. In $\psi(h_m)$, it is now players $-i$ who are one step ahead of player i.

Note that if $s_1 = s$ and $a_1^i = a_s^i$ (the state and action we fixed initially) then

$$\psi(\phi(h_m)) = h_{m-1}. \tag{9}$$

We start with the strategies for players $-i$. In view of the definition of the minmax-level, there exists a joint strategy π^{-i} of players $-i$ such that player i's reward from initial state s cannot be more than his minmax-level up to ε, i.e. $\gamma_s^i(\tau^i, \pi^{-i}) \leq v_s^i + \varepsilon$ for all strategies τ^i of player i.

Given π^{-i}, define a history-dependent strategy σ^{-i} for players $-i$ for every initial state s_1 of the form $s_1 = (t^i, s^{-i})$, for some $t^i \in S^i$, as follows. For stage 1 in state s_1 (with empty history of play) let

$$\sigma_{s_1}^{-i}(\emptyset) := \pi_s^{-i}(\emptyset),$$

where $s = (s^1, \ldots, s^n)$ is the state we fixed initially, and in general for stage $m \geq 2$ in state s_m with past history $h_{m-1} = (s_1, a_1; s_2, a_2; \ldots; s_{m-1}, a_{m-1})$ let

$$\sigma^{-i}_{s_m}(h_{m-1}) := \pi^{-i}_{(s^i_{m-1}, s^{-i}_m)}(\psi(h_{m-1})).$$

In words, in initial state s_1 at stage 1, the joint strategy σ^{-i} prescribes for players $-i$ to follow π^{-i} as if the initial state was state s, while in state s_m at stage $m \geq 2$, to follow π^{-i} as if the past history was $\psi(h_{m-1})$ and the present state was (s^{-i}_m, s^i_{m-1}).

Now we define the strategies for player i. By the definition of the minmax-level once more, there also exists a strategy π^i for player i which defends the minmax-level against σ^{-i} up to ε, i.e. $\gamma^i_t(\pi^i, \sigma^{-i}) \geq v^i_t - \varepsilon$ for all initial states t.

Given π^i, define also a strategy σ^i for player i for initial state $s_1 = s$ (for other initial states, σ^i is arbitrary) as follows. For stage 1 in state s_1 let

$$\sigma^i_{s_1}(\emptyset) := a^i_s,$$

where s and a^i_s are the state and action that we fixed initially, and in general for stage $m \geq 2$ in state s_m with past history h_{m-1} let

$$\sigma^i_{s_m}(h_{m-1}) := \pi^i_{(s^i_m, s^{-i}_{m-1})}(\phi(h_{m-1})).$$

In words, in state $s_1 = s$ at stage 1, the strategy σ^i prescribes for player i to play action a^i_s, while in state s_m at stage $m \geq 2$, to follow π^i as if the past history was $\phi(h_{m-1})$ and the present state was (s^i_m, s^{-i}_{m-1}).

Note that, by the definitions of π^{-i} and π^i, we have

$$\gamma^i_s(\sigma^i, \pi^{-i}) \leq v^i_s + \varepsilon \tag{10a}$$

and

$$\gamma^i_{(t^i, s^{-i})}(\pi^i, \sigma^{-i}) \geq v^i_{(t^i, s^{-i})} - \varepsilon \qquad \forall t^i \in S^i. \tag{11}$$

This completes step 1.

Step 2. The relation between the histories with respect to defined joint strategies (σ^i, π^{-i}) *and* (π^i, σ^{-i}). Since, σ^i and σ^{-i} are defined with mixed actions used by π^i and π^{-i} respectively, there is an important relation between the occurrence probabilities of the histories with respect to the joint strategies (σ^i, π^{-i}) and (π^i, σ^{-i}).

Take a possible history $h_m = (s_1, a_1; \ldots; s_m, a_m)$ up to stage m with initial state $s_1 = s$ and initial action $a^i_1 = a^i_s$ for player i from state s. Let $\{h_m * (s^i_{m+1}, a^i_{m+1})\}$

denote the event that the history up to stage m coincides with h_m, and additionally, player i is in state s^i_{m+1} at stage $m+1$ and he plays action a^i_{m+1} in state s^i_{m+1} at stage $m+1$. Note that $\phi(h_{m+1})$ does not include s^{-i}_{m+1} and a^{-i}_{m+1}, so it is clear which sequence we mean by $\phi(h_m * (s^i_{m+1}, a^i_{m+1}))$. Let $\left\{\phi(h_m * (s^i_{m+1}, a^i_{m+1}))\right\}$ denote the event that the history up to stage m coincides with $\phi(h_m * (s^i_{m+1}, a^i_{m+1}))$.

We will now show that for all stages m

$$\mathbb{P}_{s,(\sigma^i,\pi^{-i})}\left\{h_m * (s^i_{m+1}, a^i_{m+1})\right\} = p^i_{s^i a^i_s}(s^i_2) \cdot \mathbb{P}_{(s^i_2,s^{-i}),(\pi^i,\sigma^{-i})}\left\{\phi(h_m * (s^i_{m+1}, a^i_{m+1}))\right\} \tag{12}$$

for every possible history h_m up to stage m with initial state $s_1 = s$ and initial action $a^i_1 = a^i_s$ for player i (this is the action that σ^i prescribes with probability 1) and for every state s^i_{m+1} and action a^i_{m+1} for player i.

We use induction on m. For any strategy τ^j of any player j, let $\tau^j_t(a^j_t|h)$ denote the probability that the mixed action $\tau^j_t(h)$ puts on action a^j_t. Take first $m = 1$. By the definitions of the strategies, the lefthandside of (12) equals

$$\mathbb{P}_{s,(\sigma^i,\pi^{-i})}\left\{(s, a_1) * (s^i_2, a^i_2)\right\} = \pi^{-i}_s(a^{-i}_1|\emptyset) \cdot p^i_{s^i a^i_s}(s^i_2) \cdot \pi^i_{(s^i_2,s^{-i})}(a^i_2|\emptyset),$$

where we used that σ^i prescribes $a^i_1 = a^i_s$ with probability 1 at stage 1 in state $s_1 = s$ and also that the mixed action prescribed by σ^i in state s_2 at stage 2 equals the mixed action $\pi^i_{(s^i_2,s^{-i})}(\emptyset)$. On the other hand, the righthandside of (12) equals

$$p^i_{s^i a^i_s}(s^i_2) \cdot \mathbb{P}_{(s^i_2,s^{-i}),(\pi^i,\sigma^{-i})}\left\{((s^i_2, s^{-i}); (a^i_2, a^{-i}_1))\right\} = p^i_{s^i a^i_s}(s^i_2) \cdot \pi^i_{(s^i_2,s^{-i})}(a^i_2|\emptyset) \cdot \pi^{-i}_s(a^{-i}_1|\emptyset),$$

where for the last factor we used that the mixed action prescribed by σ^{-i} in state (s^i_2, s^{-i}) at stage 1 equals the mixed action $\pi^{-i}_s(\emptyset)$. Hence, (12) holds for $m = 1$. Suppose then that equality (12) is valid for a certain m. For $m+1$ we obtain by the definition of the strategies in a similar way that the lefthandside of (12) equals

$$\begin{aligned} &\mathbb{P}_{s,(\sigma^i,\pi^{-i})}\left\{h_{m+1} * (s^i_{m+2}, a^i_{m+2})\right\} \\ &= \mathbb{P}_{s,(\sigma^i,\pi^{-i})}\left\{h_m * (s^i_{m+1}, a^i_{m+1})\right\} \cdot p^{-i}_{s^{-i}_m a^{-i}_m}(s^{-i}_{m+1}) \cdot \pi^{-i}_{s_{m+1}}(a^{-i}_{m+1}|h_m) \\ &\cdot p^i_{s^i_{m+1} a^i_{m+1}}(s^i_{m+2}) \cdot \pi^i_{(s^i_{m+2},s^{-i}_{m+1})}(a^i_{m+2}|\phi(h_{m+1})), \end{aligned} \tag{13}$$

where for the last factor we used

$$\sigma^i_{s_{m+2}}(h_{m+1}) = \pi^i_{(s^i_{m+2},s^{-i}_{m+1})}(\phi(h_{m+1})).$$

On the other hand, the righthandside of (12) equals

$$
\begin{aligned}
&p^i_{s^i a^i_s}(s^i_2) \cdot \mathbb{P}_{(s^i_2, s^{-i}),(\pi^i,\sigma^{-i})}\left\{\phi(h_{m+1} * (s^i_{m+2}, a^i_{m+2}))\right\} \\
&= p^i_{s^i a^i_s}(s^i_2) \cdot \mathbb{P}_{(s^i_2, s^{-i}),(\pi^i,\sigma^{-i})}\left\{\phi(h_m * (s^i_{m+1}, a^i_{m+1}))\right\} \\
&\cdot p^i_{s^i_{m+1} a^i_{m+1}}(s^i_{m+2}) \cdot \pi^i_{(s^i_{m+2}, s^{-i}_{m+1})}(a^i_{m+2} | \phi(h_{m+1})) \cdot p^{-i}_{s^{-i}_m a^{-i}_m}(s^{-i}_{m+1}) \cdot \pi^{-i}_{s_{m+1}}(a^{-i}_{m+1} | h_m),
\end{aligned}
\tag{14}
$$

where for the last factor we used that in view of equalities (9) we have

$$\sigma^{-i}_{(s^i_{m+2}, s^{-i}_{m+1})}(\phi(h_m * (s^i_{m+1}, a^i_{m+1}))) = \pi^{-i}_{s_{m+1}}(h_m).$$

In conclusion, from our assumption that (12) holds for m, and from equalities (13) and (14), it follows that (12) holds for $m+1$. Consequently, equality (12) holds for all stages $m \geq 2$.

Step 3. Final conclusions. Recall that, with respect to any initial state and any joint strategy, play eventually settles, with probability 1, in a restricted game. Since the game is simple, the average reward is determined by this restricted game.

Let h^∞ denote any infinite history, with initial state s, with respect to which play eventually settles in a restricted game $\bar{G}_k$ (and the corresponding set of states $E_k = \times_{i=1}^n E^i_{k^i}$). Then, with respect to $\phi(h^\infty)$, each player j eventually settles in the same set $E^j_{k^j}$, implying that play eventually settles in $\bar{G}_k$ (and E_k) with respect to $\phi(h^\infty)$ as well. It is therefore clear by equalities (12) that the probability that this restricted game is some $\bar{G}_k$ with respect to (σ^i, π^{-i}) with initial state s equals the probability that this is $\bar{G}_k$ when player i first executes the solitary move a^i_s in state s, by which play moves to a state of the form (t^i, s^{-i}), and subsequently from state (t^i, s^{-i}), the players start playing (π^i, σ^{-i}). Hence

$$\gamma^i_s(\sigma^i, \pi^{-i}) = \sum_{t^i \in S^i} p^i_{s^i a^i_{s^i}}(t^i) \cdot \gamma^i_{(t^i, s^{-i})}(\pi^i, \sigma^{-i}).$$

Combining this with inequalities (10a) and (11), we obtain inequalities (8). As $\varepsilon > 0$ was arbitrary, the proof is complete. ■

Based on the previous lemma, we are able to derive more structural properties of the minmax-levels of simple product-games.

Lemma 3 *Let G be a simple aperiodic product-game, and $E^i_{k^i}$ a maximal communicating set for player i, for some $k^i \in K^i$.*

(1) For any two states $s^i, t^i \in E^i_{k^i}$ of player i and any joint state $s^{-i} \in S^{-i}$ of players $-i$, the minmax-level of player i satisfies $v^i_{(s^i,s^{-i})} = v^i_{(t^i,s^{-i})}$.

(2) For any two joint states $s^{-i}, t^{-i} \in E^{-i}_{k^{-i}}$ of players $-i$ and any state $s^i \in S^i$ of player i, the minmax-level of player i satisfies $v^i_{(s^i,s^{-i})} = v^i_{(s^i,t^{-i})}$.

Proof. We will show part (1); the proof of part 2 is similar. Take an arbitrary $s^{-i} \in S^{-i}$. Let F^i denote those states $s^i \in E^i_{k^i}$ for which $v^i_{(s^i,s^{-i})} \leq v^i_{(t^i,s^{-i})}$ for all $t^i \in E^i_{k^i}$. Suppose by way of contradiction that $E^i_{k^i} - F^i$ is not empty. Take a state $s^i \in F^i$ and an action $a^i_{s^i} \in \bar{A}^i_{s^i}$ which moves from state s^i to a state in $E^i_{k^i} - F^i$ with a positive probability. Then, the solitary move $a^i_{s^i}$ in state (s^i, s^{-i}) for player i would improve player i's minmax-level in expectation, which contradicts part (1) of lemma 2. Hence, $F^i = E^i_{k^i}$, and part (1) of the lemma follows. ■

Lemma 4 *Let G be a simple aperiodic product-game, and E_k a joint maximal communicating set for some $k \in K$. Then, the minmax-level v^i of any player i is constant on E_k, i.e. $v^i_s = v^i_t$ for all $s, t \in E_k$.*

Proof. Take a player i and two arbitrary states $s, t \in E_k$. Then, by applying both parts of lemma 3, we obtain

$$v^i_s = v^i_{(s^i,s^{-i})} = v^i_{(s^i,t^{-i})} = v^i_{(t^i,t^{-i})} = v^i_t,$$

hence the result. ■

The following lemma deals with the actions in the sets $\bar{A}^i_s$, which keep play in the same maximal communicating set with probability 1.

Lemma 5 *Let G be a simple aperiodic product-game. Then, for any player i the following properties hold.*

(1) Let $s = (s^1, \ldots, s^n)$ be a state such that s^i belongs to a maximal communicating set $E^i_{k^i}$. Then, regardless the mixed action x^{-i}_s played by players $-i$ in state s, all the actions in $\bar{A}^i_s$ guarantee in expectation the best possible minmax-level for player i after transition, i.e. for any actions $a^i_s \in \bar{A}^i_s$ and $b^i_s \in A^i_s$ it holds that

$$\sum_{t \in S} p_{s,(a^i_s, x^{-i}_s)}(t)\, v^i_t \geq \sum_{t \in S} p_{s,(b^i_s, x^{-i}_s)}(t)\, v^i_t.$$

(2) Let $s = (s^1, \ldots, s^n)$ be a state such that s^j belongs to a maximal communicating set $E^j_{k^j}$ for all players $j \neq i$. Then, all joint actions in $\bar{A}^{-i}_s$ for players $-i$ in state s

guarantee in expectation that player i's minmax-level cannot increase after transition, i.e. for any joint action $a_s^{-i} \in \bar{A}_s^{-i}$ and for any action $a_s^i \in A_s^i$ it holds that

$$\sum_{t \in S} p_{s,(a_s^i,a_s^{-i})}(t)\, v_t^i \leq v_s^i.$$

Proof. First we prove part (1). Take an arbitrary mixed action x_s^{-i} for players $-i$, and actions $a_s^i \in \bar{A}_s^i$ and $b_s^i \in A_s^i$ for player i in state s. Then the transition from state s according to (b_s^i, x_s^{-i}) can be decomposed into the following three subsequent steps.

Step 1. In state s, players $-i$ play x_s^{-i} while player i stays in $E_{k^i}^i$ by playing action a_s^i. By doing so, play moves to a state $\bar{s}$ with $t^i \in E_{k^i}^i$.

Step 2. From state $\bar{s}$, player i gets a sequence of solitary moves in the sense of part (1) of lemma 2, and returns back to s^i. This can be achieved in a finite number of moves, with probability 1, inside the maximal communicating set $E_{k^i}^i$. After this step, the new state is $(s^i, \bar{s}^{-i})$, and by lemma 4, player i's minmax-level remains unchanged, i.e. $v_{\bar{s}}^i = v_{(s^i,\bar{s}^{-i})}^i$.

Step 3. In state $(s^i, \bar{s}^{-i})$, player i gets a solitary move and he plays action b_s^i. By part (1) of lemma 2, player i's minmax-level cannot increase during this step.

It is obvious that these three steps together induce the same transitions from state s as the joint mixed action (b_s^i, x_s^{-i}). As player i's minmax-level cannot increase during steps 2 and 3, we conclude that step 1 with (a_s^i, x_s^{-i}) must be at least as good as the three steps together with (b_s^i, x_s^{-i}) for the minmax-level of player i. Hence, the proof of part (1) is now complete.

Part (2) of the lemma follows similarly. One can show just as in part (1), by applying part (2) of lemma 2, that for all mixed actions x_s^i of player i in state s, for all joint actions $a_s^{-i} \in \bar{A}_s^{-i}$ and $b_s^{-i} \in A_s^{-i}$

$$\sum_{t \in S} p_{s,(x_s^i,a_s^{-i})}(t)\, v_t^i \leq \sum_{t \in S} p_{s,(x_s^i,b_s^{-i})}(t)\, v_t^i.$$

Therefore, in state s, the infimum in equality (7) is attained at all $a_s^{-i} \in \bar{A}_s^{-i}$, hence we have for all $a_s^{-i} \in \bar{A}_s^{-i}$ and $a_s^i \in A_s^i$ that

$$\sum_{t \in S} p_{s,(a_s^i,a_s^{-i})}(t)\, v_t^i \leq v_s^i,$$

which proves part (2). ■

The next lemma examines the situation, for simple product-games, when player i's (unique) reward in a restricted game $\bar{G}_k$ is strictly smaller or strictly larger than his minmax-level in the original game (which is a constant by lemma 4). We refer to observation D in section 2.3.2.

Lemma 6 *Let G be a simple aperiodic product-game, and let E_k be a joint maximal communicating set for some $k = (k^1, \ldots, k^n) \in K$. Let z^i_k denote player i's unique reward in the restricted game $\bar{G}_k$, and v^i_k be player i's minmax-level on E_k in the game G (a constant, cf. lemma 4).*

(1) Suppose $z^i_k < v^i_k$. Then, there is a state $s^i \in E^i_{k^i}$ and an action $a^i_{s^i} \in A^i_{s^i} - \bar{A}^i_{s^i}$ for player i in state s^i such that if player i plays action $a^i_{s^i}$ in any state $s = (s^i, s^{-i}) \in E_k$, with $s^{-i} \in E^{-i}_{k^{-i}}$, then player i's minmax-level cannot decrease in expectation from state s, regardless the actions played by players $-i$. More precisely, for any $a^{-i}_s \in A^{-i}_s$ we have

$$\sum_{t \in S} p_{s,(a^i_{s^i}, a^{-i}_s)}(t)\, v^i_t \geq v^i_s.$$

(2) Suppose $z^i_k > v^i_k$. Then, there is a joint state $s^{-i} \in E^{-i}_{k^{-i}}$ of players $-i$ and a joint action $a^{-i}_{s^{-i}} \in A^{-i}_{s^{-i}} - \bar{A}^{-i}_{s^{-i}}$ (i.e. at least one player $j \neq i$ plays outside $\bar{A}^j_{s^j}$) such that if players $-i$ play joint action $a^{-i}_{s^{-i}}$ in any state $s = (s^i, s^{-i}) \in E_k$, with $s^i \in E^i_{k^i}$, then player i's minmax-level cannot increase in expectation from state s, regardless the action played by player i. More precisely, for any $a^i_s \in A^i_s$ we have

$$\sum_{t \in S} p_{s,(a^i_s, a^{-i}_{s^{-i}})}(t)\, v^i_t \leq v^i_s.$$

Proof. We will prove part (1); the proof of part (2) is similar.

Step 1: *Choosing state s^i and action a^i_s.* We will first argue that there must be at least one state $s \in E_k$, joint action $b^{-i}_s \in \bar{A}^{-i}_s$ and action $a^i_s \in A^i_s - \bar{A}^i_s$ such that

$$\sum_{t \in S} p_{s,(a^i_s, b^{-i}_s)}(t)\, v^i_t \geq v^i_k. \tag{15}$$

(In view of part (2) of lemma 5, even equality holds, but this is not needed for the proof.) Suppose by way of contradiction that (15) does not hold, i.e. there exists an $\alpha > 0$ such that

$$\sum_{t \in S} p_{s,(a^i_s, b^{-i}_s)}(t)\, v^i_t \leq v^i_k - \alpha$$

holds for all $s \in E_k$, $b^{-i}_s \in \bar{A}^{-i}_s$ and $a^i_s \in A^i_s - \bar{A}^i_s$. For any initial state in E_k, suppose players $-i$ play in the following way: (1) players $-i$ play an arbitrary joint strategy in

$\bar{G}_k$ as long as player i only plays actions within the restricted game $\bar{G}_k$; (2) as soon as player i plays an action outside $\bar{G}_k$, say action $a^i_s \in A^i_s - \bar{A}^i_s$ in some state $s \in E_k$, and play moves to some state t, then players $-i$ start playing a joint strategy σ^{-i} in the original game G which guarantees that player i's reward is at most $v^i_t + \alpha/2$ in G. Then, if player i only plays actions within $\bar{G}_k$, player i's reward is exactly $z^i_k < v^i_k$, while if player i decides to play such an action a^i_s outside $\bar{G}_k$, when players $-i$ play some joint action b^{-i}_s, then his reward will be at most

$$\sum_{t \in S} p_{s,(a^i_s, b^{-i}_s)}(t)\,(v^i_t + \frac{1}{2}\alpha) = \sum_{t \in S} p_{s,(a^i_s, b^{-i}_s)}(t)\,v^i_t + \frac{1}{2}\alpha \leq v^i_k - \frac{1}{2}\alpha.$$

This would mean that player i is unable to defend v^i_k from initial states in E_k in either case, which would contradict the definition of the minmax-level v^i of player i. Hence, inequality (15) holds indeed for some $s = (s^1, \ldots, s^n) \in E_k$, joint action $b^{-i}_s \in \bar{A}^{-i}_s$ and action $a^i_s \in A^i_s - \bar{A}^i_s$.

Now s^i and a^i_s are the state and action we were looking for. However, keep the whole state s and the joint action b^{-i}_s in mind, as we will use them below as well.

Step 2: Proving that state s^i and action a^i_s satisfy part (1) of the lemma, for this particular state s, i.e. for any $a^{-i}_s \in A^{-i}_s$ we have

$$\sum_{t \in S} p_{s,(a^i_s, a^{-i}_s)}(t)\,v^i_t \geq v^i_s.$$

Take an arbitrary $a^{-i}_s \in A^{-i}_s$. The transition from state s according to (a^i_s, a^{-i}_s) can be decomposed into the following three subsequent steps.

Step A. In state s, player i plays action a^i_s while players $-i$ stays in $E^{-i}_{k^{-i}}$ by playing joint action b^{-i}_s. By doing so, play moves to a state $\bar{s}$ with $\bar{s}^{-i} \in E^{-i}_{k^{-i}}$, and by inequality (15), player i's minmax-level cannot decrease in expectation during this step.

Step B. From state $\bar{s}$, players $-i$ get a sequence of solitary moves in the sense of part (2) of lemma 2, and return back to s^{-i}. This can be achieved in a finite number moves, with probability 1, inside the joint maximal communicating set $E^{-i}_{k^{-i}}$. After this step, the new state is (t^i, s^{-i}), and by lemma 3, player i's minmax-level remains unchanged during step B, i.e. $v^i_{\bar{s}} = v^i_{(t^i, s^{-i})}$.

Step C. In state (t^i, s^{-i}), players $-i$ get a solitary move and play joint action a^{-i}_s. By part (2) of lemma 2, player i's minmax-level cannot decrease during this step.

It is obvious that these three steps A, B and C together induce the same transitions from state s as the joint action (a^i_s, a^{-i}_s). As player i's minmax-level cannot decrease during all steps, we conclude that

$$\sum_{t\in S} p_{s,(a^i_s,a^{-i}_s)}(t)\, v^i_t \geq v^i_s,$$

which proves that state s^i and action a^i_s satisfy part (1) of the lemma, for this particular state s.

Step 3: Proving that state s^i and action a^i_s satisfy part (1) of the lemma (not only for state s, but for all states $(s^i,t^{-i}) \in E_k$, with $t^{-i} \in E^{-i}_{k^{-i}}$), i.e. for any $a^{-i}_s \in A^{-i}_s$ we have

$$\sum_{u\in S} p_{(s^i,t^{-i}),(a^i_s,a^{-i}_{t^{-i}})}(u)\, v^i_u \geq v^i_{(s^i,t^{-i})}.$$

Take an arbitrary $t^{-i} \in E^{-i}_{k^{-i}}$ and a joint action $b^{-i}_{t^{-i}} \in \bar{A}^{-i}_{t^{-i}}$. By lemma 3, v^i is a constant $w^i_{u^i}$ on $\{u^i\} \times E^{-i}_{k^{-i}}$, for any $u^i \in S^i$. Then, as both $b^{-i}_{t^{-i}}$ from joint state t^{-i} and b^{-i}_s from joint state s^{-i} keep play in $E^{-i}_{k^{-i}}$ with probability 1, we have

$$\sum_{u\in S} p_{(s^i,t^{-i}),(a^i_s,b^{-i}_{t^{-i}})}(u)\, v^i_u = \sum_{u^i\in S^i} p^i_{s^i a^i_s}(u^i)\, w^i_{u^i} = \sum_{u\in S} p_{s,(a^i_s,b^{-i}_s)}(u)\, v^i_u,$$

hence by inequality (15)

$$\sum_{u\in S} p_{(s^i,t^{-i}),(a^i_s,b^{-i}_{t^{-i}})}(u)\, v^i_u = \sum_{u\in S} p_{s,(a^i_s,b^{-i}_s)}(u)\, v^i_u \geq v^i_s.$$

Now similarly to step 2, it follows for all $a^{-i}_{t^{-i}}$ that

$$\sum_{u\in S} p_{(s^i,t^{-i}),(a^i_s,a^{-i}_{t^{-i}})}(u)\, v^i_u \geq v^i_{(s^i,t^{-i})},$$

which proves step 3 and part (1) of the lemma. ■

3.1.3 The minmax-levels of general product-games

Take an arbitrary product-game G. The next lemma presents a natural way of transforming G into a simple product-game $\widetilde{G}$, and claims that the minmax-levels of the players remain unchanged under this transformation.

Lemma 7 *Take an arbitrary aperiodic product-game G, with v^i_s denoting the minmax-level for every player i and for every state $s \in S$. Let $\bar{v}^i_k$ denote player i's minmax-level in any restricted game $\bar{G}_k$ (which is constant, cf. lemma 1). Let $\widetilde{G}$ denote the simple* *aperiodic product-game which is derived from G by replacing each player i's payoffs in any restricted game $\bar{G}_k$ by his minmax-level $\bar{v}^i_k$. Further, let w^i_s denote every player i's minmax-level in $\widetilde{G}$ in state s.*

Then, the minmax-levels of the product-games G and $\widetilde{G}$ are equal, i.e. $v^i_s = w^i_s$ for all players i and for all states $s \in S$.

Proof. Consider the original product-game G and take an arbitrary player i. For this game G, we will show in step 1 below that players $-i$ have a joint stationary strategy x^{-i} which guarantees that player i's reward from any initial state $s \in S$ is at most w^i_s, i.e. for all strategies π^i for player i we have

$$\gamma^i_s(\pi^i, x^{-i}) \leq w^i_s.$$

This yields $v^i_s \leq w^i_s$ for all states $s \in S$. Then, in step 2, we will prove $v^i_s \geq w^i_s$ for all $s \in S$ by showing that player i can defend w^i in G, i.e. for any initial state s and for any strategy σ^{-i} for players $-i$, player i has a strategy π^i such that $\gamma^i_s(\pi^i, \sigma^{-i}) \geq w^i_s$. Given steps 1 and 2, we will have $v^i_s = w^i_s$ for all states s, so the proof will then be complete.

Step 1: *Proving that players $-i$ have a joint stationary strategy x^{-i} such that, for all initial states s and for all strategies π^i for player i, we have $\gamma^i_s(\pi^i, x^{-i}) \leq w^i_s$.* Note first that w^i_s is also a constant w^i_k on any joint maximal communicating set E_k, by lemma 4 for the game $\widetilde{G}$. We construct the joint stationary strategy x^{-i} by distinguishing the following three mutually exclusive cases.

Case 1: States $s = (s^1, \ldots, s^n) \in S$ such that s^j is of type 1 for at least one player j (possibly $j = i$). In any such a state s, let $x^{-i}_s \in X^{-i}_s$ be a joint mixed action for players $-i$ such that for any mixed action $x^i_s \in X^i_s$ of player i we have

$$\sum_{t \in S} p_{s,(x^{-i}_s, x^i_s)}(t)\, w^i_t \leq w^i_s.$$

Obviously, by expression (7) for player i's minmax-level w^i in $\widetilde{G}$, such a joint mixed action exists.

Case 2: States in a joint maximal communicating set E_k for which $\bar{v}^i_k \leq w^i_k$. Take a joint stationary strategy y^{-i} for player i in the corresponding restricted game $\bar{G}_k$ (which is a part of the original game G) as in lemma 1. Then, let $x^{-i}_s = y^{-i}_s$ for all $s \in E_k$.

Case 3: States in a joint maximal communicating set E_k for which $\bar{v}^i_k > w^i_k$. Take a joint state $t^{-i} \in E^{-i}_{k^{-i}}$ and a joint "exit" action $a^{-i}_{t^{-i}} \in \bar{A}^{-i}_{t^{-i}}$, with respect to the game $\widetilde{G}$ and its minmax-level w^i for player i, as in part (2) of lemma 6. Then, for any $s = (s^1, \ldots, s^n) \in E_k$, let $x_s^{-i} = a_{t^{-i}}^{-i}$ whenever $s^{-i} = t^{-i}$, and let x_s^{-i} be an arbitrary joint completely mixed action on $\bar{A}_s^{-i}$ whenever $s^{-i} \neq t^{-i}$.

Take a stationary best reply x^i of player i in G against x^{-i}. We will show that $\gamma_s^i(x^i, x^{-i}) \leq w_s^i$ for any initial state $s \in S$.

First, consider an arbitrary ergodic set F for (x^i, x^{-i}). As players $-i$ will leave any set E_k considered in case 3, we conclude that $F \subset E_k$ for some E_k in case 2. Since x^i does not leave F, we also have $x_s^i \in \bar{X}_s^i$ for all $s \in F$, meaning that x^i behaves on F as a stationary strategy in the restricted game $\bar{G}_k$. Hence, by the choice of x^{-i} in case 2, we have

$$\gamma_s^i(x^i, x^{-i}) \leq \bar{v}_k^i \leq w_k^i = w_s^i \tag{16}$$

for all $s \in F$. As F was an arbitrary ergodic set, we have $\gamma_s^i(x^i, x^{-i}) \leq w_s^i$ for all states s that are recurrent for (x^i, x^{-i}).

Next, note that w^i cannot increase in expectation after transition with respect to (x^i, x^{-i}), i.e. $P(x^i, x^{-i})w^i \leq w^i$. Indeed, for cases 2 and 3 it is guaranteed by part (2) of lemma 5 and by part (2) of lemma 6 (both applied to w^i as the minmax-level of player i in $\widetilde{G}$), while it holds by construction for case 1. Consequently, we also have $P^m(x^i, x^{-i})w^i \leq w^i$ for all $m \in \mathbb{N}$, yielding $Q(x^i, x^{-i})w^i \leq w^i$.

By applying equality (5), we now obtain

$$\gamma^i(x^i, x^{-i}) = Q(x^i, x^{-i}) \cdot \gamma^i(x^i, x^{-i}) \leq Q(x^i, x^{-i}) \cdot w^i \leq w^i,$$

where the first inequality follows from inequality (16) and from the fact that entry (t, s) of the stochastic matrix $Q(x^i, x^{-i})$ is only positive if state s is recurrent for (x^i, x^{-i}). Since x^i is a best reply to x^{-i} in G, the proof of step 1 is complete.

Step 2: Proving that against any joint strategy σ^{-i} for players $-i$, player i has a strategy π^i such that $\gamma_s^i(\pi^i, \sigma^{-i}) \geq w_s^i$ for all initial states s. The proof is quite similar to step 1. Given a joint strategy σ^{-i}, player i should use a strategy π^i which prescribes to play as follows. First, in states where at least one player is in a state of type 1 (case 1*, being the counterpart of case 1 in step 1), against any joint mixed action prescribed by σ^{-i}, player i can just play a mixed action such that w^i does not decrease in expectation. Next, if a joint maximal communicating set E_k satisfies $\bar{v}_k^i \geq w_k^i$ (case 2*, being the counterpart of case 2 in step 1), then player i can defend w_k^i against σ^{-i} inside $\bar{G}_k$, whereas if E_k satisfies $\bar{v}_k^i < w_k^i$ (case 3*, being the counterpart of case 3 in step 1), then player i can leave E_k due to part (1) of lemma 6.

We remark here that, although cases 1* and 3* can be done in a stationary way, case 2* may require a history-dependent strategy for player i. As π^i is not necessarily stationary, the proof that such a π^i defends w^i in G against σ^{-i}, i.e. $\gamma_s^i(\pi^i, \sigma^{-i}) \geq w_s^i$ for all $s \in S$, differs slightly from the proof in step 1, and therefore we provide a short proof.

Consider (π^i, σ^{-i}) and take an arbitrary initial state $s \in S$. As we know, play eventually settles, with probability 1, in a restricted game. Let ξ denote the random variable for the index of this restricted game (so play settles in restricted game $\bar{G}_\xi$). Due to the construction of π^i, the corresponding set of states E_ξ falls under case 2^*, and not under case 3^*. Hence, for player i's reward we have

$$\gamma_s^i(\pi^i, \sigma^{-i}) \geq \mathbb{E}_{s,(\pi^i,\sigma^{-i})}(\bar{v}_\xi^i) \geq \mathbb{E}_{s,(\pi^i,\sigma^{-i})}(w_\xi^i).$$

(Note that this inequality is the counter-part of inequality (16) from step 1.)

Notice further that, by the construction of π^i, player i is assured that w^i cannot decrease in expectation during play with respect to (π^i, σ^{-i}) and initial state s, i.e. if W_m^i denotes the random variable for the minmax-level of player i in the state at stage m, then given any possible outcome $w' \in \mathbb{R}$ we have

$$\mathbb{E}_{s,(\pi^i,\sigma^{-i})}(W_{m+1}^i | W_m^i = w') \geq w'.$$

(Note that this inequality is the counter-part of inequality $P(x^i, x^{-i})w^i \leq w^i$ from step 1.) Hence,

$$w_s^i \leq \mathbb{E}_{s,(\pi^i,\sigma^{-i})}\left(w_\xi^i\right).$$

(This conclusion is very intuitive, and it immediately follows from basic optional stopping theorems for submartingales, as we only have finitely many states and actions. Note that this inequality is the counter-part of inequality $w^i \geq Q(x^i, x^{-i})w^i$ from step 1.) In conclusion,

$$\gamma_s^i(\pi^i, \sigma^{-i}) \geq \mathbb{E}_{s,(\pi^i,\sigma^{-i})}(w_\xi^i) \geq w_s^i,$$

proving step 2. ■

The previous lemma (and its proof) has important consequences. First, the results from section 3.1.2 are now applicable to general aperiodic product-games, providing us the necessary structural properties of the minmax-levels in the general context. This is stated next.

Corollary 8 *The results of lemmas 2 up to 5 in section 3.1.2 for simple aperiodic product-games are also valid for any general aperiodic product-game G. Lemma 6 extends as well if one interprets z_k^i as the minmax-level $\bar{v}_k^i$ of player i in the restricted game $\bar{G}_k$ (note that $\bar{v}_k^i$ is constant on E_k by lemma 1, and evidently coincides with z_k^i of $\widetilde{G}$, where $\widetilde{G}$ is the simple product-game derived in lemma 7).*

Notice that, as a consequence of the proof of lemma 7, the joint stationary strategy x^{-i} in step 1 in the proof guarantees in the original game G that player i's reward from any initial state $s \in S$ is at most $v_s^i = w_s^i$. Hence, the infimum in expression (6) of the minmax-levels is attained at stationary strategies, for all product-games. This will become important later, as we are heading towards 0-equilibria, which do not allow even small positive error terms.

Corollary 9 *(of step 1 of the proof of lemma 7) Take an aperiodic product-game G and an arbitrary player i. Then, players $-i$ have a joint stationary strategy x^{-i} which guarantees that player i's reward from any initial state $s \in S$ is at most his minmax-level v_s^i, i.e. for all strategies π^i for player i we have*

$$\gamma_s^i(\pi^i, x^{-i}) \leq v_s^i.$$

With the help of this corollary, we are now ready to prove Main Theorem 2, which claimed that, in every two-player aperiodic zero-sum product-game, both players have a stationary 0-optimal strategy.

Proof of Main Theorem 2. Take an arbitrary two-player aperiodic zero-sum product-game, and take player $i = 1$. By corollary 9, there exists a stationary strategy x^{-1} for player 2 (as $-1 = \{2\}$) which guarantees that player 1's reward is not more than v_s^1 for any initial state $s \in S$. Hence, x^{-1} is 0-optimal for player 2. One finds similarly a stationary 0-optimal strategy for player 1, which completes the proof.

3.2 The construction of 0-equilibria in product-games

In section 3.1 we achieved several results for the minmax-levels of aperiodic product-games. We will use this knowledge now to construct 0-equilibria in aperiodic product-games.

The following lemma deals with the restricted games. It states that, in any restricted game, there exists a 0-equilibrium in which, if no player deviates, the players' future expectations remain unchanged during the whole play.

Lemma 10 *Let G be an arbitrary aperiodic product-game and consider the restricted game $\bar{G}_k$, for any $k = (k^1, \ldots, k^n) \in K$. Then, there exists a 0-equilibrium π in $\bar{G}_k$ such that the corresponding rewards are independent of the initial state and all the continuation rewards remain unchanged with probability 1 during the whole play. More precisely, the reward $\bar{\gamma}_s^i(\pi[h])$ is independent of the initial state $s \in E_k$ and the history*

h, given h occurs with a positive probability with respect to π. Here $\bar{\gamma}$ denotes the average reward for the restricted game $\bar{G}_k$.

Proof. Observe the following for the game $\bar{G}_k$.

(i) The set of feasible rewards (i.e. the rewards that can be obtained by some joint strategy) is the same from any initial state in E_k. This is an immediate consequence of the fact that, as E_k is an aperiodic joint maximal communicating set, the players can move from any state in E_k to any other one in E_k, possibly in a number of steps.

(ii) The extreme points of the set of feasible rewards are induced by pure stationary strategies (cf. for example the appendix in Dutta [1995]).

(iii) Each minmax-level in $\bar{G}_k$ is a constant $\bar{v}^i_k$, by lemma 1.

Given these three observations, this game situation is almost identical to a repeated game. The following ideas and arguments are standard in Folk-theorems for repeated games. For the context of stochastic games, we refer to Dutta [1995]. Take an arbitrary feasible reward $z_k = (z^1_k, \dots, z^n_k)$ such that $z^i_k \geq \bar{v}^i_k$ for all players i. By property (ii), we may write z_k as a convex combination of rewards corresponding to pure stationary strategies a_l, $l = 1, \dots, L$, i.e.

$$z_k = \sum_{l=1}^{L} \alpha_l \cdot \bar{\gamma}(a_l).$$

Let σ be the pure joint strategy which prescribes to play as follows: play a_1 for d^1_1 stages, then a_2 for d^1_2 stages, ..., then a_L for d^1_L stages, and repeat this with lengths $d^2_1, \dots, d^2_L$, then with lengths $d^3_1, \dots, d^3_L$, and so on. The lengths d^m_l have to be chosen in such a way that, when m tends to infinity, then we have for each $l \in \{1, \dots, L\}$ that (a) d^m_l goes to infinity, so that the expected average payoff when strategy a_l is played for d^m_l stages will approach $\bar{\gamma}(a_l)$; (b) $d^m_l/(d^m_1 + \ldots + d^m_L)$ tends to α_l, so that a_l is played in the right proportion of time; (c)

$$\frac{d^m_l}{(d^1_1 + \ldots + d^m_L) + \ldots + (d^{m-1}_1 + \ldots + d^{m-1}_L) + d^m_1 + \ldots + d^m_{l-1}}$$

tends to 0, so that the average payoffs will fluctuate less and less. Due to these three properties, σ induces reward z_k, and moreover, any continuation reward is also z_k, i.e. $\bar{\gamma}_s(\sigma[h]) = z_k$ for all states $s \in E_k$ and for all histories h. Let π be the joint strategy which prescribes to play σ, unless some player i deviates from the action prescribed by σ^i. In that case, from the new state, players $-i$ should switch to a joint stationary strategy x^{-i} as in lemma 1. Since the players receive $z_k \geq \bar{v}_k$ according to σ, while if a player i deviates then his reward is not more than $\bar{v}^i_k$, the joint strategy π is a 0-equilibrium and satisfies the requirements of the lemma. ■

Now we are sufficiently prepared to prove Main Theorem 1, which claimed that, in any aperiodic product-game, there exists a 0-equilibrium.

Proof of Main Theorem 1. Take an arbitrary aperiodic product-game G. For any player i, in view of corollary 9, we may take a joint stationary strategy y^{-i} for players $-i$ such that for all initial states $s \in S$ and for all strategies τ^i for player i we have

$$\gamma^i_s(\tau^i, y^{-i}) \leq v^i_s.$$

We will below define a joint strategy π with important properties, amongst others that the rewards are individually rational. The main idea for the construction of a 0-equilibrium is then to let the players play π, unless some player i deviates from π^i and plays an action on which π^i puts probability zero. If player i deviates in such a way, then from the next state, say state s, players $-i$ switch to the joint strategy y^{-i} and push down player i's reward to a level of at most v^i_s. In fact, y^{-i} acts as a threat strategy, which forces player i to follow the prescriptions of π^i. We wish to remark that the use of such threat strategies for the construction of equilibria is standard in stochastic games.

The proof of Main Theorem 1 consists of the following steps. In step 1, we construct a joint stationary strategy x^*, which is used to reach the "right" joint maximal communicating sets. Then, in step 2 we "extend" x^* to the joint strategy π according to which the players also receive the "right" rewards in the "right" joint maximal communicating sets. Finally, in step 3, we will complete the proof by showing that π supplemented with the joint stationary strategies y^{-i}, for all i, as is described above, forms a 0-equilibrium.

Step 1: The construction of the joint stationary strategy x^* and a number of properties of x^*. As is mentioned above, x^* will "guide" the players to the "right" joint maximal communicating sets. In order to arrive at x^*, two supplementary games $\widetilde{G}$ and G^* have to be constructed. The game $\widetilde{G}$ is a simple aperiodic product-game that we derive from G, whereas G^* is a stochastic game (not necessarily a product-game) that we obtain by restricting the players in $\widetilde{G}$ to certain mixed actions. Given G^*, the joint strategy x^* will be found as a stationary 0-equilibrium in the game G^*.

Step 1.1: The simple aperiodic product-game $\widetilde{G}$. Take a 0-equilibrium σ_k in every restricted game $\bar{G}_k$ as in lemma 10. Let z_k^i denote the corresponding reward for any player i, which is independent of the initial state. Let $\widetilde{G}$ denote the simple aperiodic product-game which is derived from G by replacing each player i's payoffs in any restricted game $\bar{G}_k$ by z_k^i. Further, let w_s^i denote player i's minmax-level in $\widetilde{G}$ from initial state s. Recall that w_s^i is a constant w_k^i on E_k, by lemma 4.

We will now argue that $w_s^i \geq v_s^i$ for all players i and for all states $s \in S$. By lemma 7, v^i equals player i's minmax-level in the simple aperiodic product-game G' which is derived from G by replacing each player i's payoffs in any restricted game $\bar{G}_k$ by $\bar{v}_k^i$. Since z_k^i is a 0-equilibrium reward in $\bar{G}_k$, we have $z_k^i \geq \bar{v}_k^i$. This means that player i's payoffs in $\widetilde{G}$ are always larger or equal to his corresponding payoffs in G', hence $w_s^i \geq v_s^i$ must hold indeed, for all players i and for all states $s \in S$.

Step 1.2: The stochastic game G^*. In this step, we will define a stochastic game G^* which is derived from $\widetilde{G}$ by restricting each player i in each state $s \in S$ to a certain (non-empty) subset $X_s^{*i} \subset X_s^i$ of mixed actions. First, for every state $s = (s^1, \ldots, s^n)$ which belongs to some joint maximal communicating set E_k, fix an arbitrary completely mixed action $\bar{y}_s^i$ for every player i on $\bar{X}_s^i$. Second, suppose that $\bar{G}_k$ is a restricted game such that $z_k^i < w_k^i$ for player i and that $z_k^j \geq w_k^j$ for all $j \in \{1, \ldots, i-1\}$. Then take a state $s_k^i \in E_{k^i}^i$ and an "exit" action a_k^i for player i in state s_k^i, with respect to the game $\widetilde{G}$ and its minmax-level w^i, as in part (1) of lemma 6.

Now, given these fixed pure and mixed actions, we will now define the subset $X_s^{*i} \subset X_s^i$ of mixed actions for every player i in every state $s = (s^1, \ldots, s^n) \in S$ as follows. First, if s is a state such that s^j is of type 1 for at least one player j, then we let $X_s^{*i} := X_s^i$ for all players i. Otherwise, for states belonging to a joint maximal communicating set E_k, depending on the relation between z_k and the players' minmax-levels w_k, we distinguish the following mutually exclusive cases:

Case (a): $z_k^i \geq w_k^i$ holds for all players i. Then, we let $X_s^{*i} := \{\bar{y}_s^i\}$ for all players i.

Case (b): $z_k^i < w_k^i$ holds for player i and $z_k^j \geq w_k^j$ holds for all $j \in \{1, \ldots, i-1\}$. Then, for players $j \neq i$, we let $X_s^{*j} := \{\bar{y}_s^j\}$. As for player i, if $s^i = s_k^i$ then we let $X_s^{*i} := \{a_k^i\}$, while if $s^i \neq s_k^i$ then we let $X_s^{*i} := \{\bar{y}_s^i\}$.

Notice that, due to the construction in cases (a) and (b), joint strategies $x \in X^*$ can only differ in states s such that s^j is of type 1 for at least one player j. Moreover, the ergodic sets for all $x \in X^*$ are precisely the joint maximal communicating sets E_k belonging to case (a), due to the use of the "exit" actions which eventually make play leave each E_k belonging to case (b).

Let G^* denote the stochastic game which is derived from $\widetilde{G}$ by restricting each player i in each state $s \in S$ to the space X_s^{*i} of mixed actions. The game G^* is a well-defined stochastic game (with the extreme points of X_s^{*i}, for every $s \in S$ and for every player i, acting as the set of pure actions for player i in state s), but not necessarily a product-game.

Step 1.3: *Defining x^* as a stationary 0-equilibrium of G^* and proving a number of properties of x^*.* As the ergodic sets are the same for all $x \in X^*$, lemma 12 in the appendix yields a stationary 0-equilibrium $x^* \in X^*$ for the game G^*. Obviously, x^* is also a joint stationary strategy in the game $\widetilde{G}$ and in the original game G, but not necessarily a 0-equilibrium.

As a conclusion of step 1.3, we wish to point out three properties of x^* in the game $\widetilde{G}$, and provide a proof.

Property (1): If $s = (s^1, \dots, s^n) \in S$ is a state such that s^j is of type 1 for at least one player j, then no player i can go to better states regarding his reward by unilaterally deviating from x_s^{*i}, i.e. for every action $b_s^i \in A_s^i$ we have

$$\sum_{t \in S} p_{s,(b_s^i, x_s^{*-i})}(t)\, \widetilde{\gamma}_t^i(x^*) \le \sum_{t \in S} p_{sx_s^*}(t)\, \widetilde{\gamma}_t^i(x^*),$$

where $\widetilde{\gamma}^i$ denotes the average reward to player i in the game $\widetilde{G}$.

Property (2): If $s = (s^1, \dots, s^n) \in S$ is a state such that s^i is of type 2 for all players i, then no player i can improve on his expected minmax-level in the next state by unilaterally deviating from x_s^{*i}, i.e. for every action $b_s^i \in A_s^i$ we have

$$\sum_{t \in S} p_{s,(b_s^i, x_s^{*-i})}(t)\, w_t^i \le \sum_{t \in S} p_{sx_s^*}(t)\, w_t^i.$$

Consequently, equality (7) also yields

$$w_s^i = \sum_{t \in S} p_{sx_s^*}(t)\, w_t^i. \tag{17}$$

Property (3): x^* yields individually rational rewards in $\widetilde{G}$ for all initial states, i.e. $\widetilde{\gamma}_s^i(x^*) \ge w_s^i$ for all players i and for all initial states $s \in S$.

Now, we provide the proofs for these properties.

Proof of property (1): This property follows from the fact that x^* is a 0-equilibrium in G^*, and no player is restricted in G^* in state s.

Proof of property (2): This is due to parts (1) of lemmas 5 and 6.

Proof of property (3): This property requires a longer argument. Notice that, as $x^* \in X^*$, all ergodic sets for x^* are precisely the joint maximal communicating sets E_k belonging to case (a), as is pointed out in step 1.2. Hence, if $s \in S$ is recurrent for x^*, then s belongs to some E_k considered under case (a), and we conclude for every player i's reward corresponding to x^* from initial state s that

$$\widetilde{\gamma}^i_s(x^*) = z^i_k \geq w^i_k = w^i_s, \tag{18}$$

where $\widetilde{\gamma}^i$ denotes the average reward to player i in the game $\widetilde{G}$. This proves that x^* yields individually rational rewards in $\widetilde{G}$ for all initial states that are recurrent for x^*.

By applying equalities (7) for the game $\widetilde{G}$, in every state $s = (s^1, \dots, s^n) \in S$ where s^j is of type 1 for at least one player j, there exists a mixed action $x^i_s \in X^i_s = X^{*i}_s$ for player i which defends w^i_s against x^{*-i}_s in the sense that

$$\sum_{t \in S} p_{s,(x^i_s, x^{*-i}_s)}(t)\, w^i_t \geq w^i_s. \tag{19}$$

Given these mixed actions x^i_s in such states s, there is a unique extension (with the mixed actions prescribed by x^* in all states belonging to joint maximal communicating sets) to a stationary strategy x^i in X^{*i}. Consider the joint stationary strategy $(x^i, x^{*-i}) \in X^*$. Then, the recurrent states for (x^i, x^{*-i}) and for x^* coincide (as both belong to X^*, cf. step 1.2) and if $s \in S$ is recurrent for x^* then (as x^i equals x^{*i} on all reccurent states) we have

$$\widetilde{\gamma}^i_s(x^i, x^{*-i}) = \widetilde{\gamma}^i_s(x^*). \tag{20}$$

Then, equalities (17) together with inequalities (19) yield $P(x^i, x^{*-i})w^i \geq w^i$, which implies $P^m(x^i, x^{*-i})w^i \geq w^i$ for all $m \in \mathbb{N}$. Hence, $Q(x^i, x^{*-i})w^i \geq w^i$. By applying equality (5), we now obtain

$$\widetilde{\gamma}^i(x^i, x^{*-i}) = Q(x^i, x^{*-i}) \cdot \widetilde{\gamma}^i(x^i, x^{*-i}) \geq Q(x^i, x^{*-i}) \cdot w^i \geq w^i,$$

where the first inequality follows from inequality (18) and equality (20), and from the fact that entry (t, s) of the stochastic matrix $Q(x^i, x^{*-i})$ is only positive if state s is recurrent for (x^i, x^{*-i}), or equivalently, recurrent for x^*. Since x^{*i} is a best reply to x^{*-i} in G^* and since $x^i \in X^{*i}$, we have

$$\widetilde{\gamma}^i_s(x^*) \geq \widetilde{\gamma}^i_s(x^i, x^{*-i}) \geq w^i_s$$

for all initial states $s \in S$, which proves property (3).

Step 2. *The construction of the joint strategy π for the original game G.* Given x^* from step 1, the definition of π is easy. Let π be the joint strategy which prescribes to play as follows:

Case (1): states $s = (s^1, \ldots, s^n) \in S$ in which s^j of type 1 for at least one player j. In this case, each player i follows x^*, i.e. plays the mixed action x_s^{*i}.

Case (2): when play reaches a joint maximal communicating set E_k for which $z_k^i \geq w_k^i$ holds for all players i (cf. case (a) in step 1.2). In this case, the players switch to the joint strategy σ_k (cf. step 1.1).

Case (3): when play reaches a joint maximal communicating set E_k for which $z_k^i < w_k^i$ holds for player i and $z_k^j \geq w_k^j$ holds for all $j \in \{1, \ldots, i-1\}$ (cf. case (b) in step 1.2). In this case, players $-i$ switch to a joint stationary strategy as in lemma 1, while player i follows x^*, i.e. plays the mixed action x_s^{*i} in state $s \in E_k$.

Notice that play leaves all sets E_k in case (3), due to the exit made by player i, with the guidance of x^*. Moreover, notice also that in a set E_k in case (2), by switching to σ_k, each player i receives in expectation reward z_k^i in the game G, which is exactly what the players would receive within E_k according to x^* in the game $\widetilde{G}$. So in some sense, x^* is used to reach the "right" joint maximal communicating sets, and then σ_k takes over to induce the "right" payoffs in the original game G. Thus

$$\gamma_s^i(\pi) = \widetilde{\gamma}_s^i(x^*)$$

for all initial states $s \in S$ and for all players i, which by property (3) of step 1.3 yields that π induce rewards at least w^i for each player i. In view of this, player i will have an incentive to "exit" in any set E_k in case (2), since within $\bar{G}_k$ he can get at most $\bar{v}_k^i$, while $\bar{v}_k^i \leq z_k^i < w_k^i$.

Step 3. *Proving that π supplemented with the joint stationary strategies y^{-i}, for all players i, is a 0-equilibrium.* Let η be the joint strategy which prescribes to play π, unless some player i deviates from π^i and plays an action on which π^i puts probability zero. If player i deviates in such a way, then from the next state, players $-i$ switch to the joint strategy y^{-i} and play it for the rest of play. As is already mentioned, the role of y^{-i} is to force every player i to follow the prescriptions of π^i.

Note that the expected rewards with respect to η in the original game G equals the expected rewards with respect to π in the original game G, which is then also equal to the rewards with respect to x^* in the game $\widetilde{G}$, i.e.

$$\gamma_s^i(\eta) = \gamma_s^i(\pi) = \widetilde{\gamma}_s^i(x^*)$$

for all initial states $s \in S$ and for all players i. Notice also that if h denotes a history and $s \in S$ a state such that, with a positive probability, h can occur and s can be the present state after h with respect to η (or equivalently with respect to π), then

$$\gamma_s^i(\eta[h]) = \gamma_s^i(\pi[h]) = \gamma_s^i(\pi) = \widetilde{\gamma}_s^i(x^*), \tag{21}$$

where for the second equality we used that for σ_k the "continuation rewards" remain the same due to lemma 10. Hence, according to property (3) in step 1.3 above, we have

$$\gamma_s^i(\eta[h]) \geq w_s^i \tag{22}$$

for all players i and for such histories h and states $s \in S$. Since $w_s^i \geq v_s^i$, as is proven in step 1.1, we conclude that η yields individually rational rewards in G, i.e. $\gamma_s^i(\eta[h]) \geq v_s^i$ for all players i and for such histories h and states $s \in S$.

It remains to show that η is a 0-equilibrium in G. Notice first that no deviation which only uses actions that had a positive probability according to η can improve the expected reward of any player. Indeed, (i) within a set E_k belonging to case (2) in step 2, the players play the 0-equilibrium σ_k in $\bar{G}_k$, (ii) within a set E_k belonging to case (3) in step 2, such deviation by players $-i$ (who do not make the "exit") would not change the probability of eventually moving to another set $E_{k'}$, (iii) within a set E_k belonging to case (3) in step 2, player i has an incentive to "exit" (as is already pointed out in step 2), since within $\bar{G}_k$ he can get at most $\bar{v}_k^i$, and $\bar{v}_k^i \leq z_k^i < w_k^i$, (iv) in states belonging to case (1) in step 2, no player i can go to better states regarding his reward according to equalities (21) and to property (1) from step 1.3.

So, consider now a deviation when, for the first time, say after history h in state s, when the players should play the joint mixed action x'_s according to η, a player i deviates and plays an action b_s^i which has probability zero according to η^i, i.e. $x'^i_s(b_s^i) = 0$. This deviation is immediately noticed by players $-i$ and, according to η, they switch to the joint stationary strategy y^{-i} from the next state, say state t. Consequently, player i's reward will be at most v_t^i in expectation. Obviously, without deviation player i would receive reward $\gamma_s^i(\eta[h]) = \widetilde{\gamma}_s^i(x^*)$, in view of equalities (21). Now, observe the following.

(A) Suppose s is a state in which s^j is of type 1 for at least one player j (possibly

$j = i$). Then, $x'_s = x^*_s$, and player i's expected reward after this deviation is at most

$$\begin{aligned}\sum_{t\in S} p_{s,(b^i_s,x^{*-i}_s)}(t)\, v^i_t &\le \sum_{t\in S} p_{s,(b^i_s,x^{*-i}_s)}(t)\, w^i_t \\ &\le \sum_{t\in S} p_{s,(b^i_s,x^{*-i}_s)}(t)\, \widetilde{\gamma}^i_t(x^*) \\ &\le \sum_{t\in S} p_{sx^*_s}(t)\, \widetilde{\gamma}^i_t(x^*) \\ &= \widetilde{\gamma}^i_s(x^*) \\ &= \gamma^i_s(\eta[h]),\end{aligned}$$

where the first inequalilty follows from $v^i \le w^i$ as is pointed out in step 1.1; the second and the third inequalities follow from properties (3) and (1) in step 1.3, respectively; then the equalities follow from (4) and (21). Hence, the deviation is not profitable.

(B) Suppose $s \in E_k$ for some joint maximal communicating set E_k. Then, player i's expected reward after this deviation is at most

$$\sum_{t\in S} p_{s,(b^i_s,x'^{-i}_s)}(t)\, v^i_t \le \sum_{t\in S} p_{s,(b^i_s,x'^{-i}_s)}(t)\, w^i_t \le \sum_{t\in S} p_{sx'_s}(t)\, w^i_t \le \gamma^i_s(\eta[h]),$$

where the first inequalilty follows from $v^i \le w^i$ as is pointed out in step 1.1, the second inequality follows from parts (1) of lemmas 5 and 6 for the game $\widetilde{G}$, while the last inequality from inequalities (22). Hence, the deviation is not profitable again.

In conclusion, no deviation is profitable, and η is a 0-equilibrium in G. This completes the proof of Main Theorem 1. ■

Remark 11 *It remains unclear whether 0-equilibria always exist within the class of stationary strategies. This question is already challenging in the situation when each player i's state space S^i is just one maximal communicating set (precisely the situation we have in a restricted game), meaning that S is one joint maximal communicating set. Even though, corollary 8 would yield that all minmax-levels are constant on the whole state space S, it is still not evident how one should get a grip on the problem.*

4 Periodic product-games

The previous sections dealt with aperiodic product-games. When we allow for periodic maximal communicating sets, the situation changes. Take for example a product-game with two players in which the Markov transition structure for either player is as follows:

the state space is $\{1,2\}$, there is only one action in either state, and this action leads to the other state with probability 1. So in the product-game, depending on the initial state, play moves back and forth either between states $(1,1)$ and $(2,2)$ or between states $(1,2)$ and $(2,1)$. This game is periodic, of course. Suppose the payoffs for either player are 1 in states $(1,1)$ and $(2,2)$, while 0 in states $(1,2)$ and $(2,1)$. Then, a solitary move for player 1 in state $(1,2)$ would lead to state $(2,2)$, improving player 1's payoff. Hence, the important lemma 2 is no longer valid for periodic product-games, and the proof in the previous sections are not directly applicable. Notice also that this game has two joint maximal communicating sets, i.e. $\{(1,1),(2,2)\}$ and $\{(1,2),(2,1)\}$, but neither of them can be written as a product of the form $E^1 \times E^2$. This entails additional difficulties, and makes the analysis more technical. Nevertheless, we conjecture that the main results of this paper extend to the periodic case as well.

5 Appendix

Lemma 12 *In a stochastic game, if the ergodic sets are the same for all joint stationary strategies, then there exists a stationary 0-equilibrium.*

Proof. For a joint stationary strategy $x \in X$, consider the β-discounted reward, with $\beta \in (0,1)$, defined for player i and initial state $s \in S$ as

$$\gamma^i_{\beta s}(x) := (1-\beta) \sum_{m=1}^{\infty} \beta^{m-1} \mathbb{E}_{sx}\left(R^i_m\right),$$

where R^i_m is the random variable for the payoff for player i at stage m, and where $\mathbb{E}_{sx}$ stands for expectation with respect to initial state s and joint strategy x. Fink [1964] and Takahashi [1964] showed that, for every $\beta \in (0,1)$, there exists a stationary 0-equilibrium with respect to the β-discounted rewards.

As the ergodic sets are the same for all joint stationary strategies, it is known (cf. lemma 2.7.6 in Flesch [1998]) that for any sequence of discount factors β_m converging to 1 and joint strategies x_m converging to x we have

$$\gamma^i_s(x) = \lim_{m\to\infty} \gamma^i_{\beta_m s}(x_m) \tag{23}$$

for all states $s \in S$ and players i.

We will now work with a number of sequences in compact spaces. By taking subsequences, we may assume that all these sequences have limits. Let β_m be a sequence of discount factors converging to 1, and for any $m \in \mathbb{N}$, let x_m be a stationary β_m-discounted 0-equilibrium. Let $x = \lim_{m\to\infty} x_m$. We will show that x is a 0-equilibrium with respect to the average reward.

Take an arbitrary player i and a stationary best reply y^i to x^{-i}. Then for any initial state $s \in S$, from (23) and from the fact that x_m is a β_m-discounted 0-equilibrium, it follows that

$$\gamma^i_s(y^i, x^{-i}) = \lim_{m\to\infty} \gamma^i_{\beta_m s}(y^i, x^{-i}_m) \leq \lim_{m\to\infty} \gamma^i_{\beta_m s}(x_m) = \gamma^i(x).$$

As y^i is a best reply to x^{-i}, the joint strategy x is a stationary 0-equilibrium with respect to the average reward indeed. ■

6 References

Altman E, Avrachenkov K, Marquez R & Miller G [2005]: Zero-sum constrained stochastic games with independent state processes. *Mathematical Methods of Operations Research* 62, 375-386.

Bewley T & Kohlberg E [1978]: On stochastic games with stationary optimal strategies. *Mathematics of Operations Research* 3, 104-125.

Blackwell D & Ferguson TS [1968]: The big match. *Annals of Mathematical Statistics* 39, 159-163.

Doob JL [1953]: *Stochastic processes.* Wiley, New York.

Dutta PK [1995]: A Folk theorem for stochastic games. *Journal of Economic Theory* 66, 1-32.

Flesch J,Thuijsman F & Vrieze OJ [2007]: Stochastic games with additive transitions. *European Journal of Operational Research* (forthcoming).

Flesch J [1998]: *Stochastic games with the average reward.* PhD Thesis, University of Maastricht, the Netherlands.

Fink AM [1964]: Equilibrium in a stochastic n-person game. *Journal of Science of Hiroshima University,* Series A-I 28, 89-93.

Gillette D [1957]: Stochastic games with zero stop probabilities. In: Dresher M, Tucker AW & Wolfe P (eds.), Contributions to the theory of games III, *Annals of Mathematical Studies* 39, Princeton University Press, 179-187.

Hordijk A, Vrieze OJ & Wanrooij GL [1983]: Semi-Markov strategies in stochastic games. *International Journal of Game Theory* 12, 81-89.

Mertens JF & Neyman A [1981]: Stochastic games. *International Journal of Game Theory* 10, 53-66.

Ross KW & Varadarajan R [1991]: Multichain Markov decision processes with a sample path constraint: A decomposition approach. *Mathematics of Operations Research* 16, 195-207.

Sorin S [1986]: Asymptotic properties of a non-zerosum game. *International Journal of Game Theory,* 15, 101-107.

Takahashi M [1964]: Equilibrium points of stochastic noncooperative n-person games. *Journal of Science of Hiroshima University,* Series A-I 28, 95-99.

Thuijsman F & Vrieze OJ [1991]: Easy initial states in stochastic games. In: Raghavan TES, Ferguson TS, Vrieze OJ & Parthasarathy T (eds.), *Stochastic Games and Related Topics*, Kluwer, Dordrecht, 85-100.

Vieille N [2000-a]: Equilibrium in 2-person stochastic games I: A reduction. *Israel Journal of Mathematics,* 119, 55-91.

Vieille N [2000-b]: Equilibrium in 2-person stochastic games II: The case of recursive games. *Israel Journal of Mathematics,* 119, 93-126.