

Stochastic games on a product state space.

Citation for published version (APA):

Flesch, J., Schoenmakers, G. M., & Vrieze, K. (2008). Stochastic games on a product state space. *Mathematics of Operations Research*, 33, 403-420. <https://doi.org/10.1287/moor.1070.0304>

Document status and date:

Published: 01/01/2008

DOI:

[10.1287/moor.1070.0304](https://doi.org/10.1287/moor.1070.0304)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

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

[Link to publication](#)

General rights

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

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

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

www.umlib.nl/taverne-license

Take down policy

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

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.



Mathematics of Operations Research

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

Stochastic Games on a Product State Space

János Flesch, Gijs Schoenmakers, Koos Vrieze,

To cite this article:

János Flesch, Gijs Schoenmakers, Koos Vrieze, (2008) Stochastic Games on a Product State Space. Mathematics of Operations Research 33(2):403-420. <https://doi.org/10.1287/moor.1070.0304>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2008, INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Stochastic Games on a Product State Space

János Flesch

Department of Quantitative Economics, Maastricht University, 6200 MD Maastricht, The Netherlands,
J.Flesch@ke.unimaas.nl

Gijs Schoenmakers, Koos Vrieze

Department of Mathematics, Maastricht University, 6200 MD Maastricht, The Netherlands
 {GM.Schoenmakers@micc.unimaas.nl, kvrieze@ictvrieze.nl}

We examine so-called product-games with an aperiodic transition structure, with respect to the average reward, for which we present an approach based on communicating states. For the general n -player case, we establish the existence of 0-equilibria. 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.

Key words: noncooperative games; stochastic games; Markov decision problems; equilibria

MSC2000 subject classification: Primary: 91A15

ORMS subject classification: Primary: stochastic games

History: Received March 16, 2007; revised May 7, 2007 and November 10, 2007.

1. Introduction.

Stochastic games and product-games. An n -player *stochastic game* is given by (1) a set of players $N = \{1, \dots, n\}$, (2) a nonempty and finite set of states S , (3) for each state $s \in S$, a nonempty and finite set of actions A_s^i for each player i , (4) for each state $s \in S$ and each joint action $a_s \in \times_{i \in N} A_s^i$, a payoff $r_s^i(a_s) \in \mathbb{R}$ to each player i , (5) for each state $s \in S$ and each joint action $a_s \in \times_{i \in N} A_s^i$, 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 $i \in N$ has to choose an action a_1^i from his action set $A_{s_1}^i$. These choices have to be made independently. The chosen joint action $a_1 = (a_1^1, \dots, a_1^n)$ induces an immediate payoff $r_{s_1}^i(a_1)$ 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 $a_2^i \in A_{s_2}^i$ has to be chosen by each player i in state s_2 . Then, given action combination $a_2 = (a_2^1, \dots, a_2^n)$, player i receives payoff $r_{s_2}^i(a_2)$ 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 $A_{s^i}^i$ for each state $s^i \in S^i$; (3) a transition probability distribution $p_{s^i a^i}^i$ over the state space S^i for each state $s^i \in S^i$ and for each action $a^i \in A_{s^i}^i$. 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, called product-games, in which the transition structure is derived by taking the product of n Markov transition structures. 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 $S = S^1 \times \dots \times S^n$; (3) the action set for each player $i \in N$ in each state $s = (s^1, \dots, s^n) \in S$ is $A_s^i = A_{s^i}^i$; (4) the transition probability distribution p_{sa_s} , for each state $s = (s^1, \dots, s^n) \in S$ and for each joint action $a_s = (a_s^1, \dots, a_s^n) \in \times_{i \in N} A_s^i$, is

$$p_{sa_s}(\bar{s}) = \prod_{i \in N} p_{s^i a_s^i}^i(\bar{s}^i)$$

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_s^i chosen by player i ; i.e. for any $\bar{s}^i \in S^i$ we have

$$p_{sa_s}(S^1, \dots, S^{i-1}, \bar{s}^i, S^{i+1}, \dots, S^n) = p_{s^i a_s^i}^i(\bar{s}^i).$$

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 $\Gamma^1, \dots, \Gamma^n$, which are linked by payoff functions r^1, \dots, 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. [1], although in a somewhat different fashion. They only examine two-player games in which the sum of the payoffs is always equal to zero (zero-sum games), and drop 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 show 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 that 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. [8]); 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 $\times_{s \in S} A_s^i$. 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_s^i for player i in state $s \in S$ is a probability distribution on A_s^i . The set of mixed actions for player i in state s is denoted by X_s^i . 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 $\pi_s^i(h) \in X_s^i$ 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 $\pi_s^i[h](h')$ in any present state s for any history h' as if h had happened before h' ; i.e., $\pi_s^i[h](h') = \pi_s^i(h \oplus h')$, 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 $X^i = \times_{s \in S} X_s^i$. We use the notation x^i for stationary strategies for player i , while x_s^i 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 $A^i = \times_{s \in S} A_s^i$.

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 $p_{s\bar{s}}(x)$ 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 to that state 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, 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

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

the limit is known to exist (cf., Doob [4], Theorem 2.1, p. 175). Entry (s, \bar{s}) of the stochastic matrix $Q(x)$, denoted by $q_{s\bar{s}}(x)$, 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 [4]) that

$$Q(x) = Q(x)P(x) = P(x)Q(x) = Q^2(x). \quad (2)$$

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

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

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 . We wish to remark that all our further results hold for the limsup as well.

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

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

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

Every player i has a pure stationary best reply against any fixed joint stationary strategy of his opponents (cf., Blackwell [2], Hordijk et al. [11]); 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 min max 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 . It is thus also the lowest reward that the other players can inflict on player i . Note that, against different joint strategies of players $-i$, player i may have to use different strategies to defend his min max level. It is known that the min max 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 and Vrieze [23] (their proof is given for only two players but directly extends to the n -player case in combination with Neyman [14], who showed that the min max levels equal the limit of the discounted min max levels in n -player stochastic games), there always exists an initial state s in the set $\{t \in S \mid v_t^i = \min_{t' \in S} v_{t'}^i\}$ 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

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

which means that no player can gain more than ε by a unilateral deviation. 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 min max 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 [10] and solved by Blackwell and Ferguson [3], and the game in Sorin [21] 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 [24, 25] managed to establish the existence of ε -equilibria, for all $\varepsilon > 0$. However, little is known about n -player stochastic games, and it is unresolved whether they always have ε -equilibria, for all $\varepsilon > 0$. This is probably the most challenging open problem in the field of stochastic games.

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.

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 and Neyman [13] 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 . We also refer to Maitra and Sudderth [12] for the existence of the value in a very general setup.

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 [13] proved (even in a stronger form) that both players have ε -optimal strategies for any $\varepsilon > 0$.

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 §2, we will discuss preliminary concepts and results. Then, in §3, we will present and prove the two main theorems, and provide illustrative examples. Finally, §4 concludes with a short discussion on the case of periodic product-games.

2. Preliminary concepts and results. Some of the content of this section is very similar to the decomposition presented in Ross and Varadarajan [15] for Markov decision problems (i.e., stochastic games with only one player).

Classification of states. As we know, in a product-game, each player i has a Markov transition structure Γ^i of his own. We start by analysing each Γ^i separately. We would like to emphasise that such a separate analysis is only possible because, in a product-game, each player affects the transitions of his own states only. We distinguish between two basic types of states in the state space S^i of Γ^i . A state $s^i \in S^i$ belongs to type 1 if it is transient for each stationary strategy x^i of player i . Otherwise, s^i belongs to type 2, in which case player i has a stationary strategy for which s^i is recurrent.

Maximal communicating sets. Two states s_1^i and s_2^i of type 2 are said to communicate with each other, if there exists a stationary strategy x^i of player i such that s_1^i and s_2^i belong to the same ergodic set. We note that communication between states have been used extensively in the literature of stochastic games (cf., Vieille [24, 25], Solan and Vieille [19], Solan [18]).

This relationship of communication is an equivalence relation on the set of states of type 2. As such, it induces equivalence classes, which for obvious reasons are called maximal communicating sets. 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 of 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 of the initial state and player i 's strategy; (4) there is always at least one among the maximal communicating sets which player i is unable to leave; i.e., there are no transitions to states outside; (5) for any strategy of player i , regardless of the initial state, 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 $E_{k^i}^i$, where $k^i \in K^i$, denote the maximal communicating sets for player i . Further, let $K := \times_{i=1}^n K^i$. For any $k = (k^1, \dots, k^n) \in K$, the product $E_k := \times_{i=1}^n E_{k^i}^i$ is called a joint maximal communicating set.

In every state s^i of the communicating set $E_{k^i}^i$, for every $k^i \in K^i$, let $\bar{A}_{s^i}^i$ denote the set of those actions $a_{s^i}^i \in A_{s^i}^i$ which keep play in $E_{k^i}^i$ with probability 1. The sets $\bar{A}_{s^i}^i$ are clearly nonempty. For every state $s = (s^1, \dots, s^n) \in S$, we also let $\bar{A}_s^i := \bar{A}_{s^i}^i$.

Aperiodicity. A maximal communicating set $E_{k^i}^i$ of player i is called aperiodic, if the Markov chain associated to a stationary strategy x^i of player i that only uses completely mixed actions on $\bar{A}_{s^i}^i$ for all $s^i \in E_{k^i}^i$, is aperiodic on $E_{k^i}^i$. (Obviously, this definition is independent of the particular choice of x^i .) The importance of the notion of aperiodicity lies in the fact that for such a completely mixed stationary strategy, after sufficiently many stages, the system can be in any state $s^i \in E_{k^i}^i$ with positive probability, regardless of the initial state in $E_{k^i}^i$.

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 §4.

Restricted games. Take an aperiodic product-game and some $k = (k^1, \dots, k^n) \in K$. 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}_s^i , we obtain a restricted game \bar{G}_k . Obviously, \bar{G}_k is an aperiodic product-game itself.

These restricted games play a key role in the analysis of product-games because of the following observation. As is pointed out above, for any initial state and strategy of the players, each player i eventually settles in one of his maximal communicating sets $E_{k^i}^i$, 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.

For a restricted game \bar{G}_k , let $\bar{v}_{k,s}^i$ denote the minmax level of player i in \bar{G}_k for initial state $s \in E_k$. If, for some player i , the inequality $\bar{v}_{k,s}^i \geq v_s^i$ holds for all initial states $s \in E_k$, then we call \bar{G}_k *satisfactory* to player i . Otherwise, \bar{G}_k is called *unsatisfactory* to player i . In words, if \bar{G}_k is satisfactory to player i , then player i weakly prefers \bar{G}_k to G , as far as his minmax level is concerned on E_k . Let K^* denote the set of indices $k \in K$ such that \bar{G}_k is satisfactory to all players. Further, let $K[i]$ denote the set of indices $k \in K$ such that \bar{G}_k is unsatisfactory to player i but satisfactory to all players $j \in \{1, \dots, i-1\}$. Obviously, $K^*, K[1], \dots, K[n]$, forms a partition of K .

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, and the actions of player 2 by the columns. Therefore, 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 $A_1^1 = A_2^1 = \{1, 2\}$, $A_3^1 = \{1\}$, and transitions

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

Therefore, 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 $E_I^1 := \{1, 2\}$ and $E_{II}^1 := \{3\}$ are maximal communicating sets, with index-set $K^1 = \{I, II\}$. Moreover, both E_I^1 and E_{II}^1 are aperiodic. As for the actions which keep play in these maximal communicating sets, we obtain $\bar{A}_1^1 = \{1, 2\}$, $\bar{A}_2^1 = \{1\}$, $\bar{A}_3^1 = \{1\}$.

The underlying Markov transition structure for player 2 is given by state space $S^2 = \{1, 2\}$, action sets $A_1^2 = \{1, 2\}$, $A_2^2 = \{1\}$, and transitions $p_{11}^2 = (1, 0)$, $p_{12}^2 = (0, 1)$, $p_{21}^2 = (0, 1)$. Further, both $E_I^2 := \{1\}$ and $E_{II}^2 := \{2\}$ are aperiodic maximal communicating sets, with $K^2 = \{I, II\}$ and $\bar{A}_1^2 = \bar{A}_2^2 = \{1\}$.

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).

Later we will examine which restricted games are satisfactory to the players.

3. The main results and the proof.

3.1. The main theorems. For the class of aperiodic 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.*

The idea of the construction of an equilibrium η , for an aperiodic product-game G , is as follows. The equilibrium η will prescribe to follow a joint strategy π , unless some player i deviates from π^i by playing an action outside the support of π^i (i.e., 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 a joint stationary strategy y^{-i} and push down

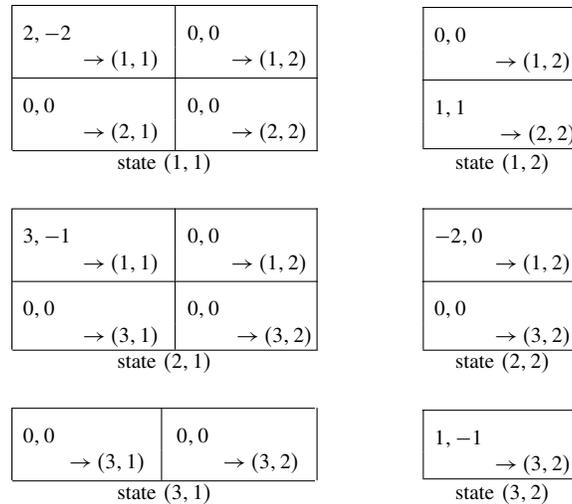


FIGURE 1. Game of Example 1.

player i 's reward to his min max level v_s^i . In fact, y^{-i} acts as a threat strategy, whose task is to force player i to follow the prescriptions of π^i . Punishment with y^{-i} will be shown to be severe enough. Finally, our construction will guarantee that no deviation inside the support of π^i (such deviations are hard to detect) is profitable for any player i .

Now we briefly describe the construction of π , which shows a number of similarities with the construction in Vieille [24, 25]. The joint strategy π prescribes to play in the following way:

(1) When entering some E_k , with $k \in K^*$ (i.e., the restricted game \bar{G}_k is satisfactory to all players): In this case, π will prescribe to stay on E_k and play a certain equilibrium in \bar{G}_k . Here, the players collect “high” payoffs (cf., solvable sets in Vieille [24, 25].)

(2) When entering some E_k , with $k \in K[i]$ (i.e., the restricted game \bar{G}_k is unsatisfactory to player i): In this case, π will prescribe player i to exit E_k^i (and thereby to leave E_k), while all other players patiently wait for player i 's exit. It will be taken care of that no player's min max level drops in expectation. Payoffs in E_k are disregarded (cf., controlled sets in Vieille [24, 25]).

(3) Outside all joint maximal communicating sets: In this case, π will let the players play for their future perspectives. Payoffs in these states are disregarded.

Note that, according to π , play will surely settle in a restricted game belonging to case (1).

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.*

3.2. The proofs of Main Theorems 1 and 2. In this section, we provide a proof for Main Theorems 1 and 2. We will focus on Main Theorem 1, as Main Theorem 2 will follow (cf., the end of §3.2.3) along the way without major additional difficulties. Since an equilibrium reward for player i , from any initial state s , is at least his min max level v_s^i , we start by studying the min max levels of product-games.

3.2.1. The structure of the min max levels of the restricted games.

LEMMA 3.1. *Let G be an arbitrary aperiodic product-game and consider the restricted game \bar{G}_k , for any $k = (k^1, \dots, k^n) \in K$, and an arbitrary player i . Then, the min max level \bar{v}_k^i of any player i in \bar{G}_k is constant; i.e., $\bar{v}_{k,s}^i = \bar{v}_{k,t}^i$ ($=: \bar{v}_k^i$) 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 min max level $\bar{v}_{k,s}^i$; i.e., for all strategies π^i for player i in \bar{G}_k we have*

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

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 $\alpha^i := \min_{t \in E_k} \bar{v}_{k,t}^i$. As is mentioned in the introduction, by applying Thuijsman and Vrieze [23] together with Neyman [14] for the game \bar{G}_k , there exists a state $s' \in \{t \in E_k \mid \bar{v}_{k,t}^i = \alpha^i\}$ 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

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

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

$$\bar{\gamma}_s^i(\pi^i, x^{-i}) \leq \alpha^i. \quad (8)$$

Let x^i be a completely mixed stationary strategy in \bar{G}_k for player i . Consider the joint stationary strategy (x^i, x^{-i}) . By the definitions of Z and x^{-i} , the set Z is closed for (x^i, x^{-i}) (i.e., play does not leave Z). Hence, there must exist an ergodic set $F \subset Z$ for (x^i, x^{-i}) . Due to the choice of x^i and the aperiodicity of E_{ki}^i , if $u \in F$, then $(t^i, u^{-i}) \in F$ for all states $t^i \in E_{ki}^i$. Thus, F must be of the form $F = \bar{F} \times E_{ki}^i$ for some nonempty $\bar{F} \subset E_{k^{-i}} = \times_{j \in N - \{i\}} E_{kj}^j$.

Define a joint stationary strategy y^{-i} for players $-i$ in \bar{G}_k as follows: let $y_t^{-i} = x_t^{-i}$ for all $t \in F$ and let y_t^{-i} be an arbitrary completely mixed action on \bar{A}_t^{-i} for all $t \in (E_k - F)$. Now, take an arbitrary strategy π^i for player i in \bar{G}_k and consider (π^i, y^{-i}) with any initial state in E_k . Based on the aperiodicity of E_{kj}^j for all $j \neq i$ and the choice of y^{-i} outside F , players $-i$ eventually visit \bar{F} , and hence play eventually visits F . Then, as F is closed with respect to (x^i, x^{-i}) and y^{-i} equals x^{-i} on F , once play reaches F , it will never leave it and, in view of (8), player i 's reward will be at most α^i . Consequently, the min max level \bar{v}_k^i of player i in \bar{G}_k equals the constant α^i , and y^{-i} satisfies the second part of the lemma. \square

As an illustration, we now revisit the game in Example 1. First, take 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 min max level $\bar{v}_{(I,I)}^1$ 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), giving 3/2 on average, before returning to state (1, 1). Hence, player 1's min max level is $\bar{v}_{(I,I)}^1 = 2$, whereas, for similar reasons, player 2's min max level is $\bar{v}_{(I,I)}^2 = -2$.

For the other three restricted games, we have $\bar{v}_{(I,II)}^1 = \bar{v}_{(I,II)}^2 = 0$, $\bar{v}_{(II,I)}^1 = \bar{v}_{(II,I)}^2 = 0$, and $\bar{v}_{(II,II)}^1 = 1$, $\bar{v}_{(II,II)}^2 = -1$.

3.2.2. The min max 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 $r_s^i(a_s) = r_t^i(b_t)$ for any states $s, t \in E_k$ and for any joint actions $a_s \in \bar{A}_s$, $b_t \in \bar{A}_t$.

By a *solitary move* of player i in state $s \in S$ we mean a step where only player i chooses an action, say $a_s^i \in A_s^i$, and where the process moves to a new state of the form (t^i, s^{-i}) , $t^i \in S^i$, with probability $p_{s^i a_s^i}^i(t^i)$. A solitary move for players $-i$ is defined similarly.

The following lemma claims that player i is unable to improve on his min max level v^i in expectation by a solitary move. Similarly, solitary moves by players $-i$ cannot decrease player i 's min max level v^i in expectation.

LEMMA 3.2. *Let G be a simple aperiodic product-game. Take an arbitrary player i and a state $s = (s^1, \dots, s^n) \in S$.*

(1) *For any action $a_s^i \in A_s^i$ of player i , it holds that*

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

(2) *For any joint action $a_s^{-i} \in A_s^{-i}$ of players $-i$, it holds that*

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

PROOF. We only show part (1) of the lemma; part (2) can be proven similarly. Let $\varepsilon > 0$. The idea of the proof is as follows. We compare two specific games Ω and $\tilde{\Omega}$. The game Ω is simply the original game G with initial state s , while $\tilde{\Omega}$ is the game that results from G after the execution of the solitary move a_s^i by player i

in state s ; i.e., $\tilde{\Omega}$ starts in initial state (t^i, s^{-i}) , $t^i \in S^i$, with probability $p_{s^i a_s^i}^i(t^i)$. We define two joint strategies π for Ω and $\tilde{\pi}$ for $\tilde{\Omega}$ in such a way that the following properties hold:

Property (a) for Ω : $\gamma_s^i(\pi) \leq v_s^i + \varepsilon$.

Property (b) for $\tilde{\Omega}$: $\gamma_{(t^i, s^{-i})}^i(\tilde{\pi}) \geq v_{(t^i, s^{-i})}^i - \varepsilon$, for any $t^i \in S^i$.

Property (c): π and $\tilde{\pi}$ yield the same expected rewards in Ω and $\tilde{\Omega}$, respectively; i.e.,

$$\gamma_s^i(\pi) = \sum_{t^i \in S^i} p_{s^i a_s^i}^i(t^i) \gamma_{(t^i, s^{-i})}^i(\tilde{\pi}).$$

It will then follow from properties (a), (b), and (c) that

$$\begin{aligned} v_s^i + \varepsilon &\geq \gamma_s^i(\pi) = \sum_{t^i \in S^i} p_{s^i a_s^i}^i(t^i) \gamma_{(t^i, s^{-i})}^i(\tilde{\pi}) \\ &\geq \sum_{t^i \in S^i} p_{s^i a_s^i}^i(t^i) \cdot (v_{(t^i, s^{-i})}^i - \varepsilon) = \sum_{t^i \in S^i} p_{s^i a_s^i}^i(t^i) \cdot v_{(t^i, s^{-i})}^i - \varepsilon. \end{aligned} \quad (9)$$

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

Step 1. The construction of π for Ω and $\tilde{\pi}$ for $\tilde{\Omega}$. For notational purposes, in this proof we assume that the histories also contain the present state, so a history up to stage m is in the form $h_m = (s_1, a_1; \dots; s_{m-1}, a_{m-1}; s_m)$. When h_m is used in the sequel, h_{m-1} will denote the restriction of h_m up to stage $m-1$. It is useful to separate each h_m into the coordinates belonging to player i and to players $-i$, respectively; i.e., let $h_m^i = (s_1^i, a_1^i; \dots; s_{m-1}^i, a_{m-1}^i; s_m^i)$ and let $h_m^{-i} = (s_1^{-i}, a_1^{-i}; \dots; s_{m-1}^{-i}, a_{m-1}^{-i}; s_m^{-i})$. Let H_m denote the set of histories for Ω up to stage m . Define \tilde{H}_m similarly for $\tilde{\Omega}$. By separating the coordinates, we obtain the sets $H_m^i, H_m^{-i}, \tilde{H}_m^i, \tilde{H}_m^{-i}$. As players $-i$ begin in s^{-i} in both Ω and $\tilde{\Omega}$, we have $H_m^{-i} = \tilde{H}_m^{-i}$ for all m .

Now we define a transformation $\phi: \tilde{H}_m^i \rightarrow H_m^i$. For $h_m^i \in \tilde{H}_m^i$, with $m \geq 2$, let

$$\phi(h_m^i) = (s^i, a_s^i; s_1^i, a_1^i; \dots; s_{m-2}^i, a_{m-2}^i; s_{m-1}^i).$$

Naturally, we define $\phi(h_1^i) = s^i$.

We start with the strategies for players $-i$. By the definition of the minmax level v_s^i , there exists a joint strategy π^{-i} of players $-i$ such that player i 's reward cannot be more in Ω than $v_s^i + \varepsilon$; 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 $\tilde{\pi}^{-i}$ in $\tilde{\Omega}$ for players $-i$ by $\tilde{\pi}^{-i}(h_m^i, h_m^{-i}) = \pi^{-i}(\phi(h_m^i), h_m^{-i})$ for all $(h_m^i, h_m^{-i}) \in \tilde{H}_m$.

Now we define the strategies for player i . By the definition of the minmax level again, there exists a $\tilde{\pi}^i$ for player i in $\tilde{\Omega}$ such that $\gamma_u^i(\tilde{\pi}^i, \tilde{\pi}^{-i}) \geq v_u^i - \varepsilon$ for all initial states $u = (t^i, s^{-i})$, $t^i \in S^i$, in $\tilde{\Omega}$. Finally, let π^i be defined for Ω as $\pi^i(h_m^i, h_m^{-i}) = \tilde{\pi}^i(\psi(h_m^i), h_m^{-i})$, for all $(h_m^i, h_m^{-i}) \in H_m$ with $m \geq 2$, where

$$\psi(h_m^i) = (s_2^i, a_2^i; \dots; s_{m-1}^i, a_{m-1}^i; s_m^i).$$

Naturally, we define $\pi^i(h_1^i, h_1^{-i}) = a_s^i$. Note that

$$\pi^i(\phi(h_m^i), h_m^{-i}) = \tilde{\pi}^i(h_{m-1}^i, h_{m-1}^{-i}) \quad (10)$$

for all histories $(h_m^i, h_m^{-i}) \in \tilde{H}_m$.

Step 2. Proving Properties (a), (b), and (c) for π and $\tilde{\pi}$. Note that Properties (a) and (b) are satisfied, by the definitions of π^{-i} and $\tilde{\pi}^i$. So, it remains to verify Property (c).

We will show that the stochastic processes induced by π in Ω and by $\tilde{\pi}$ in $\tilde{\Omega}$ are strongly related, namely

$$\mathbb{P}_{\tilde{\pi}}(h_m^i, h_{m+1}^{-i}) = \mathbb{P}_{\pi}(\phi(h_{m+1}^i), h_{m+1}^{-i}) \quad (11)$$

for all $(h_{m+1}^i, h_{m+1}^{-i}) \in \tilde{H}_{m+1}$. Here, the left-hand side should be interpreted as the probability, with respect to $\tilde{\pi}$ in $\tilde{\Omega}$, that the i -coordinate of the history up to stage m coincides with h_m^i and the $-i$ -coordinate of the history up to stage $m+1$ coincides with h_{m+1}^{-i} . The interpretation of the right-hand side is analogous.

The proof goes by complete induction.

For $m = 1$, with $h_1^i = (t^i)$ and $h_2^{-i} = (s^{-i}, a_1^{-i}, s_2^{-i})$, we have

$$\mathbb{P}_{\tilde{\pi}}(h_1^i, h_2^{-i}) = p_{s^i a_s^i}^i(t^i) \cdot \tilde{\pi}^{-i}(h_1^i, h_1^{-i})(a_1^{-i}) \cdot p_{s^{-i} a_1^{-i}}^{-i}(s_2^{-i})$$

and also with $\phi(h_2^i) = (s^i, a_s^i, t^i)$,

$$\mathbb{P}_{\pi}(\phi(h_2^i), h_2^{-i}) = p_{s^i a_s^i}^i(t^i) \cdot \pi^{-i}(s^i, s^{-i})(a_1^{-i}) \cdot p_{s^{-i} a_1^{-i}}^{-i}(s_2^{-i}).$$

Since $\tilde{\pi}^{-i}(h_1^i, h_1^{-i}) = \pi^{-i}(\phi(h_1^i), h_1^{-i}) = \pi^{-i}(s^i, s^{-i})$, the equality (11) holds for $m = 1$.

Suppose that equality (11) is valid for a certain m . For $m + 1$ we obtain

$$\mathbb{P}_{\tilde{\pi}}(h_{m+1}^i, h_{m+2}^{-i}) = \mathbb{P}_{\tilde{\pi}}(h_m^i, h_{m+1}^{-i}) \cdot \tilde{\pi}^i(h_m^i, h_m^{-i})(a_m^i) \cdot p_{s_m^i a_m^i}^i(s_{m+1}^i) \cdot \tilde{\pi}^{-i}(h_{m+1}^i, h_{m+1}^{-i})(a_{m+1}^{-i}) \cdot p_{s_{m+1}^{-i} a_{m+1}^{-i}}^{-i}(s_{m+2}^{-i})$$

and

$$\begin{aligned} \mathbb{P}_{\pi}(\phi(h_{m+2}^i), h_{m+2}^{-i}) &= \mathbb{P}_{\pi}(\phi(h_{m+1}^i), h_{m+1}^{-i}) \cdot \pi^i(\phi(h_{m+1}^i), h_{m+1}^{-i})(a_m^i) \cdot p_{s_m^i a_m^i}^i(s_{m+1}^i) \\ &\quad \cdot \pi^{-i}(\phi(h_{m+1}^i), h_{m+1}^{-i})(a_{m+1}^{-i}) \cdot p_{s_{m+1}^{-i} a_{m+1}^{-i}}^{-i}(s_{m+2}^{-i}). \end{aligned}$$

From our assumption that (11) holds for m , from equality (10), and from the definition of $\tilde{\pi}^{-i}$, it follows that (11) holds for $m + 1$. Consequently, (11) holds for all m .

Recall that, with respect to any initial state and any joint strategy, play eventually settles, with probability 1, in a restricted game. By equality (11), the probability that this restricted game is some \tilde{G}_k with respect to π in Ω equals the probability that this is \tilde{G}_k with respect to $\tilde{\pi}$ in $\tilde{\Omega}$. Since the game is simple, the average reward is determined by this restricted game, and hence the expected reward for π in Ω equals the expected reward for $\tilde{\pi}$ in $\tilde{\Omega}$. This proves Property (c), and therefore the proof is complete. \square

LEMMA 3.3. *Take a simple aperiodic product-game G . Let E_{ki}^i denote a maximal communicating set for player i , for some $k^i \in K^i$, and $E_{k^{-i}}^{-i}$ denote a joint maximal communicating set for players $-i$, for some $k^{-i} \in K^{-i}$.*

(1) *For any two states $s^i, t^i \in E_{ki}^i$ of player i and any joint state $s^{-i} \in S^{-i}$ of players $-i$, the min max level of player i satisfies $v_{(s^i, s^{-i})}^i = v_{(t^i, s^{-i})}^i$.*

(2) *For any two joint states $s^{-i}, t^{-i} \in E_{k^{-i}}^{-i}$, of players $-i$ and any state $s^i \in S^i$ of player i , the min max level of player i satisfies $v_{(s^i, s^{-i})}^i = v_{(s^i, t^{-i})}^i$.*

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 $s^i \in E_{ki}^i$ for which $v_{(s^i, s^{-i})}^i \leq v_{(t^i, s^{-i})}^i$ for all $t^i \in E_{ki}^i$. Suppose by way of contradiction that $E_{ki}^i - F^i$ is not empty. Take a state $s^i \in F^i$ and an action $a_{s^i}^i \in \bar{A}_{s^i}^i$ which moves from state s^i to a state in $E_{ki}^i - F^i$ with a positive probability. Then, the solitary move $a_{s^i}^i$ in state (s^i, s^{-i}) for player i would improve player i 's min max level in expectation, which contradicts part (1) of Lemma 3.2. Hence, $F^i = E_{ki}^i$, and part (1) of the lemma follows. \square

COROLLARY 3.1. *Let G be a simple aperiodic product-game, and let E_k be a joint maximal communicating set for some $k \in K$. Then, the min max level v^i of any player i is constant on E_k ; i.e., $v_s^i = v_t^i (=: v_k^i)$ for all $s, t \in E_k$.*

LEMMA 3.4. *Let G be a simple aperiodic product-game. Then, for any player i the following properties hold.*

(1) *Let $s = (s^1, \dots, s^n)$ be a state such that s^i belongs to a maximal communicating set E_{ki}^i . Then, regardless of the joint mixed action x_s^{-i} played by players $-i$ in state s , all the actions in \bar{A}_s^i guarantee in expectation the best possible min max level for player i after transition; i.e., for any actions $a_s^i \in \bar{A}_s^i$ and $b_s^i \in A_s^i$ it holds that*

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

(2) *Let $s = (s^1, \dots, s^n)$ be a state such that s^j belongs to a maximal communicating set E_{kj}^j for all players $j \neq i$. Then, all joint actions in \bar{A}_s^{-i} for players $-i$ in state s guarantee in expectation that player i 's min max 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, by part (1) of Lemma 3.3 and part (1) of Lemma 3.2, we have for any $t^{-i} \in S^{-i}$ that

$$\sum_{t^i \in S^i} p_{s^i a_s^i}^i(t^i) v_{(t^i, t^{-i})}^i = v_{(s^i, t^{-i})}^i \geq \sum_{t^i \in S^i} p_{s^i b_s^i}^i(t^i) v_{(t^i, t^{-i})}^i.$$

Hence

$$\begin{aligned} \sum_{t \in S} p_{s, (a_s^i, x_s^{-i})}(t) v_t^i &= \sum_{t^{-i} \in S^{-i}} p_{s^{-i} x_s^{-i}}^{-i}(t^{-i}) \left[\sum_{t^i \in S^i} p_{s^i a_s^i}^i(t^i) v_{(t^i, t^{-i})}^i \right] \\ &\geq \sum_{t^{-i} \in S^{-i}} p_{s^{-i} x_s^{-i}}^{-i}(t^{-i}) \left[\sum_{t^i \in S^i} p_{s^i b_s^i}^i(t^i) v_{(t^i, t^{-i})}^i \right] \\ &= \sum_{t \in S} p_{s, (b_s^i, x_s^{-i})}(t) v_t^i. \end{aligned}$$

As for part (2) of Lemma 3.4, take arbitrary joint actions $a_s^{-i} \in \bar{A}_s^{-i}$ and $b_s^{-i} \in A_s^{-i}$ in state s . It follows similarly to part (1) that for all mixed actions x_s^i of player 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

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

which proves part (2). \square

Consider a restricted game \bar{G}_k , within a simple product-game, which is unsatisfactory to player i (i.e., $i \in K[i]$). The next lemma proposes a way for player i to exit \bar{G}_k . A similar result holds for players $-i$.

LEMMA 3.5. *Let G be a simple aperiodic product-game, and let E_k be a joint maximal communicating set for some $k = (k^1, \dots, k^n) \in K$. Let z_k^i denote player i 's unique reward in the restricted game \bar{G}_k , and let v_k^i be player i 's min max level on E_k in the game G (a constant, cf., Corollary 3.1).*

(1) *Suppose $z_k^i < v_k^i$. Then, there is a state $s^i \in E_{k^i}^i$ and an action $a_{s^i}^i \in A_{s^i}^i - \bar{A}_{s^i}^i$ for player i in state s^i such that if player i plays action $a_{s^i}^i$ in any state $s = (s^i, s^{-i}) \in E_k$, with $s^{-i} \in E_{k^{-i}}^{-i}$, then player i 's min max level cannot decrease in expectation from state s , regardless of the actions played by players $-i$. More precisely, for any $a_s^{-i} \in A_s^{-i}$ we have*

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

(2) *Suppose $z_k^i > v_k^i$. Then, there is a joint state $s^{-i} \in E_{k^{-i}}^{-i}$ of players $-i$ and a joint action $a_{s^{-i}}^{-i} \in A_{s^{-i}}^{-i} - \bar{A}_{s^{-i}}^{-i}$ (i.e., at least one player $j \neq i$ plays outside $\bar{A}_{s^{-i}}^j$) such that if players $-i$ play joint action $a_{s^{-i}}^{-i}$ in any state $s = (s^i, s^{-i}) \in E_k$, with $s^i \in E_{k^i}^i$, then player i 's min max level cannot increase in expectation from state s , regardless of the action played by player i . More precisely, for any $a_s^i \in A_s^i$ we have*

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

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

Due to $z_k^i < v_k^i$, when starting in E_k , player i can only defend his min max level v_k^i in G if he leaves E_k . Therefore, there must be at least one state $s^* \in E_k$, joint action $b_{s^*}^{-i} \in \bar{A}_{s^*}^{-i}$, and action $a_{s^*}^i \in A_{s^*}^i - \bar{A}_{s^*}^i$ such that

$$\sum_{t \in S} p_{s^*, (a_{s^*}^i, b_{s^*}^{-i})}(t) v_t^i \geq v_k^i.$$

Now we will show that state s^{*i} and action $a_{s^{*i}}^i$ satisfy part (1) of the Lemma 3.5.

Take an arbitrary state $s \in E_k$ with $s^i = s^{*i}$ and a joint action a_s^{-i} of payers $-i$ in state s . Note that for any $t^i \in S^i$,

$$\sum_{t^{-i} \in S^{-i}} p_{s^{-i} a_s^{-i}}^{-i}(t^{-i}) v_{(t^i, t^{-i})}^i \geq v_{(t^i, s^{-i})}^i = v_{(t^i, s^{*-i})}^i = \sum_{t^{-i} \in S^{-i}} p_{s^{*-i} b_{s^{*-i}}^{-i}}^{-i}(t^{-i}) v_{(t^i, t^{-i})}^i,$$

where the inequality follows from part (2) of Lemma 3.2 and the equalities from part (2) of Lemma 3.3. Therefore,

$$\begin{aligned} \sum_{t \in S} p_{s, (a_{s^*}^i, a_{s^*}^{-i})}(t) v_t^i &= \sum_{t^i \in S^i} p_{s^i a_{s^*}^i}(t^i) \left[\sum_{t^{-i} \in S^{-i}} p_{s^{-i} a_{s^*}^{-i}}(t^{-i}) v_{(t^i, t^{-i})}^i \right] \\ &\geq \sum_{t^i \in S^i} p_{s^i a_{s^*}^i}(t^i) \left[\sum_{t^{-i} \in S^{-i}} p_{s^{*-i} b_{s^*}^{-i}}(t^{-i}) v_{(t^i, t^{-i})}^i \right] \\ &= \sum_{t \in S} p_{s^*, (a_{s^*}^i, b_{s^*}^{-i})}(t) v_t^i \\ &\geq v_k^i, \end{aligned}$$

which completes the proof of part (1). \square

EXAMPLE 2. Consider the simple product-game G with two players given in Figure 2. 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 remains $(0, 0)$, while in $\bar{G}_{(II,II)}$, it remains $(1, -1)$. So, the game is simple, indeed.

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

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

Player 1's minmax level is clearly 0 for initial state $(3, 1)$, in view of player 2's first action. Now consider another arbitrary 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

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

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

With the help of this example, we will now illustrate some lemmas in this section.

Regarding Lemma 3.2: Consider, for instance, a solitary move for player 1 in state $(2, 1)$. Then, player 1's first action, as a solitary move, yields state $(1, 1)$, while the second action yields state $(3, 1)$. As $v_{(2,1)}^1 = v_{(1,1)}^1 = 1$ and $v_{(3,1)}^1 = 0$, player 1 is indeed unable to improve on his minmax level by such a solitary move.

Regarding Lemma 3.4: Consider 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 of the action chosen by player 2.

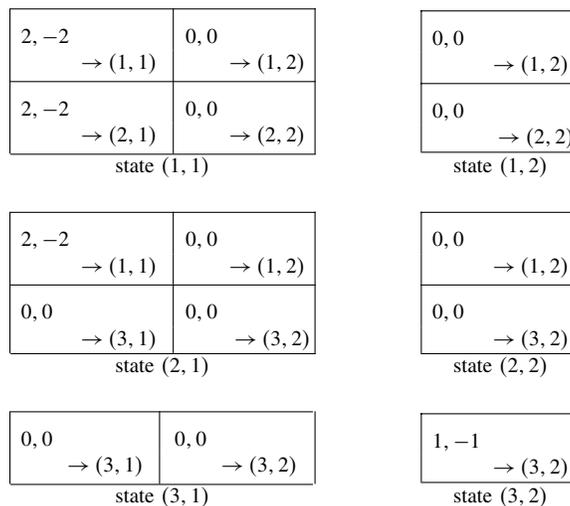


FIGURE 2. Game of Example 2.

Therefore, action 2 can never be better than action 1 for player 1, with respect to player 1's expected min max level after transition.

Regarding Lemma 3.5: Consider the restricted game $\bar{G}_{(I,II)}$, in which player 1's unique reward is 0 while his min max level v^1 is 1. Note that player 1 can leave $\bar{G}_{(I,II)}$ by moving to state (2, 2) and playing his second action there. By doing so, his min max level v^1 remains unchanged.

3.2.3. The min max 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 \tilde{G} , by replacing payoffs by min max levels, and claims that the min max levels of the players remain unchanged under this transformation. The idea to replace payoffs by min max levels also appeared in Solan [17] and in a more sophisticated way in Solan and Vohra [20].

LEMMA 3.6. *Take an arbitrary aperiodic product-game G , with v_s^i denoting the min max level for every player i and for every state $s \in S$. Let \bar{v}_k^i denote player i 's min max level in any restricted game \bar{G}_k (which is constant; cf., Lemma 3.1). Let \tilde{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 min max level \bar{v}_k^i . Further, let w_s^i denote every player i 's min max level in \tilde{G} in state s .*

Then, the min max levels of the product-games G and \tilde{G} are equal; i.e., $v_s^i = w_s^i$ 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 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_s^i ; i.e., for all strategies π^i for player i we have

$$\gamma_s^i(\pi^i, x^{-i}) \leq w_s^i. \quad (12)$$

This yields $v_s^i \leq w_s^i$ for all states $s \in S$. Since $v_s^i \geq w_s^i$ for all $s \in S$ follows in a similar fashion, we will have $v_s^i = w_s^i$ for all states s , and the proof will then be complete.

Now we prove (12). Note first that w_s^i is also a constant w_k^i on any joint maximal communicating set E_k , by Corollary 3.1 for the game \tilde{G} . We construct the joint stationary strategy x^{-i} by distinguishing the following three mutually exclusive cases.

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

$$\sum_{t \in S} p_{s, (x_s^{-i}, x_s^i)}(t) w_t^i \leq w_s^i.$$

Obviously, by expression (7) for player i 's min max level w^i in \tilde{G} , such a joint mixed action exists.

Case 2: States in a joint maximal communicating set E_k for which $\bar{v}_k^i \leq w_k^i$. 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 3.1. Then, let $x_s^{-i} = y_s^{-i}$ for all $s \in E_k$.

Case 3: States in a joint maximal communicating set E_k for which $\bar{v}_k^i > w_k^i$. Take a joint state $t^{-i} \in E_k^{-i}$ and a joint "exit" action $a_{t^{-i}}^{-i} \in \bar{A}_{t^{-i}}^{-i}$, with respect to the game \tilde{G} and its min max level w^i for player i , as in part (2) of Lemma 3.5. Then, for any $s = (s^1, \dots, s^n) \in E_k$, 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 , the mixed action x_s^i only uses actions from \bar{A}_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

$$\gamma_s^i(x^i, x^{-i}) \leq \bar{v}_k^i \leq w_k^i = w_s^i \quad (13)$$

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 3.4 and by part (2) of

Lemma 3.5 (both applied to w^i as the min max level of player i in \tilde{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 (13) 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 (12) is complete. \square

For an illustration of the above lemma, we refer to the games in Examples 1 and 2. Indeed, the product-game in Example 1 (which is now game G with min max levels v) leads to the simple product-game in Example 2 (which is now game \tilde{G} with min max levels w). Just as in the proof of the above lemma, we can construct a stationary strategy y^1 for player 1 (y^1 being x^{-2} for players $-i$ with $i = 2$) which guarantees in G that player 2's reward is not more than w_s^2 for all initial states $s \in S$. Recall that $w_{(3,1)}^2 = 0$ while $w_s^2 = -1$ for all $s \in S - \{(3, 1)\}$, and that player 2's min max levels within the restricted games are $\bar{v}_{(I,I)}^2 = -2$, $\bar{v}_{(I,II)}^2 = \bar{v}_{(II,I)}^2 = 0$, $\bar{v}_{(II,II)}^2 = -1$. Following the proof, as the joint maximal communicating sets $E_{(I,I)}$, $E_{(II,I)}$, and $E_{(II,II)}$ all belong to Case 2 (i.e., $\bar{v}_k^2 \leq w_k^2$ for $k = (I, I), (II, I), (II, II)$), the strategy y^1 has to guarantee in the corresponding restricted games that player 2's reward is not more than \bar{v}_k^2 . Also, y^1 has to leave $E_{(I,I)}$, belonging to Case 3 (i.e., $\bar{v}_{(I,I)}^2 > w_{(I,I)}^2$). It is easy to see that

$$y_{(1,1)}^1 = (1, 0), \quad y_{(1,2)}^1 = (0, 1), \quad y_{(2,1)}^1 = (1, 0), \quad y_{(2,2)}^1 = (0, 1), \quad y_{(3,1)}^1 = y_{(3,2)}^1 = (1)$$

satisfies all these requirements.

The previous lemma (and its proof) has useful consequences.

COROLLARY 3.2. *The results of Lemmas 3.2 to 3.4 in §3.2.2 for simple aperiodic product-games are also valid for any general aperiodic product-game G . Lemma 3.5 extends as well if one interprets z_k^i as the min max level \bar{v}_k^i of player i in the restricted game \bar{G}_k .*

Also, the infimum in expression (6) of the min max levels is attained at stationary strategies for all product-games. This is stated next.

COROLLARY 3.3 (OF THE PROOF OF LEMMA 3.6). *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 min max 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 3.3, 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. \square

3.2.4. The construction of 0-equilibria in product-games. In the previous sections we achieved several results for the min max levels of aperiodic product-games. We will use this knowledge now to construct 0-equilibria in aperiodic product-games.

The following lemma 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. The reasoning remains valid for all irreducible stochastic games. Note, however, that Flesch et al. [7] (with 3 players) and Simon [16] (with only 2 players) constructed examples proving that such a result would not hold for all stochastic games in general.

LEMMA 3.7. *Let G be an arbitrary aperiodic product-game and consider the restricted game \bar{G}_k , for any $k = (k^1, \dots, k^n) \in K$. 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. Indeed, for any extreme point z , Dutta [5] (cf., the appendix) shows that there exists a joint pure stationary strategy x which induces reward z for some fixed initial state $s \in E_k$. As $\gamma(x) = Q(x)\gamma(x)$ (as mentioned in (5)), $\gamma_s(x)$ is a convex combination of the rewards for x within the ergodic sets which can be reached from state s with a positive probability. Take such an ergodic set $F \subset E_k$. Then, because of property (i) and the choice of z , the joint strategy x must yield reward z within F . Since E_k is a joint maximal communicating set, there exists a joint pure stationary strategy y such that y coincides with x on F and, moreover, play reaches F from any initial state in E_k . Clearly, y induces reward z from all initial states in E_k .

(iii) Each min max level in \bar{G}_k is a constant \bar{v}_k^i , by Lemma 3.1.

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 [5]. Take an arbitrary feasible reward $z_k = (z_k^1, \dots, z_k^n)$ such that $z_k^i \geq \bar{v}_k^i$ 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_2^1 stages, \dots , then a_L for d_L^1 stages, and repeat this with lengths d_1^2, \dots, d_L^2 , then with lengths d_1^3, \dots, d_L^3 , and so on. The lengths d_l^m 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_l^m goes to infinity, so that the expected average payoff when strategy a_l is played for d_l^m stages will approach $\bar{\gamma}(a_l)$; (b) $d_l^m / (d_1^m + \dots + d_L^m)$ tends to α_l , so that a_l is played in the right proportion of time; (c)

$$\frac{d_l^m}{(d_1^1 + \dots + d_L^1) + \dots + (d_1^{m-1} + \dots + d_L^{m-1}) + (d_1^m + \dots + d_L^m)}$$

tends to 0, so that the average payoffs will fluctuate less and less. Because of 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 3.1. 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}_k^i , the joint strategy π is a 0-equilibrium and satisfies the requirements of the lemma. \square

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

PROOF OF MAIN THEOREM 1. For the idea of the construction, we refer to §3.1.

Let S° denote the set of states $s = (s^1, \dots, s^n) \in S$ in which s^i is of type 1 for at least one player i . Recall that, on each E_k , $k \in K$, the min max values in G and in \bar{G} are both a constant v_k and \bar{v}_k , respectively, due to Lemma 3.1 and Corollary 3.1.

In every restricted game \bar{G}_k with $k \in K^*$, take a 0-equilibrium σ_k as in Lemma 3.7. Let z_k^i denote the corresponding reward for any player i , which is independent of the initial state on E_k . Then, as $k \in K^*$, we have $z_k^i \geq \bar{v}_k^i \geq v_k^i$ for all players i .

In every restricted game \bar{G}_k with $k \in K[i]$, 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 as in part (1) of Lemma 3.5 (with Corollary 3.2).

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 joint maximal communicating sets E_k with $k \in K^*$. Then, in Step 2 we “extend” x^* to the joint strategy π according to which the players also receive rewards z_k in the sets E_k , $k \in K^*$. Finally, in Step 3, we will complete the proof by showing that π supplemented with some joint stationary strategies y^{-i} , for all i , forms a 0-equilibrium.

Step 1: *The construction of the joint stationary strategy x^* and a number of properties of x^* .* First, two supplementary games \tilde{G} and G^* have to be constructed. Then, given G^* , the joint strategy x^* will be found as a stationary 0-equilibrium in G^* .

Step 1.1: The simple aperiodic product-game \tilde{G} . For $k \in K - K^*$, let $z_k^i := \bar{v}_k^i$ for all players i . Then, let \tilde{G} denote the simple aperiodic product-game which is derived from G by replacing each player i 's payoffs by z_k^i in any restricted game \tilde{G}_k , $k \in K$.

Step 1.2: The stochastic game G^ .* In this step, we will derive a stochastic game G^* (not necessarily a product-game) from \tilde{G} . First, for every state s which belongs to some joint maximal communicating set E_k , $k \in K$, fix an arbitrary completely mixed action \bar{y}_s^i for every player i on \bar{A}_s^i . Now, we define a subset $X_s^{*i} \subset X_s^i$ of mixed actions for every player i in every state $s = (s^1, \dots, s^n) \in S$ as follows:

Case (1): $s \in E_k$ with $k \in K^*$. In this case, we let $X_s^{*i} := \{\bar{y}_s^i\}$ for all players i .

Case (2): $s \in E_k$ with $k \in K[i]$. In this case, 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 $X_s^{*i} := \{\bar{y}_s^i\}$.

Case (3): $s \in S^\circ$. In this case, we let $X_s^{*i} := X_s^i$ for all players i .

Note that $X_s^{*i} = X_s^i$ or X_s^{*i} is a singleton, for any player i in any state $s \in S$, and also that joint strategies in X^* can only differ in states $s \in S^\circ$. Moreover, the ergodic sets for all $x \in X^*$ are precisely E_k , $k \in K^*$, due to the use of the “exit” actions, which eventually make play leave each E_k belonging to Case (2).

Let G^* denote the stochastic game which is derived from \tilde{G} by restricting each player i in each state $s \in S$ to the space X_s^{*i} of mixed actions.

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 A.1 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 \tilde{G} and in the original game G , but not necessarily a 0-equilibrium.

We wish to point out three properties of x^* , with a proof.

Property (1): If $s \in S^\circ$, then no player i can go to better states regarding his reward in \tilde{G} by unilaterally deviating from x_s^{*i} , i.e., for every action $b_s^i \in A_s^i$ we have

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

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

Property (2): If $s \in E_k$ for some $k \in K$, then no player i can improve on his expected min max level in G by unilaterally deviating from x_s^{*i} , i.e., for every action $b_s^i \in A_s^i$ we have

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

Consequently, equality (7) also yields

$$v_s^i \leq \sum_{t \in S} p_{s, x_s^*}(t) v_t^i. \quad (14)$$

Property (3): x^* yields rewards in \tilde{G} that are at least the min max levels in G ; i.e., $\tilde{\gamma}_s^i(x^*) \geq v_s^i$ 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 property is a consequence of parts (1) of Lemmas 3.4 and 3.5 (with Corollary 3.2).

Proof of Property (3): This property requires a longer argument. By applying Equation (7) for the game G , in every state $s \in S^\circ$, there exists a mixed action $x_s^i \in X_s^i = X_s^{*i}$ for player i such that

$$\sum_{t \in S} p_{s, (x_s^i, x_s^{*-i})}(t) v_t^i \geq v_s^i. \quad (15)$$

Given these mixed actions x_s^i in all $s \in S^\circ$, there is a unique extension (with the mixed actions prescribed by x^* in all E_k , $k \in K$) to a stationary strategy x^i in X^{*i} . Then, by inequality (14), we obtain $P(x^i, x^{*-i})v^i \geq v^i$. Hence, $P^m(x^i, x^{*-i})v^i \geq v^i$ for all $m \in \mathbb{N}$, and therefore $Q(x^i, x^{*-i})v^i \geq v^i$.

Note that the ergodic sets are precisely E_k , $k \in K^*$, for both (x^i, x^{*-i}) and x^* , as both belong to X^* (cf., Step 1.2). Hence, if s is recurrent for (x^i, x^{*-i}) , then (as x^i equals x^{*i} on all recurrent states) we have

$$\tilde{\gamma}_s^i(x^i, x^{*-i}) = \tilde{\gamma}_s^i(x^*) = z_k^i \geq \bar{v}_k^i \geq v_k^i = v_s^i, \quad (16)$$

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

By applying equality (5), we now obtain

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

where the first inequality follows from (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^* and since $x^i \in X^{*i}$, we have

$$\tilde{\gamma}_s^i(x^*) \geq \tilde{\gamma}_s^i(x^i, x^{*-i}) \geq v_s^i$$

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): when play enters a set E_k with $k \in K^*$. In this case, the players switch to the joint strategy σ_k (cf., begin of the proof).

Case (2): when play enters a set E_k with $k \in K[i]$. In this case, players $-i$ switch to a joint stationary strategy as in Lemma 3.1, while player i follows x^* ; i.e., plays the mixed action x_s^{*i} in state $s \in E_k$.

Case (3): in any state $s \in S^\circ$. In this case, each player i follows x^* ; i.e., plays the mixed action x_s^{*i} .

Note that play according to π leaves all sets E_k belonging to Case (2), with the guidance of x^{*i} . On the other hand, in a set E_k in Case (1), by switching to σ_k , each player i receives 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 \tilde{G} . So in some sense, x^* is used to reach the sets E_k , $k \in K^*$, and then σ_k is used to induce the “right” payoffs.

Thus, by Property (3) of Step 1.3

$$\gamma_s^i(\pi) = \tilde{\gamma}_s^i(x^*) \geq v_s^i \tag{17}$$

for all initial states $s \in S$ and for all players i , which means that π induces individually rational rewards in G .

Step 3. Proving that π supplemented with the joint stationary strategies y^{-i} , for all players i , is a 0-equilibrium. For any player i , in view of Corollary 3.3, 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

$$\gamma_s^i(\tau^i, y^{-i}) \leq v_s^i.$$

Let the joint strategy η be defined as in §3.1 Note that the expected rewards are equal with respect to η and with respect to π , hence by (17),

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

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

$$\gamma_s^i(\eta[h]) = \gamma_s^i(\pi[h]) = \gamma_s^i(\pi) = \tilde{\gamma}_s^i(x^*) \geq v_s^i, \tag{18}$$

where for the second equality we used that for σ_k the “continuation rewards” remain the same due to Lemma 3.7. We may thus conclude that η yields individually rational rewards in G , 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 .

Deviations inside the support of η : Here, we only consider deviations by playing actions with a positive probability according to η . We show that such deviations by a player cannot improve his expected reward. Indeed, (i) within a set E_k , $k \in K^*$ (cf., Case (1) in Step 2), the players play the 0-equilibrium σ_k in \tilde{G}_k , (ii) within a set E_k , $k \in K[i]$ (cf., Case (2) in Step 2), such a 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 , $k \in K[i]$ (cf., Case (2) in Step 2), player i has an incentive to “exit,” since within \tilde{G}_k he can get at most \bar{v}_k^i , while $\bar{v}_k^i < v_k^i$, (iv) in states $s \in S^\circ$ (cf., Case (3) in Step 2), no player i can go to better states regarding his reward according to the equalities in (18) and to Property (1) from Step 1.3.

Deviations outside the support of η : Consider now a deviation when, for the first time, say after history h in state s , while the players should play a joint mixed action x_s' according to η , some player i deviates and plays an action b_s^i which has probability zero according to η^i ; i.e., $x_s^i(b_s^i) = 0$. 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 $\gamma_s^i(\eta[h]) = \tilde{\gamma}_s^i(x^*)$, in view of (18). Now, observe the following.

(A) Suppose $s \in S^\circ$. Then, $x_s^i = x_s^*$, and player i 's expected reward after this deviation is at most

$$\sum_{t \in S} p_{s, (b_s^i, x_s^{*-i})}(t) v_t^i \leq \sum_{t \in S} p_{s, (b_s^i, x_s^{*-i})}(t) \tilde{\gamma}_t^i(x^*) \leq \sum_{t \in S} p_{s, x_s^*}(t) \tilde{\gamma}_t^i(x^*) = \tilde{\gamma}_s^i(x^*) = \gamma_s^i(\eta[h]),$$

where the inequalities follow from properties (3) and (1) in Step 1.3, respectively, while the equations follow from (4) and (18). Hence, the deviation is not profitable.

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

$$\sum_{t \in S} p_{s, (b_s^i, x_s^{*-i})}(t) v_t^i \leq \sum_{t \in S} p_{s, x_s^*}(t) v_t^i \leq \gamma_s^i(\eta[h]).$$

Here, the first inequality follows from part (1) of Lemma 3.4 for $k \in K^*$, and from Property (2) in Step 1.3 if $k \in K - K^*$. The second inequality is a consequence of (18). 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. \square

REMARK 3.1. 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). Even though, Corollary 3.2 would yield that all min max 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 min max levels of this game coincide with the min max levels of the game in Example 2, hence

$$v_{(1,1)}^1 = v_{(1,2)}^1 = v_{(2,1)}^1 = v_{(2,2)}^1 = v_{(3,2)}^1 = 1, \quad v_{(3,1)}^1 = 0,$$

while $v^2 = -v^1$. For the min max values of the restricted games, we refer to §3.2.1 Hence, as to which restricted games are satisfactory, we obtain $K^* = \{(II, I), (II, II)\}$, while $K[1] = \{(I, II)\}$ and $K[2] = \{(I, I)\}$.

Now consider the pure stationary strategy x^1 for player 1 defined as

$$x_{(1,1)}^1 = (1, 0), \quad x_{(1,2)}^1 = (0, 1), \quad x_{(2,1)}^1 = (1, 0), \quad x_{(2,2)}^1 = (0, 1), \quad x_{(3,1)}^1 = x_{(3,2)}^1 = (1),$$

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

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

This pair (x^1, x^2) actually could play the role of π in this example. Indeed, in each restricted game \bar{G}_k , $k \in K^*$, the pair (x^1, x^2) lets the players play a 0-equilibrium, while x^1 leaves $E_{(I, II)}$ and x^2 leaves $E_{(I, I)}$. Notice that no threat strategies are needed here, so (x^1, x^2) is a 0-equilibrium.

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 $(2, 2)$, and 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 3.2 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. Nevertheless, we conjecture that the main results of this paper extend to the periodic case as well.

Appendix.

LEMMA A.1. *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

$$\gamma_{\beta s}^i(x) := (1 - \beta) \sum_{m=1}^{\infty} \beta^{m-1} \mathbb{E}_{sx}(R_m^i),$$

where R_m^i 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 [9] and Takahashi [22] 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 [6]) that for any sequence of discount factors β_m converging to 1 and joint strategies x_m converging to x we have

$$\gamma_s^i(x) = \lim_{m \rightarrow \infty} \gamma_{\beta_m s}^i(x_m) \quad (19)$$

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 $x = \lim_{m \rightarrow \infty} x_m$. 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 (19) and from the fact that x_m is a β_m -discounted 0-equilibrium, it follows that

$$\gamma_s^i(y^i, x^{-i}) = \lim_{m \rightarrow \infty} \gamma_{\beta_m s}^i(y^i, x_m^{-i}) \leq \lim_{m \rightarrow \infty} \gamma_{\beta_m s}^i(x_m) = \gamma_s^i(x).$$

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. \square

Acknowledgments. The authors thank Eilon Solan and an anonymous referee for their constructive suggestions, which helped improve the presentation and relate the paper to existing literature.

References

- [1] Altman, E., K. Avrachenkov, R. Marquez, G. Miller. 2005. Zero-sum constrained stochastic games with independent state processes. *Math. Methods Oper. Res.* **62** 375–386.
- [2] Blackwell, D. 1962. Discrete dynamic programming. *Ann. Math. Statist.* **33** 719–726.
- [3] Blackwell, D., T. S. Ferguson. 1968. The big match. *Ann. Math. Statist.* **39** 159–163.
- [4] Doob, J. L. 1953. *Stochastic Processes*. John Wiley & Sons, New York.
- [5] Dutta, P. K. 1995. A Folk theorem for stochastic games. *J. Econom. Theory* **66** 1–32.
- [6] Flesch, J. 1998. Stochastic games with the average reward. Unpublished doctoral dissertation, University of Maastricht, The Netherlands.
- [7] Flesch, J., F. Thuijsman, O. J. Vrieze. 1997. Cyclic Markov equilibria in a cubic game. *Internat. J. Game Theory* **26** 303–314.
- [8] Flesch, J., F. Thuijsman, O. J. Vrieze. 2007. Stochastic games with additive transitions. *Eur. J. Oper. Res.* **179** 483–497.
- [9] Fink, A. M. 1964. Equilibrium in a stochastic n -person game. *J. Sci. Hiroshima Univ., Ser. A-I* **28** 89–93.
- [10] Gillette, D. 1957. Stochastic games with zero stop probabilities. M. Dresher, A. W. Tucker, P. Wolfe, eds. *Contributions to the Theory of Games*, Vol. III. *Ann. Math. Stud.* **39**. Princeton University Press, Princeton, NJ, 179–187.
- [11] Hordijk, A., O. J. Vrieze, G. L. Wanrooij. 1983. Semi-Markov strategies in stochastic games. *Internat. J. Game Theory* **12** 81–89.
- [12] Maitra, A., W. Sudderth. 1998. Finitely additive stochastic games with Borel measurable payoffs. *Internat. J. Game Theory* **27** 257–267.
- [13] Mertens, J. F., A. Neyman. 1981. Stochastic games. *Internat. J. Game Theory* **10** 53–66.
- [14] Neyman, A. 2003. Existence of the minmax. A. Neyman, S. Sorin, eds. *Stochastic Games and Applications, NATO Science Series*, Vol. 570. Kluwer Academic Press, Norwell, MA, 173–193.
- [15] Ross, K. W., R. Varadarajan. 1991. Multichain Markov decision processes with a sample path constraint: A decomposition approach. *Math. Oper. Res.* **16** 195–207.
- [16] Simon, R. 2006. Value and perfection in stochastic games. *Israel J. Math.* **156** 285–309.
- [17] Solan, E. 1999. Three-player absorbing games. *Math. Oper. Res.* **24** 669–698.
- [18] Solan, E. 2003. Perturbations of Markov chains with applications to stochastic games. A. Neyman, S. Sorin, eds. *Stochastic Games and Applications, NATO Science Series*, Vol. 570. Kluwer Academic Press, Norwell, MA, 265–280.
- [19] Solan, E., N. Vieille. 2002. Correlated equilibrium in stochastic games. *Games Econom. Behav.* **38** 362–399.
- [20] Solan, E., R. Vohra. 2002. Correlated equilibrium and public signalling in absorbing games. *Internat. J. Game Theory* **31** 91–121.
- [21] Sorin, S. 1986. Asymptotic properties of a non-zero-sum game. *Internat. J. Game Theory* **15** 101–107.
- [22] Takahashi, M. 1964. Equilibrium points of stochastic noncooperative n -person games. *J. Sci. Hiroshima Univ., Ser. A-I* **28** 95–99.
- [23] Thuijsman, F., O. J. Vrieze. 1991. Easy initial states in stochastic games. T. E. S. Raghavan, T. S. Ferguson, O. J. Vrieze, T. Parthasarathy, eds. *Stochastic Games and Related Topics*. Kluwer Academic Press, Dordrecht, The Netherlands, 85–100.
- [24] Vieille, N. 2000. Equilibrium in 2-person stochastic games I: A reduction. *Israel J. Math.* **119** 55–91.
- [25] Vieille, N. 2000. Equilibrium in 2-person stochastic games II: The case of recursive games. *Israel J. Math.* **119** 93–126.